



Journal of Service Science and Management

Chief Editor: Samuel Mendlinger



Journal Editorial Board

ISSN: 1940-9893 (Print), 1940-9907(Online)

<http://www.scirp.org/journal/jssm>

Editor-in-Chief

Prof. Samuel Mendlinger

Boston University, USA

Editorial Board (According to Alphabet)

Dr. Sónia Bentes	Institute of Accounting and Management of Lisbon, Portugal
Dr. Hengjin Cai	Wuhan University, China
Dr. Chia-Fen Chi	National Taiwan University of Science & Technology, Taiwan, China
Dr. Tsan-Ming Choi	Hong Kong Polytechnic University, Hong Kong, China
Dr. Yong Deng	Shanghai Jiao Tong University, China
Dr. Jerry Fjermestad	New Jersey Institute of Technology, USA
Dr. Leung Kit-nam Francis	City University of Hong Kong, Hong Kong, China
Dr. Javier Sánchez García	Universitat Jaume I, Spain
Dr. Dai-Ping Hu	Shanghai Jiao Tong University, China
Dr. Guang-Wei Hu	Nanjing University, China
Dr. Jongsu Lee	Seoul National University, Korea
Dr. Patrick L. Leoni	University of Southern Denmark, Denmark
Dr. Patrick C. H. Ma	Hong Kong Polytechnic University, Hong Kong, China
Dr. Mehron Sepehri	Sharif University of Technology, Iran
Dr. Chun-Shan Shen	Chinese Academy of Sciences, China
Dr. Kan-Liang Wang	Xi'an Jiao Tong University, China
Dr. Shu-Liang Wang	Wuhan University, China
Dr. Xu-Song Xu	Wuhan University, China
Dr. Fengqi You	Carnegie Mellon University, USA
Dr. Tingsheng Zhao	Huazhong University of Science and Technology, China
Dr. Hong-Ping Zhu	Huazhong University of Science and Technology, China

Editorial Assistant

Xiao-Qian Qi

Scientific Research Publishing, USA

Guest Reviewers (According to Alphabet)

Stephan Aier	Anders Henten	Masanori Ono
Marco Bertoni	Junfeng Jiang	Bo van der Rhee
Paolo Boccaredelli	Daniel Klapper	Alberto Sánchez
Ruay-Shiung Chang	Stefan Lessmann	Olaf Schneider
Sohail Chaudhry	Kin Fun Li	Dimitri Schuurman
Arvind Chel	Kunpeng Li	Alessandra Vecchi
Gang Chen	Chung-Chu Liu	Robert Winter
Shin-Guang Chen	Steven Lynden	Meilong Xie
Pi-Yu Chen	Charles Møller	Jie Zhang
Charlotte Christiansen	Masoud Zare Naghadehi	
Torsten J. Gerpott	Ulf Nielsson	

CONTENTS

Volume 2 Number 4

December 2009

Culture and Organizational Improvisation in UK Financial Services

S. A. Leybourne.....237

Service System Decoupling for Mass Customization: A Case Study in Catering Services

J. Chen & Y. H. Hao.....255

Periodicity of Cycle Time in a U-Shaped Production Line with Heterogeneous Workers under Carousel Allocation

M. Hiraiwa & K. Nakade.....265

The Exploration of Implementation of Emissions Trading System in China

M. Hu.....270

The Research on the Appraisal of Market Opportunity Based on AHP

D. H. Yang, Y. J. Tan & Y. B. Sun.....276

Dalian High-Tech SMEs Growth Evaluation Based on Catastrophe and Principal Component Projection Method

L. Li, P. F. Zhou & Z. H. Li.....282

How to React to the Subprime Crisis? - The Impact of an Interest Rate Freeze on Residential Mortgage Backed

J. Hein & T. Weber.....289

Flight Attendants' Emotional Labor and Exhaustion in the Taiwanese Airline Industry

C.-P. Chang & J.-M. Chiu.....305

Towards a More Business-Oriented Definition of Corporate Social Responsibility: Discussing the Core Controversies of a Well-Established Concept

M. S. Fifka.....312

Combining Personal Ontology and Collaborative Filtering to Design a Document Recommendation System

D.-N. Chen & Y.-C. Chiang.....322

The Dynamic Multi-Task Supply Chain Principal-Agent Analysis

S. L. Li, C. H. Wang & D. L. Zhu.....329

Narrating National Geo Information Infrastructures: Balancing Infrastructure and Innovation

H. Koerten & M. Veenswijk.....334

The Transmission of Pricing Information of Dually-Listed between Hong Kong and New York Stock Exchange

S. F. Li & S. Chen.....348

Problems Facing Small and Medium Size Contractors in Swaziland

W. D. Thwala & M. Mvubu.....353

A Projection Clustering Technique Based on Projection

X. Y. Liu, X. J. Xie & W. P. Wang.....362

Revenue and Duration of Oral Auction

J. M. Shi & A.-C. Chang.....368

Informal Financing of Small – Medium Enterprise Sector: The Case of Greece

P. Petrakis & K. Eleftheriou.....378

Knowledge Innovative Organization: The Effect of Constant Organization Renewal

A. S. M. Sohel-Uz-Zaman & U. Anjalin.....384

Study on Option Price Model of the Transaction of Information Commodities

C. P. Hu & X. J. Qi.....394

A Personalized Recommendation Algorithm Based on Associative Sets

G. R. Jiang, H. Qing & T. Y. Huang.....400

Gender Differences in Satisfaction with the Type of Work University Employees Do: Evidence from the University of Botswana

T. T. Fako, S. R. T. Moeng & N. Forchheh.....404

A Study of the Joint Advertising Channels

M. Lei, S. G. Sun & D. Yang.....418

Grey Incidence Relation Analysis and Granger Causality Tests of the Income Level and Economic Growth – Case Study on Gansu Province, China

B. Xue, X. P. Chen, W. W. Zhang, J. Wang, X. J. Guo & Y. Geng.....427

The Development of Services in Customer Relationship Management (CRM) Environment from ‘Technology’ Perspective

G. K. Agrawal & D. Berg.....432

Journal of Service Science and Management (JSSM)

SUBSCRIPTIONS

The *Journal of Service Science and Management* (Online at Scientific Research Publishing, **www.SciRP.org**) is published quarterly by Scientific Research Publishing, Inc., USA.

E-mail: **service@scirp.org**

Subscription rates: Volume 2 2009

Print: \$50 per copy.

Electronic: freely available on **www.SciRP.org**.

To subscribe, please contact Journals Subscriptions Department, E-mail: **service@scirp.org**

Sample copies: If you are interested in subscribing, you may obtain a free sample copy by contacting Scientific Research Publishing, Inc. at the above address.

SERVICES

Advertisements

Advertisement Sales Department, E-mail: **service@scirp.org**

Reprints (minimum quantity 100 copies)

Reprints Co-ordinator, Scientific Research Publishing, Inc., USA.

E-mail: **service@scirp.org**

COPYRIGHT

Copyright© 2009 Scientific Research Publishing, Inc.

All Rights Reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as described below, without the permission in writing of the Publisher.

Copying of articles is not permitted except for personal and internal use, to the extent permitted by national copyright law, or under the terms of a license issued by the national Reproduction Rights Organization.

Requests for permission for other kinds of copying, such as copying for general distribution, for advertising or promotional purposes, for creating new collective works or for resale, and other enquiries should be addressed to the Publisher.

Statements and opinions expressed in the articles and communications are those of the individual contributors and not the statements and opinion of Scientific Research Publishing, Inc. We assume no responsibility or liability for any damage or injury to persons or property arising out of the use of any materials, instructions, methods or ideas contained herein. We expressly disclaim any implied warranties of merchantability or fitness for a particular purpose. If expert assistance is required, the services of a competent professional person should be sought.

PRODUCTION INFORMATION

For manuscripts that have been accepted for publication, please contact:

E-mail: jssm@scirp.org

Culture and Organizational Improvisation in UK Financial Services

Stephen A. Leybourne

Boston University, Boston, USA.
Email: sleyb@bu.edu

Received August 21, 2009; revised October 6, 2009; accepted November 10, 2009.

ABSTRACT

This paper considers certain aspects of a four-year program of research, and addresses the changing cultural requirements to support the rise of improvisational working practices within the UK financial services sector. Specifically, it reports on some of the outcomes of a study encompassing over 100 hours of interviews, together with a variety of other primary and secondary data. The outcomes of the full study are documented elsewhere, and they identify a number of key factors that contribute to the successful use and control of improvisational working practices. One of these factors is a supportive organizational culture, and this specific area is dealt with in this paper. A particular focus is how the sample of organizations has attempted to identify and create supportive cultural conditions for improvisational work to take place. In order to bring clarity to the outcomes of this study, a matrix of the case study organizations is also offered, which segregates those organizations according to their cultural support for improvisation and apparent improvisation effectiveness. Some comment on the current difficulties in the Financial Services sector has also been included, as it could be argued that improvisation may have contributed to shortcomings in control processes by members of that sector.

Keywords: Improvisation, Culture, Financial Services

1. Introduction

Over the last ten years or so, evidence has emerged suggesting that more progressive organizations are moving away from a slavish adherence to agreed processes and procedures. Rather, they are exploring different, more radical ways of competing, which depend on allowing trusted and empowered employees to experiment with more creative and less predictable ways of achieving. This developing trend has been labeled organizational improvisation. The literature on organizational improvisation has matured, building on early philosophical ideals from Ryle (1979), and more organizationally-oriented work such as Weick's (1979) early insights into sense-making. A number of later contributions have organized the evolving output (notably Cunha *et al.*, 1999), and placed it in an appropriate managerial and theoretical context. As a result of this activity, organizational improvisation has progressed from being seen as a dysfunction resulting from poor planning (Quinn, 1980), to participating in, and becoming more recognized, within the lexicon of management theory (Leybourne, 2005). As our understanding of the antecedents, influencing factors, and outcomes of improvisation becomes more compre-

hensive, it is time to focus on the supporting framework that allows successful improvisation to flourish in organizations. Arguably, one of the more influential of those pre-existing conditions for effective organizational improvisation is a supportive organizational culture and climate.

This raises the question of how culture and climate support effective improvisational working practices. This paper will therefore examine various dimensions of organizational improvisation, and the way in which it is used within a range of organizations operating within a specific business sector, the intention being to isolate and analyze identifiable components of organizational culture and climate that may encourage or negate effective improvisation.

Although at the superficial level some practitioners perceive little difference between culture and climate, the academic definitions are quite distinct. Deal and Kennedy (1982) talk of culture in terms of: "*the way things are done around here*", although a more exact definition is "*the collection of traditions, values, policies, beliefs, and attitudes that constitute a pervasive context for everything we do and think in an organization*" (McLean & Marshall, 1993). Organizational climate is a rather more ephemeral concept, and one which Mullins (1999: 810)

suggests, when applied to organizations “*can be said to relate to the prevailing atmosphere surrounding the organization, to the level of morale, and to the strength of feeling or belonging, care and goodwill among members*”. This review will deal with organizational culture first, examining it at both the sectoral and organizational level, before turning to the notion of climate.

The concept of culture has “*...been borrowed from anthropology, where there is no consensus on its meaning*” (Smircich, 1983: 339). Jelinek *et al.* (1983: 331) however suggest that culture is “*...another word for social reality*”, and that it is “*the shaper of human interaction and the outcome of it, continually created and recreated by people’s ongoing interactions*” (Jelinek *et al.*, 1983: 331). Although this description of culture is by no means universally adopted, it does suggest that cultural norms are constantly changing. We are however interested in culture from an organizational and also from a sectoral viewpoint, and there have been a number of attempts to define culture in these contexts. Corporate or organizational culture is defined by Gordon (1991: 397) as “*...an organization-specific system of widely shared assumptions and values that give rise to typical behavior patterns*.” Whipp *et al.* (1989: 565) suggested that “*...the concept of culture, at the level of the firm, refers to the collection of beliefs, values and assumptions held by the members of an organization*.” Schein (1985) went further, defining three levels of cultural phenomena in organizations: at the surface level, behaviors; at the middle level, values; and at the deepest level, basic assumptions. He considered that the deeper basic assumptions were the essence of culture.

The emergence of organizational culture as an area of academic and managerial interest stems from a number of historic circumstances, including the changes in the nature of work brought about by flexible working and the breakdown of the “theory X” approach to management. As employees responded to organizational desire for “multi-skilled” capability, a desire has emerged to align or develop a “sense of belonging” to the organization. This has resulted in an increasing focus on organizational culture and climate, reflected in sets of differing values, norms, and beliefs, embedded in different structures and systems (Handy, 1993). It is also evident that different sets of these elements emerge at different loci, resulting in fragmented cultures.

The distinction between culture and climate can be a fragile one, and is often disputed. It is suggested that organizational climate promotes a psychological approach (Denison, 1990; Linstead, 2004), whilst neglecting the cultural and symbolic forms that inform culture. This indicates that climate can be short-lived, in that it reflects employees “feelings” towards an organization, which are prone to change (Schneider, 1983). Although both culture and climate are linked to the value system of organ-

izational members (Rollinson & Broadfield, 2002), the traditional view is that culture generates the values that are a component of how people act within an organization, whilst climate reflects the alignment between organizational and personal values. An alternative view, propounded by Denison (1996), is that the differences between culture and climate are minimal, and that the differences manifest themselves in terms of measurement.

Brookes & Dawes (1999) consider the issue of merger or consolidation as a trigger for cultural change within organizations, suggesting that merger is an ideal opportunity to “*...achieve a cultural change on a broader footing, in particular, embedding a much more customer-driven philosophy which could generate significant competitive advantage*” (Brookes & Dawes, 1999: 197). They also highlight the fact that they are espousing the “*...human aspects of organization*”, or to apply another label, the socio-cultural systems which comprise organizations (Brookes & Dawes, 1999: 195).

It is apparent that much environmental change has taken place in the management of the organization, and one of the ways that these adjustments are manifesting themselves is in a relaxation of processes and procedures, and a move towards allowing trusted and empowered employees to experiment with more creative, and less controlled, ways of achieving. Organizational improvisation is an example of one way that such creativity is being applied, although there are other established debates that contribute to the dismantling of organizational bureaucracy and the rise of autonomous working styles. Notably, the principles of responsible autonomy (Friedman, 1977), professional autonomy (Freidson, 2001; Faulconbridge & Muzio, 2008), post-bureaucracy (Hecksher, 1994), and employee empowerment itself have all assisted in or argued the benefits of the relaxation of bureaucratic control.

However, responsible autonomy, where the aim is to integrate the workers to the goals of the organization by giving them independence and by encouraging them to adapt to possible changes at work and in work organization, is linked to manipulative aspects of employee “control” and to career hierarchies and progression, and to job security (Sturdy, Knights & Willmott, 1992), whereas improvisation is enabled by “consent” to experiment with new ways of achieving organizational tasks and activities. Functional autonomy, which is defined by Friedson (1970: 53) as “*the degree to which work can be carried out independently of organizational or medical supervision and can attract its own clientele independently*” has elements that align usefully with improvisational work, but is essentially embedded in medical or professional, or client-based practice (Faulconbridge & Muzio, 2008). Post-bureaucracy is the antithesis of the principles of bureaucratic organization (Hecksher, 1994), involving rationality and the allocation of defined and planned

work. Although this mirrors the ethos of improvisation, the creative and intuitional components of improvisation (Moorman & Miner, 1998a) add significantly to its effectiveness as an enabler of change, and as a lens for the analysis of cultural change.

There have been a number of comprehensive reviews of organizational improvisation (Cunha *et al.*, 1999; Leybourne, 2006), much of which has evolved from Weick's (1979) work on sense-making, and Moorman and Miner's (1998a; 1998b) output that assists in identifying the early key constructs of creativity, intuition, and bricolage. Later work by Miner, Bassoff and Moorman (2001) posited adaptation, innovation, compression (in the temporal sense), and learning as additional constructs of improvisational activity within organizations. The development of management practices since the turn of the millennium has embraced a number of these constructs as important outcomes and antecedents of organizational performance, leading to an appreciation of improvisation as a lens for the analysis of organizational activity.

From a philosophical stance improvisation relates to how thoughts develop. Ryle (1979: 125) suggests that; *"the vast majority of things that happen [are] unprecedented, unpredictable, and never to be repeated"*, and that *"the things we say and do... cannot be completely pre-arranged"*. To a partly novel situation the response is necessarily partly novel, else it is not a response. His assertion is that however much an activity is planned, there will always be a novel set of circumstances to deal with, and that improvisation requires using resources that are to hand to resolve unforeseen circumstances. This assertion also resonates with the tenets of autonomous working, which is under adoption by modern organizations, and which is also the essence of bricolage (Lehner, 2000), which in turn is an essential component of improvisation (Moorman & Miner, 1998a).

From the mid-1990s onwards much of the literature on improvisational work practices within organizations took this stance and applied it to organizational routines and processes. Some of the outcomes from these debates apply metaphors to explain the way improvisation is used, for example adopting and applying ideas from jazz performance (Barrett, 1998a & 1998b; Eisenhardt, 1997; Hatch, 1998 & 1999), and from improvisational theatre (Crossan, 1997; Kanter, 2002; Yanow, 2001). Later work used grounded theory approaches to consider the temporal aspects of improvisation, and particularly pressure to achieve complex tasks to a demanding or compressed timetable (Brown & Eisenhardt, 1997; Moorman & Miner, 1998a & 1998b). This work is building the foundations to allow empirical research of a more positivist nature—for example: Akgun & Lynn's (2002) work on the links between improvised new product development and speed-to-market. Latterly, consideration has also been given to the interactions between improvisation and

learning (Chelariu *et al.*, 2002, Miner *et al.*, 2001), improvisation and entrepreneurial activity (Baker *et al.*, 2003; Hmieleski & Corbett, 2003) and the ways in which tacit knowledge (upon which intuition, and therefore improvisation, may draw) is acquired (Koskinen, Pihlanto & Vanharanta, 2003), and the role of experience in the acquisition of tacit knowledge (Cooke-Davis, 2002). This is in turn feeding in to improvisation as a tool for strategic decision-making within turbulent environments (Velez-Castrillon, Vera & Kachra, 2008).

The outcomes of this emerging literature base include an appreciation of the benefits and effectiveness of improvisational working practices, both as a tool and an appreciated skill for managers, and as a lens for the analysis of organizations. This stems from early work by Orlikowski and Hofman (1997), who suggest that organizational transformation (which is inherent in all modern organizations) is an ongoing improvisation enacted by organizational actors trying to make sense of and act coherently in the world, offering strong links to Weick's sense-making model. Mendonça, Cunha, Kaivo-oja and Ruff (2004: 213) suggest that *"a crucial element for improvisation to occur is the existence of a 'safe' environment"* which sees that the errors are not only inevitable, but also potential sources of learning, linking strongly with the requirement for a supportive organizational culture.

2. Samples and Methods

The study that underpins this research, and provides much of the primary data upon which the findings articulated in this paper have been based, was located in a sub-sector of the U.K. financial services sector. Six retail lending institutions, ranging from a major quoted bank, through building societies and ex-building societies (U.K.-based mutually-owned organizations originally formed specifically to supply housing finance), to smaller retail lending organizations, were used. This sample was chosen taking into account the relative populations of organizations in each of the sub-sectors, the required number of cases required to provide an opportunity to develop theory (Eisenhardt, 1989: 537; Stake, 1994: 237), and the need to include cases with differing characteristics, or polar types (Pettigrew, 1988).

The data collection and analysis involved a number of visits to each participant organization. The larger organizations received up to eight visits, with the smaller and more compactly organized organizations receiving from three to six visits. Over ninety employees and organizational stakeholders were interviewed, using a semi-structured interview framework derived from a rigorous operationalisation process, underpinned by the literature review. This resulted in the collection of in excess of 100 hours of highly relevant and focused interview data. Observational and secondary data, including project documentation, was also incorporated into this

qualitative study. This data was analyzed, using the Huberman and Miles (1998) Interactive Data Analysis model as a guiding principle for data relevance, inclusion and reduction, and each of the six organizations was written up as a case study. Cross-case analysis was then undertaken, and themes, trends, and modes of operation were identified.

The use of computer aided qualitative data analysis software (CAQDAS) was considered, but was rejected, as there is a perceived danger that in adopting CAQDAS tools to analyze qualitative data, the researcher is forced to adapt to requirements imposed by the software (Woolgar, 1991; Lonkila, 1991). Lonkila suggests that such software programs “...could also be misused as purely rhetorical weapons to convince the readers or academic community of the scientific nature of ones research” (Lonkila, 1991: 46), and that it can “...prevent an interactive and easy movement between emerging conceptual structures and the data” (Lonkila, 1991: 49). Agar also sees a potentially destructive step in the process of analyzing qualitative data using software, notably, that there is a danger that a researcher may “...conduct an ethnography to maximize fit between the process and the available software” (Agar, 1991: 193). Fielding and Lee (1998: 68-84) also synthesise a number of disadvantages of CAQDAS, including accessibility and availability (Russell & Gregory, 1993), the exaggeration of possible benefits (Wietzman & Miles, 1995: 335), lack of closeness to the data (Agar, 1991: 185), and unintended consequences (Seidel, 1991: 109).

The study identified a number of processes, mechanisms and routines that the six organizations used (formally and informally) to develop and manage improvisational working practices. Often this improvisational activity is informal and surreptitious, and as such causes problems for the organizations, especially as they operate in a sector that is traditionally risk-averse (Brooks & Dawes, 1999: 197; Trethowen & Scullion, 1997: 62). The prime focus of this paper is however the movement away from standardized and documented processes and mechanisms, towards more improvisational modes of working, and how culture and climate within organizations may assist or hinder this shift.

3. Case Study Findings

Inevitably, the six case study organizations displayed varying levels of maturity, sophistication, and effectiveness in their understanding, control and management of improvisational working practices, and the culture that has emerged or been cultivated to support that activity. To some extent, this was linked to confidence in the ability of employees to improvise effectively, within a given framework that limited the extent of improvisational activity. The cultural norms and values of the organizations also significantly affected the way in which improvisa-

tional working practices were used, or in some cases, abused. Each organization that contributed to the study is examined separately, the various issues examined, and a summarizing analysis made. A comparison across the six organizations will then be considered as a part of the wider conclusions.

3.1 BigBank

At *BigBank*, a major U.K.-based bank with over 77,000 employees, there was evidence of improvisational activity within a fragmented organizational culture, brought about by challenges in managing the merging of disparate acquisitions and merger partners. Notably, the management core of *BigBank* is formed from the senior managers of two merged banks with a centralized structure dominating, resulting in significant resistance from managers used to managing their own organizational domains.

One senior manager considered that it could possibly take a generation to remove the ingrained loyalties to one or other component bank, and suggested that “*I could find you an ‘Atilla the Hun’ culture... and at the other end, a ‘Mahatma Gandhi’ culture. There is so much opportunity for individuals to build a culture how they want it to be in their area...*”. The branch network has its own cultural problems. Branches often appear to take their cultural lead from the branch manager, and two of the branches visited had tight-knit, customer-focused teams, in each case led by a young manager who had built their own supportive branch culture. In direct contradiction to this approach, the I.T. and operations areas of the bank are traditionally acknowledged to have a more confrontational work environment, with greater personal accountability, a feeling that is noticeably less prevalent in other divisions of the organization. This may be compounded by the fact that, particularly within the systems areas of the bank, there is a strong focus and emphasis on delivery against the project timetable. It was also suggested that this delivery focus is present regardless of whether the need for the change has been superseded.

Within the merged bank, the part of the merged organization where managers were originally employed is referred to as “heritage”. This term refers to the constituent part of the merged bank in which a person was originally employed, and there is an inference that employees still have loyalty to managers from their own heritage. It is however suggested that 70% of the junior staff employed by the merged organization have joined since 1995. These junior staff have little loyalty to either historic heritage within the enlarged bank. Another issue that has affected employees, particularly in the lending areas of *BigBank*, is the centralization of decision-making, and the removal of lending discretion from line managers within the branch network. This shift is familiar to employees from one half of the merged entity, but alien to employees from the other. Indeed, one Branch

Manager in her 30's, who has been with one of the newly merged banks for her entire working life, suggested that prior to the 1995 merger, she *"used to have an awful lot of discretion. It used to be a proper bank"*. Amongst more established middle managers, attitudes do appear to be linked to the working arrangements within the component bank where career progression was achieved.

The indicators are that although *BigBank* is moving forward in its re-focusing of the organization from a product to a customer focus, there is much work still to do. The change program run by the University of *BigBank* is however only experienced by about 200 middle managers each year, comprising less than 0.25% of the employee base, and the clash of heritage, which closely equates to historic culture within *BigBank*, is still causing concern.

There is significant support for improvisation within *BigBank*, with many operational areas, all change management respondents, and all project managers confirming that much of their work is improvised. The virtual university disseminates research from, amongst other sources, the Cranfield School of Management on informal pockets of best practice, which is seen by *BigBank* as closely linked to improvisation. There is evidence that although the bank has formal structures and processes, another set of activities achieves much of the real progress. These activities are based on informal networks and relationships, and they appear to operate outside of the procedures and guidelines documented by the bank.

There is recognition within *BigBank* that this can assist in managing change. One manager linked this to the organizational climate of the bank, arguing:

"I think the climate... is more one of, if you are working within the broad framework of... management and you can find a better way of achieving the goal more effectively, then I think you have got the freedom to [do that]. Therefore, improvise or innovate"

This attitude indicates strong support for improvisational work routines. Also, evidence from within the organization underpins the fact that at many levels and functional specializations, improvisation is supported.

During the discussions within *BigBank* many instances of these informal routines and improvisational activities were identified. A senior manager in the Management Development area of the bank suggested that:

"Actually most of the work that does happen in the organization is done through informal structures, loose collaborations of people who have similar views or initiatives... Actually you just get on the phone and make something happen informally, which fits with the idea of improvisation for me".

Improvisation is therefore recognized as a valid, and in some areas an essential, component in change. However, there is evidence that the opportunity to improvise is being deliberately removed from branch systems and processes, which may run counter to *BigBank's* intention to focus on customers, many of whom use a branch as their principal channel to access products offered by the bank. Table 1 displays the cultural and improvisational attributes of *BigBank*.

3.2 MutualCo

MutualCo is one of the few surviving Mutually-owned mortgage lending organizations in U.K., with a workforce of approximately 1300. It devotes significant time and energy to the development of its employees, and a recent staff satisfaction survey showed that matters were improving. Senior managers within the society are taking more interest in the skills and development of employees. Managers suggest that the focus has moved from delivery of benefits, to a more behavioral focus, where training and the effect of changes on a developing workforce is appreciated.

Table 1. Cultural and improvisational attributes of *BigBank*

<i>BigBank</i> – Cultural and Improvisational Attributes	
<u>Culture – Positive Attributes</u>	<u>Culture - Negative Attributes</u>
<ul style="list-style-type: none"> • Strong linkages between projects, H.R., and training • Dedicated area change managers monitoring impact of change on staff • Dedicated department to communicate major change • Use of training maps for career development • Attempts to measure behavioral change 	<ul style="list-style-type: none"> • Excessive use of political influence by senior managers • Project procedures and methods have little coverage of people/cultural issues • Many layers of authority/hierarchy • Lack of time for self-training • Cultural differences between merging banks and between operational areas causing entrenched resistance
<u>Improvisation - Positive Attributes</u>	<u>Improvisation - Negative Attributes</u>
<ul style="list-style-type: none"> • Support for improvisation implicit within the organization • Training on aspects of improvisation • Informal networks to assist with improvisation • Improvisation accepted as contributing to project-managed change 	<ul style="list-style-type: none"> • Rigidity in customer-facing procedures negates improvisation • Perception that senior managers are deeply entrenched in traditional banking mentality

Recently there has been a more focused approach towards culture and employee behaviors. However, the Group I.T. Manager attempted to put this into perspective, suggesting that:

“the current man [new chief executive] is trying to change the culture of MutualCo, and I think he is going to have a hard job... the words that come out are, he wants us to work, but he wants us to be happy, and he wants us to have fun. They are light words, and everybody wants to be happy and have fun, but it is no good when you have an avalanche coming down on top of you with work loads all the time, and you work the weekend, and you work nights. If they want people to be happy in their work, they have got to stop putting people under so much pressure, and try to reward them in the best way that they can.”

At a less influential level, a junior Customer Services Helpdesk Operator reinforced this view, saying:

“When I first started with the society, I thought they were a very caring society and they were interested in your thoughts. We seemed to drift away from that about 2 years ago, although now we seem to be getting back into it. Before, it was very much, you either do these targets, or maybe the job that you are doing isn’t suitable for you anymore.”

These comments point to a dual standard in the society, with the rhetoric leaning towards a focus on employee relations, with activity directed towards improving skills and behaviors. However, the reality is that there is still an emphasis on progress, delivery, and the achievement of targets.

Communication is also recognized as a problem, especially in the branch network. One employee pointed out *“Once you start going outside of Head Office, then I think that perhaps the communication is not that strong”*. There is however evidence that the situation is improving, and the society intranet and E-mail is acknowledged to

be assisting in this area. Under the new senior management team, managers particularly talk about leaving the blame culture behind, moving away from the risk-averse culture of the mid 1990s, and a “work hard-play hard” environment.

One manager recognized that within the projects and I.T. area, attempts were being made to address some of these cultural issues, saying *“they are trying to move from a blame culture to one of, OK, you can make a mistake, and that is learnful... fault would be apportioned, but not in a negative way...”*. It is however apparent that the culture of *MutualCo* is fragmented. This is resulting in at least one department attempting to address its cultural shortcomings independently. There is also a widely held view that the move to a new, modern Head Office building will remedy cultural shortcomings within the organization, and assist with healing cultural fragmentation. This is however unlikely to happen without considerable complementary activity.

The positive and negative cultural attributes of *MutualCo* are displayed in Table 2. It is apparent from these attributes that much effort is being directed towards the development of employees, and the creation of a sympathetic environment to allow them to thrive and produce results for the society. However, it appears that at the Executive Director level, pressure is still applied to employees to produce improvements in effort, performance, and output, without sufficient attention to the social and behavioral issues that impinge upon such improvement activity.

MutualCo has an approach to the creation and planning of change that is moderately rigid, using tools and techniques drawn from methodologies such as PRINCE¹. There is however recognition at all levels within the managerial hierarchy that forms of improvisation are used, but the degree is disputed. The Head of I.T. describes this type of activity in terms of “work-arounds”

Table 2. Cultural and improvisational attributes of *MutualCo*

<i>MutualCo</i> –Cultural and Improvisational Attributes	
<u>Culture – Positive Attributes</u>	<u>Culture - Negative Attributes</u>
<ul style="list-style-type: none"> • New senior management emphasis on people aspects of the organization • Starting to manage tacit knowledge • New emphasis on retraining and stimulating staff • Progress on building skill sets and socio-behavioral norms • Decision-making devolved to lower levels 	<ul style="list-style-type: none"> • Geographical scattering of departments hampering communication • Major growth has led to focus on product development projects rather than Behavioral Projects • Head office relocation caused skills losses • Project management standards do not include people/cultural issues • Problems with communication, especially to branch network/staff
<u>Improvisation - Positive Attributes</u>	<u>Improvisation - Negative Attributes</u>
<ul style="list-style-type: none"> • Willingness to accept improvisation as an accelerator of change • Program/project managers admit to the regular use of improvisation 	<ul style="list-style-type: none"> • Conflict with internal audit department • Risk aversion negatively influences ability to improvise • Pressure to deliver may encourage reckless improvisation

¹‘PRINCE’ (Projects IN a Controlled Environment) and PRINCE2’ are project management methodologies widely used in the UK Public Sector.

and “*management of the issues*”. Middle managers articulate improvisational processes more strongly, suggesting that “*I improvise to get things done*”, “*I will get it done any way I can*”, and “*you couldn’t get a project ‘live’ without improvisation*”. The group I.T. manager reinforced this view, suggesting:

“If our chief exec wants something done, it is going to happen, and if he wants it done in a month, we will say, I doubt if we can do it. We will try for it, but one of the first things that goes is following all the rules and regulations, crossing the ‘t’s, dotting the ‘i’s, doing the documentation, that goes out.”

It appears that this is accepted within the organization, and there is clear evidence of high levels of improvisational activity within *MutualCo*. There is however a conflict between this attitude and the auditing of processes and project plans, and project managers complain of pressure to deliver projects quickly, and the lack of involvement of internal auditors in project planning and project progress meetings.

It appears that, notwithstanding a real sense of risk aversion within the organization embedded by 150 years of tradition, and a stable market for the single loan product offered for much of that time, areas of *MutualCo* are changing. As a result of the relative lack of development historically, the change that is now being achieved is happening in spite of the many problems that exist within the organization. Managers using improvisational practices appear to be the catalyst for such change. The positive and negative attributes relating to culture and improvisation within *MutualCo* are displayed in Table 2.

3.3 ExSociety

ExSociety was a mutually-owned mortgage lending organization, but is now part of a very large banking and finance Plc. As an organization, it appears to have a well developed and well communicated vision and values, widely adopted and championed across the organization. The organizational culture of *ExSociety* is well developed, coherent, and appears to be homogeneous and respected across the organization. Cultural norms are reinforced by the recognition and provision of single status facilities, and are built around two key messages; “*doing the right thing*”, and a continuing focus on customer excellence. The adoption of customer satisfaction as a key performance indicator within *ExSociety* has had a significant effect the evolving culture and on performance since its adoption in the late 1990s.

Initial impressions of *ExSociety* are favorable, with harmonization initiatives and the inclusion of employee issues having a positive effect on employee morale, commitment, and enthusiasm. There is strong evidence of adoption of the vision and values, which include employee development, at all levels within the organization. Indeed, in discussing the values of the organization, one

manager from the compliance and internal audit area who had been employed with *ExSociety* for about a year stated:

“Something I have not come across before in a company is, it is almost taken to heart by all the staff, and they recognize it, and they understand it, and they effectively live by it. They try to ‘do the right thing’”.

There have been various descriptions of the culture of *ExSociety*, including entrepreneurial, changing, innovative, people oriented, welcoming, open, and honest, friendly, and caring. These all appear to be positive attributes, and the evidence collected within and around *ExSociety* is generally supportive of and consistent with these labels, notwithstanding the amount of change that has allegedly taken place in the last decade. *ExSociety* also shares knowledge. If “best practice” is identified by a work team, forums are used to disseminate such information throughout the organization. Identification of and exploitation of these pockets of best practice that develop within *ExSociety* is seen as a key activity in the improvement of working practices.

Improvisation is accepted as a desirable skill across the whole of *ExSociety*. One member of the branch staff, when asked whether she had the opportunity to improvise, said “*ExSociety encourage taking ownership of everything. I think that is a good thing really, because it gives people confidence to think, ‘well, if there is a problem, how do I solve it’, rather than passing it on, and thinking, ‘well, this is not really to do with me’*”. This statement links neatly with the concept of “ownership” that is enshrined in the corporate values of *ExSociety*. Also, there is a very strong alignment between these values and the attitudes of employees, and improvisation is encouraged if it contributes to the delivery of excellence to the customer.

Discussions with the Strategic Change Manager also revealed strong support for improvisational activity, which he sees as the way in which most work is managed and achieved. A Compliance Manager did however point out that, given the volume of regulation within the sector, it was important to set boundaries, and that those boundaries would have to be “*embedded in training*”. This suggests a dualist view. Improvisation is used extensively to allow customer service employees to deliver customer excellence. The evidence does however indicate that some areas of change would benefit from a less rigorous approach, provided that a framework was defined for such improvisation. Within *ExSociety* managers have tended to identify and link improvisation with other initiatives such as continuous improvement, which provide the framework within which employees can be encouraged to improvise in order to enhance the delivery of job outputs. There is also an acceptance that the desires of the current parent organization may inhibit improvisation.

Table 3. Cultural and improvisational attributes of *ExSociety*

<i>ExSociety</i> – Cultural and Improvisational Attributes	
<u>Culture - Positive Attributes</u>	<u>Culture - Negative Attributes</u>
<ul style="list-style-type: none"> • Vision and values that are inclusive of employees • Open communication at all levels • Lack of status badges • Investment in employees • Relative lack of blame apportionment within the mechanisms of change 	<ul style="list-style-type: none"> • Some evidence of minor political activity linked with the parent • A perceived need by some managers that more emphasis is needed on knowledge and behaviors
<u>Improvisation – Positive Attributes</u>	<u>Improvisation - Negative Attributes</u>
<ul style="list-style-type: none"> • Improvisation recognized as contributing towards speed and innovation within the organization • Senior management support for improvisational activity • Audit and Compliance support, within agreed boundaries 	<ul style="list-style-type: none"> • Perceived rigidity of program management tools and techniques • Possible difficulties with the volume of regulation within the sector

Table 3 displays the key positive and negative attributes of *ExSociety* that impinge upon its culture, and upon the ability of the organization to improvise within its change initiatives. However, the apparently open and innovative culture of this organization should assist in setting a framework within which improvisation can be used to support such change, and there appears to be a keen desire to achieve this within *ExSociety*.

3.4 FinanceCo

This company is a relatively autonomous Finance subsidiary of a major U.K. bank. There is an admission from *FinanceCo* employees at all levels that throughout the 1990s the overarching focus was on developing and nurturing an aggressive sales culture. This is changing with integration with the parent, with the Governor of the parent bank stating “*we reinforce our ability to fulfill the professional and personal aspirations of our staff with a learning culture which permeates the whole of our business. Everyone benefits from the infectious enthusiasm this culture breeds...*”. However, it is acknowledged that much work is needed to meet the expectations of that statement, particularly within *FinanceCo*, where a politically oriented and sales driven culture has prevailed.

One manager confirmed that training in the behavioral aspects of managing is lacking, saying “*I don’t think anything specific is being done on those kinds of soft skills that are required*”, even though “*they are probably the key parts for the manager*”. This contrasts with the opinion of his manager, who suggested when talking about the behavioral aspects of managing teams that there will be “*tasks that actually make inference to the fact that you should be doing that [managing team behaviors]*”. The importance of addressing such skills has already been considered, and evidence reinforces and supports those findings.

The learning culture espoused by the senior executives of the parent bank is now beginning to diffuse throughout *FinanceCo*, and a new director, part of whose role is to

manage this culture change, has been appointed to the H.R. area of the newly integrated organization. Initiatives are currently being considered to access the considerable body of industry-specific knowledge that has built up within *FinanceCo* over many years, and to make it available on a wider basis. However, this work is in the very early stages of development, and may run counter to the sales and performance related historic culture.

The integration of *FinanceCo* into the parent bank has been rationalized on the basis of “*sharing best work practices, harmonizing*” although one manager suggested that “*the way it was sold was... that it would be working together, very collaborative, and I think initially that did not happen. It was very much, ‘we are the parent, you are the sibling’... and we are in charge.*” It is certainly apparent from responses across all parts of the organization that the parent bank has taken control, and that it was always their intention to limit what they perceived as cost and authority excesses within *FinanceCo*.

Rather naturally, the culture within the newly organized set of businesses is fragmented, and appears to be moving from a power culture towards a composite of role culture and task culture (Handy, 1993: 183-192). Culturally, *FinanceCo* is seen as being more sales and customer-oriented than the parent bank, and part of the rationale for the integration of *FinanceCo* into the parent bank is to exercise more control over these management shortcomings. The company is evolving from a traditional hierarchical structure, and has attempted to move from this model to a more reactive and flexible way of working. In order to achieve this, the organization is attempting to embrace less structured processes and mechanisms within those areas of the business where such practices are appropriate. In reality, this desire seems to be stimulating more improvisational working within the sales-oriented areas, and also in those areas where creativity and free thinking can contribute to perceived improvement.

It is however apparent that there are areas that are ac-

tively resisting a move towards more improvisational practices. Many areas of *FinanceCo* are becoming more rigid in their working practices as a result of the increasing influence of the parent bank, which is recognized as having a long history of cautious growth, and an aversion to risk taking. This caution is manifesting itself in the imposition of controls on areas of *FinanceCo*. Specifically, the I.T. area and the collections and litigation areas operate under service level agreements with their internal client departments, and these agreements specify processes and mechanisms that are mandatory. This restrains the ability and inclination of employees in those areas to improvise, notwithstanding the desire of the parent bank to learn from the allegedly more entrepreneurial and improvisational practices of *FinanceCo*.

There is however a feeling within *FinanceCo* that improvisational practices will assist in the future development of the organization. Within the lending areas of *FinanceCo*, the management interest is on improvisation in order to win business and to meet customer expectations. The management of risk is equally important, but the organizational focus within *FinanceCo* is on ensuring that loans are profitable, and that flexibility exists to satisfy customer demands. One manager within Group Training and Development suggested that some people felt more secure within a comfort zone that acted as a framework for their actions, whereas others were happy with improvisational techniques. He was also of the opinion that it was up to a manager to manage these preferences within his team. However, the data suggests that *FinanceCo* respondents have mixed views about improvisational work practices, and evidence indicates that there is doubt about their use. Specifically, the PRINCE-based routines imposed on the I. T. area of the organization impart a rigour that negates improvisation. It is however apparent that some managers still improvise to achieve tasks, and that there is greater use of improvisation within customer service areas.

The key positive and negative attributes relating to cultural aspects, and also to the use of improvisation

within *FinanceCo* are detailed in Table 4. It is however apparent that improvisation is used extensively across the various functional specializations within the organization, albeit that more rigidity and structure appears to exist within the I.T. development and support areas of the business.

3.5 NewCo

NewCo is a small lending organization, with a staff of around 150, delivering loan products via a call centre, an internet accessed applications systems, and a small broker network. The overarching philosophy within *NewCo* is the importance of the culture of the organization, and the emphasis that is placed upon staff development, training, and the social and behavioral aspects of working. The atmosphere within the offices is one of quiet professionalism, but with no overbearing sense of authority or discipline. Workers are smart, and senior employees are mainly dressed in suits, whilst more junior staff tend to be dressed more casually. *NewCo* has a well-defined and well-publicized vision and values, which is prominently displayed in most open areas and manager's offices.

Employees work in teams, and team members are paired up with a "buddy"-new staff with an experienced one in order to share expertise and learning. This system is used across most areas of the organization. The senior management is very visible, and informal contact with management is a normal occurrence. Directors and managers are addressed and referred to on a first name basis, and *NewCo* organizes regular social events, which senior managers attend.

HR development issues are well defined, with incentives, and mechanisms for assessing performance against agreed targets and core skills. This focus on staff development is a major influence in shaping the culture of *NewCo*, and it links closely with the aforementioned vision and values. In addition, the organization works hard to build an atmosphere of cooperation and caring amongst staff at all levels, using a combination of suggestion schemes, staff incentives, targets and rewards, and social

Table 4. Cultural and improvisational attributes of *FinanceCo*

<i>FinanceCo</i> – Cultural and Improvisational Attributes	
<u>Culture - Positive Attributes</u>	<u>Culture - Negative Attributes</u>
<ul style="list-style-type: none"> • Willingness to recognize pockets of internal expertise • Curbing of cost excesses • Appointment of new senior H.R. executive • Parent-imposed move to a learning culture 	<ul style="list-style-type: none"> • Limited training activity in behavioral aspects of managing • Evidence of weaknesses in socio-behavioral skills training
<u>Improvisation - Positive Attributes</u>	<u>Improvisation - Negative Attributes</u>
<ul style="list-style-type: none"> • Allows quick responses to external changes within projects • Allows flexibility in structuring lending products • Allows employees to meet customer expectations • Support from Group Internal Audit 	<ul style="list-style-type: none"> • Loss of control over processes by senior management • Not a substitute for poor planning and execution • Less accepted within the I.T. area

Table 5. Cultural and improvisational attributes of *NewCo*

<i>NewCo</i> – Cultural and Improvisational Attributes	
<u>Culture - Positive Attributes</u>	<u>Culture - Negative Attributes</u>
<ul style="list-style-type: none"> • Strong focus on quality of staff • Strong focus on development of staff • BQF model used to manage and measure socio-behavioral attributes • The scale of the organization allows a homogeneous culture to prevail • Good communication across the organization 	<ul style="list-style-type: none"> • Speed of growth may adversely affect current culture • Effect of new parent company not yet known, leading to uncertainty
<u>Improvisation - Positive Attributes</u>	<u>Improvisation - Negative Attributes</u>
<ul style="list-style-type: none"> • Recognition of improvisation as a positive force for change • Strong management support for improvisation • Mechanisms for formalizing successful improvised work processes • Most respondents admit to the use of improvisation 	<ul style="list-style-type: none"> • Anticipation that new parent company will require a more structured approach

events. Visual evidence of this abounds within *NewCo*. Staff at all levels appreciate this activity, and numerous interview respondents suggested that *NewCo* was not only the best organization they had worked for, but also that they were well paid, and that good work was encouraged, recognized, and rewarded. Introducers and suppliers have also benefited from this approach, with a *NewCo* “open day” and golf tournament called “One Great Day”, indicating the confidence of the organization.

Currently, *NewCo* is of a size where employees are aware of each other and their different skill sets, although as the organization grows, this will cease to be the case. Idea generators have differing levels and mixes of skills, and utilize varying ways of planning and executing these process changes. Communication is also seen as vital across *NewCo*, and there are effective formal and informal networks of communication in operation within the business. The Marketing Director of *NewCo* stated “*communications is something we work very, very hard at...*” and junior staff admit that the communication at *NewCo* is “*much better*” than previous employers. This focus on communication is assisting in the relatively effective management within the organization.

Within *NewCo* innovative thinking, rapid decision-making, and a speedy transition from idea generation to implementation are encouraged, and there was an attempt to reflect this in the way in which the organization developed in its first years of operation. As the organization has grown, a degree of formality has been introduced, but the senior management encourages innovation by staff, together with creativity, which is one of the important components of improvisation. This encouragement comes from a willingness to allow staff the freedom to try new ways of achieving work tasks, and informal networks within the organization act as conduits for the dissemination of new working practices. There is also an informal forum to discuss these issues, and a mechanism, albeit also informal, to feed emerging best practice back into new written procedures.

Improvisational work practices are encouraged by de-

liberately not surrounding employees with documented routines and formalized sets of expectations about how work is to be carried out. Team members and their managers are allowed to decide on work flows, responsibilities, and job design, and the sharing of pockets of best practice that emerge from this loose structure is encouraged. A set of limits to the framework within which work can be improvised is encapsulated into induction training, and is also passed on to new employees via the “buddying” system, which also provides an element of informal control.

There are many examples of improvisation within the data. Most notably, the whole area of user-led change to internal processes could be said to involve improvisation, as the majority of these initiatives appear to arise as a result of team members undertaking tasks or meeting customer expectations outside of formal organizational procedures. If these initiatives are proven to work, they are informally adopted, and disseminated across the organization using formal and informal employee networks. Ultimately, they are formally adopted into organizational routines. There is evidence that this improvisational activity is seen by *NewCo* as “*a way of staying ahead of the competition*” by compressing the time needed to react to market signals.

It can be seen that improvisation is encouraged across the organization. However, *NewCo* is growing quickly, and there is evidence that although controls have evolved, they are relaxed in favor of a more improvisational style of working when problems arise or when systems, processes, or routines need to be altered quickly. The positive and negative attributes relating to culture and to improvisational activity within *NewCo* are noted in Table 5. There is however considerable evidence to suggest that improvisation has been accepted as an important component of work carried out within *NewCo*, and that senior managers within the organization see strong links between improvisation, innovation, and the ability to meet customer requirements.

3.6 DivestCo

This organization is a long-established consumer finance and loan company, with an extensive branch network, which has gone through much change, and now needs to stabilize. One analyst from the strategy team talked of “*building a robust operational platform from which to pursue strategies*”, and the Senior Manager-Organizational Development, who is responsible for people and excellence issues, talked of the need for “*a period of stability and consolidation*”. However, currently *DivestCo*, having divested of its core business, is attempting to survey the opportunities that exist within its chosen sector, in order to make choices as to its future direction. Concurrently with this activity, work is progressing to develop a set of organizational values that will assist in enabling the company to achieve its aims within such a future strategy. This activity appears somewhat premature, as the required values are likely to differ according to the chosen direction. There is also evidence of acceptance within *DivestCo* that, regardless of the chosen strategy, important issues to address in the future will revolve around employee behaviors and the culture of the organization.

Notwithstanding this need to address cultural issues, *DivestCo* has already moved some way towards a more open and blame free culture. It is also apparent that different cultures prevail in different parts of the organization, with a macho, performance and volume-driven culture within the sales and branch departments, and evidence of a more intimate, almost paternal, yet politicized culture within most parts of the head office. The Senior Manager-Organizational Development also said “*we pride ourselves on having a can-do culture*”, although there are perceived frictions between such a statement and freedom from blame within the culture of the organization. It appears that there are a number of contradictions within the data as it relates to the culture of *DivestCo*. There is no doubt that the culture is not homogeneous, and this is demonstrated by the sales-based, performance-driven culture of the branch employees, contrast with cursory attempts to generate a more sympathetic and people-centered culture within parts of head office. Additionally, this divergence is complicated by an outwardly distant team of senior managers, who do not visibly interact with other employees.

There is a natural tension between the rigidity and robustness of some processes imposed upon *DivestCo* by its parent, and the desire to improvise in order to shorten delivery times for change initiatives. This has resulted in a desire to improvise, which is tempered by adopted standards and procedures. At various levels within the organization, the recognition of improvisation as a positive factor is accepted, with the Head of H.R. saying “*it probably doesn't happen enough...*”, and the Strategic Change Manager saying “*I do it all the time.*” It is how-

ever evident that some functional areas are less inclined to improvise, as evidenced by the Finance Manager, who stated “*improvisation sounds a bit scary to me, I have to say, from an accountants point of view*”.

Evidence supports the use of improvisation within *DivestCo*, but it is apparent that there are considerable differences of opinion as to use. The rhetoric surrounding improvisation relates to flexibility, serving the customer, and being innovative in the way lending can be structured. However, the reality seems to be closer to achieving the “can do” philosophy by doing whatever is necessary to deliver agreed outputs, regardless of the current need for them. Against this background, there is an appreciation within the company of the positive effects of improvisation. Notwithstanding this appreciation, with the current exercise to develop a new vision and values for the organization, together with the cultural changes that may be required in their adoption, there is a view that improvisational tendencies may need to be suppressed in the short term. It is however recognized that controlled improvisation is good for the organization, as it contributes to the generation of ideas, to the creation of new and innovative products, and to initiatives that will ultimately lead to the development and success of *DivestCo*.

There is strong evidence of the use of improvisational processes within *DivestCo*, and there is a perception that a key element surrounding improvisational processes is the control of risk. *DivestCo* addresses this link between risk and control formally, running workshops on balancing risk with control. One manager suggested that about fifty percent of his work was improvised, and a number of respondents saw improvisation as an extension of the continuous improvement process that is used to manage more informal change. Table 6 summarizes the positive and negative attributes relating to culture and to improvised routines within *DivestCo*.

4. Comparative Findings and Discussions

The desire to manage organizational culture is at the forefront of many managers' aspirations (Ogbonna & Harris, 1998: 273). Organizational culture is also an important component in the management of strategic change. Pant and Lachman (1998: 196) suggest that strategies contain implicit values that need to align with the values of the organization. *ExSociety* and *NewCo* are comfortable with the values implicit in their organizational culture, and attempt to ensure that change initiatives are coordinated with those values. Organizational culture does however embrace a wider set of components. Schein (1985) identified three levels of culture “surface manifestations”, which are behavior patterns that can be seen and heard; the aforementioned “values”, which are located below surface manifestations, and underpin them, and “basic assumptions”, those things that individuals

hold about the organization and how it functions. These assumptions can be difficult to access and identify. He suggests that culture consists of basic assumptions, and that surface manifestations and values are generated by and support such assumptions. Each is however valid in considering culture within organizations, and will be applied to the cases in turn.

The surface manifestations of the case study organizations differ greatly. *BigBank* has many divisions, and over 2500 branches and other locations, with limited cultural homogeneity across them, notwithstanding an attempt to create this. *MutualCo* attempts to offer a consistent message across branch premises, and is hopeful that its new Head Office building will allow it to create a homogeneous culture. *ExSociety* also has a positive Head Office image, and there is evidence that it also labors to make the culture inclusive of the branch network. *FinanceCo* and *DivestCo* project a fragmented set of surface manifestations, with significant structural problems, and *NewCo* has by far the most consistent set of surface manifestations, being a small organization located in a single building.

Each organization has a documented organizational vision, and a set of values that underpin the vision. In *ExSociety* and *NewCo*, employees have adopted the values of the organization very positively. *DivestCo* was developing a new vision and a new set of organizational values during the period that data for this study was being

collected. There was however an assumption within the organization that it would “*end up with more old DivestCo values*”. In *BigBank*, there is a contradiction between the vision and values, which articulate matters relating to customer relationships and products, and the over-riding organizational focus of the bank, which is “enhancing shareholder value”. In *MutualCo* and *FinanceCo*, there appears to be little connection between the attitudes of the employees and the organizational vision and values, other than a commitment to mutuality within *MutualCo*. This however appears to be driven by a desire to survive as an independent organization, rather than an articulation of organizational values to be adopted by employees.

The basic assumptions within Schein’s (1985) three level schema of organizational culture are more difficult to assess, notwithstanding the fact that Schein sees them as the essence of organizational culture. Two important aspects of such assumptions are an organization’s relationship with its environment, and the way in which the organization perceives itself. These aspects can be considered in the context of the persona that the organization projects to its stakeholders. This can be drawn from a synthesis of data contained in organizational mission statements and similar artefacts, albeit that this persona can fragment across different parts of an organization. A summary of the three levels of culture for each case study organization is displayed in Table 7.

Table 6. Cultural and improvisational attributes of *DivestCo*

<i>DivestCo</i> – Cultural and Improvisational Attributes	
<u>Culture - Positive Attributes</u>	<u>Culture - Negative Attributes</u>
<ul style="list-style-type: none"> • Activity to change cultural norms • Attempts to move away from a “blame” culture 	<ul style="list-style-type: none"> • Non-visibility of senior managers • Lack of effective communication within specific areas of the business
<u>Improvisation - Positive Attributes</u>	<u>Improvisation - Negative Attributes</u>
<ul style="list-style-type: none"> • Recognition that improvisation is effective in certain areas within the organization • An admission that many people use improvisation • A desire to improve improvisation skills • Formalization of some areas relating to improvisation. • Workshops to address the link between risk and control 	<ul style="list-style-type: none"> • Risk aversion within the organization • Tensions between improvisation and the need to address regulatory constraints

Table 7. Assessment of cultural levels

Assessment of Cultural Levels			
<i>Organization</i>	<i>Surface Manifestations</i>	<i>Values</i>	<i>Basic Assumptions</i>
<i>BigBank</i>	Fragmented	Contradictory	Enhancing Shareholder Value
<i>MutualCo</i>	Fragmented but Improving	Weak but Focused	Customer Value through Mutuality
<i>ExSociety</i>	Homogeneous and Positive	Strong	Customer Excellence
<i>FinanceCo</i>	Fragmented and Troubled	Weak	Proactive Product Innovation
<i>NewCo</i>	Homogeneous and Very Positive	Strong	Employee and Customer Focus
<i>DivestCo</i>	Fragmented and Uncertain	Developing	Growing Profitable Businesses

Many of the practices and mechanisms used to manage change in the case study organizations conform to the “informally formal” description in Bacon *et al.* (1996: 95). A degree of formality is provided by documented standards and procedures, and informality is evidenced by the relative freedom of managers to interpret the execution phase of change projects as they wish, including the use of improvisation. This seems to point again to the need for hybrid managers. Perhaps patterns of practice which exhibit an informal formality may provide a key to resolving the tension between the need for systematic practices and consistency on one hand, and intuitive organic practices that nurture innovation and flexibility on the other.

It would therefore assist organizations to ensure that employees have the skills to work in an informally formal way, and so contribute positively. In order to recruit and develop such skills, the case study organizations are attempting to develop or acquire employees with suitable skill-based and socio-behavioral profiles, and support those people to grow with the organization, and to change as it changes. *ExSociety* and *NewCo* are achieving the most success in this area, and *BigBank* and *DivestCo* are improving. *FinanceCo* is poor, but could improve as a circumstance of integration into its parent, and *MutualCo* appears to be saddled with a number of long-serving middle managers who are holding this process back.

Evidence within this study indicates that *ExSociety* and *NewCo*, who devote much time and effort to the recruitment and development of employees with flexible skill-sets and attitudes, achieve change better. Additionally, respondents opinions and secondary data from project plans and post implementation reviews strongly suggests that these two organizations have enthusiastically adopted team-based project structures to implement change, and devote considerable time and resources to the effective use of team-based processes and mechanisms. This is important, as ideally within change activities, disparate groups of third-party and permanent employees are melded into effectively functioning project teams, capable of delivering the required change. There is a strong trend in the data towards a link between socio-behavioral and cultural factors, and the satisfactory performance of change initiatives. The superior performance of *ExSociety* and *NewCo* in this area has already been discussed. The relatively poor performance of *FinanceCo*, and the poor but improving performance of *MutualCo*, tend to reinforce this linkage, given that *FinanceCo* is weak in socio-behavioral areas, and *MutualCo* is improving under its new, more people-centered developing culture.

There is however overwhelming evidence of extensive use and acceptance of improvisation in the management of change within the respondent organizations. Some of this activity is conscious, and some emerges because of

circumstances. Chelariu *et al.* (2002: 141) suggest that this organizational activity is “...a reflection of the pressures of an environment characterized by unprecedented fast change.” Stacey (1996) maintains that these environmental conditions are uncertainty, complexity (described in terms of interdependent environments), and dynamism (described in terms of short-lived opportunities and threats to survival). U.K. retail financial service is such an environment. Improvisation assists in dealing with this volume of change, and *BigBank* senior managers and project managers are keen exponents. *MutualCo* managers suggest that they could not operate without improvising. *ExSociety* have the most rigidly applied project standards and procedures, but at senior levels there is a belief that there is insufficient improvisation, and that this is stifling innovation. Interestingly, there is less improvisation by project managers within *ExSociety* than in any other case study organization, with the exception of *FinanceCo*, which produced the least evidence of improvisation. *NewCo* uses such techniques across the organization, and is also developing effective mechanisms to capture the emerging best practice that accrues. Evidence from *DivestCo* indicates that there is a willingness to improvise, but this is bounded by the memory of major failures in the mid 1990s, where an element of blame was attached to insufficiently rigid processes. It is apparent from the data that there are significant differences across the organizations, and across levels of seniority within those organizations. Some improvisation is also surreptitious, avoiding accountability and the scrutiny of senior managers.

Given such environments, it is understandable that the case study organizations may wish to use improvisational practices. There are however doubts as to its effectiveness, and none of the organizations are able to support empirically an assertion demonstrating that improvisation aids the management of strategic change. *BigBank* is attempting this through benchmarking initiatives. *MutualCo* has problems, particularly with the increase in risk that has to be managed. *ExSociety* senior managers cite an assumed link between improvisation and innovation, and are intentionally allowing employees the organizational and temporal space to learn from experimentation and improvisation. *FinanceCo* and *DivestCo* display little evidence of effective improvisational activity, and *NewCo* is at the opposite end of the continuum, encouraging improvisational activity at all levels within the organization. This is an area where the other organizations, with the exception of *ExSociety*, which has a set of forums that could assist in this area, tend to be weak. Table 8 maps the acceptance, application, and effectiveness of improvisation practices across the six organizations.

Findings from this study indicate that project and change managers embrace improvisation almost universally as a means of achieving change. Many managers,

Table 8. Use of improvisation within the case study organizations

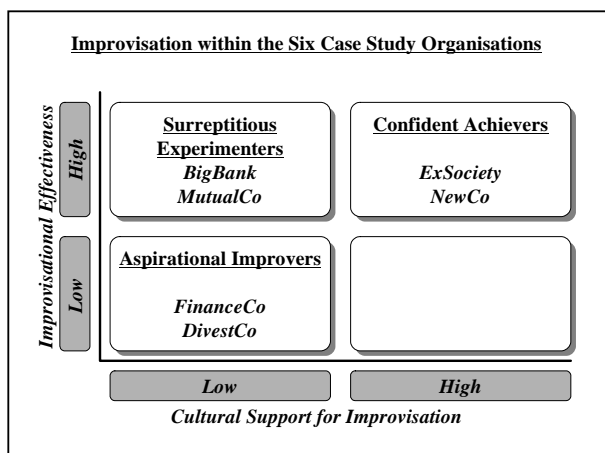
	Use of Improvisation		
	<i>Acceptance</i>	<i>Application</i>	<i>Effectiveness</i>
<i>BigBank</i>	Strong	Widespread	Poor
<i>MutualCo</i>	Strong	Widespread	Poor
<i>ExSociety</i>	Growing	Growing	Improving
<i>FinanceCo</i>	Weak	Limited	Poor
<i>NewCo</i>	Strong	Widespread	Improving/Good
<i>DivestCo</i>	Growing	Limited	Poor

across all six organizations, have strongly articulated opinions about the need to move away from agreed plans in order to execute that change. Indeed, managements may “...make a conscious decision to improvise as a means of creating more flexibility of behavior and more spontaneous decision making” (Chelariu *et al.*, 2002: 141). Crossan and Sorrenti (1997: 155) see this as “...intuition guiding action in a spontaneous way.” This is especially true within *BigBank* and *FinanceCo*, where improvisation is also seen as a means of circumventing intra and inter-organizational political resistance. It is also apparent that improvisation often takes place without senior management knowledge, especially within *BigBank* and *FinanceCo*.

Managers are however better able to support improvisation if it is bounded by some kind of limiting framework. This is supported by the literature; Brown & Eisenhardt's (1997: 16) “limited structure”, e Cunha *et al.*'s (1999: 318) “minimal structure”, and Weick's (1998: 545) “guidelines”. As the Financial Services sector is highly regulated, and tends to be risk averse (Brooks & Dawes, 1999: 197), such a framework is usually based around the management of risk. It is also recognized that improvisation is more effective if mechanisms exist to

share successful improvisational activity (Moorman & Miner, 1998b: 713; Chelariu *et al.*, 2002: 142), and to communicate lessons learned from it to relevant parts of the organization that can benefit from such activity (Moorman & Miner, 1998b: 713). This requires the development of organizational memory (Moorman & Miner, 1998b: 713-714). Respondents within all the case study organizations voiced concerns about the ability of their organization to capture good improvisational practice and encapsulate it within such a memory for future use. Both *ExSociety* and *NewCo* have mechanisms to assist with this. In *ExSociety* improvisational activity is an enabler, and the management identifies and implements new improvisational processes as part of the streamlining of work processes. There is however evidence that the I.T.-based change initiatives do not include as much improvisational activity as initiatives to change operational processes. This is because more rigorous and defined procedures surround the implementation and testing of new IT-based systems, which provide the core account processing for most financial services organizations. *NewCo*, a significantly smaller and younger organization, uses improvisation at all levels, and has informal forums to identify and disseminate improvisational practices that have the potential to become “best practice” within the organization. There appears to be little evidence of such initiatives in the other four organizations, notwithstanding the rhetoric surrounding this activity in *BigBank* and *MutualCo*.

Within this study, the six case study organizations achieved different degrees of competence with their improvisational effectiveness, and their cultural support for improvisational activity. Figure 1 plots the case study organizations on a matrix that uses these two factors as the axes, using the evidence within the data as a basis for positioning. This distributes the organizations into three groups of two. The organizations where the data demonstrates a lower level of support for improvisation within the organizational culture and a low improvisational effectiveness quotient have been labeled “aspirational improvers”, for their desire to improvise, albeit that this desire is moderated by lack of tangible support. *Fi-*

**Figure 1. Matrix of the six case study organizations**

nanceCo and *DivestCo* fall into this category.

Those organizations where the data indicates a lower level of support within the organizational culture for improvisational activity and a high improvisational effectiveness quotient appear to have token management support for experimentation and improvisation, but are not fully supportive of employees who fail to improvise effectively. Blame is often attached to failure, and this makes employees cautious in exposing their improvisation to management scrutiny. An element of surreptitiousness therefore also appears within this group of organizations, which comprises *BigBank* and *MutualCo*. These companies have been labeled “surreptitious experimenters”. It is inevitable that an element of surreptitiousness prevails here, as improvisation requires participants to step away from the shared responsibility embedded in the “plan, then execute” paradigm, and to embrace individual unplanned activity, where failure to improvise effectively is very visible.

The third pair of organizations produced data indicating that they possess a high level of support for improvisation within their organizational culture and a high improvisational effectiveness quotient. Employees are supported and effective in their improvisational activity, and the culture allows learning from mistakes, and the capture of effective improvisational activity, both formally and informally. This group, which comprises *ExSociety* and *NewCo*, has been labeled “confident achievers”. They manage change effectively using innovative processes, many of which are generated by the use of improvisation, and they have mechanisms to capture and disseminate such successful innovation to other areas of the organization.

5. Conclusions

There is very little literature pertaining to the use of improvisation within the implementation of project-managed change. There is compelling evidence that improvisation is used in this area, and project managers in all six organizations provide overwhelming support for improvisational activities as a means of executing change. Given the significant shift away from “command and control” based hierarchies, and towards trusted and empowered employees seeking opportunities to maximize profit, it is inevitable that partial reliance on intuition and creativity will encourage improvised solutions and interventions to assist with organizational transformation and change.

There is however a negative implication to improvising, particularly when applied to recent growth and failure within the financial services sector. Albrecht (1979) sees five significant areas of change in lifestyle notably: a move from rural living to urban living; a move from a stationary to a mobile society; a move from self-suffi-

ciency to consumption; a move from isolation to inter-connectivity; and a move from physically active to sedentary. This research centered around factors which cause stress in managers, informs the way people live, work, and manage their lives, and the way in which producers and service industries have to evolve in order to meet the needs of consumers. Specifically, sectors have had to evolve and change to service altering consumption patterns.

The social context of the U.K. retail financial services sector has been affected by Albrecht’s factors, and through the 1990s and since the turn of the millennium this change has been accelerating, particularly in technological terms. It is however the move towards consumption driven by increased affluence in the advanced economies which has driven much social and cultural change and seen the emergence of consumerism, concern for equal opportunities, environmentalism, and other social movements which impinge upon the sector. This social change has also been instrumental in the development of new cultural norms.

The culture of an industry or sector is more difficult to define, but Gordon (1991: 398) suggests that “...it is possible for differences in values, or even assumptions, to exist within a company, as long as they do not undermine the basic assumptions on which the industry depends.” This suggests that a sector or industry culture does exist, and it is considered that industry and sector are effectively interchangeable descriptors in this context. The basic assumptions upon which an industry or sector depends are the embedded core values and implicit rules to which component organizations within a given sector adhere. An example of a basic assumption that relates to U.K. retail financial services is fiscal responsibility. In the life insurance sub-sector this could be represented by an ability to pay claims when they fall due, and in the banking sub-sector it could be represented by an expectation that deposits could be repaid on demand. Whipp *et al.* (1989: 565) refer to these basic assumptions as logics of action. This view of logics of action at the sector level can be seen as an extension of the work of Bacharach and Lawler (1980), which has been developed and refined in Bacharach *et al.* (1996: 478). They argue that:

“...in essence, a logic of action may be seen as the implicit relationship between means and ends underlying the specific actions, policies and activities of organizational members. While the logic of action is for the most part taken for granted, it becomes manifest when parties try to explain themselves or justify to others the selection of specific means, ends, and the linkage between the two.”

There is some common ground here with macrocultures (Abrahamson & Fombrun, 1994), which link the cultural ambitions and expectations of managers within a sector, the existence of which has already been recog-

nized. From the discussion of culture, there is a logical step to the consideration of culture change.

There are of course reasons why a change in culture within a sector may be desirable. Competitive issues have been at the forefront of change within many sectors, including U.K. retail financial services. Ezzamel *et al.* (1994: 22) document the move by U.K. financial services organizations from:

"...the established, highly bureaucratic administrative control that has long characterised the industry, engendering new forms of control based upon the twin ideals of empowerment and heightened accountability. Managers and staff alike are expected to become 'multiskilled' to facilitate organizational flexibility in the pursuit of 'quality'."

They suggest that much of this activity, which for public consumption is being carried out to improve customer service, is actually being carried out in the name of efficiency and cost saving to drive down the cost base.

However, in the Banking and Finance sector, these changes have resulted in evidence that the developing culture of personal gain over financial prudence has caused the sector to over-reach itself, ignoring the need to manage risk, and driving a desire for growth and short-term profit. This has been to the detriment of the aforementioned fiscal responsibility and financial prudence, and it is reasonable to suggest that the relaxation of controls and the reliance on, often misplaced, trust in traders and managers within the sector is largely to blame. During 2008 and 2009 there has been a degree of attrition, with stronger members of the sector absorbing weaker ones, and in U.K., this has been driven by vigorous government intervention. Unfortunately, many of the principles that underpin improvisational activity are likely to have contributed to the current situation.

REFERENCES

- [1] E. Abrahamson and C. J. Fombrun, "Macrocultures: Determinants and consequences," *Academy of Management Review*, Vol. 19, No. 4, pp. 728–755, 1994.
- [2] M. Agar, "The right brain strikes back in fielding," N.G. & Lee R.M. (eds.), *Using Computers in Qualitative Research* London: Sage Publications, pp. 181–194, 1991.
- [3] A. E. Akgun and G. S. Lynn, "New product development team improvisation and speed-to-market: An extended model," *European Journal of Innovation Management*, Vol. 5, No 3, pp. 117–129, 2002.
- [4] K. Albrecht, "Stress and the Manager Englewood Cliffs," NJ: Prentice-Hall, 1979.
- [5] S. B. Bacharach, P. Bamberger, and W. J. Sonnenstuhl, "The organizational transformation process: The micropolitics of dissonance reduction and the alignment of logics of action," *Administrative Science Quarterly*, Vol. 41, No. 3, pp. 477–506, 1996.
- [6] S. B. Bacharach and E. J. Lawler, "Power and politics in organizations," San Francisco, CA: Jossey-Bass, 1980.
- [7] N. Bacon, P. Ackers, J. Storey, and D. Coates, "It's a small world: Managing human resources in small businesses," *International Journal of Human Resource Management*, Vol. 7, No. 1, pp. 82–101, 1996.
- [8] T. Baker, A. S. Miner, and D. T. Eesley, "Improvising firms: Bricolage, account giving and improvisational competencies in the founding process," *Research Policy*, Vol. 32, pp. 255–276, 2003.
- [9] F. J. Barrett, "Creativity and improvisation in Jazz and organizations: Implications for organizational learning," *Organization Science*, Vol. 9, Vol. 5, pp. 605–622, 1998.
- [10] F. J. Barrett, "Managing and improvising: Lessons from Jazz," *Career Development International*, Vol. 3, No. 7, pp. 283–286, 1998.
- [11] Brooks and J. Dawes, "Merger as a trigger for cultural change in the retail financial services sector," *Service Industries Journal*, Vol. 19, Vol. 1, pp. 194–206, 1999.
- [12] S. L. Brown and K. M. Eisenhardt, "The art of continuous change: Linking complexity theory and time-paced evolution in relentlessly shifting organizations," *Administrative Science Quarterly*, Vol. 42, No. 1, pp. 1–34, 1997.
- [13] A. Bryman, "Quantity and quality in social research," London: Unwin Hyman, 1988.
- [14] C. Chelariu, W. J. Johnston, and L. Young, "Learning to Improvise, Improvising to Learn: A process of responding to complex environments," *Journal of Business Research*, Vol. 55, No. 1, pp. 141–147, 2002.
- [15] T. Cooke-Davis, "The 'real' success factors on projects," *International Journal of Project Management*, Vol. 20, pp. 185–190, 2002.
- [16] M. Crossan, "Improvise to innovate," *Ivey Business Quarterly*, pp. 36–42, Autumn 1997.
- [17] M. Crossan, "Improvisation in action," *Organization Science*, Vol. 9, No. 5, pp. 593–599, 1998.
- [18] M. Crossan and M. Sorrenti, "Making sense of improvisation in J. P. Walsh and A. S. Huff (eds.)," *Advances in Strategic Management*, Vol. 14, pp. 155–180, Greenwich, CT: JAI Press, 1997.
- [19] M. P. Cunha e, J. V. Cunha da, and K. Kamoche, "Organizational improvisation: What, when, how and why," *International Journal of Management Reviews* 1, pp. 299–341, 1999.
- [20] T. E. Deal and A. A. Kennedy, "Corporate cultures: The rites and rituals of corporate life," London: Penguin Books, 1982.
- [21] D. R. Denison, "Corporate culture and organizational effectiveness," New York: John Wiley, 1990.
- [22] D. R. Denison, "What is the difference between organizational culture and organizational climate? – A native's view on a decade of paradigm wars," *Academy of Management Review*, Vol. 21, No. 3, pp. 619–654, 1996.
- [23] P. R. Dickson, "Marketing management (2nd Edition)," New York: The Dryden Press, 1997.

- [24] M. Eisenhardt, "Building theories from case study research," *Academy of Management Review*, Vol. 14, pp. 532–550, 1989.
- [25] M. Eisenhardt, "Strategic decisions and all that jazz," *Business Strategy Review*, Vol. 8, No. 3, pp. 1–3, 1997.
- [26] Ezzamel, H. Willmott, and S. Lilley, "Changing management practices in financial services," *Journal of General Management*, Vol. 20, No. 1, pp. 22–28, 1994.
- [27] J. R. Faulconbridge and D. Muzio, "Organizational professionalism in globalizing law firms," *Work, Employment and Society*, Vol. 22, No. 1, pp. 7–25, 2008.
- [28] G. Fielding and R. M. Lee, "Computer Analysis and Qualitative Research," London: Sage Publications, 1998.
- [29] E. Freidson, "Profession of medicine: A study of the sociology of applied knowledge," New York, NY: Harper & Row, 1970.
- [30] A. Friedman, "Responsible autonomy versus direct control over the labor process," *Capital and Class* 1, 1977.
- [31] P. Frost and R. Stablein, [Eds.] "Doing exemplary research," London: Sage Publications Ltd, 1992.
- [32] E. Gardener, B. Howcroft, and J. Williams, "The new retail banking revolution," *Service Industries Journal*, Vol. 19, No. 2, pp. 83–100, 1999.
- [33] C. J. G. Gersick, "Time and transition in work teams: Toward a new model of group development," *Academy of Management Journal*, Vol. 31, pp. 9–41, 1988.
- [34] G. G. Gordon, "Industry determinants of organizational culture," *Academy of Management Review*, Vol. 16, No. 2, pp. 396–415, 1991.
- [35] C. Handy, "Understanding organizations: managing differentiation and integration," New York, Oxford University Press, 1993.
- [36] J. Hatch, "Jazz as a metaphor for organizing in the 21st century," *Organization Science* Vol. 9, No. 5, pp. 556–557, pp. 565–568, 1998.
- [37] M. J. Hatch, "Exploring the empty spaces of organizing: How improvisational jazz helps redescribe organizational structure," *Organization Studies*, Vol. 20, No. 1, pp. 75–100, 1999.
- [38] C. Heckscher, "Defining the post-bureaucratic type in C. Heckscher and A. Donnellon, (eds)," *The Post-bureaucratic Organization: New Perspectives on Organizational Change* Sage, Thousand Oaks, CA, pp. 14–62, 1994.
- [39] K. M. Hmieleski and A. C. Corbett, "Improvisation as a framework for investigating entrepreneurial action," Unpublished Paper presented to the American Academy of Management Conference, Seattle, 1st–6th August, 2003.
- [40] A. M. Huberman and M. B. Miles, "Data management and analysis methods in K. M. Denzin & Y. S. Lincoln (eds)," *Collecting and Interpreting Qualitative Materials* London: Sage Publications, pp. 179–201, 1998.
- [41] M. Jelinek, L. Smircich, and P. Hirsch, "Introduction: A code of many colours," *Administrative Science Quarterly*, Vol. 28, pp. 331–338, 1983.
- [42] R. M. Kanter, "Strategy as improvisational theater," *MIT Sloan Management Review*, pp. 76–81, Winter 2002.
- [43] K. U. Koskinen, P. Pihlanto, and H. Vanharanta, "Tacit knowledge acquisition in a project work context," *International Journal of Project Management*, Vol. 21, pp. 281–290, 2003.
- [44] J. M. Lehner, "Bricolage during implementation of strategies: Effects on flexibility," Unpublished Paper presented to the American Academy of Management Conference, Toronto, 4th–9th August, 2000.
- [45] S. A. Leybourne, "Project management and the implementation of strategic change within the UK financial services sector," Unpublished Doctoral Dissertation: Cardiff Business School, 2002.
- [46] S. A. Leybourne, "Improvising within management: Oxy-moron, paradox, or legitimate way of achieving?" Unpublished Paper presented at the Academy of Management Conference, Honolulu, HI-5th-10th August 2005.
- [47] S. A. Leybourne, "Managing improvisation within change management: Lessons from UK financial services," *Service Industries Journal*, Vol. 26, No. 1, pp. 1–23, 2006.
- [48] S. Linstead, L. Fulop, and S. Lilley, "Management and organization: A critical text," Basingstoke, Hants: Palgrave Macmillan, 2004.
- [49] Lonkila, "Grounded theory as an emerging paradigm for computer-assisted qualitative data analysis in U. Kelle (ed.)," *Computer-aided Qualitative Data Analysis: Theory, Methods and Practice*, London: Sage Publications pp. 41–51, 1991.
- [50] A. McLean and J. Marshall, "Intervening in cultures," University of Bath Working Paper, 1993.
- [51] S. Mendonça, M. P. Cunha, J. Kaivo-oja, and F. Ruff, *Wild Cards, Weak Signals and Organizational Improvisation Futures*, Vol. 36, pp. 201–218, 2004.
- [52] A. S. Miner, P. Bassoff, and C. Moorman, "Organizational improvisation and learning: a field study," *Administrative Science Quarterly*, Vol. 46, pp. 304–337, 2001.
- [53] C. Moorman and A. S. Miner, "The convergence of planning and execution: improvisation in new product development," *Journal of Marketing*, Vol. 62, No. 3, pp. 1–20, 1998.
- [54] C. Moorman and A. S. Miner, "Organizational improvisation and organizational memory," *Academy of Management Review*, Vol. 23, No. 4, pp. 698–723, 1998.
- [55] L. J. Mullins, "Management and organizational behavior (5th ed.)," Harlow, England: Financial Times/Prentice Hall, 1999.
- [56] E. Ogbonna and L. C. Harris, "Managing organizational culture: Compliance or genuine change?" *British Journal of Management*, Vol. 9, No. 4, pp. 273–288, 1998.
- [57] W. J. Orlikowski and J. D. Hofman, "An improvisational model for change management: The case of groupware technologies," *Sloan Management Review*, Vol. 38, No. 2, pp. 11–21, 1997.

- [58] N. Pant and R. Lachman, "Value incongruity and strategic choice," *Journal of Management Studies*, Vol. 35, No. 2, pp. 195–212, 1998.
- [59] A. M. Pettigrew, "Longitudinal field research on change: Theory and practice," Paper presented at the National Science Foundation Conference on Longitudinal Research Methods in Organizations: Austin, TX, 1988.
- [60] J. B. Quinn, "Strategies for change: Logical incrementalism," Homewood, IL, Richard D. Irwin, 1980.
- [61] D. Rollinson and A. Broadfield, "Organizational behavior and analysis: An integrated approach (2nd Ed.)," Harlow: Pearson Education, 2002.
- [62] C. K. Russell and D. M. Gregory, "Issues for consideration when choosing a qualitative data management system," *Journal of Advanced Nursing*, Vol. 18, pp. 1806–1816, 1993.
- [63] G. Ryle, "On thinking," Oxford: Basil Blackwell, 1979.
- [64] E. H. Schein, "Organizational culture and leadership," San Francisco, CA: Jossey-Bass, 1985.
- [65] J. Seidel, "Method and madness in the application of computer technology to qualitative data analysis in fielding N. G. & Lee R. M. (eds.)," *Using Computers in Qualitative Research* London: Sage Publications pp. 107–116, 1991.
- [66] L. Smircich, "Concepts of culture and organizational analysis," *Administrative Science Quarterly*, Vol. 28, No. 3, pp. 339–358, 1983.
- [67] D. Stacey, "Complexity, creativity and management," San Francisco, CA: Berrett-Koehler, 1996.
- [68] E. Stake, "Case studies," In K. M. Denzin & Y. S. Lincoln [eds.], *Handbook of qualitative research*, pp. 236–247, London: Sage Publications, 1994.
- [69] A. Sturdy, D. Knights, and H. Willmott, "Introduction: Skill and consent in the labor process in sturdy A., Knights D. & Willmott H. (eds.)," *Skill and Consent: Contemporary Studies in the Labour Process*, Routledge: London, 1992.
- [70] J. Trethowan and G. Scullion, "Strategic responses to change in retail banking in the U.K. and the Irish Republic," *International Journal of Bank Marketing*, Vol. 15, No. 2, pp. 60–68, 1997.
- [71] M. L. Tushman, "Winning through innovation," *Strategy & Leadership*, Vol. 25, No. 4, pp. 14–19, 1997.
- [72] Velez-Castrillon, D. M. Vera, and A. Kachra, "An improvisational model of strategic decision making," Unpublished paper presented at The Academy of Management Conference, Anaheim, CA-8th-13th August 2008.
- [73] K. E. Weick, "The social psychology of organizing (2nd Ed.)," Reading, Mass: Addison-Wesley Publishing Company, 1979.
- [74] K. E. Weick, "Improvisation as a mindset for organizational analysis," *Organization Science*, Vol. 9, No. 5, pp. 543–555, 1998.
- [75] E. A. Weitzman and M. B. Miles, "Computer programs for qualitative data analysis," Thousand Oaks, CA: Sage Publications, 1995.
- [76] R. Whipp, R. Rosenfeld, and A. Pettigrew, "Culture and competitiveness: Evidence from two mature UK industries," *Journal of Management Studies*, Vol. 26, No. 6, pp. 561–585, 1989.
- [77] Woolgar, "Configuring the user: The case of usability trials in Law J. (ed.)," *A Sociology of Monsters: Essays on Power, Technology and Domination*, London: Routledge pp. 57–99, 1991.
- [78] D. Yanow, "Learning in and from improvising: lessons from theater for organizational learning," *Reflections: Journal of the Society for Organizational Learning*, Vol. 2, No. 4, pp. 58–65, 2001.
- [79] R. K. Yin, "Case study research: Design and methods," London: Sage Publications Ltd, 1984.
- [80] R. K. Yin, "Case study research: Design and methods [2nd Ed.]," London: Sage Publications Ltd, 1994.

Service System Decoupling for Mass Customization: A Case Study in Catering Services

Jue CHEN¹, Yunhong HAO²

¹The tourism school, Zhejiang Gongshang University, Hangzhou, China; ²School of Business Administration, Zhejiang Gongshang University, Hangzhou, China.
Email: chenjue8@yahoo.com

Received April 17, 2009; revised June 19, 2009; accepted July 28, 2009.

ABSTRACT

The paper examines the key issues on system decoupling in service operations of mass customization by conducting a case study in catering services. It firstly justifies the effectiveness of applying concept of mass customization into service system decoupling to deal with the operation dilemma and then reveals the nature of decoupling decisions for mass customization purpose after discussions on the importance of modularization and the role of technologies including IT in the decoupling process. Based on these analyses, a Judgment-Matrix-based model on how to make the decoupling decisions in balancing the multiple operation objectives is then proposed and further research directions are finally suggested.

Keywords: Mass Customization, System Decoupling, Customer Contact, Modularization

1. Introduction

System decoupling is an effective approach to cope with the influences caused by customer contact by dividing the service system into two components: back-stage and front-stage. However the relevant researches often take a dichotomous perspective, assuming that the objectives of service operations focus on either efficiency (costs and related issues) or services (customization and responsiveness). Mass customization (MC) has been regarded as an innovative way of doing business by putting together these seemingly contradictory operation objectives and hence one of the most advanced operation model in 21st century. The paper intends to discuss the system decoupling issue in the context of service mass customization which requires a comprehensive consideration of the various operation objectives rather than dichotomous thinking. This in addition will also lead to an extended application of manufactory-originated MC theory into the service sector.

The remainder of this paper is structured as follows. To begin with, we review current literatures addressing both service system decoupling and MC, which leads to the suggestion of research questions. Then the propositions of basic consideration on system decoupling for mass customization and a relevant model for decoupling decision-making are presented. Next, we conduct a case study in a restaurant, based on which implications are

drawn for justification of the previous propositions. Finally, the paper ends with conclusions and suggestions for further research.

2. Literature Review and Research Questions

2.1 Customer Contact and Service System Design

One of the most distinct features for services (services in this paper refer to general “business services” in “real world” rather than the services in virtual world such as web or software services) is customer contact, which means that customer should be physically present in the service delivery system [1]. The customer contact provides a source of complexity that is not generally found in manufacturing operations [2]. Contacts with the customers and their involvement in the service delivery process affect the effectiveness and efficiency of services operations. Customer contact is a double-edge sword to the service system, bringing up both risks and opportunities. It on one hand introduces uncertainties and variation in the service delivery system and makes demands on the design of facilities, staff and technology in the production system [3]. On the other hand, customer contact provides valuable opportunities for responding to a customer’s needs and cross-selling other products [4] and involving customer self-services which would help increasing efficiency for the organization [5].

2.2 Service System Decoupling

2.2.1 Decoupling for Efficiency

To deal with the influences caused by customer contact some researchers introduced decoupling approach breaking the service system into its component back and front-office stages. According to the customer contact approach proposed by Chase and Tansik [1,6,7], customer contact activities should be decoupled from non-contact activities to do justice to their different design requirements and maximize the efficiency of the service delivery system. Consequently, service organizations consist of a front stage and a back stage. In the back office customer contact is passive or nonexistent, service processes may be designed with manufacturing-like principles in mind [6]. The process may take advantage of standardization and automation to enhance the efficiency and effectiveness of operations. While in the front stage customer contact is high and active, it brings about unavoidable inefficiencies, and the human relation skills become a dominant factor in designing the process [6]. Study [8] adds that moving some of the back-stage functions from developed areas (or even countries) to underdeveloped regions may increase greatly efficiency because of the reduction in labor costs and taxation.

2.2.2 Decoupling for Speed, Quality and Sales Opportunities

However, back-stage tasks need not only be viewed as tasks where efficiency is the sole purpose. Research [9] and [10] propose using the back office to provide superior results in non-cost strategies by linking the back office to the same marketing directives that the front office is organized around. This represents a change in focus from managing for cost containment and transaction management to managing for revenue enhancement and customer loyalty by aligning the back office for speed or high quality (eg. flexibility, customization and similar features of services).

Chase and Hayes realize the limitation of previous study that they overlooked the fact that there are positive benefits to both the customer and the organization by having the customer contact [4]. Afterwards Chase and

Aquilano propose a matrix for discovering the tradeoff between sales opportunities and efficiency in service system design. More back stage decoupled from the front stage means more system efficiency but less sales opportunities [11].

Metters and Vargas hold that in several situations coupling front office and back office jobs can be a viable strategy and both coupled and decoupled system designs can support efficiency and high-quality service strategies [12].

2.2.3 Summary on the Decoupling Approach

Based on the literature reviewed above we could summarize the tradeoffs (or dilemma) involved between contact benefits of the front-stage operations and the potential efficiency of the limited or no contact back-stage operations. In another word, the low decoupling (focusing on front-stage) or high decoupling (focusing on back-stage) in service system brings different advantages for service operations as summarized in Table 1. The decoupling approach should reduce costs and increase productivity, but these advantages may be off-set by losses in the key competitive strategies of sales opportunities and service quality (delivery speed, flexibility and customization). The method of system decoupling is in essence trying to alleviate this operational dilemma but can not provide a complete solution.

2.3 Mass Customization and Enablers

MC refers to the ability of production of customized goods and services on a mass basis and has recently aroused more and more academic concern. The visionary concept of MC was first coined by Stan Davis [13] in Future Perfect. And the concept of MC was first fully expound by Pine [14] who implied a view of MC as in some sense a historically inevitable successor to mass production, the principal in which to complete in the future. The MC is essentially an oxymoron since it puts together seemingly contradictory notions --- the production and the distribution of customized goods and services on a mass basis. In another word, the requirements of

Table 1. Advantages form focusing on front-stage and back-stage

Advantages form focusing on front-stage (lowly decoupled system)	Advantages from focusing on back-stage (Highly decoupled system)
<ol style="list-style-type: none"> 1. Improving service delivery (including flexibility, customization, speed and responsiveness) 2. Increasing sale opportunities (cross selling) 3. Improving efficiency by involving customer participations 	<ol style="list-style-type: none"> 1. Improving efficiency by adopting industrial principles 2. Reducing costs on labor and taxation by favorably locating the back-stage 3. Developing expertise for the staff and thus improving service quality and speed

MC reside with three aspects: quick responsiveness, customization and economy of scale (mass efficiency) [15]. To meet these operational objectives simultaneously, three major technical challenges in mass customization system are identified, namely maximizing reusability, fast production responding to customers' needs and integrated product life-cycle [15].

To cope with the operation dilemma and overcome the challenges in achieving MC, the following enablers have been mentioned in literature: modular design in product and process [16]; postponement and supply chain management [17]; efficient information system [18]. Among them the modularization is regarded as the most fundamental approach to reach the goal of MC. Both [14] and [16] held that modular design in product and process could facilitate MC since it helped to achieve both scale and scope economy required by MC.

2.4 Research Questions

As discussed above the system decoupling has become a vital decision in service system design and an important approach (strategy) to cope with both the benefits and risks caused by customer contact in service delivery. However researches in this domain often take a dichotomous perspective, assuming that service organizations focus on either efficiency or cross-selling, either low costs or high-quality services [19]. This could not meet the requirements of MC in which companies perform well at multiple operation criteria (eg, cost efficiency and good services) simultaneously and provide complete solutions to operation dilemma. Then the questions could be raised if we apply the concept of MC into the service system decoupling:

- Is it effective to apply MC concept into service system decoupling?
- What is the nature of system decoupling if the concept of MC is integrated in?
- How to strike a balance in between the multiple operation objectives required by MC when making a decoupling decision for service system?

3. System Decoupling for Service Mass Customization: Conceptual Propositions

3.1 The Nature of System Decoupling for Mass Customization

System decoupling serves as an important decision in service system design to cope with both the benefits and uncertainties caused by customer contact and an approach to alleviate to some extent the contradiction between efficiency and service provision (including customization). However it could not provide the complete solution to this operation dilemma. The concept of MC provides a key to solve the problem. Most of the literatures as mentioned previously agreed that modularization would be

the fundamental solution to MC. Therefore system decoupling for MC purpose should involve modularization in service system and make decoupling decisions based on it. Modularization is the use of modules to facilitate assembly and configuration of finished products [20]. It can be used to develop complex products or services using similar component which may lead to high efficiency. Therefore in the condition of modularization the service system decoupling issue is in nature to determine the allocation of service modules to back or front stage rather than a geographical or physical separation of the system.

3.2 Decoupling Decisions in Service Mass Customization

As mentioned above the service functions to be provided in service MC system should be rationalized into service modules. Thus the decoupling decisions in service MC system can be converted into the front or back stage decisions for each service module. Here we propose a two-step method for these decisions.

Firstly the decoupling decisions should be made for each service module according to the extent of customer contact since the latter is the decisive factor for the decisions as discussed previously. The decoupling decisions for high contact service modules (modules can not be provided without customer contact) are quite simple since they "have to" be arranged in front-stage while the middle and low contact service modules (modules can be provided without customer contact) can be located either in front or back office. The decoupling decisions therefore in essence should only be made for these modules with middle and low customer contact.

Secondly there are also exceptions in high contact modules whose degree of customer contact could be changed by introducing technologies since the nature of customer contact is changing nowadays [21] due to the developments in technology including IT. Those modules together with the middle and low contact modules are here named as "free modules", meaning that they are "non-high-contact" and hence can be allocated to either front office or back office. The "free modules" are left for a further comprehensive decoupling decision-making as discussed in the following section.

3.3 The Model of Comprehensive Decoupling Decision for "Free Modules"

3.3.1 The Model and Explanation

Here we propose a basic model demonstrating the comprehensive consideration for decoupling decisions of the "free modules" and method to conduct these decisions. The model is based on the understanding that decoupling decisions need comparative evaluation of the benefits obtained from allocating service module to front-stage and back-stage respectively, which are summarized in Table 1. The model is shown as Formula 1 below.

$$s = \frac{\sum_{i=1}^3 F_i \gamma_{F_i}}{\sum_{j=1}^3 B_j \gamma_{B_j}} \quad (1)$$

$$\text{and } \sum_{i=1}^3 \gamma_{F_i} = 100\%, \sum_{j=1}^3 \gamma_{B_j} = 100\%$$

In Formula 1, S refers to the final result of the decoupling decision for a “free module”. If it is greater than 1, the service module should be allocated to the front stage and to the back stage if it is less than 1.

$F_i \gamma_{F_i}$ denotes the total benefit (or grade) that can obtain from allocating the module to front stage; F_i denotes the benefit (or grade) in each specific aspect that can obtain from allocating the module to front stage, as demonstrated in Table 2 which is based on the analysis in Table 1. F_1 , for example, means the benefit that can get from improving service delivery by putting the module in front-stage. γ_{F_i} denotes the weights associated with the benefit in each specific aspect.

Similarly, $B_j \gamma_{B_j}$ denotes the total benefit (or grade) that can obtain from allocating the module in back stage; B_j denotes the benefit (or grade) in each specific aspect that can obtain from putting the module in front stage, also explained in Table 2. γ_{B_j} denotes the weights associated with the benefit in each specific aspect.

Managers and experts' evaluation may help to get the values of F_i and B_j . The values can be estimation of the direct economic benefits, or subjective grading based on their experiences and knowledge. Therefore the key of the decoupling decision-making in Formula 1 is the assessment of weight values associated with the benefit in each specific aspect, namely the value of γ_{B_i} and γ_{F_i} . We can obtain them by calculating the relative importance with Judgment Matrix as described in the following sections.

3.3.2 The Judgment Matrix

We build two Judgment Matrixes respectively by using two groups of judgment objects with three each. These judgment objects come from the benefit in each specific aspect that can obtain from allocating the module to front or back stage, namely the F_1, F_2, F_3 and B_1, B_2, B_3 as explained in Table 2.

In the judgment matrix $F=[f_{ij}]$, f_{ij} denotes the relative importance for F_i versus F_j , and $i, j=1,2,3$. Similarly in judgment matrix $B=[b_{ij}]$, b_{ij} denotes the relative importance of B_i over B_j , and $i, j=1,2,3$. The judgment matrix is symmetrical and consistent, that is: $f_{ik} \cdot f_{kj} = f_{ij}$.

It is therefore obvious that the greater the value is, the more important the factor i is. A judgment scale with five properties can be applied: 1 to 5 respectively denotes that factor i and j are the same important, more important, important, very important and extremely important. We also need to do single-sort of hierarchy in the factor values and consistency test.

3.3.3 Consistency Test

The indicators for consistency test is

$$C.I. = \frac{\lambda_{\max} - n}{n - 1},$$

$$\text{and } \lambda_{\max} \approx \frac{1}{n} \sum_{i=1}^n \frac{(F\gamma)_i}{\gamma_i} = \frac{1}{n} \sum_{i=1}^n \frac{\sum_{j=1}^n f_{ij} \gamma_j}{\gamma_i},$$

and $(F\gamma)_i$ denotes the i_{th} factor of vector $\vec{F\gamma}$.

Then we sort for the corresponding indicator named R.I., and calculate the ratio of consistency:

$$C.R. = \frac{C.I.}{R.I.} < 0.1,$$

by which to determine whether it is consistent.

3.3.4 Calculation on the Weights of Factors

We calculate it by means of the Root Square Method, formulas are as follow:

Table 2. Indications of F_i and B_j

F_i (Benefits for modules being allocated in front-stage)		B_j (Benefits for modules being allocated in back-stage)	
F_1	Improving service delivery (including flexibility, customization, speed and responsiveness)	B_1	Improving efficiency by adopting industrial (rationale) principles
F_2	Increasing sale opportunities (cross selling)	B_2	Reducing costs on labor and taxation by favorably locating the back-stage
F_3	Improving efficiency by involving customer participations	B_3	Developing expertise for the staff by decoupling and thus improve service quality and speed

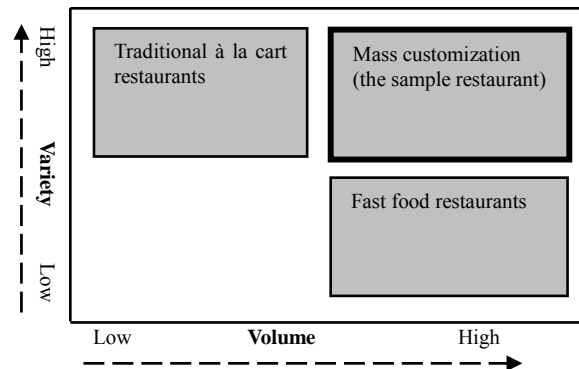


Figure 1. Volume-variety matrix for process type

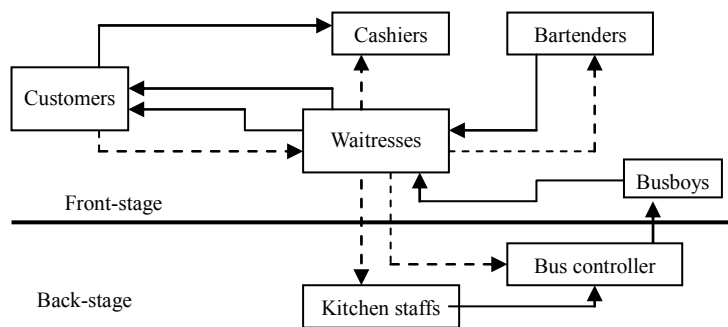


Figure 2. The service process mapping

$$\text{Weight is: } (\gamma)_i = \left(\prod_{j=1}^n f_{ij} \right)^{\frac{1}{n}}$$

$$\text{Relative weight is: } (\gamma)_i^0 = \frac{\gamma_i}{\sum_i \gamma_i}$$

The relative weight $(\gamma)_i^0$ is just the percentage of weighted average, namely the γ_{Fi} or γ_{Bi} we want to obtain. Then with the values of F_i and B_j we can calculate the final value of S in Formula 1, thus completing the decision-making of front or back stage for the module.

4. The Case and the Problem Analysis

4.1 The Sample Organization and Problems

The research sample was a private à la cart restaurant located in eastern China. It had 800 dinning capacity and could provide around 300 dish items according to its menu. However the restaurants had been suffering revenue decline since the enlargement of its operation scale and received plenty of complaints showing customers' dissatisfactions. As the members of Consultancy Services we were invited to diagnose the problems and propose possible improvement measures. Based on a statistic analysis we found that the major problems should be put

priority on the low speed and inefficiency and inadequacy of service delivery such as too long and unfair waiting time, service staffs "absence" from customers, slow response to customer's additional order or special requirements, etc. In addition, the complaints come mostly from the "walk in" customers who had experienced the à la cart services.

4.2 Identification of Process Type and Service Process Mapping

We began with the identification of the operation type of the restaurant based on volume-variety analysis, giving us a general understanding of the nature of the business. There existed obvious differences in between the sample restaurant and traditional à la cart restaurants and fast food restaurants. The traditional à la cart restaurant provided wide rang of customized food and services but in very small volume while the fast food restaurant produced a large volume of dishes within very limited variety. Whereas the sample restaurant could provide wide variety of food and services on a relatively mass basis, it therefore should be positioned in the mass customization area in the volume-variety matrix as shown in Figure 1.

The à la cart service process of the restaurant with a clear decoupling of the front-stage and back-stage could be described in Figure 2. The à la cart services were provided through a paper-based information flow. There

were two sorts of order sheets for waitresses to keep down customers' orders. One was the order sheet for dishes, the other was for drinks. After having taken customers' orders, the waitresses would send 3 copies of the order sheets for dishes respectively to kitchen staffs and bus controller and cashiers, and send 2 copies of the sheet for drinks to bartenders and cashiers. Receiving the order sheet for dishes, the kitchen staff would cook to orders and then send the cooked dishes to the bus controllers. The latter would check the dishes with the order sheets and then tell busboys to carry them to the corresponding tables with helps from the waitresses. For the drinks services, the bartenders would prepare drinks according to the order sheets and the waitresses would then fetch and take the drinks to the corresponding customers. After meal the customers would go to the check-out counter and pay the bill through the cashiers' services. In Figure 1, the dotted line arrows refer to information flows, the line arrows refer to materials flows and monetary flows.

4.3 Identification of the Fail Points in the Service System

4.3.1 The Information Flows and the Service Delivery

The information system in the restaurant remained as the same as they had been when the operation scale had been very small. It was a paper-based system and unable to deal with the greatly enlarged business. Firstly the waitresses had to spend 3-5 minutes to deliver 3 copies of dish order sheets and 2 drinks order sheets to 4 partners: kitchen staffs, bus controller, bartenders and cashiers. This meant that the dish-cooking process and drinks preparation had to be postponed at least 3-5 minutes, which seemed not a short period of time to a waiting customer. Secondly, the paper-based information system affected negatively the dish production in back-stage --- the kitchen. The kitchen staffs should determine the priority of cooking as doing "scheduling" in the manufacturing. Otherwise unscheduled production would lead to not only low efficiency but also unfairness to customers. However the kitchen staffs who were in charge of the final scheduling always got "lost" in work because the paper-based information was not reliable and clear enough for them to arrange the scheduling when dealing with so many orders. Thirdly the cashiers had to collect a lot of copies of the order sheets for each table (customer) and be prone to make mistakes if too many customers check out at the same time (unfortunately it is common in catering services). We could also find similar problem among bartenders and bus controllers.

4.3.2 The Fail Points Related to System Decoupling

Many decoupling-related fail points could be also found in both the front-stage and back-stage. Full service catering meant longer dinning time and hence many customers would always require re-processing (such as re-heating,

re-shaping and re-flavoring) some already served dishes. These requirements would be settled by sending the dishes to the kitchen and "inserting" into the routine schedule for dish production. But they had been always delayed if the kitchen had been busy and the schedule had been too tight. There had been considerably some re-processing claims from the customers. Pooling these orders together and processing them in the back-stage--- kitchen seemed very efficient. However this had led to delayed services and greatly reduced the responsiveness, causing customers' dissatisfaction.

Another fail point could be found in the cashier services which had been designed as front-stage activities. Cashiers had been arranged sitting behind the counter near to the exit of the dinning hall and providing check-out services in the face of customers who approached the counter after meal. The cashier had to collect all the order sheets for the customer and figured out the sum of payment immediately while they had been in many cases disrupted by the customers' further inquiry even some irrelevant questions such as "how about the weather tomorrow?" In the peak time customers had to in many cases queue before the cashier counter and mistakes in payment frequently occurred.

Drinks preparation and delivery had been inefficiently arranged, assuming the services as front office activities. According the service procedure the waitresses would pass the order sheets for drinks to the bartender who would prepare the drinks to orders one after another (it was very time-consuming). Then the waitresses would carry the prepared drinks to the corresponding table and serve to customers. This procedure would lead to the waitresses "absence from the customers" for 5-8 minutes and most of customers had been unhappy about this period of time "without services".

Cooking to fresh is the requirement for seafood and hence the seafood should be kept alive before they are ordered by customers. Most of the customers preferred ordering the seafood after seeing them alive. The waitresses would always lead the customers into the seafood-keeping area for ordering. However the area had been located as a back-stage unit far into the kitchen and hence the customers' ordering activities had inevitably caused some disturbances to the dish production in the kitchen.

5. System Improvement and Results

5.1 Rationalization of the Service Functions

Under the agreement from the top managements in the restaurant we started the re-engineering plan with the rationalization of all the service functions (or units) in the service system. After a comprehensive analysis of all the service functions and customer demands and interrelationship in between them, some of the service units were

split and some were merged to deal with the problems mentioned above. The drinks-carrying services originally conducted by waitresses, for example, were taken away and merged with function of drinks-preparation to form a new service unit—"drinks service centre", which aimed to increase greatly the efficiency in drinks service. Another example was the split of the original reservation unit into two: the "reservation" and the "Planning and decoration" for enhancement of the growing banqueting services. Similarly many service functions were rationalized based on the utilization of commonality within service units and facilitation of service provision. More shared units were identified after the rationalization and hence the service functions in the restaurant decreased from 33 to 25.

5.2. Re-Engineering of the Information System

After a deep consideration the restaurants decided finally to purchase and set up a computer-based dish-order system and re-design the information flow for the service process. In the new information flow the waitresses would just input the customers' orders into their wireless PDA and press the "confirm" button, the order information would then be automatically sent into the terminals in kitchen, bus controller, bartenders and cashiers. The back-stage production thus was triggered within seconds and saved at least 3-5minutes for the order-receiving process. Furthermore with the accurate and easy-processing information the relevant staffs (including kitchen staffs, bartenders, bus controllers and cashiers) could readily arrange their job efficiently (eg. the computer could automatically provide schedule for the kitchen staffs to organize efficient dish production).

5.3 Re-Decoupling the Service System

5.3.1 Switching from Back-Stage to Front-Stage

As mentioned early re-processing the already served dishes was arranged in the back-stage for high efficiency but incurred customers' complaints due to the slow responsiveness. The restaurant was persuaded to move this function to front-stage. Some re-heating equipments (e.g. microwave oven) and cutting appliances (e.g. knives) were purchased and placed in the dinning hall, and attractive instruction brochures of simple cooking methods in microwave oven and some seasoning materials were put in the visible place of the dining hall. Thus the re-heating and shaping requirements by customers could be undertaken right in the front stage with helps from the waitresses without bringing disturbance to in-process production in the kitchen by "inserting" the these orders into the already rigid schedule. These measures also encouraged customers' self-services and alleviated production pressure in the kitchen so as to improve the service efficiency while maintaining a quick responsiveness to customers' special needs.

Other functions in back-stage were also reviewed and some of them were re-deployed to the front-offices for a comprehensive and holistic consideration. Seafood-keeping area, for an instance, was decoupled into two parts. One was the storage area for seafood-keeping; the other was the display part for customer to make orders. The display part was removed from the kitchen to the front-stage, thus minimizing the disturbance caused by customers to the back-stage production.

5.3.2 Switching from Front-Stage to Back-Stage

The cashier services were re-designed as a back-stage function rather than front-stage activities as it previously had been. Thus customers needed not go to the cashier counter to pay the bill themselves after dining. Rather they could just inform the waitresses and the latter would help them to conduct the payment activities. With the help of computer-based order system, the cashiers could readily process the bills. When the waitresses come to inform a payment, they could finish processing the bill within seconds since the bills could be pre-processed in the computer before the final payment, and very importantly, they could do the job without contacting the customers who might bring disturbance, which also eliminated the queue before the cashier counter.

The services for drinks preparation and delivery were also switched from a front office function to a back-stage one. Equipped with the computer terminal receiving drinks orders from the PDA held by the waitresses, the bartenders could efficiently arrange the drink preparation just like a store-keeper. To facilitate service delivery the dining hall was divided into several dining areas. The bartenders just put into a cart all the drinks needed in all the order sheets within a specific dinning area. Then the bar assistants would push the carts to corresponding dinning areas and deliver the drinks to the waitresses along the route just as the postmen deliver their mails. The back-stage-style services of drinks preparation and delivery could shorten the lead time greatly and provide more support for the front-line staff since the waitresses needed not fetch the drinks themselves and would not be "absent" from the customer, thus service quality could be improved.

However one reform was not very successful. The restaurant had a dish display area where samples of some master piece of dishes (or dishes in promotion) were displayed for order. The managements of the restaurant held that it had taken up too much space in the front-stage. After the installation of the computer system they transformed it into dinning place and provided customers as a substitution with the picture-based menu and a big-screen projector introducing new dishes. However this seemed not very effective and more customers preferred ordering by seeing the "real" sample dishes rather than the paper or electronic image.

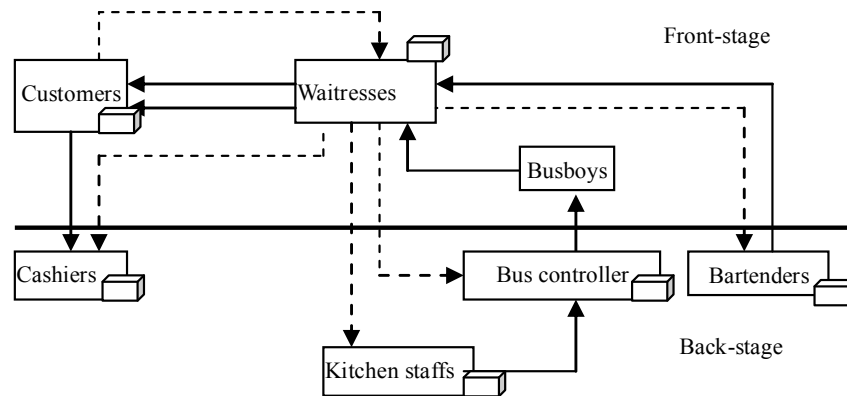


Figure 3. The improved decoupling system

Table 3. Judgment matrix for being in front stage and the calculation of its weight

	F ₁	F ₂	F ₃	γ_i	γ_i^0	λ_i	$\lambda_{\max} \approx 3.054$ C.I.=0.027<0.1 R.I.=0.52 C.R.=0.052<0.1
F ₁	1	4	4	2.52	0.661	3.054	
F ₂	1/4	1	2	0.794	0.208	3.053	
F ₃	1/4	1/2	1	0.500	0.131	3.054	
Judgment matrix				3.814			

Table 4. Judgment matrix for being in back stage and the calculation of its weight

	B ₁	B ₂	B ₃	γ_i	γ_i^0	λ_i	$\lambda_{\max} \approx 3.094$ C.I.=0.047<0.1 R.I.=0.52 C.R.=0.009<0.1
B ₁	1	5	4	2.714	0.687	3.095	
B ₂	1/5	1	2	0.737	0.186	3.093	
B ₃	1/4	1/2	1	0.5	0.127	3.094	
Judgment matrix				3.951			

The improved decoupling service system is shown in Figure 3 (Referring to the computer terminals). Most of the measures took effective finally and the managements of the restaurant were satisfied with our efforts with the regain of revenue and customer satisfaction.

5.4 Justification of the Judgment Matrix Proposed

The decoupling decision for cashier services is taken as an example to illustrate the application of the model proposed previously. The cashier service module belonged to “free modules” without high customer contact, and hence could utilize Formula 1 to make decoupling decision. γ_{Bi} and γ_{Fi} were the focuses of the calculation. Based on Table 2 summarizing the possible benefits gained from putting this module in front or back stage, we take an analysis of the importance of these benefits and give weight to each.

We began with building the Judgment Matrixes. We got the first matrix by comparing the importance in be-

tween the three benefits being put in front stage and then conducted the following calculation and consistency test (the normalizing result indicated that this matrix met the requirements and was consistent). The relevant data and the calculation results are shown in Table 3. Similarly, we obtained Table 4 by building and calculating the Judgment Matrix for putting the module in back stage.

From these two tables we got the values for γ_{Fi} and γ_{Bi} , respectively 0.661, 0.208, 0.131 and 0.687, 0.186, 0.127.

With the helps from the managers and experts, we got the values of F_i (7,7,3) and B_j (7,3,9). Thus we could finally figure out:

$$\sum_{i=1}^3 F_i \gamma_{Fi} = 6.476 \quad \text{and} \quad \sum_{j=1}^3 B_j \gamma_{Bj} = 6.591$$

By using Formula 1 we got S less than 1. Therefore the cashier service module should be put in back stage according to proposed model, which coincided with the actual decoupling decision.

6. Findings and Discussions

6.1 The Effectiveness of Applying Mc Concepts into Service System Decoupling

System decoupling is regarded as an effective way to deal with the influences caused by customer contact according to previous literatures. But these researches have been based on a dichotomous assumption that service organizations focus on either low costs or high services. The case put the MC concepts and methods into the decoupling of catering service system and tested the effectiveness of those methods in achieving multiple business objectives simultaneously. These include modularization of service system, introduction of technologies and comprehensive consideration on decoupling decisions. These issues and their effectiveness could be justified by most of the successful re-decoupling measures adopted by the sample restaurant.

6.2 Modularization as a Basis for System Decoupling in Mc

Modularization is one of the most effective enablers to facilitate mass customization and the premise for system decoupling in mass customization. The case proved this viewpoint in the rationalization of service functions by splitting or merging them. The rationalization in the restaurant was in essence a process of modularization since it had been conducted by utilizing commonalities within service functions and facilitating service provision. The “drinks-service centre”, for example, was set up for efficient service delivery by utilizing the commonality among service functions which required cooperation from the drinks service function. The rationalization finally led to identification of more shared service units and decrease of service units over all, which helped a lot to achieve economy of scale while maintaining service quality, reflecting the features of modularization. Additionally the following-up decoupling measures was then based on the service function rationalization or modularization, which also explained the nature of decoupling for MC proposed previously.

6.3 Balance-Striking among the Multiple Objectives in Decoupling Decision for Mc

Literatures tell that tradeoffs involved between contact benefits of the front-office operations and the potential efficiency of the back-office with limited or no customer contact. In decoupling process for MC, balance thinking among different operation objectives should be adopted to deal with the tradeoffs involved in decoupling decision for each service module. Namely the system efficiency, service customization, responsiveness and relevant issues should be taken into consideration simultaneously for

allocating each individual service module to front or back office. The switching of re-processing job from the kitchen to front-stage, for example, was based on the consideration that the efficiency brought by putting the function in back-stage could not offset the customers' dissatisfaction due to the slow responsiveness (service speed). The re-design of the cashier services into a back-stage function also justified this viewpoint since the benefits from being a front-stage activity could not match the increased efficiency being in the back-stage. This also showed that the operation tradeoffs involved in the decoupling process could be evaluated by carefully comparing the relative importance of the benefits gained from being in front stage and back stage respectively as demonstrated in the model based on Judgment Matrix.

6.4 Technologies and System Decoupling in Mc

The vital role of information technology in mass customization operations could be justified by the re-engineering of the information system in the case. Besides, the case address more issues of technologies related to decoupling. The most important criterion for decoupling decision is whether the service function needs customer contact or not. It seems that the nature of the service decide whether customer contact is required. However this has been changed by the development of technologies especially IT which could remove the customer from the service system without affecting service quality, which gives us more flexibility to utilize the decoupling approach. We could find that most of the measures for system re-decoupling in the case linked closely to the use of technologies. The introducing of microwave oven and related equipments enabled the re-processing function moving from back stage to the front stage. The big-screen projector took customers out of the display area and thus saved much space for front stage use. The computer-based order system played very vital role in transferring the cashier and drink preparation services from front stage to back stage. However the use of technologies is not always effective and they can not substitute the customer contact in all cases. The unsuccessful reform in display area of sample dishes in the case justified the point.

7. Conclusions

The paper discusses the system decoupling issues in mass customization operation in the context of restaurant services. It justifies that integration of the concept and methods of MC into system decoupling is effective to deal with the operation dilemma and that modularization plays vital role in the decoupling decision-making process. It also explains that the nature of decoupling decision for MC purpose is the allocation of the service modules to front or back stage rather than the physical or geographical separation of service system. Furthermore a

Judgment-Matrix-based model is proposed on how to make decoupling decisions for mass customization by balancing the multiple operation objectives. Moreover the technologies, especially the IT, are proved to be very important for system decoupling in delivering mass customized services. However the proposed model and implications in the paper need further test in various service contexts for better theoretical generalities, giving the limitations of the case study methodology used in this research.

REFERENCES

- [1] R. B. Chase, "The Customer Contact approach to services: Theoretical based and practical extensions," *Operation Research* 29, pp. 698–705, 1981.
- [2] W. Nie and D. L. Kellogg, "How professors of operations management view service operations?" *Production and Operations Management* 8, pp. 339–355, 1999.
- [3] M. H. Safizadeh, J. M. Field, and L. P. Ritzman, "An empirical analysis of financial services processes with a front-office or back-office orientation," *Journal of Operations Management* 21, pp. 557–576, 2003.
- [4] R. B. Chase and R. H. Hayes, "Beefing up operations in service firms," *Sloan Management Review* 33, pp. 15–26, 1991.
- [5] K. D. Hoffman and J. E. G. Bateson, "Essentials of service marketing," The Dryden Press, Orlando, 1997.
- [6] R. B. Chase, "Where does the customer fit in a service operation?" *Harvard Business Review* 56, pp. 137–142, 1978.
- [7] R. B. Chase and D. A. Tansik, "The customer contact model for organization design," *Management Science* 29, pp. 1037–1050, 1983.
- [8] P. Breathnach, "Globalization, information technology and the emergence of Nich transnational cities: The growth of the call centre sector in Dublin," *Geoforum* 324, pp. 477–485, 2000.
- [9] J. L. Heskett, "Lessons in the service sector," *Harvard Business Review* 65, pp. 118–126, 1987.
- [10] J. L. Heskett, W. E. Sasser, and L. A. Schlesinger, "The service profit chain," The Free Press, New York, 1997.
- [11] R. B. Chase and N. J. Aquilano, "A matrix for linking marketing and production variables in service system design," In: Richard, D. (eds), *Production and Operations Management*, 6th edition Irwin, Homewood, 1992.
- [12] R. D. Metters and V. Vargas, "A typology of de-coupling strategies in mixed services," *Journal of Operations Management* 18, pp. 663–682, 2000.
- [13] S. M. Davis, "Future perfect," Addison-Wesley, New York, 1987.
- [14] B. J. Pine II, B. Victor, and A. C. Boynton, "Making mass customization work," *Harvard Business Review* 71, pp. 108–119, 1993.
- [15] M. M. Tseng and J. Jiao, "Design for mass customization," *CIRP Annals* 45, pp. 153–156, 1996.
- [16] R. Duray, P. Ward, G. Milligan, and W. Berry, "Approaches to mass customization: configurations and empirical validation," *Journal of Operation Management* 18, pp. 605–625, 2000.
- [17] G. D. Silveira, D. Borenstein, and F. S. Fogliatto, "Mass customization: Literature review and research directions," *International Journal of Production Economics* 72, pp. 1–13, 2001.
- [18] K. Turowski, "A virtual electronic call centre solution for mass customization," *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences*, Vol. 8, pp. 8027–8036, 1999.
- [19] L. G. Zomerdijk and J. D. Vries, "Structuring front office and back office work in service delivery systems," *International Journal of Operations & Production Management* 27, pp. 108–131, 2007.
- [20] E. D. Arnheiter and H. Harren, "A typology to unleash the potential of modularity," *Journal of Manufacturing Technology Management* 16, pp. 699–711, 2005.
- [21] C. A. Voss, "Rethinking paradigms of service: Service in a virtual environment," *International Journal of Operations & Production Management* 23, pp. 88–104, 2003.

Periodicity of Cycle Time in a U-Shaped Production Line with Heterogeneous Workers under Carousel Allocation

Mikihiko Hiraiwa, Koichi Nakade

Department of Architecture, Civil Engineering and Industrial Management Engineering, Nagoya Institute of Technology, Japan.
Email: 16518507@stn.nitech.ac.jp, nakade@nitech.ac.jp

Received September 3, 2009; revised October 12, 2009; accepted November 17, 2009.

ABSTRACT

A U-shaped production line with multiple machines and multiple workers is considered under carousel allocation in which all workers take charge of all machines in the same order. Nakade and Ohno (2003) show that, when the processing, operation and walking times are constant, the overall cycle time, which is a time interval between successive outputs of finished goods, is the greatest value of the maximal sum of the processing and operation times among machines and the time required for a worker to operate and walk around the production line without waiting for processing divided by the number of workers. In this paper, it is considered that operation times at each machine may be different between workers. If processing time is short, it is expected that the overall cycle time will be equal to the time for a worker to operate and walk around the line divided by the number of workers. However, under some specified cases, the overall cycle time is longer than that of this time, and the overall cycle time changes periodically. From numerical examples, it is shown that the order of arrivals of workers at machine 1 affects the overall cycle time. We give some properties on the periodicity of cycle times and discuss about cycle times.

Keywords: U-Shaped Production Line, Carousel Allocation, Operation Times, Cycle Time, Periodicity of Cycle Times

1. Introduction

In these days, U-shaped layout is commonly used in many production lines. In a U-shaped production line, there are two types of worker allocations, which are separate type and carousel type. In the separate allocation, each worker deals with a unique set of machines. In the carousel allocation, workers deal with all machines in the same route. In this paper, the carousel allocation is considered. An advantage of U-shaped layout is that it is easy to adjust the throughput of finished products from the last machine for fluctuation of demand by changing the number of workers [1,2].

Nakade and Ohno [3] have considered a U-shaped production line with multiple multi-function workers, and show the overall cycle time which is a time interval between successive outputs of finished products. When the processing, operation, and walking times are constant, the overall cycle time is the greatest value of the maximal sum of the processing and operation times among machines and the time required for a worker to operate and walk around the production line without waiting for

processing divided by the number of workers. They assume that operation times at each machine are the same among workers.

Recently, many temp staffs are employed as workers and they are committed into production lines because of cost reduction. For example, many foreigners and part time workers are employed for a short span. Therefore it becomes difficult to maintain the workers well-skilled in the long time, so the worker's skills remain mutually different.

Nakade and Nishiwaki [4] have considered a U-shaped production line with multiple heterogeneous multi-function workers and propose an algorithm for computing an optimal allocation of workers to machines which minimizes the overall cycle time.

The formula of cycle time derived in Nakade and Nishiwaki [4] is simple and understandable, where it is assumed that different workers are assigned to the different machines. In the other allocation scheme, all workers have operations at all machines and as a carousel workers go around machines. We say this scheme carousel allocation.

In this paper we consider the case that operation times at each machine are different among workers under carousel allocation. In this case, if the processing times are very small, then it may be guessed that the overall cycle time is equal to the greatest value of times required for each worker to operate and walk around the line divided by the number of workers. However, under some specified cases, the overall cycle time is larger than this time, and the overall cycle time changes periodically. We discuss the periodicity of cycle times and observe some properties on cycle times.

In the next section we describe a U-shaped production line. In Section 3 we show that the overall cycle time changes periodically by numerical examples and discuss the periodicity of cycle times. In Section 4 we conclude and discuss future research.

2. A U-Shaped Production Line

We consider a U-shaped production line with K machines and J workers, which is shown in Figure 1. It is assumed that $J \leq K$. There is enough raw material in front of machine 1. The material is processed at machines 1 to K sequentially, and customers receives finished products from machine K . Workers deal with all machines in the same route and operate an item at each machine. Operation times at each machine may be different among workers, and the operation time of worker j at machine k is denoted by $s_{j,k}$, which is deterministic. Worker j starts his first operation at machine k_j and visits machines $k_j+1, k_j+2, \dots, K, 1, 2, \dots$ sequentially. The deterministic walking time from machine k to $k+1$ ($k=1, 2, \dots, K$) is denoted by $r_{k,k+1}$, where it is assumed that $K+1=1$. The processing time at machine k is denoted by i_k , which is deterministic. When a worker visits a machine, if the preceding worker is operating at the machine or the

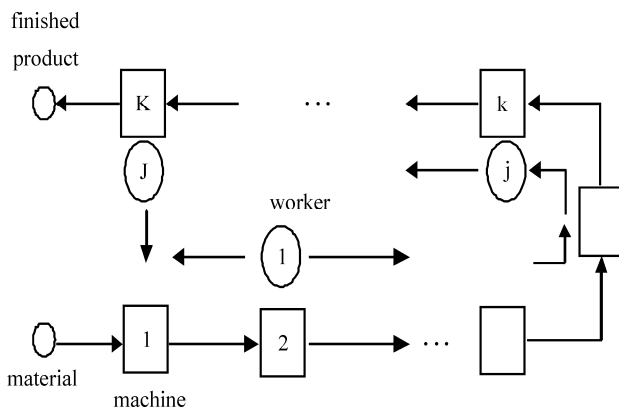


Figure 1. U-shaped production line

machine is processing the preceding item, then the worker must wait for the completion of operating or processing. The overall cycle time is defined as a time interval between successive outputs of finished products. We also define the cycle time of worker j as the time interval between the successive arrivals of worker j at machine 1. It is noted that the overall cycle time is the cycle time divided by the number of workers. The first cycle time of worker j is the time interval between a start of first operation and the first arrival of worker j at machine 1. The n th ($n \geq 2$) cycle time of worker j is the time interval between the $(n-1)$ th and n th arrivals for worker j at machine 1.

3. Cycle Time

3.1 Case of the Same Operation Times among Workers

In the case that operating times at each machine are the same among workers, from Nakade and Ohno (2003), when processing, operation and walking times are constant, the overall cycle time $C^{overall}$ is expressed as

$$C_c^{overall} = \max \left\{ \max_{k \in \hat{K}} (s_k + i_k), \frac{1}{J} \left(\sum_{k \in \hat{K}} s_k + \sum_{k \in \hat{K}} r_{k,k+1} \right) \right\}, \quad (1)$$

where s_k denotes the operation time at machine k and $\hat{K} = \{1, 2, \dots, K\}$. The overall cycle time is the greatest value of the maximal sum of the processing and operation times among machines and the time required for a worker to operate and walk around the production line without waiting for processing divided by the number of workers.

3.2 Case of Different Operation Times among Workers

We consider the case that operating times at each machine may be different among workers. In this case, from Equation (1), the overall cycle time $C^{overall}$ may be guessed as

$$C^{overall} = \max \left\{ \max_{k \in \hat{K}} \left(\frac{1}{J} \sum_{j \in \{1, 2, \dots, J\}} s_{j,k} + i_k \right), \frac{1}{J} \max_{j \in \{1, 2, \dots, J\}} \left(\sum_{k \in \hat{K}} s_{j,k} + \sum_{k \in \hat{K}} r_{k,k+1} \right) \right\}. \quad (2)$$

In what follows, we consider the cycle time instead of the overall cycle time. Since the overall cycle time is the cycle time divided by the number of workers, the cycle time C may be guessed as

$$C = \max \left\{ \max_{k \in \hat{K}} \left(\sum_{j \in \{1,2,\dots,J\}} s_{j,k} + J i_k \right), \max_{j \in \{1,2,\dots,J\}} \left(\sum_{k \in \hat{K}} s_{j,k} + \sum_{k \in \hat{K}} r_{k,k+1} \right) \right\}. \quad (3)$$

From Equation (3), when processing times are zero, it is guessed that the cycle time is equal to the greatest value among times required for each worker to operate and walk around the line. However, under some specified cases, the cycle time is greater than this value, and the cycle time changes periodically. The examples are shown in the next section.

4. Periodicity of Cycle time

4.1 Numerical Examples

The set of operation times of each worker at each machine is denoted by

$$S_l = \begin{pmatrix} s_{1,1}, \dots, s_{1,K} \\ \vdots \\ s_{J,1}, \dots, s_{J,K} \end{pmatrix},$$

where subscript l is a number which distinguishes sets of operation times.

Let $J = 3$ and $K = 4$. It is assumed that processing times and walking times are zero, that is, $i_k = 0$ and $r_{k,k+1} = 0$ for $k \in \hat{K}$. Worker 1 starts his first operation at machine 1, worker 2 at machine 2 and worker 3 at machine 3. The first cycle is until the first arrival of a worker at machine 1. The n th ($n \geq 2$) cycle is the

Table 1. Cycle times for S_1

cycle\worker	1	2	3
1	8.0	7.0	6.0
2	8.0	8.0	8.0
3	8.0	8.0	8.0
4	8.0	8.0	8.0
5	8.0	8.0	8.0
6	8.0	8.0	8.0

Table 2. Cycle times for S_2

cycle\worker	1	2	3
1	8.0	7.0	2.0
2	11.0	11.0	11.0
3	8.0	8.0	8.0
4	11.0	11.0	11.0
5	8.0	8.0	8.0
6	11.0	11.0	11.0

interval between $(n-1)$ th and n th arrivals of a worker at machine 1.

Cycle times, when operation times of each worker at each machine are $S_1 = \begin{pmatrix} 5, 1, 1, 1 \\ 1, 5, 1, 1 \\ 1, 1, 5, 1 \end{pmatrix}$, are shown in Table 1.

Cycle times of all workers are 8 at all cycles except for the first cycle. This is the case that the cycle time does not change periodically. Then the cycle time 8 in Table 1 is equal to the value which is derived from Equation (3).

Cycle times, when operation times of each worker at each machine are $S_2 = \begin{pmatrix} 5, 1, 1, 1 \\ 1, 1, 5, 1 \\ 1, 5, 1, 1 \end{pmatrix}$, are shown in Table 2.

In this case, the cycle time changes periodically. Except for the first cycle, cycle times of each worker take values of 8 and 11 alternately. The cycle time 8 in Table 2 is equal to the value which is derived from Equation (3). The cycle time 11 is more than the value which is derived from Equation (3).

4.2 Discussion

We investigate many examples including the above, and we have the following properties on the cycle time.

1) In many cases, in which the difference is not so large among workers, the cycle time is the same as Equation (3).

2) If the cycle time changes periodically, then the period of the cycle time is two.

3) If the cycle time changes periodically, then the smaller cycle time is equal to the one which is derived from Equation (3), and the greater cycle time is more than the one which is derived from Equation (3).

4) Initial machines of each worker do not affect the cycle time if an order of arrivals of workers at machine 1 does not change, where, for example, the orders “4,3,2,1” and “2,1,4,3” are regarded as the same order.

5) An order of arrivals of workers at machine 1 affects the cycle time.

The movements of each worker when operation times are equal to S_2 are shown in Figure 2. When operation times are equal to S_2 , for example, worker 2 tends to wait for the completion of operation of worker 3 at machine 2 since machine 2 is bottleneck for worker 3. And also, since machine 3 is bottleneck for worker 2, worker 2 operates at his bottleneck machine (machine 3) after the waiting at machine 2. This makes the cycle time of worker 2 amplified. If a worker has his bottleneck machine after the bottleneck machine of his preceding worker, then the cycle time of the worker is amplified. In the next cycle, worker 3 finished his operation at his bottleneck machine (machine 2) while worker 2 operates at

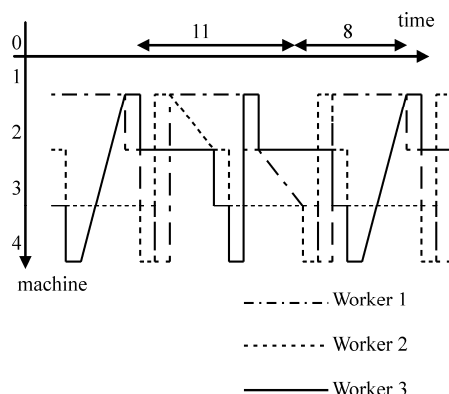
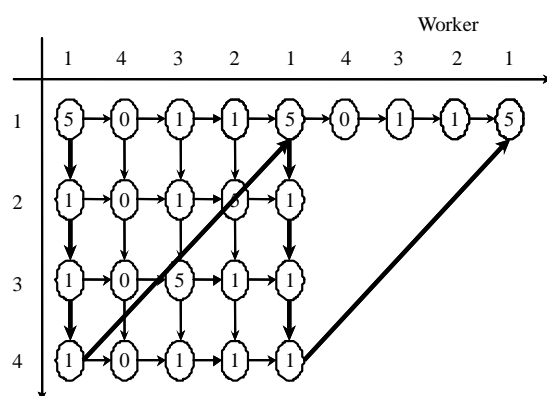
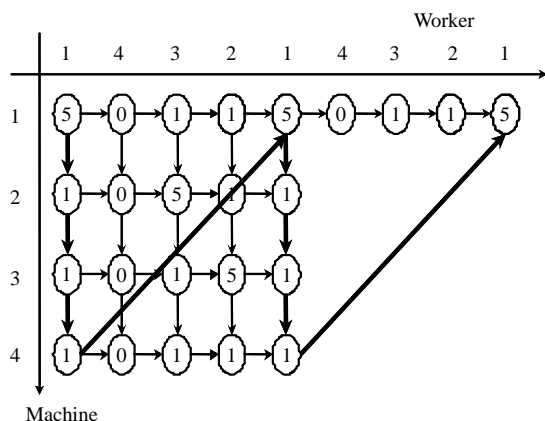


Figure 2. Movements of workers

Figure 3. Arrow diagram for S_1 Figure 4. Arrow diagram for S_2

his bottleneck machine (machine 3) and the succeeding machines (machines 4 and 1). Therefore in this cycle worker 3 does not need to wait for the completion of the operation of worker 3 at machine 2. Hence the cycle time of worker 2 does not increase in this cycle. These are why the cycle time of worker 2 changes periodically in Figure 2.

On the other hand, when operation times are equal to S_1 , machine 3 is bottleneck for worker 3, machine 2 for worker 2 and machine 1 for worker 1. When worker 3 operates at machine 3, worker 2 tends to operate at machine 2, and when worker 2 operates at machine 2, worker 1 tends to operate at machine 1. That is, when a certain worker operates at his bottleneck machine, the other workers tend to operate at their bottleneck machines. Therefore, cycle times are the same value in every cycle.

Let us change initial machines of each worker under the condition that the order of arrivals of workers at machine 1 does not change. Worker 1 starts his first operation at machine 1, worker 2 machine 3 and worker 3 machine 4. Operation times of each worker at each machine is equal to S_2 . Then cycle times of each worker are equal to the results which are shown in Table 2. In all examples which we investigate, initial machines of each worker do not affect cycle times.

Compare sets of operation times S_1 with S_2 . The set S_2 is the case that operation times of workers 2 and 3 at each machine are exchanged in the set S_1 . From results of Table 1 and 2, the order of arrivals of workers at machine 1 affects the cycle time.

Figures 3 and 4 are arrow diagrams which show cycle times of worker 1 for successive two cycles. Numbers on vertical axis denote machines and numbers on horizontal axis denote workers. Nodes denote operations of each worker at each machine and numbers on nodes denote operation times of each worker at each machine. Figure 3 shows the case of S_1 and Figure 4 shows the case of S_2 . Worker 4 is a dummy worker and the operation times are zero. When we follow arrows from the most upper left node, which is referred as to the initial node, to the most upper right node, which is referred as to the last node, if the sum of cycle times on the route is maximal, then the sum is equal to the sum of cycle times of worker 1 for successive two cycles. In Figure 3, on the route that the sum of cycle times is maximal, the sum of cycle times is 16. This value is equal to the sum of cycle times for successive two cycles in Table 1. Similarly, in Figure 4 the sum is 19. This value is equals to the sum of cycle times for successive two cycles in Table 2. With Comparing Figures 3 with 4, in Figure 3 there are two nodes at which the operation time is equal to 5 on all routes, while in Figure 4 there are two nodes on which the operation time is equal to 5 on some routes, and there are three nodes on which the operation time is equal to 5 on others.

5. Conclusions

In this paper we consider a U-shaped production line with multiple machines and multiple workers under car-

ousel allocation. In the case that operation times at each machine are different among workers, we investigate the cycle time. Under some specific cases, the cycle time changes periodically. We have some insights into the periodicity of the cycle time and discuss cycle times. For future research the condition that the cycle time changes periodically is derived.

REFERENCES

- [1] Y. Monden, "Toyota production system," An integrated Approach, 3rd edition, Engineering and Management Press, Georgia, 1998.
- [2] G. J. Miltenburg, "One-piece flow manufacturing on u-shaped production lines: A tutorial," IIE Transactions, Vol. 33, pp. 303–321, 2001.
- [3] K. Nakade and K. Ohno, "Separate and carousel type allocations of workers in a U-shaped production line," European Journal of Operational Research, Vol. 145, pp. 403–424, 2003.
- [4] K. Nakade and R. Nishiwaki, "Optimal allocation of heterogeneous workers in a U-shaped Production Line," Computers & Industrial Engineering, Vol. 54, pp. 432–440, 2008.

The Exploration of Implementation of Emissions Trading System in China¹

Min HU

Economics and management school of Sichuan Normal University, Chengdu, China.
Email: humin998@yahoo.com.cn

Received June 10, 2009; revised July 23, 2009; accepted September 1, 2009.

ABSTRACT

This paper studies the route, steps and measures of the implementation of the emissions trading system in China. It combines desk research and case study. The desk research aims to explore the inherent discipline of emissions trading system, so as to disclose its nature and features while the case study involves the one-year field research conducted by the author, and provides the pilot emission systems in some provinces, together with the difficulties in its implementation. There are possibilities of comprehensive coverage of emission trading system, which can only be protected by a law relevant to environmental property protection (issued by the Standing Committee of National People's Congress). And there is necessity for China to set up a quota system for different pollutions. After the initial allocation among provinces, the quota will be divided among enterprises, partial paid and partial for free, under the principle of "Giving priority to efficiency with due consideration to fairness". It is useful to establish an efficient secondary market of emission trading system.

Keywords: Emissions Trading, System Innovation, Sustainable Development

1. Introduction

China's economic and social development, resources saving and ecological environment protection have become increasingly sharp, as a constraint of the bottleneck in achieving sustainable development. Exploring environmental tools for the conservation of resources, environmental protection and effective achievement of the economic and social comprehensive development, is a significant major for China right now. Theoretical analysis and international experience have shown that emissions trading system as a system of market-based incentives designed to effectively achieve these objectives, it is worth drawing on the use of China. However, it is also necessary to fully integrate the implementation of China's emissions trading system with China's economic and social reality, We need to do some research in the specific systematic construction and operate mechanism as well as make adaptive adjustments, so can we ensure to

turn it into the emissions trading system which suitable for the national needs, as well as a powerful tools to prevent pollutions. This article seeks to make some explorations.

2. The Inspirations of China's Experience in the Emissions Trading Pilot Areas

In recent years, Tianjin, Jiangsu, Zhejiang, Shanghai, Shanxi, Henan, Guangxi and other provinces and municipalities of China have launched a pilot emissions trading, Jiaxing Zhejiang in November 10, 2007 established the first national emissions trading reserves center, indicating that China has taken a substantive step for implementation of the right of sewage system. We can draw some experience in the following inspiration by summing up the experience of the pilot areas.

2.1 Legal Matter

The thought of "Total Control" has not been embodied in laws, the property of environmental resources are still lacking the evidence, and the entire emission trading system is extremely fragile. In September, 2002, Xiu Zhou district of Jia Xing began the emission trading system pilot, but the outcome wasn't that good, one of the most crucial reasons is the "Total Control" doesn't have

¹Fund Project: This article is the project "Sustainable Development of the Sichuan-based emissions trading mechanisms research" of Philosophy and Social Planning of "Eleventh Five" in Sichuan Province, "211 Project" features three research projects of Southwest University of Finance and Economics, the results of "the implementation of health-intensive countries consumption patterns and promoting the study of ecological civilization" (item number: 211D3T10).

any bottom lines. For example, Shen Yue-ping, the deputy director of the Environmental Protection Agency of Jia Xing pointed out: "The rights of Sewage emission means add more administrative permit, if there is no requirement for the volume. It is not the scarcity of resources, so there is no need for transactions [1]. When the total control becoming a rigid target, the environment capacity is able to become a scarce resource, the market pricing mechanism for emissions trading have the conditions. However, the thought of "total control" must be established in the law before it can authoritative. In recent years, although China has made significant achievements in environmental protection legislation and building several levels of environmental legal systems, the trading legislation in national level is still empty. Under the existing environmental protection laws and regulations, only the individual provisions for special pollutants embodied in the "total control", the main laws and regulations do not have the specific rules for "total control", which directly led to the shortage of legal rules for the property rights of environmental resources, so that the emissions trading system itself is very fragile.

2.2 Inadequate Liquidity

Emissions trading market does not have much of the attractiveness of business, which result in the limitation of trading volume and liquidity shortage. For instance, the establishment of the first national emissions trading reserve center in Jiaxing, Zhejiang, the annual emissions of COD was about 118,000 tons, sulfur dioxide emissions was 273,000 tons in 2007, but the transaction into the trading center was only hundreds of tons of the emission reductions in scale, which is still minimal; And in 2004, emissions of sulfur dioxide and a pilot of emissions trading permit system of sulfur dioxide had been carried out in Henan Province for 4 years, and suffered from the embarrassment of a zero-transaction [2].

2.3 The Market Risk of "Bad Money Drives out Good Money"

Now the emissions trading mainly carried out in the pilot areas, and this may have the market risk of "bad money drives out good". Sewage emission enterprises in the emissions trading pilot need to spend a huge amount of money to spend in buying emission rights, which will result in increased operating costs of enterprises, business competitions in the surrounding areas without the rights of emissions trading is in a "cost disadvantage", brought about unfair competition objectively. As a result, the full implementation of China's emissions trading system is imperative.

2.4 The Uncontrolled Situation in the Market

Economic system is still in the transition period, the market economy is not that perfect, which result in low

efficiency of market of emission rights, and even the market failure. Emissions trading system is a policy instruments to control pollution and protect the environment by market-based instruments, nature is the product of the full development of market economy. However, the time for establishing Chinese socialist market economic system is still short. We have made great achievements in all aspects, but as a whole we are still in the transition period, we are not perfect and mature enough in many ways, including well-defined property rights, the effective transfer of property rights and the details in the legal protection of property rights. Such an Imperfect market system will inevitably have an adverse effect on the emissions trading system.

2.5 A Awareness of Environmental Protection

People do not have a strong awareness of environmental protection, emissions trading system still lack of social supervision. The economic development of China in recently years proves that, whether it is the national or local government, they just chase the economic interests but ignore the environmental protection, people lack of environmental awareness in general. The environmental awareness, however, as a non-governmental social force, has supervised, supported and promoted the making and practicing of laws concerning environmental protection. However, with the weak public awareness of environmental protection, it's difficult to keep a firm hand on the supervision of trade on the disposal of waste.

2.6 Social Credit System

The unsound social credit system increases the possibility of emissions trading market failure. Social credit system is the fundamental basis of the effective operation of the market economy. If this system collapses, a small part of people will benefit a little from it temporarily. However, if lack legal sanctions and public condemnation, it may make some impacts on other economic entities, so that more companies can benefit from the collapse of the credit system, then, finally, the government's regulatory system and the rules of the market become useless, which will lead to an immeasurable loss.

3. The Difficulties of Implementation of Emissions Trading System in China

3.1 A Vast Territory

A vast territory, the ecological environment and the level of economic development have shown significant regional characteristics, the national conditions in China result in the complexity of emissions trading.

3.1.1 Regional Characteristics of the Environment

Regional characteristics of the environment are quite significant. Ecologically speaking, human activities are

affected by geographical location, atmospheric circulation, landscape, China can be divided into three nature zone macroscopically. The existence of three nature zones decide the diversities of the climate resources, wild flora and fauna resources, and agriculture materials, so that it further decide the nature of environmental capacity, amount and differences of protection. Due to the unique characteristics of Chinese ecological environment, the regional pollutants needed to be control is quite different, and there is also some differences in pollutants control standards, the subject of emissions trading, measurements and supervision of the actual emissions. Thus, it is quite difficult for China to implement the emission trading system comprehensively.

3.1.2 The Clear Differences in the Level of Economy

The clear differences in the level of economy. Economically speaking, due to the differences between the nature resources and opportunities of the historical development, China has formed the economic and technological development level of the gradient distribution of the three major economic areas in central and western. The existence of the three economic zones reflect the diversities of the market economic development and people's income as well as the diverse demands of economic means in the future and the level of environmental quality. In economically developed areas of eastern and central areas, with the improvements of people's requirements to the environmental quality, there is a trend that the high-polluted enterprises are becoming less, this reflects the implementation of the "total control" is more easily. However, in the western areas, the companies with polluted emission account for a large proportion of the economic development and the employments, the promotion of emission trading system is harder.

3.2 A Balance between the Pollutants Control and Further Development of the Economy

In the unchanged situation of the traditionally extensive economic development means, how to find a balance between the pollutants control and further development of the economy is another difficult in establishing emission trading system.

The population of China, trend of the economy, the level of the technology and management are the three basic factors for damaging the environment. According to the forecast of related departments, among those three reasons, the zero increase in China's future population will need 30 to 40 years, the rapid increase in the future economy will continue its trend in the next 10 to 20 years, the level of technology and management still need a long time, so that it is impossible for China's economic growth means transferring from extensive one to intensive one in a short term [3–7]. The existence of the extensive growth will give an obvious impact on the emission trading policy with the goal of "total control".

3.3 China's Small Sized Companies

The number of China's small sized companies is great, and the spread of them is wild. However, the pollutants they produced account for half of the total pollutants. The basic national situation decide that on one hand, the cost of information collection is quite expensive in China's emission trading market, the cost of supervision and implementation of the environmental protection department is also expensive, so that it decreases its efficiency and results in market failure. On the other hand, if it is quite common for companies with heavy pollutants can't effectively deal with it in a short time because of the limitation of fund and technology, it will do harm on the practice of this system.

3.4 The Goal of the Solution of Environment Pollution

There is a difference between the goal of local government and central government in the solution of environment pollution, and those differences will add some difficulties in the promotion of the emission trading system. Pollutants can't be eliminated totally, so any companies will produce pollutants during its production. The local government will focus on the economy, including the development of local economy, the unemployment and the GDP which is a measure of the political performance, rather than the impact on the environment. However, by the "total control", it will definitely make more difficulties in the promotion of emission trading system. According to the report from the Xinhua news agency, on 29th of May 2007, the explorations of Cyan bacteria caused shortage of water which make the habitants' lives becoming more difficult. More ironically, the "iron solution" caused hundreds enterprises with high consumption, high pollution, high danger and low efficiency shut down. Unexpectedly, this news attracted lots of organizations to make investment, including the towns nearby in Subei, as well as eastern-west, Sichuan and Jiangxi. The essence of this phenomenon reflects that in the eyes of some local governments in China, the basic policy still remains the pursuit of increase of GDP, lacking of recognition of the limitation of resources, and haven't pay attention on the environmental protection on mind, so that it continuously creates hidden obstacles to the emission trading system.

3.5 The Power of Deciding the Total Emission Controlled

Due to the characteristics of the China's ecological and economic zones, the power of deciding the total emission controlled by the central government or the local government still play a significant role in the effectiveness of the emission trading system. If the power belongs to the local government, the environment can't take the total emission finally due to the differences between the goals

of central government and local government; if it belongs to the central government, it will have a fierce competition with the local government.

3.6 The Market Participates

It is a tough question for the market participates to reach the goal of guaranteeing the vitality of the market, as well as upgrade the format of production by emission trading system. Therefore, how to choose formats, such as: involving in this program compulsively, volunteer, opening comprehensively and directly, etc, is another problem facing China.

4. Discussion and Conclusions

4.1 The Path of Emission Trading System in China

The emission trading system derived from the United States, widely used in the western countries, and some part of China. As the most effective way of treating emission, China will put this system in a general manner in future. So, in essence, it is a problem of transplant system. According to the experience of developed countries, and combination of the social situation of China, the means of establishing emission trading system can be showed in the Figure 1.

4.2 The Steps and Specific Measures of Establishing Emission Trading System in China

4.2.1 Legal Framework

Firstly to confirm the regulation of environmental resource priority system by the NPC to better the laws and to embody the “overall control” of the intension of environmental protection and prevention with a view to establishing such system on China’s environment in legal framework.. there is no specific rules of “total control” in the laws and legislations in China, the shortage can de-

cide the fragility of emission trading system, and difficulties in the emission control, the extensive growth of the economy hasn’t changed. Therefore, the first step of establishing emission trading system in China is to confirm the environmental resource system that it can embody the intension of “total control”.

4.2.2 The Initiative Distribution

In the level of a state, it confirms the initiative distribution of the total control of emission. Because China is a country with wide areas, distinguished ecological environment, the huge economic gap among different areas, and the different goals of local government and central government, that means local government have the trend of focusing the speed of economic growth but ignoring the environmental protection. If the power of confirming the total emission belongs to the central government, it will reduce the welfare of the society, so the power should belongs to the central government. one reasonable choice for the national overall emission of pollutants as well as for provincial and municipal levels is that the central government confirms national emission of certain pollutants in a given period of time in accordance with domestic development, and that local governments enforce local standards in accordance with relevant principles; that the central government confirms the overall emissions in different regions and regulates them with respective standards in accordance with provincial and municipal economic and social development and in the principle of keeping the balance between national demand and local features.

The initiative distribution of every single province and area. Under the guidance of “efficiency combine with justice”, and in the background of China’s economic and social situation, the initiative distribution of emission trades need to combine with two modes: distribute for free, and distribute for charge. The key of this mode is: in the given amount of emission, how to confirm the proportion of free distribution and charged distribution, and

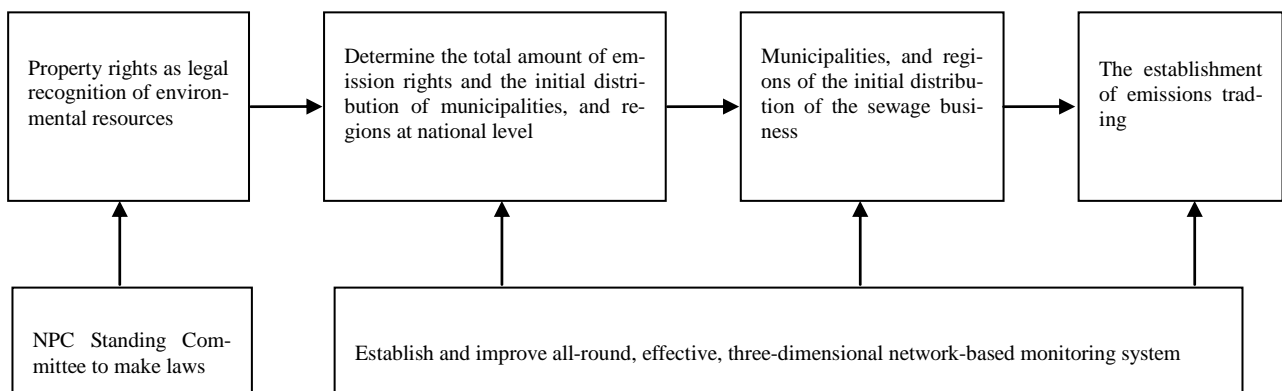


Figure 1. The path of emission trading system in china

how to distribute it to different corporations in accordance with standard allocation after the confirmation of the total amount. Free distribution of the sewage is the main reason for the contribution of enterprises and the employment created by a net increase of new value-added, therefore, the ratio of free distribution of the main considerations are the impact of these two indicators. Environmental protection department of the provinces and municipalities can trade the remaining parts with sewage companies through auctions, bidding and pricing, after free allocation to sewage companies. And local governments can preserve part of the emission permits in accordance with their own needs, in an attempt to provide support for the future companies with high-efficiency.

4.2.3 The Secondary Market

The establishment of effective transfer of ownership of emission rights in the secondary market. Secondary market for emissions trading as a supplier of emission rights and the right to demand of sewage through the use of tools for trading emission rights trading and financing market of production factors, in a sense, to cultivate, develop and standardize the emission trading market, is more important than the well-defined administrative powers sewage.

4.3 The Relevant Recommendations of China's Emissions Trading System

4.3.1 Environmental Protection Consciousness

Cultivate national awareness of environmental protection, strengthening the right of the sewage system of external oversight. National environmental awareness as non-governmental social forces, which play an important role in the oversight, support and promote of the formulation and implementation of environmental policies. Therefore, we must cultivate national awareness of environmental protection, strengthening the right of the sewage system of external oversight.

4.3.2 Integrity of Social Environment

Establishing integrity of social environment and enhancing the development and implementation of emissions trading system. Establishing trust system is the public goods provided by the government. Firstly, the government should maintain absolute authority of national credit as well as break local protectionism. After emissions trading system is in place, the government should strongly give a guidance and expectation in accordance with the amount of sewage, time, space, and trading rules. Instead of change so fast that companies feel that this is a short-term policy, passively wait and see. At the same time, establish and improve the credit system and the relevant legal system to ensure that after emissions trading, the normal performance of the contract. Through the establishment of integrity of society, guide these businesses to abide by the laws, and can discharge the sewage

age consciously.

4.3.3 Relevant Laws and Provide Adequate Legal Basis

Establish and improve relevant laws and provide adequate legal basis for the emission trading system. With the legal basis, the national environmental administration departments should be issued the specific rules and regulations relating to emissions trading, Emissions trading a clear approval to the transaction from the uniform standard, in order to facilitate the actual operation of China's emissions trading. Make sure the uniform standard from approval to trade, in order to facilitate the actual operation of China's emissions trading.

4.3.4 The Power of Rent-Seeking

Strengthen the initial allocation of emission rights in the management of the power of rent-seeking. Initial allocation of emission rights for the rental of power - the mainly two causes of the phenomenon of rent-seeking is, on the one hand, "soft power," on the other hand is the space rent-seeking in the design of allocation mechanism. One of the performances of the previous reason for government officials to determine the total amount of sewage, to control the randomness in the initial distribution in municipalities. Therefore, it is necessary to establish a complete set of evaluation systems, to define the total target scientifically and reasonably, instead of free decision by the officials. Enterprises in all the initial emission rights on the specific share, in strict accordance with the principle of "open, fair, just and transparent", and the method of calculation in advance. Finally, to further improve the legal system, strengthen supervision, to establish a high degree of integrity and efficiency as the hallmark of the political and administrative system. From the second reason which has caused rent-seeking --- power-renting, in the model of free allocation of emission rights, government officials have greater rights of discretion, which has left a space for rent-seeking. Therefore, with a large scale of emission trading system, the initial model of allocation of emission rights, transition to paid transfer should be gradually put into practice. As result --- relatively speaking --- rent-seeking scales of auctions, bidding, pricing and trading can be reduced in system -designing. Thus more rent-seeking activities can be regulated.

4.3.5 Take Proactive Measures to Tackle the Problem of the Market failure

Take proactive measures to tackle the problem of the failure on emissions trading market. With a view to preventing the inefficiency on emissions trading market, greater-forced policies ought to be put in place, in order to heavily punish opportunism activities such as exceeding and hiding emissions; regulate monopolizing activities; create a fully competitive market; for those companies who actively reduce the overall emissions and sell credits on reduction of emissions, the government should

support them in tax, technology and capitals and guarantee them with price priority, in an attempt to activate the market; confirm the deadlines on emission rights to reduce the uncertainty of emission rights; break barriers on emission trading market, and take actions to eliminate local protectionism.

4.3.6 Enhance the Construction of Supervision

Enhance the construction of supervision on the emission permits market and the legal force on environmental protection. In the construction of supervision on the emission permits market, we should establish the “three systems” in pollution reduction; establish effective, overall and comprehensive supervision network to create a new type of supervision model; promote managerial, scientific and technological standards, in order to create preferential conditions for the emission permits trade system.

4.3.7 Reform the Counting System of Environmental Costs

Reform the counting system of environmental costs. The newly published Counting Principles requests that companies should include costs such as disposal of pollution, in an attempt to make themselves accountable, thus to inspire other companies to reduce pollution, and thus to have them strike a balance between profit making and environmental protection and take part in the emission permits trade system.

4.3.8 Put Green GDP into Practice

Better local performance checking index system; put Green GDP into practice; and to those areas which have a weak eco-system, the government should allow zero growth in local GDP in a given period of time.

REFERENCES

- [1] Zhang Yue, “Jiaxing, Zhejiang, the first test of emissions trading, the initial allocation of fair,” China Environment News,” November 13, 2007.
- [2] http://paper.people.com.cn/scb/html/2007-12/28/content_35730363.htm.
- [3] http://www.cas.ac.cn/html/Books/O61BG/c1/2002/1/5/1.5_3.htm.
- [4] S. D. Li, “A research on china’s establishment of an emission permits trading system,” China Soft Science Magazine, Vol. 8, pp. 19–23, 2002.
- [5] J. Xu and W. W. Wan, “Designing of trading cost and emission permits trading system,” China Soft Science Magazine, Vol. 7, pp. 38–41, 2002.
- [6] D. H. Chen and F. Jiang, “Theory and practice on china’s emission permits,” Soft Science Magazine, Vol. 2, pp. 12–15, 2004.
- [7] D. Pierce, “There is no end of the world: Economics, environment and sustainable development,” Beijing, China Finance and Economy Press, pp. 116–118, 1996.

The Research on the Appraisal of Market Opportunity Based on AHP

Donghong YANG¹, Yujie TAN¹, Yanbin SUN¹

¹School of Economics and Management, Daqing Petroleum Institute, Daqing, China.
Email: dqyangdonghong@126.com

Received April 21, 2009; revised June 25, 2009; accepted August 2, 2009.

ABSTRACT

In this article, through analyzing the inner and external environments of the manufacture enterprise, the hierarchy model of the market opportunity has been constructed. Based on the AHP, the mathematics evaluation model of the market opportunity has been established. Take a simulative enterprise as example, according to the hierarchy model and the mathematics evaluation model of the market opportunity, the judgment matrixes of every hierarchy have been established and the weight of each market opportunity opposite to the target hierarchy is obtained. The reasonable decision-making method has been provided to choose the market opportunity for the manufacture enterprise.

Keywords: Analysis Hierarchy Process, Market Opportunity, Environment Analysis

1. Introduction

With the change and development of the market demand and the intense competition environment in the market, any manufacture enterprise can not always rely on the existing products and services. If the manufacture enterprise wants to survive and develop for a long time, it must look for new market opportunities.

In the internal and external enterprise environment, market opportunities are situation and opportunities which can make the manufacture enterprise get tremendous benefit and develop rapidly. Market opportunities existing in the internal and external environment are the customer demand which have not yet been met or fully met [1].

Market opportunities generally exist in two situations. First, to find the new or potential customers for the existing products and services and to start and develop a new market, or to find the new features and new uses of existing products and guide people to use them. Second, to create, develop, design and produce new products or new services which have new features to meet the changing needs of people.

Along with the changes of market environment, demand also will be changed. Therefore, market opportunities will emerge or disappear. Any market opportunity objectively exists and any enterprise can find and take advantage of the market opportunities. There are no exclusive rights [2].

However, just in the specific conditions one market opportunity can success. Because of the different condi-

tions of the manufacture enterprises, in the use of a market opportunity, the different manufacture enterprises will gain different interests and competitive advantages. Therefore, manufacture enterprises must accurately analyze and evaluate the market opportunities to find one that is suitable for the manufacture enterprise's development.

2. The Appraisal of Market Opportunity

2.1 The Appraisal Process of Market Opportunity

Market opportunities come from customer demand, and customer demand changes with the internal and external environment of the manufacture enterprise. When looking for market opportunities, the manufacture enterprises analyze the enterprise's suppliers, customers and competitors, as well as the internal organizational structure, financial status, professional skills, human resources and network conditions [3]. The optimal market opportunity can be identified after analysis and evaluation. (See Figure 1)

2.2 The Hierarchy Model of Market Opportunity

The market gravitation, the enterprise resource, the competitors and the customers are considered to analyze and evaluate a particular market opportunity or some market opportunities which is whether conform to manufacture enterprise development needs or not.

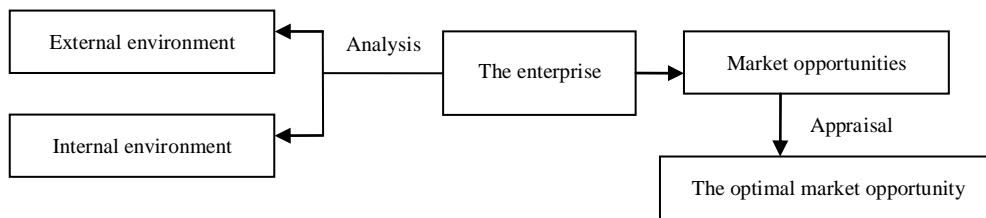


Figure 1. The appraisal process of market opportunity

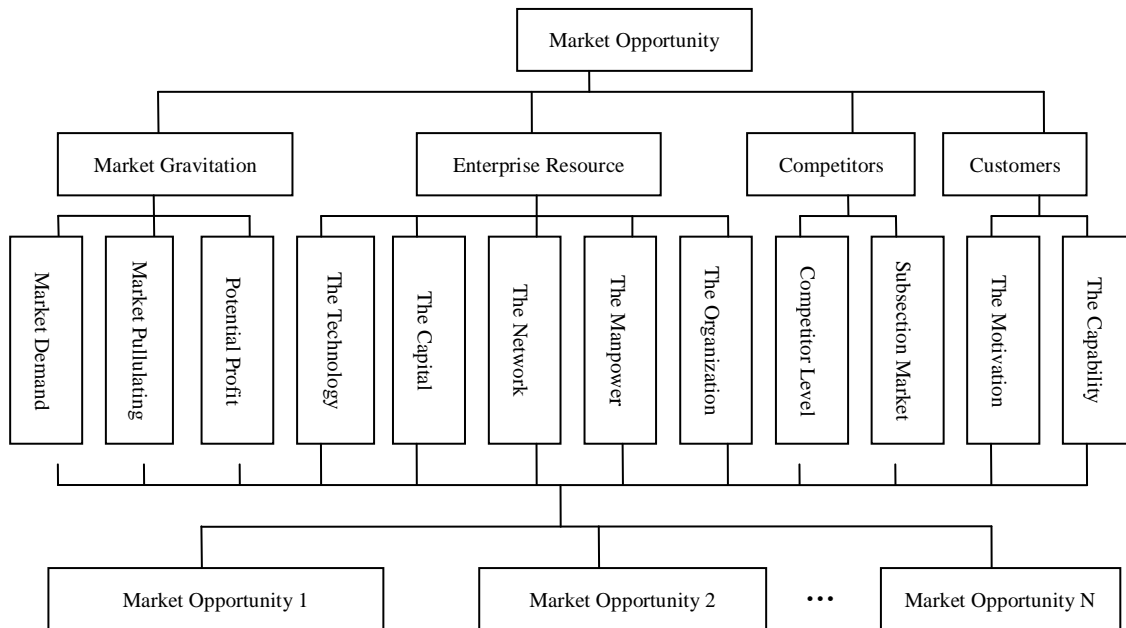


Figure 2. The hierarchy model of the market opportunity

The market gravitation refers to market attract ability to specific products which are produced by the enterprise. The market gravitation includes market demand, market pullulating and potential profit. The more the market demand of products and services, the market more attract the products and services of enterprise. The market profitability of products and services is high, the capital profitability is high, which attracts enterprises to product the products and services.

The enterprise resource is the useful thing which is worth for the survival and development of enterprise. The enterprise resource is an important foundation conditions for production and operation of enterprise. The enterprise resources include the technology, the capital, the network, the manpower and the organization. When the market can effectively and rationally use resources of enterprise, in accordance with the advantages of enterprise resources into products and management, and get benefit through sale, the market opportunity is fit for the survival and development of enterprise.

In a certain degree of market competition environment, when the enterprises analyze and evaluate the market

opportunities, the competitor level and subsection market should be considered.

The products and services will be sold to the customers. So when the enterprises analyze and evaluate the market opportunities, the consumption motivation and the consumption capability should be considered.

The hierarchy model of the market opportunity is established as follow. (See Figure 2)

2.3 The Mathematics Evaluation Model of Market Opportunity

2.3.1 Constructing Judgment Matrixes

Through paired comparing the factors, the judgment matrix $A = (a_{ij})_{n \times n}$ is constructed. In determining the value of a_{ij} , the 1-9 and their reciprocals are quoted [4].

2.3.2 Single-Level Ranking and Consistency Test

1) Single-level ranking

The eigenvector W normalized corresponding to the judgment matrix is the Sort weights of the factors corresponding to the factor on the above hierarchy. When

calculating the maximum eigenvalue λ_{\max} and W , the method is as follows [5].

(a) Each list of matrix elements is standardized. The general element is

$$\bar{a}_{ij} = \frac{a_{ij}}{\sum_{k=1}^n a_{kj}} \quad (i, j = 1, 2, \dots, n) \quad (1)$$

(b) The summation of elements in each row of the judgment matrix standardized is gained.

$$\bar{W}_i = \sum_{j=1}^n \bar{a}_{ij} \quad (i = 1, 2, \dots, n) \quad (2)$$

(c) The vector $\bar{W} = (\bar{W}_1, \bar{W}_2, \dots, \bar{W}_n)^T$ is standardized.

$$W_i = \frac{\bar{W}_i}{\sum_{i=1}^n \bar{W}_i} \quad (i = 1, 2, \dots, n) \quad (3)$$

The vector $W = (W_1, W_2, W_3, \dots, W_n)^T$ is the Single-level sort weights, and the maximum eigenvalue is

$$\lambda_{\max} = \sum_{i=1}^n \frac{(AW)_i}{nW_i} \quad (4)$$

2) Consistency test

(a) The consistency target is

$$CI = \frac{\lambda_{\max} - n}{n - 1} \quad (5)$$

(b) Find the corresponding average random consistency target RI which is shown in Table 1.

(c) The consistency proportion is

$$CR = \frac{CI}{RI} \quad (6)$$

When $CR < 0.10$, the consistency of judgment matrix is acceptable, otherwise, the judgment matrix is modified, and then its consistency is tested.

2.3.3 Sort of Level Ranking and Consistency Test

1) Sort of level ranking

The hierarchical model includes target layer A , guideline B and project C . The relative weights of target layer A corresponding to guidelines B is $W^{(1)} = (W_1^{(1)}, W_2^{(1)}, W_3^{(1)}, \dots, W_k^{(1)})^T$

The relative weights of the various criteria of guideline B corresponding to project C is $W_j^{(2)} = (W_{1j}^{(2)}, W_{2j}^{(2)}, W_{3j}^{(2)}, \dots, W_{nj}^{(2)})^T$ ($j = 1, 2, \dots, k$).

The relative weights of projects corresponding to target layer A is calculated in the Table 2.

2) Consistency test

$$CR = \frac{\sum_{j=1}^m CI_j a_j}{\sum_{j=1}^m RI_j a_j} \quad (7)$$

When $CR < 0.10$, the consistency of judgment matrix is acceptable, otherwise, the judgment matrix is modified, and then its consistency is tested.

The vector $W = (W_1, W_2, W_3, \dots, W_n)^T$ is the relative weights of the projects. The policy-makers choose the project whose weight is the greatest as the optimal project.

3. Empirical Analysis

The market opportunities are identified by application of the model. The simulative manufacture enterprise is now facing three chances to choose, to be known as A, B and

Table 1. The average random consistency target of judgment matrix [6]

Rank	1	2	3	4	5	6	7	8	9
RI	0.00	0.00	0.58	0.90	1.12	1.24	1.32	1.41	1.45

Table 2. Sort of level ranking

H.C	H. B	B1	B2	...	Bk	Sort of level Ranking W
		W1	W2	...	Wk	
	C1	W11	W12	...	W1k	
	C2	W21	W22	...	W2k	
	⋮	⋮	⋮	...	⋮	
	Cn	Wn1	Wn2	...	Wnk	

$$W_i = \sum_{j=1}^k W_j^{(1)} * W_{ij}^{(2)} \quad (i = 1, 2, \dots, n)$$

C. The basic steps of identifying and evaluating the market opportunities are as follows.

First, the hierarchy model of the market opportunity is drawn. (See Figure 2)

Then, the judgment matrixes are constructed. (See Table 3 to Table 19)

Table 3. The judgment matrix of A-B

A	B ₁	B ₂	B ₃	B ₄	W
B ₁	1	1/3	3	2	0.246
B ₂	3	1	4	3	0.497
B ₃	1/3	1/4	1	2	0.142
B ₄	1/2	1/3	1/2	1	0.115

$$\lambda_{\max} = \sum_{i=1}^n \frac{(AW)_i}{nWi} = 4.215, \quad CI = \frac{\lambda_{\max} - n}{n-1} = 0.072, \quad CR = \frac{CI}{RI} = 0.080 < 0.1, \text{ the consistency is satisfactory.}$$

Table 4. The judgment matrix of B₁-C

B ₁	C ₁₁	C ₁₂	C ₁₃	W
C ₁₁	1	1/3	1/4	0.120
C ₁₂	3	1	1/3	0.272
C ₁₃	4	3	1	0.608

$$\lambda_{\max} = 3.074, \quad CI = 0.037, \quad CR = 0.064 < 0.1, \text{ the consistency is satisfactory.}$$

Table 5. The judgment matrix of B₂-C

B ₂	C ₂₁	C ₂₂	C ₂₃	C ₂₄	C ₂₅	W
C ₂₁	1	1/2	3	4	5	0.301
C ₂₂	2	1	3	4	5	0.397
C ₂₃	1/3	1/3	1	3	2	0.146
C ₂₄	1/4	1/4	1/3	1	3	0.097
C ₂₅	1/5	1/5	1/2	1/3	1	0.059

$$\lambda_{\max} = 5.280, \quad CI = 0.070, \quad CR = 0.062 < 0.1, \text{ the consistency is satisfactory.}$$

Table 6. The judgment matrix of B₃-C

B ₃	C ₃₁	C ₃₂	W
C ₃₁	1	1/2	0.333
C ₃₂	2	1	0.667

$$\lambda_{\max} = 2, \quad CI = 0, \quad CR = 0 < 0.1, \text{ the consistency is satisfactory.}$$

Table 7. The judgment matrix of B₄-C

B ₄	C ₄₁	C ₄₂	W
C ₄₁	1	1/3	0.250
C ₄₂	3	1	0.750

$$\lambda_{\max} = 2, \quad CI = 0, \quad CR = 0 < 0.1, \text{ the consistency is satisfactory.}$$

Table 8. The judgment matrix of C₁₁-D

C ₁₁	D ₁	D ₂	D ₃	W
D ₁	1	2	3	0.539
D ₂	1/2	1	2	0.297
D ₃	1/3	1/2	1	0.164

$$\lambda_{\max} = 3.009, \quad CI = 0.005, \quad CR = 0.008 < 0.1, \text{ the consistency is satisfactory.}$$

Table 9. The judgment matrix of C₁₂-D

C ₁₂	D ₁	D ₂	D ₃	W
D ₁	1	1/3	1/4	0.123
D ₂	3	1	1/2	0.320
D ₃	4	2	1	0.557

$$\lambda_{\max} = 3.018, \quad CI = 0.009, \quad CR = 0.016 < 0.1, \text{ the consistency is satisfactory.}$$

Table 10. The judgment matrix of C₁₃-D

C ₁₃	D ₁	D ₂	D ₃	W
D ₁	1	1/3	1/3	0.141
D ₂	3	1	2	0.525
D ₃	3	1/2	1	0.334

$$\lambda_{\max} = 3.054, \quad CI = 0.027, \quad CR = 0.046 < 0.1, \text{ the consistency is satisfactory.}$$

Table 11. The judgment matrix of C₂₁-D

C ₂₁	D ₁	D ₂	D ₃	W
D ₁	1	1/3	1/2	0.164
D ₂	3	1	2	0.539
D ₃	2	1/2	1	0.297

$$\lambda_{\max} = 3.009, \quad CI = 0.005, \quad CR = 0.008 < 0.1, \text{ the consistency is satisfactory.}$$

Table 12. The judgment matrix of C₂₂-D

C ₂₂	D ₁	D ₂	D ₃	W
D ₁	1	1/4	1/3	0.120
D ₂	4	1	3	0.608
D ₃	3	1/3	1	0.272

$$\lambda_{\max} = 3.074, \quad CI = 0.037, \quad CR = 0.064 < 0.1, \text{ the consistency is satisfactory.}$$

Table 13. The judgment matrix of C₂₃-D

C ₂₃	D ₁	D ₂	D ₃	W
D ₁	1	1/3	2	0.252
D ₂	3	1	3	0.589
D ₃	1/2	1/3	1	0.159

$$\lambda_{\max} = 3.054, \quad CI = 0.027, \quad CR = 0.046 < 0.1, \text{ the consistency is satisfactory.}$$

Table 14. The judgment matrix of C_{24} -D

C_{24}	D_1	D_2	D_3	W
D_1	1	3	2	0.525
D_2	1/3	1	1/3	0.142
D_3	1/2	3	1	0.334

$\lambda_{\max} = 3.054$, $CI = 0.027$, $CR = 0.046 < 0.1$, the consistency is satisfactory.

Table 15. The judgment matrix of C_{25} -D

C_{25}	D_1	D_2	D_3	W
D_1	1	1/3	1/2	0.159
D_2	3	1	3	0.589
D_3	2	1/3	1	0.252

$\lambda_{\max} = 3.054$, $CI = 0.027$, $CR = 0.046 < 0.1$, the consistency is satisfactory.

Table 16. The judgment matrix of C_{31} -D

C_{31}	D_1	D_2	D_3	W
D_1	1	1/5	1/3	0.106
D_2	5	1	3	0.633
D_3	3	1/3	1	0.261

$\lambda_{\max} = 3.039$, $CI = 0.019$, $CR = 0.033 < 0.1$, the consistency is satisfactory.

Table 17. The judgment matrix of C_{32} -D

C_{32}	D_1	D_2	D_3	W
D_1	1	4	3	0.620
D_2	1/4	1	2	0.224
D_3	1/3	1/2	1	0.156

$\lambda_{\max} = 3.109$, $CI = 0.055$, $CR = 0.094 < 0.1$, the consistency is satisfactory.

Table 18. The judgment matrix of C_{41} -D

C_{41}	D_1	D_2	D_3	W
D_1	1	4	5	0.665
D_2	1/4	1	3	0.231
D_3	1/5	1/3	1	0.104

$\lambda_{\max} = 3.087$, $CI = 0.043$, $CR = 0.075 < 0.1$, the consistency is satisfactory.

Table 19. The judgment matrix of C_{42} -D

C_{42}	D_1	D_2	D_3	W
D_1	1	4	3	0.620
D_2	1/4	1	2	0.224
D_3	1/3	1/2	1	0.156

$\lambda_{\max} = 3.109$, $CI = 0.055$, $CR = 0.094 < 0.1$, the consistency is satisfactory.

Table 20. Sort of level ranking

H. B	B1				B2				B3		B4		W
	0.246				0.497				0.142		0.115		
H. D	C ₁₁ 0.120	C ₁₂ 0.272	C ₁₃ 0.608	C ₂₁ 0.301	C ₂₂ 0.397	C ₂₃ 0.146	C ₂₄ 0.097	C ₂₅ 0.059	C ₃₁ 0.333	C ₃₂ 0.667	C ₄₁ 0.250	C ₄₂ 0.750	
D ₁	0.539	0.123	0.141	0.164	0.120	0.252	0.525	0.159	0.106	0.620	0.665	0.620	0.278
D ₂	0.297	0.320	0.525	0.539	0.608	0.589	0.142	0.589	0.633	0.224	0.231	0.224	0.453
D ₃	0.164	0.557	0.334	0.297	0.272	0.159	0.334	0.252	0.261	0.156	0.104	0.156	0.269

The single-level ranking is done. The consistency of all the judgment matrixes is satisfactory. (See Table 20)

The relative weights of projects corresponding to target layer A is calculated in the Table 6. The corresponding average random consistency proportion of hierarchy B is $CR = 0.063 < 0.10$, the consistency is satisfactory.

According to the comprehensive analysis of market opportunities, the ranking of the three market opportunities is Market Opportunity2, Market Opportunity1, Market Opportunity3.

4. The Significance of Market Opportunity Evaluation

Through analyzing and evaluating market opportunities, the manufacture enterprise can expand reproduction to promote the enterprise to a higher stage. The manufacture enterprise can maintain the normal production and operation, but also can seeks to growth and progress.

The survival and development of manufacture enterprises can not be separated from environment. From the macro-environment, the survival and development of

manufacture enterprises is restricted mainly by the changes of the demographic, economic, political and legal, scientific and technological, natural, social and cultural environment. From the micro-environment, the survival and development of manufacture enterprises is closely influenced and affected by their own internal factors, place affiliated companies, customers, competitors, the public and so on. To adapt to the environment is the most basic issues of manufacture enterprises survival. Only concerning any change from the environment at any time, finding out the problems timely, evaluating the favorable market opportunities timely and avoiding the risk, the manufacture enterprise can get better living space to enhance the capacity of adapt to the environment.

The development assurances can not be separable from the relatively stable profits and revenue. Through continuously analyzing and evaluating market opportunities, the manufacture enterprise can excavate new growth point to demand the products which can fully meet needs and desires for market. Combined with reasonable prices, convenient place and appropriate promotion, the manufacture enterprise will be able to expand revenues, en-

hance its effectiveness, so as to stabilize the profitability of manufacture enterprises.

REFERENCES

- [1] L. Q. Meng, "Distinguishing and evaluating market opportunity of enterprise," *Journal of Lianyungang College of Chemical Technology*, Vol. 3, No. 2, pp. 57, 1997.
- [2] J. C. Li, "The study and analysis of national market chance of printing enterprise," *Wuhan University*, China, 2005.
- [3] J. Gao and M. Zhang, "Marketing skill for enterprise to recognize market opportunities," *Northern Economy and Trade*, Vol. 12, pp. 109, 2001.
- [4] J. Tang and J. Tian, "Application of AHP in the evaluation system on teaching laboratories of fundamental courses," *Research and Exploration in Laboratory*, Vol. 25, No. 3, pp. 282, 2006.
- [5] Q. F. Geng, "Research on the selection of the supply chain cooperators based on the AHP-Fuzzy evaluation," *Value Engineering*, Vol. 10, pp. 39, 2005.
- [6] H. Q. Zhang and Z. Q. Li, "The application of AHP to network plan evaluation," *Construction and Design for Project*, Vol. 10, pp. 84, 2004.

Dalian High-Tech SMEs Growth Evaluation Based on Catastrophe and Principal Component Projection Method

Lin LI¹, Pengfei ZHOU², Zhenghe LI³

¹School of Management, Dalian Jiaotong University, Dalian, China; ²School of Hydraulic Engineering, Dalian University of Technology, Dalian, China; ³Headmaster Office, Dalian Jiaotong University, Dalian, China.
Email: linli@djtu.edu.cn, pfzhou@dlut.edu.cn, lzh@djtu.edu.cn

Received July 21, 2009; revised September 9, 2009; accepted October 16, 2009.

ABSTRACT

In the course of rapid economic development, high-tech small and medium enterprises (SMEs) are gradually playing an important role, which become important support to regional economic growth and science and technology development. So SMEs growth becomes a universal problem. And how to evaluate the SMEs growth becomes an important step, especially to high-tech SMEs growth. In this paper, catastrophe theory and improved principal components projection method are used and a mutation series of high-tech SMEs growth evaluation index system is built. Taking Dalian high-tech SMEs as an example, high-tech SMEs growth is evaluated, which contributes to high-tech SMEs growth forecast and government to formulate policies to support high-tech SMEs development.

Keywords: Catastrophe Theory, Improved Principal Components projection Method, Enterprises Growth, High-tech Small and Medium Enterprises (SMEs), Principal Component Projection Method

1. Introduction

In recent years, with the development of high technology, enormous high-tech SMEs are build, but in which fast growth small and medium high-tech enterprises only account for about 5% [1]. Though the number is low, yet they make a large contribution to job enlargement, wealth increase, new start industry promotion, etc. What factor influences the growth of high-tech SMEs? That becomes a problem concerned by industrialist, governor, policymaker, banker (especially venture investor) and scholar together. About 20 or 30 years ago, many scholars have discussed this problem from different viewpoints. These researches help understand the high-tech SMEs growth process. But now the research achievement can't be used to design a model, which can preview the growth potential of high-tech SMEs. In addition, like hypotheses and methods used to different national small and medium-sized enterprises, sometimes unlike empirical analysis outcome is got. The internationalization process of high-tech SMEs is often different from that of more mature industries [2].

Although there is no single agreed definition of high-tech SMEs, there are generally characterized by

high-tech SMEs with advanced knowledge and capabilities in technology, and educated workforce, and the ability to adapt quickly to fast changing environments. These characteristics facilitate the internationalization of high-tech SMEs, which have been known to act quickly when windows of opportunity in foreign markets present themselves [3–7].

In China there are many vibrant SMEs, especially the representation of high-tech SMEs, whose survival are the basic form of expression in the contemporary conditions of socialization of production and specialization, and is an important part of the modern collaboration division of labor system. They have become the most active elements to China's economy, which contribute much to China's economic development, technology innovation, and so on.

For the strength, stability and development of the anti-risk capability, the current situation on the high-tech SMEs growth is worrisome. High-tech SMEs growth is fraught with difficulties and hardships, which make the growth become a universal problem [8]. How to evaluate the high-tech SMEs growth is very important. At present, domestic and foreign scholars focus on high-tech SMEs growth study from two aspects [9–10]. That is, how to

establish indicators and which methods are used to evaluate. In this paper, catastrophe theory and improved principal components projection method are used and a mutation series of high-tech SMEs growth evaluation index system is built. Taking Dalian high-tech SMEs as an example, high-tech SMEs growth is evaluated, which contributes to their growth forecast and government to formulate policies to support SMEs developments.

2. High-Tech SMEs Evaluation Based on Combination of Catastrophe Theory and Principal Component Projection Method

2.1 High-Tech SMEs Evaluation Based on Catastrophe Theory

Catastrophe theory is a mutation research (qualitative) founded by French mathematician Majorelle Thom, which is about the characteristics of system variables to control the variables following the mathematical theory and taken as 'a revolution after mathematical calculus [11]. Catastrophe theory is actually a multi-dimensional fuzzy membership function.

To evaluate the complexity and abstract goals, the systematic goals of multi-level contradictions are decomposed. And combining of fuzzy math with the Catastrophe theory, the fuzzy membership function mutation is got.

Using normalization formula to a comprehensive quantitative computing, the final as a parameter is got [12]. The main advantage of the method is not used for indicators of weight. However, it is considered the relative importance of various evaluation indicators, and qualitative integrates and quantitative analysis, which reduce the subjectivity and yet scientific, and is a reasonable, simple and accurate calculation. So Catastrophe theory is a worthwhile evaluation method. To a dynamic system, the influence function of Catastrophe system is $f(x)$. According to Catastrophe theory, all of its critical point congregates into balance-surface. Through the equation to get the first derivative zero That is, $f'(x)=0$, and the singular point can be got through solving $f(x)$ the second derivative, and got $f''(x)=0$. $f'(x)=0$ and $f''(x)=0$ eliminate x , the differences point set equation of Catastrophe can be got. Differences point set equation shows that when all control variables meet this equation, the system will mutate. Through the differences point set equation that in decomposition form, a formula can be got. By a formula, hanging different variables with different states are quality into the same state quality.

At present, the total Catastrophe system types is seven, the most common are three. That is, cusp catastrophe system, dovetail catastrophe system and butterfly catastrophe system

The model of cusp catastrophe system is:

$$f(x) = x^4 + ax^2 + bx$$

The model of dovetail catastrophe system is:

$$f(x) = \frac{1}{5}x^5 + \frac{1}{3}ax^3 + \frac{1}{2}bx^2 + cx$$

The model of butterfly catastrophe system is:

$$f(x) = \frac{1}{6}x^6 + \frac{1}{4}ax^4 + \frac{1}{3}bx^3 + \frac{1}{2}cx^2 + dx$$

$f(x)$ means a influence function of a state variable x in a system. And the coefficient of x is a, b, c, d , which represents the control variables of the state variables. If an indicator only is decomposed into two sub-indexes, the system can be seen cusp catastrophe system. If an indicator is decomposed into three sub-indexes, the system can be seen as dovetail catastrophe system. And if an indicator is only decomposed into four sub-indexes, the system can be seen butterfly catastrophe system [13].

According to catastrophe theory, in a cusp catastrophe system, the differences point set equation of $f'(x)=0$ and $f''(x)=0$ can be got $\{a=-6x^2, b=8x^3\}$. Then the catastrophe fuzzy membership function formula can be gotten.

$$x_a = a^{\frac{1}{2}}, x_b = b^{\frac{1}{3}}.$$

In the formula, x_a means x corresponding to a , x_b means x corresponding to b . Similarly, the formula of dovetail catastrophe system is:

$$x_a = a^{\frac{1}{2}}, x_b = b^{\frac{1}{3}}, x_c = c^{\frac{1}{4}}$$

The butterfly catastrophe system is:

$$x_a = a^{\frac{1}{2}}, x_b = b^{\frac{1}{3}}, x_c = c^{\frac{1}{4}}, x_d = d^{\frac{1}{5}}$$

In essence, it is a multi-dimensional fuzzy membership function, any state variables and control variables are within the scope of 0-1. If the raw data is not the number of 0-1, according to the principle 'add, subtraction, multiplication and division' the decision-making results unchanged.

2.1.1 The Establishment of the Hierarchical Structure Mode

Firstly, according to the purpose of evaluation, the evaluation purposes are decomposed into multi-level sub-primary and secondary indicators. It just needs to know the bottom original data, and the high-tech SMEs evaluation data can be got.

It not only analyzes the business growth impacting indicators and the classification indicators, but also needs

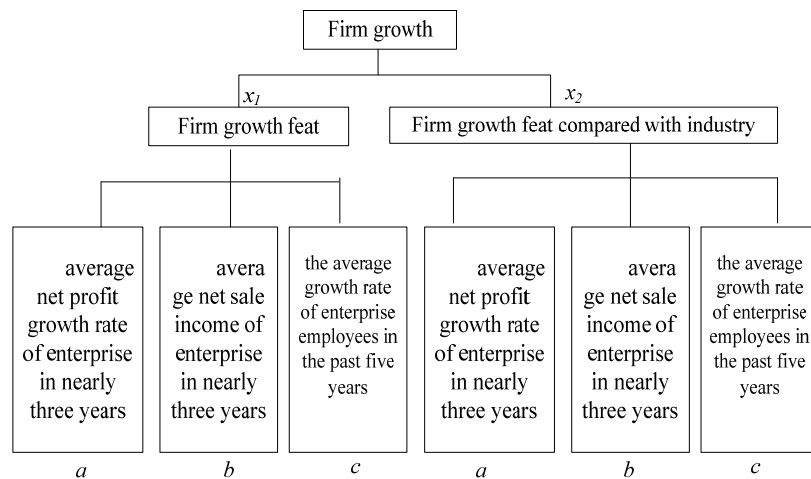


Figure 1. Hierarchical structure model of SMEs growth

to establish evaluation index system according to Catastrophe theory. In this paper, the index system is shown in Figure 1.

2.1.2 Confirming the Catastrophe Type of

Hierarchical Structure Model of SMEs Growth

According to Catastrophe theory, the second-level indicators for the type of Catastrophe system is complementary dovetail catastrophe system and the control variables are a, b, c. High-tech SMEs growth feat compared with industry is complementary dovetail catastrophe system and the control variables are a, b, c. The first level high-tech SMEs growth is complementary cusp catastrophe system and the control variables are x_1, x_2 .

2.1.3 Calculating the Catastrophe Data and Evaluation

Calculate the bottom control variables of each evaluation units, according to Catastrophe theory, control variables must be taken from the number of 0-1. Therefore, using the following formula:

$$y_{ij} = \frac{p_{ij} - \min_{1 \leq j \leq n} p_{ij}}{\max_{1 \leq j \leq n} p_{ij} - \min_{1 \leq j \leq n} p_{ij}} \quad (1 \leq i \leq n)$$

In the formula, y_{ij} is the j th control variable number of the i th objective evaluation system, p_{ij} is the j th index number of the i th objective evaluation system, n is the number of evaluation units, m is the target system for a number of indicators. Subsystem catastrophe number is taken, which can be evaluated for the control variable evaluation of upper system.

2.2 High-Tech SMEs Evaluation Based on Improved Principal Components Projection Method

Unlike traditional principal components projection meth-

od, the structure constructed by the improvement of the principal component projection is composed of two parts: a regional assessment and an evaluation criteria object. The steps of calculation are as follows:

2.2.1 Determine the Evaluation Matrix

According to constructing the methods of the evaluated regions, we can determine evaluation matrix. If there are n evaluated objects, every object is described by m indexes, the matrix will be got. It is:

$$X = \begin{pmatrix} x_{11} & \dots & x_{1m} \\ \dots & \dots & \dots \\ x_{n1} & \dots & x_{nm} \end{pmatrix}$$

In this case, we choose b1 as the average sales growth rate of enterprise in recent three years, b2 as the average employees growth rate of enterprise in recent five years, b3 as the average net profit growth rate of enterprise, and we choose b4, b5, b6 to represents situation of b1, b2, b3 compared to the average level of the same industry respectively, so the total index number is six, that is $m=6$.

2.2.2 Standardizing the Data

For a decision problem with many indexes, given decision matrix, because of different dimensions, the amount scales among indexes will be too different to be compared, and in order to make comprehensive evaluation before using the improved principal component projection, we should eliminate the two kinds of differences above, that is to say, standardizing the matrix X .

Then, standardized evaluation matrix $Y = (y_{ij})_{n \times m}$ will be got $y_{ij} \in [0,1]$. To all values of the standardized indexes, the greater the index is, the better it is. As there are different styles of evaluation indexes, different methods are chosen to deal with. The following method

is used to normalize the values of indexes regarding evaluation matrix.

$$y_{ij} = \frac{x_{ij} - \min_{1 \leq j \leq n} x_{ij}}{\max_{1 \leq j \leq n} x_{ij} - \min_{1 \leq j \leq n} x_{ij}} \quad (1 \leq i \leq n)$$

2.2.3 Determine the Index Weight and Building up Empower-Decided Matrix

There are many methods to determine the index weight, such as subjective method, objective method, empower combination method. In this section, the method of level analysis will be chosen to determine the index weight ω_{ij} , then, weighting the standardized matrix according to weight determined, and ordering that level vectors of empower-decision matrix

$$Z \{ Z = (z_{ij})_{n \times m}, z_{ij} = \omega_{ij} \times x_{ij} \}$$

Correspond to the evaluated objects, vertical vectors correspond to the indexes are empowered.

The steps on confirming weight with improved principal component projection method: Firstly, building up the hierarchical structure in terms of the characteristics of evaluation indexes. As the table 1 described, indexes of upper levels corresponding to b1, b2 and b3 are represented by B1, indexes of upper levels corresponding to b4, b5 and b6 are represented by B2. Then, comparison and judge matrix are built up, which are used to judge one to one. After the establishment of the target level-structure, the relations of indexes between up and down will be determined. As indexes are in the same level, comparison between one and one according to 9 scales are made (Table 1).

We will get judge-matrix $A = \{a_{ij}\}$.

The values in A should meet the following conditions:

$$a_{ij} > 0, a_{ij} = \frac{1}{a_{ji}}, a_{ii} = 1.$$

Table 1. The meaning of 9 scales

Number	The importance of grades	a_{ij}
1	i, j are equally important	1
2	i is slightly important	3
3	i is obviously important	5
4	i is strongly important	7
5	i is extremely important	9
6	i is slightly less important	1/3
7	i is obviously less important	1/5
8	i is strongly less important	1/7
9	i is extremely less important	1/9

The results are as follows:

$$A_{B_1, B_2} = \begin{pmatrix} 1 & 2 \\ 1/2 & 1 \end{pmatrix}$$

$$A_{b_1, b_2, b_3} = A_{b_4, b_5, b_6} = \begin{pmatrix} 1 & 2 & 1/2 \\ 1/2 & 1 & 1/3 \\ 2 & 3 & 1 \end{pmatrix}$$

Finally, the weight is determined.

$$\omega_i = \frac{M_i}{\sum_{i=1}^n M_i}$$

In this equation, $M_i = \sqrt[n]{\prod_{j=1}^n a_{ij}}$.

The results are as follows:

$$W_{B_1, B_2} = (\omega_{B_1}, \omega_{B_2}) = (0.667, 0.333);$$

$$W_{b_1, b_2, b_3} = W_{b_4, b_5, b_6} = (\omega_{b_1}, \omega_{b_2}, \omega_{b_3})$$

$$= (\omega_{b_4}, \omega_{b_5}, \omega_{b_6}) = (0.315, 0.236, 0.449)$$

$$W_{b_1, b_2, b_3, b_4, b_5, b_6} = (\omega_{B_1} \omega_{b_1}, \omega_{B_1} \omega_{b_2}, \omega_{B_1} \omega_{b_3},$$

$$\omega_{B_2} \omega_{b_4}, \omega_{B_2} \omega_{b_5}, \omega_{B_2} \omega_{b_6})$$

$$= (0.210, 0.157, 0.299, 0.106, 0.079, 0.149)$$

Judging the consistency ratio of the matrix,

$$CR = \frac{CI}{RI} = \frac{0.0045}{0.52} < 0.1, \text{ which meets the require-}$$

ments of consistency, and we will find that the largest characteristic value.

$$\lambda_{\max} = 3.009$$

$$CI = \frac{\lambda_{\max} - n}{n - 1} = \frac{3.009 - 3}{3 - 1} = 0.0045.$$

2.2.4 Orthogonal Transformation of Indexes

Due to the large number of evaluation indexes, interrelated relationship among them will result in overlap evaluation information, and it will interfere with the determination of relatively important positions about indexes. But if the values of the indexes are handled by means of orthogonal transformation, the overlap information among indexes will be filtered. Supposing that $W = Z'Z$, we could calculate the result with Matlab, then, characteristic values of matrix W will be got.

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n.$$

Corresponding to unit characteristic, vectors are a_1, a_2, \dots, a_n , named $A = (a_1, a_2, \dots, a_n)$. If we order that $U = ZA$, orthogonal decision-making matrix $U = (u_{ij})_{n \times m}$ can be got, decision-making vector is $U = (u_1, u_2, \dots, u_n)$.

2.2.5 Constructing an Ideal Decision-Making Object and Calculating the Values about Projection

Treating each decision-making object as a dimensional vector and constructing an ideal decision-making object

$$d^* = (d_1, d_2, \dots, d_m)$$

Then uniting d^* with $d_j = \max_{1 \leq i \leq n} (u_{ij})$, we could get the result as follows:

$$d_0^* = \frac{d^*}{\|d^*\|} = \frac{d^*}{\sqrt{d_1^2 + d_2^2 + \dots + d_m^2}}$$

d_0^* represents the ideal decision-making object, calculating projection values of every evaluated vector about ideal decision-making object:

$$D_i = \bar{u}_i \times d_0^* = \frac{\sum_{j=1}^m d_j u_{ij}}{\sqrt{d_1^2 + d_2^2 + \dots + d_m^2}}$$

2.2.6 Ordering the Projection Values and Determining the Level

By the magnitude of projection values calculated with the equation above, we can judge the closing degree between every evaluated object and ideal object, projection values $0 \leq D_i \leq 1$, and to D_i , the greater, the better. The greater value represents that the evaluated object \bar{u}_i is better.

If we choose D_i as comprehensive evaluation values of the n evaluated objects, according to the greater-better principle, the final results of evaluation order could be got.

The lastly new k objects also have k projection results, it is that the values of the endpoints about evaluation criteria intervals also have projection values, these k values form $k-1$ intervals, every interval corresponds to one kind of evaluation criteria level. And $n-k$ projection values of evaluated objects must fall into these $k-1$ intervals, that is to say, projection values of evaluated objects fall on projection values intervals of endpoints about evaluation criteria intervals. Eventually, the interval fell on indicates that evaluated objects are in the corresponding situation.

2.3 Growth Evaluation of High-Tech SMEs Based on Combination

There are many methods to evaluate the growth of the enterprise, such as Catastrophe Theory, Principal Component Projection Method. If we combine the evaluate results of two methods just mentioned to play their own strengths, it is clear that the effect of the evaluation will be improved. And we combine the two kinds of results to make the weighted evaluation. The basic processes are as follows.

Firstly, using ...to deal with the two kinds of results, secondly, using the formula $y = \beta y_1 + (1 - \beta) y_2$ to get the results of combined evaluation, in the formula: y_1 represents evaluation results which are dealt with Catastrophe Theory after normalization, y_2 represents evaluation results which are dealt with Principal Component Projection Method after normalization, β represents the weighted factor. Due to a number of expert advisory, and the specific circumstances of data collection, we select 0.5 as the reasonable value of β .

3. Empirical Analysis on Dalian High-Tech SMEs Growth Evaluation

3.1 Source of the Data and Description of the Data Structure

All data of this study result from surveying enterprises. The region where questionnaires are released is Dalian, the person in charge of the business or senior managers knowing about the enterprise very well are chose to fill in the questionnaires. The number of the questionnaires released is 167, among them, 112 recovered are valid, recovery rate is 67.1%. The enterprises selected in this time meet the needs of the research basically.

By analyzing the data of 112 valid sample, we get the Ownership situation: Collective enterprises account for 2.66%, private enterprises account for 50.67%, stock companies limited account for 5.33%, limited liability companies account for 21.33%, joint-stock cooperative enterprises account for 1.33%, foreign-funded enterprises account for 5.33%, equity joint venture companies account for 10.67%, others account for 2.67%.

Situation of the business type: Production-oriented enterprises account for 24%. Trade Enterprises account for 20%; Consulting, service-oriented enterprises account for 29.33%. Financial investment companies account for 1.33%. Outsourcing service accounts for 5.33%. Other types account for 14.67%; others account for 5.33%.

3.2 The Effectiveness and Reliability of High-Tech SMEs Growth Evaluation

Before data analysis, to test the effectiveness and reliability of data, data reliability test is needed. Cronbach α is a major test of the inherent reliability, through which the α of six grow feat is got, that is 0.910. That means the results of the questionnaire have high internal consistency coefficient. And the performance of the overall business growth correlation coefficient shows in Table 2.

Before a large sample investigation, 20 high-tech SMEs are randomly selected to be interviewed and pre-investigated in Dalian high-tech park. Table 3 lists the evaluation results of 20 SMEs based on Catastrophe the-

Table 2. Overall high-tech SMEs growth performance coefficient

Number	Content	Overall SMEs growth performance coefficient
Y 1	Average net profit growth rate of enterprise in nearly three years	.725
Y 2	Average net sale income of enterprise in nearly three years	.713
Y 3	The average growth rate of enterprise employees in the past five years	.781
Y 4	Average net profit growth rate of enterprise compared with other firm in a industry in nearly three years	.786
Y 5	Average net sale income of enterprise compared with other firm in a industry in nearly three years	.727
Y 6	The average growth rate of enterprise employees compared with other firm in a industry in the past five years	.773

ory and principle component projection method. According to the 112 SMEs research results, the statistics on high-tech SMEs growth are got in Table 4.

High-tech SMEs growth is a process that make enterprises resource value-added. To individual enterprise, it always grows up with the scale expansion and corresponding complexity increase companions. According to the evaluation findings of SMEs growth, we take that the enterprises score between 0.6-1.0 the enterprises as the fast-growing enterprises score between 0.0-0.4. The enterprises are as the low-growing enterprises.

Evaluating on the high-tech SMEs growth based on catastrophe theory and principle component projection method, in addition to provide a new tool for the effective analysis and comprehensive evaluation of high-tech SMEs, but also have the following practical value:

The government can judge the overall development situation according to the evaluation results, which contributes to develop targeted strategies, policies and measures to promote high-tech SMEs development.

High-tech SMEs can have a comprehensive understanding of the overall development of other High-tech SMEs, which help to judge enterprises in the industry performance level. Analyzing on the firm growth and potential sources can help to strengthen the growth of enterprise management clearly in the specific direction.

Table 3. The growth evaluation of high-tech SMEs

Firm Number	Catastrophe theory		Principle component projection method		Combination
	Before Normalization	After Normalization	Before Normalization	After Normalization	
1	0.794	0.835	0.206	0.587	0.711
2	0.773	0.813	0.204	0.581	0.697
3	0.390	0.410	0.077	0.219	0.315
4	0.951	1.000	0.351	1.000	1.000
6	0.624	0.656	0.171	0.487	0.572
7	0.390	0.410	0.077	0.219	0.315
8	0.856	0.900	0.249	0.709	0.805
9	0.411	0.432	0.050	0.142	0.287
10	0.913	0.960	0.308	0.877	0.919
11	0.781	0.821	0.195	0.556	0.688
12	0.446	0.469	0.060	0.171	0.320
13	0.841	0.884	0.233	0.664	0.774
14	0.902	0.948	0.292	0.832	0.890
15	0.841	0.884	0.233	0.664	0.774
16	0.841	0.884	0.233	0.664	0.774
17	0.841	0.884	0.233	0.664	0.774
18	0.781	0.821	0.195	0.556	0.688
19	0.781	0.821	0.195	0.556	0.688
20	0.841	0.884	0.233	0.664	0.774
21	0.438	0.461	0.058	0.165	0.313

Table 4. Study on the evaluation results of SMEs growth

Growth evaluation	0.0-0.4	0.4-0.6	0.6-1.0	Total
Sample Quantity	55	28	29	112
Sample percent	50.8%	23.5%	25.7%	100%

The evaluation results contribute to social research institutions for more in-depth analysis of empirical and theoretical research. Study on the high-tech SMEs development is still in the initial stage. All the available research data, information is also scarce. Evaluation on high-tech SMEs growth can provide a lot of empirical analysis and data support, which help to deepen the high-tech SMEs research.

4. Conclusions

In China there are many vibrant SMEs, especially the representative of high-tech SMEs, whose survival are the basic form of expression in the contemporary conditions of socialization of production and specialization, and is an important part of the modern collaboration division of labor system. Catastrophe theory is used and a mutation series of high-tech SMEs growth evaluation index system is built. Evidence shows the growth of Dalian high-tech SMEs is relatively slow, nearly 50% of Dalian high-tech SMEs grows slowly, which need to be regarded by the government and Dalian high-tech SMEs. Some counter-measures should be put forward by the industry or the government to support the fast growth of Dalian high-tech SMEs.

5. Acknowledgements

This research was supported by National Social Science Fund Project (09BJY055), Liaoning Financial Research Fund Project (08D008); Social Science Project of Dalian (09DLSK193) and Dalian S&T Plan Fund Project (2008D12ZC102.2009D11ZC099).

REFERENCES

- [1] L. Li, "High-tech firm growth mechanism and strategy based on knowledge alliance (Periodical style)," *Technology Management Research*, Vol. 12, pp. 68–73, 2007.
- [2] S. Saarenketo, K. Puimalainen, O. Kuivalainen and K. Kylaheiko, "Dynamic knowledge-related learning processes in internationalizing high-tech SMEs (Periodical style)," *Production Economics*, Vol. 89, pp363–378, 2004.
- [3] L. G. Weinzimmer, P. C. Nystrom, and S. J. Freeman, "Measuring organizational growth: Issues, consequences and guidelines (Periodical style)," *Journal of Management*, Vol. 24, No. 2, pp 235–262, 1998.
- [4] J. Baldwin and G. Gellatly, "Are there high-tech industries or only high-tech firms? Evidence from new technology-based firms," *Analytical Studies Branch, Micro-Economic Analysis Division, Statistics Canada (Research report style)*, pp. 24–32, 1998.
- [5] M. Lindqvist, "Infant multinationals: Internationalisation of small technology-based firms (Research report style)," In D. Jones, & M. Kloststen (Eds.), *Technology, innovation and everprise: The European Experience*, pp. 111–117, 2002.
- [6] N. Karagozoglu and M. Lindell, "Internationalisation of small and medium-sized technology-based firms: An exploratory study (Periodical style)," *Journal of Small Business Management*, Vol. 36, No. 1, pp. 59–65, 1998.
- [7] M. Lindell and N. Karagozolu, "Global strategies of US and Scandinavian R&D intensive small and medium sized companies (Periodical style)," *European Management Journal*, Vol. 15, No. 1, pp. 92–101, 1997.
- [8] D. Crick and M. Spence, "The internationalisation of 'high performing' UK high-tech SMEs: A study of planned and unplanned strategies (Periodical style)," *International Business Review*, Vol. 14, pp. 167–185, 2005.
- [9] N. Coviello and H. Munro, "Growing the entrepreneurial firm: Networking for international market development (Periodical style)," *European Journal of Marketing*, Vol. 29, No. 7, pp. 49–61, 1999.
- [10] J. Bell, "The internationalization of small computer software firms--A further challenge to 'stage' theories (Periodical style)," *European Journal of Marketing*, Vol. 29, No. 8, pp. 60–75, 1995.
- [11] J. Wiklund, "Small firm growth and performance: Entrepreneurship and beyond," *Jonkoping International Business School*, pp. 138–139, 1998.
- [12] L. G. Weinzimmer, P. C. Nystrom, and S. J. Freeman, "Measuring organizational growth: Issues, consequences and guidelines," *Journal of Management*, Vol. 24, No. 2, pp. 235–262, 1998.
- [13] X. H. Chen, "Evaluation of SME growth model based on Catastrophe theory," *Financial Research*, Vol. 11, pp. 13–17, 2006.

How to React to the Subprime Crisis? - The Impact of an Interest Rate Freeze on Residential Mortgage Backed

Julia Hein, Thomas Weber

Department of Economics, University of Konstanz, Konstanz, Germany.
Email: Thomas.A.Weber@uni-konstanz.de

Received July 24, 2009; revised August 29, 2009; accepted October 3, 2009.

ABSTRACT

Several policy options have been discussed to mitigate the current subprime mortgage crisis. This paper analyses an interest rate freeze on adjustable rate mortgages as one possible reaction. In particular, the implications on Residential Mortgage Backed Securities (RMBS) are studied. We examine shifts in the underlying portfolio's discounted cash flow distributions as well as changes in the payment profile of RMBS-tranches. We show that the positive effects of a rate freeze, e.g. less foreclosures and a stabilizing housing market, can outweigh the negative effect of lower interest income such that investors might be better off.

Keywords: Interest Rate Freeze, Subprime Mortgages, Residential Mortgage, Backed Securities (RMBS)

1. Introduction

Starting in mid 2007, rising delinquency and foreclosure rates in the US subprime mortgage market triggered a severe financial crisis which spread around the world. Although subprime mortgages, that were granted to borrowers with weak credit record and often require less documentation, only account for about 15 percent of all outstanding US mortgages, they were responsible for more than 50 percent of all mortgage loan losses in 2007 [1]. Most of the subprime losses were caused by high foreclosure rates on hybrid adjustable rate mortgages (ARM). These loans offer fixed initial interest rates at a fairly low level, which are replaced by higher rates linked to an interest rate index after two or three years¹. Thus, borrowers face a significant payment shock after the interest reset which increases the probability of delinquencies. In previous years, rising real estate prices and, thus, increasing home owner equity enabled mortgage associations to waive part of delinquent interest payments in exchange for an increase in nominal value of the mortgage or to renegotiate the mortgage. But during the last year the trend in real estate prices has reversed in

many regions of the United States leading to “negative equity” of many borrowers, i.e. to real estate values that are lower than their outstanding debt. Consequently, default rates increased².

Several policy options have been discussed to tackle this crisis. The primary concern of policy makers was to lower the financial burden of subprime borrowers and, thus, to avoid further delinquencies and foreclosures which in turn may stabilize house prices. The first policy option is to provide direct financial support by disbursing money to borrowers. In fact, this has been done in February 2008 by means of the *Economic Stimulus Act 2008*, which included tax rebates amounting from \$300 to \$600 per person. Whereas this policy action benefited every tax payer and was not directly linked to the mortgage loans, the *Housing Bill* of July 2008 was especially targeted to subprime borrowers. Here a second policy option was taken by providing state guarantees for mortgage loans. Thus, borrowers, who are close to foreclosure, can refinance their loans at lower interest rates. Although both policy actions certainly help to improve the situation of borrowers, the big drawback of these instruments is that they are mainly financed by the tax payer who cannot be blamed for the crisis. In contrast, mortgage banks, who have been criticized for lax lending standards [2], benefit from less defaults without accepting a responsibility.

¹According to the IMF [1], \$ 250 billion subprime mortgages are due to reset in 2008.

²Mortgage loan contracts in the United States often exclude personal liability such that borrowers do not face any further financial burden when they default.

A third policy option, which takes the banks' failure into account and which was proposed by the US government on December 6th, 2007, is an interest rate freeze. This means that banks agree to waive (part of) the interest rate step up on their ARMs. Although this proposal did not become effective it raises the question whether such an instrument may be better suited to mitigate the current crisis. The aim of this paper is therefore to investigate the implications of an interest rate freeze.

Of course, subprime debtors will benefit from this measure through reduced interest obligations. In contrast, the effect on the lenders is not a priori clear. On the one hand, they receive lower interest on a significant portion of mortgage loans. On the other hand, they might benefit in a twofold way. First, the number of defaults potentially declines. Second, the average loss given default (LGD) might be lower when house prices stabilize. Consequently, there will be a shift in the repayment distribution of the mortgage loan portfolio, which will be examined in the paper.

But the impact of an interest-rate freeze is not limited to borrowers and lenders. More than half of all subprime mortgages that were granted in recent years were sold in residential mortgage backed securities (RMBS). In these RMBS transactions cash flows from the underlying mortgage pool are allocated to tranches with different seniority: several rated tranches and an equity tranche. Due to a priority of payments scheme the equity tranche absorbs most of the losses whereas the senior tranche exhibits only low risk. Part of the RMBS tranches were purchased by outside investors, i.e. foreign banks, non-mortgage banks, insurance companies. As (part of) the mortgage interest payments is used to cover losses that otherwise might hit the rated tranches, RMBS investors lose part of their loss protection following an interest rate freeze. However, they also benefit from potentially lower default losses. The combined effects lead to a reallocation of cash flows and losses among investors which will also be analysed in this paper.

Throughout the paper we look at three sample portfolios: two subprime RMBS portfolios, from which one is well diversified across regions and the other is geographically concentrated in regions that are later hit hardest by the crisis, and one portfolio representing the US mortgage market as a whole. First, we simulate the stochastic repayments of each single mortgage loan using Monte Carlo simulation. In particular, we use the regional house price index as the systematic factor driving the default rate as well as the loss given default. Additionally, we assume that each increase in the payment obligations of a debtor, e.g. through an interest rate step up, raises the default probability. Taking regional diversification into account, we aggregate the single payments to get the repayment distributions of the mortgage loan portfolios underlying the different RMBS transactions.

For all three portfolios we further assume a true-sale RMBS transaction with four differently rated tranches and an equity piece. We use a benchmark scenario without crisis elements to calibrate the sizes and loss protection of the tranches to the respective rating. This scenario includes an interest rate step-up after year two for non-prime mortgage loans.

Subsequently, we derive the portfolio repayment distributions and the resulting tranche characteristics in a crisis scenario that reflects the current situation in the United States. In particular, the average house prices are assumed to have decreased by six percent in the second year of the RMBS transaction. As we show, the crisis leads to a significant reduction of the expected discounted cash flows of the respective portfolios ranging from five percent for the US market portfolio to more than 15 percent for the non-diversified subprime portfolio. Hence, the equity piece often does not suffice to cover the losses which means that the rated tranches need to absorb a significant share of the portfolio loss. Consequently, the risk characteristics of all tranches worsen as compared to the benchmark case which makes severe downgrades necessary as observed in the markets.

Starting from this crisis scenario we investigate the impact of an interest rate freeze. We assume all scheduled interest rate step-ups to be waived which decreases the claims on the RMBS portfolio. As subprime borrowers evade this payment shock, foreclosure rates decrease. First, we study only this direct effect of an interest rate freeze. Our simulation results show that the net change in expected portfolio payments is negative as is the effect on most tranches. The consequences are not uniform for all tranches however: the better the tranche, the less its characteristics deteriorate. The senior tranches of the two subprime RMBS even improve.

In the second case, we additionally include a positive 'second round' effect on house prices. In particular, the lower number of foreclosures takes pressure off the housing market resulting such that the negative downward trend is stopped. Given this combined impact, our results indicate that the positive effects are able to (over-) compensate for the loss due to the interest rate freeze. In particular, all rated tranches benefit in this scenario as compared to the crisis scenario. Therefore we conclude that an interest rate freeze on mortgage loans does not only improve the debtor situation, but might also render investors in RMBS tranches better off at the expense of the equity tranche which takes most of the crisis losses.

The remainder of this paper is structured as follows. First we comment on related literature. Section 3 describes the set-up and calibration of our simulation model. In Section 4, we analyse the effects of a mortgage crisis on our sample mortgage portfolios and also on RMBS-tranches backed by these portfolios. Furthermore, we investigate the consequences of an interest rate freeze on

portfolio and tranche characteristics. Section 5 concludes.

2. Literature Review

Our paper is closely related to the empirical study by *Cagan* [3] analysing the impact of an interest rate reset in adjustable rate mortgages (ARM). Based on a dataset of ARMs originated between 2004 and 2006, he estimates that 59% of these mortgages face a payment increase of more than 25% after the initial period with low rates. He anticipates that in total approximately 13% of adjustable-rate mortgages will default due to the interest rate reset, which corresponds to 1.1 million foreclosures over a total period of six to seven years. This increase in default rates is not equally distributed across all mortgages but depends on the size of the interest rate step-up and the loan-to-value ratio. Additionally, the author estimates that each one-percent fall in national house prices causes an additional 70,000 loans to enter reset-driven foreclosure. Given a house price drop of 10% he projects that more than 22% of ARMs will default due to the interest rate reset. This underlines the impact of a policy reaction to scheduled interest rate step-ups in the present market environment.

Ashcraft and Schuermann [4] discuss the securitization of subprime mortgages. First they provide a detailed analysis of the key informational frictions that arise during the securitization process and how these frictions contributed to the current subprime crisis. They also document the rating process of subprime mortgage backed securities and comment on the ratings monitoring process. They conclude that credit ratings were assigned to subprime RMBS with significant error which has led to a large downgrade wave of RMBS tranches in July 2007³.

Several further research articles provide general information about subprime loans and the current mortgage crisis. *Chomsisengphet/Pennington-Cross* [5] comment on the evolution of the subprime market segment. In particular, some legal changes in the beginning of the 1980s, which allowed to charge higher interest rates and higher fees on more risky borrowers and which permitted to offer adjustable rate mortgages, enabled the emergence of subprime loans. The Tax Reform Act in 1986 allowing interest deductions on mortgage loans made high loan-to-value (LTV) ratios financially more rewarding and, thereby, sub-prime mortgages more attractive. In the beginning of the 1990s the increasing use of securitizations as funding vehicles triggered rapid growth in the subprime mortgage market. Between 1995 and 2006 the volume in this market segment increased from \$ 65 billion to more than \$ 600 billion and also the share on the total mortgage market significantly increased [6]. At the same time the percentage of the outstanding subprime loan volume being securitized went up from about 30%

to around 80% [2]. *Dell' Ariccia et al.* [7] show that the rapid expansion of the subprime market was associated with a decline in lending standards. Additionally, they find that especially in areas with higher mortgage securitization rates and with more pronounced housing booms lending standards were eased. Lower lending standards can thus be identified as one reason for the subprime mortgage crisis.

According to the *IMF* [1], subprime borrowers typically exhibit one or more of the following characteristics at the time of loan origination: weakened credit histories as indicated by former delinquencies or bankruptcies, reduced repayment capacities as indicated by low credit scores or high debt-to-income ratios and incomplete credit histories. Given this very broad definition subprime borrowers are not a homogeneous group. For example, Countrywide Home Loans, Inc. distinguishes four different risk categories of subprime borrowers⁴. These subcategories may depend on the borrower's FICO (Fair Isaac Corporation) credit score, which is an indicator of the borrowers credit history, the Loan-To-Value (LTV) ratio of the mortgage loan, and the debt-to-income ratio⁵. Analysing a data set of securitized loans from 1995 to 2004, *Chomsisengphet/Pennington-Cross* [5] find strong evidence for risk-based pricing in the subprime market. In particular, interest rates differ according to credit scores, loan grades and loan-to-value ratios.

Using a dataset of securitized subprime mortgages from 2001 to 2006, *Demyanyk/Hemert* [8] compare the characteristics of different loan vintages in order to identify reasons for the bad performance of mortgages originated in 2006, which triggered the subprime mortgage crisis. Their sample statistics show that the average FICO credit score increased from 620 in 2001 to 655 in the 2006 vintage, which corresponds to the observation that the market expanded in the less risky segment. During the same period average loan size increased from \$ 151,000 to \$ 259,000 whereas the average loan-to-value (LTV) ratio at origin stayed approximately the same at 80%. Applying a logit regression model to explain delinquencies and foreclosure rates for the vintage 2006 mortgages, *Demyanyk/Hemert* [8] identify the low house price appreciation as the main determinant for the bad performance. Also *Kiff/Mills* [6], who comment on the current crisis, see the slow down in house prices as the main driver for the deterioration in 2006 vintage mortgage loans. Furthermore they emphasize that although the average subprime borrower credit score increased during the last years, also LTV and debt-to-income ratios increased, which made the mortgages more risky.

Gerardi et al. [9] analyse a dataset of homeownership experience in Massachusetts. They find that the 30 day

³In fact there was a second downgrade wave in the beginning of 2008 on which the authors do not comment.

⁴See www.cwbc.com or *Chomsisengphet/Pennington-Cross* [5].

⁵*Kiff/Mills* [6] classify a mortgage as subprime if the LTV is above 85% and/or the debt-to-income ratio exceeds 55%.

delinquency rate shows rather limited variance as it fluctuates between 2 and 2.8 % of borrowers. Further, there is no significant correlation to the change in house prices. In contrast, they find a strong negative correlation between foreclosure rates and the house price index over the whole sample period from 1989 to 2007. In particular, *Gerardi et al.* [9] point out that the house price decline starting in summer 2005 was the driver of rising foreclosure rates in 2006 and 2007. These findings show that the house price index drives the portion of delinquent mortgages that are foreclosed rather than the number of delinquencies themselves.

Estimating cumulative default probabilities they further find that subprime borrowers default six times as often as prime borrowers. This corresponds to *Pennington-Cross* [10] who also compares the performance of subprime to prime mortgage loans and finds that the latter are six times more likely to default and 1.3 times more likely to prepay. Analysing the determinants of default he concludes that for both - prime and non-prime loans - decreasing house prices as well as increasing unemployment rates increase credit losses.

All these empirical studies indicate a strong relationship between mortgage loan defaults and house price appreciation in the subprime market. This corresponds to the theoretical literature on mortgage loan default. According to option pricing theory a borrower, who is not personally liable, should default when the associated put option is in the money, e.g. when the mortgage debt exceeds the house value. Therefore we will use the house price index as the main determinant of default in our simulation model.

3. Model Set-Up

Our analysis is based on a cashflow simulation model. Mortgage loans are more likely to default when they are in “negative equity”, i.e. when the current real estate value is lower than the outstanding debt. This event is usually triggered by a downturn in the house price. Therefore we use a macro factor representing the regional house price index as the systematic determinant of default. We assume the regional house price index to have a nationwide and a regional component. The house price at default further determines the loss incurred in a distressed sale following a foreclosure.

Payment shocks due to interest rate resets can cause additional foreclosures, especially when house prices have already declined. We account for this by adding a

function depending on changes in payment obligations to the idiosyncratic debtor component of our model.

3.1 Simulation Model

RMBS are usually backed by mortgage loans from different regions. This regional diversification reduces the variance of the repayment distribution of the mortgage portfolio and thereby helps to make the rated tranches less risky. For each region we assume the regional house price index (HPI) to be the main driver of the foreclosure rate. Further, for each region k we decompose the percentage change of the HPI in year t into an overall positive long-term trend c and a deviation from this trend driven by a nationwide factor M_t and an orthogonal regional factor $B_{k,t}$:

$$\begin{aligned}\Delta HPI_{k,t} &= c + a \cdot (\sqrt{\rho_M} M_t + \sqrt{\rho_k} B_{k,t}) \\ &= f(M_t, M_{t-1}, B_{k,t}, B_{k,t-1})\end{aligned}\quad (1)$$

Unconditionally, M_t and $B_{k,t}$ are assumed to be standard normally distributed. Empirical evidence suggests, however, that house price changes display a strong positive autocorrelation⁶. Therefore we incorporate a first-order autocorrelation of 0.5 for each factor. Thus, conditional on M_{t-1} , M_t has a mean of $0.5 \cdot M_{t-1}$ and a standard deviation of $\sqrt{0.75}$. The same holds for the regional factors.

ρ_M and ρ_k account for correlations of house price changes across and within regions. We calibrated the nationwide and regional correlations to $\rho_M = 0.1$ and $\rho_k = 0.2$ and the scaling factor to $a = 0.1$. This implies unconditional standard deviations of 5.5% (3.7%) for annual regional (nationwide) house price changes which is in line with empirical evidence⁷. The unconditional mean annual change of the HPI equals the long-term trend c on both, the regional as well as the national level.

For the loans in the underlying mortgage pool we distinguish five debtor groups by credit quality: Prime, Alt-A, Subprime 1, Subprime 2 and Subprime 3. These groups can be interpreted as representing different ranges of the FICO score and further borrower characteristics like payment history and bankruptcies⁸.

The assumed expected default probabilities for the different debtor groups and maturities are shown in the credit curves in Table 1 in the appendix. The numbers correspond to empirical evidence [9].

In each simulation run a path of annual group migrations is computed for each loan in the portfolio. For debtor i located in region k this path depends on a series of latent migration variables $L_{i,k,t}$, $t = 1, \dots, 7$. In this respect our simulation model resembles a migration

⁶In an empirical study based on 15 OECD countries *Englund and Ioannidis* [11] estimate an average first-order autocorrelation coefficient of 0.45.

⁷There exist different house price indices for the US. For example, Freddie Mac's Conventional Mortgage Home Price Index (CMHPI-Purchase Only) shows a standard deviation of 3.8% (nationwide) and 5.2% regionally, since 1975.

⁸There exist no general classification scheme of mortgage loans except for the distinction between Prime, Alt-A and Subprime. Nevertheless it is common to further subdivide the subprime category into several grades [5].

Table 1. Assumed credit curve and one-year migration matrix

Panel A: Standard Case
Credit Curve

t	1	2	3	4	5	6	7
Prime	0.20%	0.52%	0.94%	1.47%	2.07%	2.75%	3.50%
Alt-A	0.50%	1.11%	1.80%	2.57%	3.41%	4.30%	5.23%
Sub1	1.50%	2.98%	4.44%	5.87%	7.29%	8.69%	10.06%
Sub2	2.50%	4.88%	7.15%	9.31%	11.36%	13.32%	15.19%
Sub3	3.50%	6.71%	9.67%	12.41%	14.94%	17.30%	19.51%

Derived One-Year Migration Matrix

Debtor	Prime	Alt-A	Sub1	Sub2	Sub3	D
Prime	88.0%	6.5%	3.0%	1.5%	0.8%	0.2%
Alt-A	9.0%	82.0%	5.0%	2.0%	1.5%	0.5%
Sub1	3.0%	6.0%	82.0%	5.0%	2.5%	1.5%
Sub2	0.5%	2.5%	6.0%	82.0%	6.5%	2.5%
Sub3	0.2%	0.8%	3.0%	7.5%	85.0%	3.5%
D	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%

Panel B: Stressed Migration (Due to Interest Rate Step-Up)

Stressed One-Year Migration Matrix in $t = 3$

Rating	Prime	Alt-A	Sub1	Sub2	Sub3	D
Prime	88.00%	6.50%	3.00%	1.50%	0.80%	0.20%
Alt-A	6.80%	81.51%	6.22%	2.63%	2.08%	0.76%
Sub1	0.66%	1.96%	74.44%	10.45%	6.67%	5.82%
Sub2	0.07%	0.58%	1.96%	84.44%	14.25%	8.69%
Sub3	0.03%	0.15%	0.77%	2.65%	85.13%	11.28%
D	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%

Resulting Credit Curve

t	1	2	3	4	5	6	7
Prime	0.20%	0.52%	0.94%	1.47%	2.07%	2.75%	3.50%
Alt-A	0.50%	1.11%	2.81%	3.62%	4.49%	5.41%	6.36%
Sub1	1.50%	2.98%	8.29%	9.82%	11.32%	12.78%	14.2%
Sub2	2.50%	4.88%	12.67%	14.84%	16.89%	18.83%	20.68%
Sub3	3.50%	6.71%	16.36%	19.00%	21.43%	23.69%	25.80%

Panel A gives the *credit curve* for different debtor groups. Each entry in the credit curve describes the average probability of default for a given initial debtor group and maturity t . The numbers are chosen in accordance with empirical results (see e.g. *Gerardi et al.* 2007). The standard one-year migration matrix is subsequently matched to this credit curve. Panel B displays our assumed stressed migration matrix for year 3. It is assumed that the interest rate step-up causes significant payment shocks which increase the downgrade probabilities of all non-prime loans in year 3. Even though the migration probabilities in the following years return to the normal level, the expected cumulative default probabilities in every subsequent year are increased as shown in the resulting credit curve in Panel B.

model for the assessment of collateralized loan obligations where debtors can “migrate” between different rating groups⁹.

At each annual payment date t , we derive the latent variable

$$L_{i,k,t} = \frac{1}{a} (\Delta HPI_{k,t} - c) + \sqrt{1 - \rho_M - \rho_k} \cdot \varepsilon_{i,t} \quad (2)$$

with $\varepsilon_{i,t}$ iid $N(0,1)$.

If the value of the latent variable $L_{i,k,t}$ lies above (below) a certain threshold, which corresponds to the quantile of the standard normal distribution associated with the migration probabilities in the so-called migration matrix, the mortgage is upgraded (downgraded) to the respective debtor category. Panel A of Table 1 shows the unconditional expected annual migration probabilities for years

⁹In general, either migration models or factor models are used to model loan defaults. E.g. in the literature on securitization, *Franke/Krahn* [12] simulate rating transitions whereas *Hull/White* [13] use a one-factor model and *Duffie/Garleanu* [14] as well as *Longstaff/Rajan* [15] apply multi-factor models in their analysis. We use a mixture of these two approaches.

without changes in interest obligations as well as the corresponding multi-year cumulative default probabilities. The numbers are chosen to match the empirical findings on prime and subprime default rates of *Geradi et al.* [9]. Since these numbers are estimated from a time series between 1987 and 2007, they already incorporate the positive long-term trend in house prices. Consequently we subtract the long-term trend c from our house price changes such that only the deviation from the expected (positive) long-term growth during the last years enters the latent variable.

Due to our assumption of positive autocorrelation in house price changes, our latent variable is not necessarily standard normally distributed, but only normally distributed and the mean depends on the previous realizations. As the thresholds for $L_{i,k,t}$ stay unchanged this yields higher (lower) downward migration and default probabilities in years where negative (positive) house price changes are expected¹⁰.

As can be seen in Equations (2), ρ_M and ρ_k also account for correlation of loan defaults across and within regions. Given our calibrated numbers, 30% ($= 0.1 + 0.2$) of the default risk is due to systematic risk in house price changes and 70% is due to idiosyncratic risks. The idiosyncratic component is given by $\varepsilon_{i,t}$, which includes borrower specific shocks like unemployment, illness or divorce. A payment shock resulting from an increase in interest obligations of a debtor adds to the idiosyncratic risk. We capture this by subtracting a deterministic term from the latent variable in the year of an interest rate step-up. In total,

$$L_{i,k,t} = \frac{1}{a} (\Delta HPI_{k,t} - c) + \sqrt{1 - \rho_M - \rho_k} \cdot \varepsilon_{i,t} - b_i (r_{i,t} - r_{i,t-1}) \quad (3)$$

where $r_{i,t}$ denotes the contractual interest rate of loan i in year t . The impact factor b_i determines the magnitude of the payment shock and is calibrated for each debtor group separately: We chose b_i such that the number of additional defaults due to an interest rate reset matches the forecast made in Cagan [3] for the corresponding percentage interest rate step-up and loan-to-value ratio.

¹⁰Hence, our migration model implicitly accounts for endogenous migration thresholds which is an extension to classical migration models where always standard normally distributed migration variables are drawn.

¹¹According to *Ashcraft/Schuermann* [4] only about 20 percent of mortgage loans in MBS pools are interest only. Other loans mostly pay annuities, which mainly comprise interest payments in the first years, and may even contain a grace period of two to five years in which only interest is paid. Since we only consider a seven year RMBS transaction, our assumption seems to be reasonable.

In our simulations we assume an interest step-up in year three by 1% for all Alt-A loans and by 2% for all Subprime loans. Together with the impact factors (see Table 3) this implies higher downgrade and also higher default probabilities in year 3 as shown in the stressed one-year migration matrix given in Panel B of Table 1. Applying this stressed migration matrix in the year of the interest rate freeze, significantly increases multi-year default probabilities even though migration probabilities are assumed to return to the 'normal' case in the following years.

For simplicity we consider interest only mortgages, i.e. in each year, in which the mortgage stays in one of the five debtor categories, only interest payments are made whereas the total nominal value is due at final maturity¹¹. The interest rate consists of a variable base rate and a spread component. The amount of the spread is determined by the debtor category of the mortgage at the beginning of the transaction. In case of default we assume the real estate to be sold in a distress sale with a discount of q percent of the current market value. Given the HPI of date t defined as:

$$HPI_{k,0} = 1$$

$$HPI_{k,t} = \prod_{\tau=1}^t (1 + \Delta HPI_{k,\tau}) \quad (4)$$

the percentage loss given default of a mortgage in region k at date t is then derived as

$$LGD_{i,k,t} = 1 - \underbrace{(1-q)}_{\text{percentage proceed in distressed sale}} \cdot \underbrace{\frac{1}{LTV} HPI_{k,t}}_{\text{1/LTV at date } t} \quad (5)$$

Thus, we implicitly account for a positive correlation between foreclosure rates and loss given defaults. Due to the definition of our latent variable, a decline in HPI triggers higher default rates and at the same time implies higher loss given defaults.

Having derived the annual portfolio cash flows we calculate the sum of discounted cash flows net of transaction costs (DC_n) for each simulation run n :

$$DC_n = \sum_{t=1}^T \frac{CF_{n,t}}{(1+r_f)^t} - PV(TC) \quad (6)$$

where $CF_{n,t}$ denotes the portfolio cash flow at date t and $PV(TC)$ the present value of annual transaction costs. Dividing this figure by the initial portfolio volume we get a proxy for the relative value of the underlying portfolio. We perform 10,000 simulation runs and determine the

distribution of this portfolio value as well as several statistics like mean, standard deviation and 99%-quantile.

Given the simulated portfolio cash flows at each annual payment date we subsequently derive tranche payments. We assume that all losses (interest and principal) are first covered by the excess spread of the transaction, i.e. the difference between the interest income from the underlying portfolio and the interest payments to the rated tranches net of transaction costs, and then by reducing the nominal value of the equity tranche. Further, we assume the existence of a reserve account which means that if the excess spread of one period is not wiped out by period losses, the excess cash flow is collected in this account earning the risk-free rate and providing a cushion for future losses¹². The holder of the equity tranche does not receive any payments until maturity when he receives the remaining cash flow of the transaction. If the equity tranche has been reduced to zero due to previous losses, the face value and subsequently the in-

terest claim of the lowest rated tranche is reduced to cover the losses. If this tranche claim has already been reduced to zero, the next tranche is used to cover the losses, etc.

3.2 Sample Transactions

Throughout our analysis we consider three illustrative sample portfolios: two subprime mortgage portfolios and one representing the US mortgage market as a whole. The former only include Alt-A and subprime mortgage loans and differ with respect to their regional diversification. The latter predominantly consist of prime (60%) and Alt-A loans (25%). Five percent each fall in the three subprime classes giving a total subprime share of 15% for the portfolio which roughly resembles the subprime portion in the US mortgage market. The explicit portfolio compositions are given in Table 3 in the appendix.

Each mortgage is assumed to pay the risk-free rate, which is assumed to be constant at 4%, plus a spread

Table 3. Portfolio characteristics and model assumptions

	('Pacific') Subprime Portfolio	US Mortgage Market Portfolio	
Portfolio:			
Initial Volume	\$ 100,000,000	\$ 100,000,000	
Number of Mortgages	500	500	
initial LTV	90%	90%	
<i>Share in Region</i>			
Pacific	20% (40%)	20%	
New England	20% (40%)	20%	
North Central	20% (20%)	20%	
Atlantic	20% (-)	20%	
South Central	20%(-)	20%	
<i>Share of</i>			<i>Spreads (bps)</i>
Prime	-	60%	150
Alt-A	20%	25%	225
Subprime 1	30%	5%	300
Subprime 2	30%	5%	350
Subprime 3	20%	5%	400
<i>Interest Rate ($t = 0$)</i>	7.2%	6.0%	
<i>Interest Rate Step-Up (after 2 Years)</i>			<i>Impact Factor (b)</i>
Prime		0%	0%
Alt-A		1%	15
Subprime 1-3		2%	30
<i>RMBS-Structure:</i>			<i>Spreads (bps)</i>
Tranches		AAA	30
		AA	50
		A	80
		BBB	150
		Equity	-
Transaction Costs		1% p.a.	
Maturity		7 years	

This table presents the assumed portfolio compositions of our three sample portfolios as well as the assumed tranche structure. The two subprime portfolios only differ in their regional diversification. The regional composition of the 'Pacific' Subprime portfolio is given in brackets. The depicted spreads are paid in addition to the risk-free rate of 4%.

¹²According to *Ashcraft/Schuermann* [4] excess spread is at least captured for the first three to five years of a RMBS deal, which justifies the assumption of a reserve account.

ranging between 150 and 400 basis points differentiated by debtor category as shown in Table 3. Further we assume that mortgage loans with an initial subprime (Alt-A) rating include an interest rate step-up of 2% (1%) after two years, i.e. all spreads are increased by 200 bps (100 bps) after this initial period¹³. The long-term trend in house prices is assumed to be $c = 3\%$, the loan-to-value ratio at origin is 90% for each mortgage¹⁴ and the discount in case of a distressed sale is $q = 30\%$ ¹⁵. Geographically we differentiate five regions¹⁶: Pacific, North Central (including Mountain), South Central, Atlantic (middle and south) and New England. The first subprime portfolio is concentrated in Pacific (40%) and New England (40%), the regions to perform worst in the crisis, the remaining 20% are North Central and Mountain. The other two portfolios are well diversified across all regions.

First we simulate payments for the portfolios in the benchmark case, i.e. without any crisis. In year 3 the latent variable $L_{i,k,t}$ for each loan is stressed by the impact factor of the current debtor category times the scheduled interest rate step-up which causes an increase in expected cumulative default rates as shown in Panel B of Table 1. Since there is no step-up for prime loans, the expected default rates of these loans stay the same.

Columns 3 in Tables 4, 5 and 6 present some statistics describing the portfolios' repayment distribution. In the benchmark case the expected value of discounted cash flows (net of transaction costs) clearly exceeds the nominal value for all three portfolios. The exceedance equals more than two times the standard deviation of discounted cash flows. For the subprime portfolios the average value of the discounted portfolio payment stream after deducting all fees is 113.4% of the initial face value. Since we use the risk-free rate for discounting, this number corresponds to a yearly average premium of 1.9% on top of the risk-free rate. The standard deviations over seven years are 5% for the well diversified portfolio and 5.6% for the concentrated one. In case of the representative portfolio the expected discounted value in the benchmark case is 105.5%, yielding an average premium of 0.8% p.a., with a standard deviation of 2.4% over seven years.

Subsequently, we simulate payments of three residential mortgage backed security (RMBS) transactions wh-

ich are backed by the sample portfolios and have a maturity of seven years. We assume that four rated tranches AAA, AA, A and BBB are issued that earn the usual market spreads as shown in Table 3. Additionally, we assume annual transaction costs of 1%, which are paid before any interest payment to the tranches.

We calibrate tranche sizes such that their default probabilities in the bench-mark scenario are roughly in line with the historical averages given by Standard & Poor's for the respective rating classes and a seven year maturity. The resulting tranche sizes are also shown in Tables 4, 5 and 6¹⁷. The calibrated tranche structures correspond to typical RMBS structures observed in the market [4]. As can be seen the AAA tranches is smallest and the equity tranche is highest for the 'Pacific' subprime portfolio, which is due to a worse regional diversification.

4. Analysis of Mortgage Crisis

4.1 Crisis Scenario

Having calibrated our model to the benchmark case we now turn to modeling the crisis scenario. In particular we assume that the sample transaction was set-up two years ago (e.g. second quarter of 2006) with tranche sizes as derived before. The nationwide and regional house price index changes are set to match Freddie Mac's Conventional Mortgage Home Price Index¹⁸. In the first year of the transaction, the increase in national house prices slowed down to 2.6%, in the second year there was a downturn of 6%. Regionally, Pacific developed worst with a cumulated two year decrease of 15% in house prices, whereas South Central saw an appreciation of 6%. Table 2 shows all regional trends and the corresponding nationwide and regional macro factors. Figure 1 shows

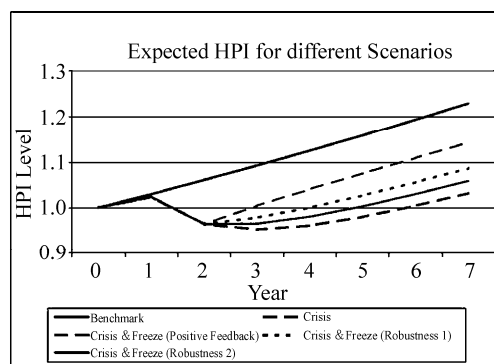


Figure 1. Expected HPI for different scenarios

This figure shows the expected development of the House Price Index (HPI) averaged over our five regions for different simulation scenarios: In the benchmark case a long-term house price growth of 3% p.a. is assumed. The crisis scenario is set by fixing macrofactor realisations in the first two years (see Table 1) which cause a severe decline in house prices. Departing from this crisis scenario the positive feedback scenario assumes that there is no autocorrelation between years 2 and 3, such that house prices recover faster from the crisis. The scenarios Robustness 1 and 2 assume a weaker house price stabilization as the autocorrelation is only reduced by a half or one fourth, respectively.

¹³This step-up is assumed to be fixed at loan origination and is independent of possible downward migrations until the reset date.

¹⁴Gerardi *et al.* [9] report a mean LTV ratio of 83% and a median of 90% in the last three years.

¹⁵Pennington-Cross [16] provides a survey study on the discount in case of a distressed sale and finds that foreclosed property appreciates on average 22% less than the area average appreciation rate. Given that foreclosures also lead to additional costs, we will assume a discount of 30% on the current market value in our simulation analysis. Cagan [3] also states that foreclosure discounts of 30% are quite usual.

¹⁶We followed the regions defined in Freddie Mac's Conventional Mortgage Home Price Index but pooled some neighbouring regions.

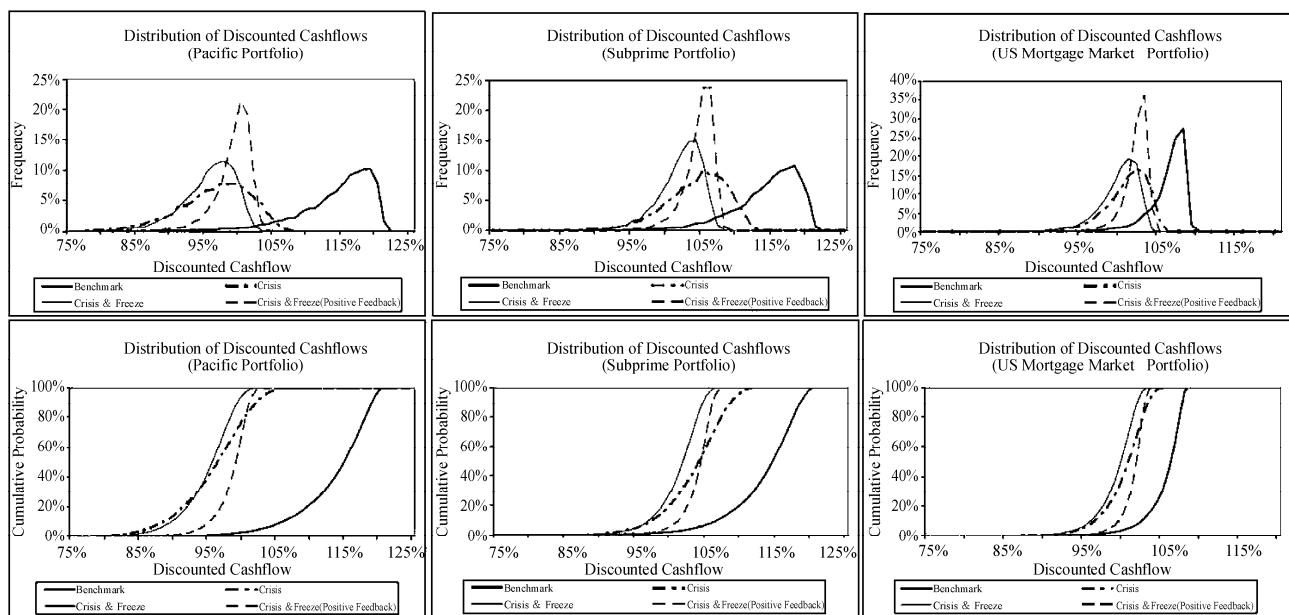
¹⁷A slightly different tranche structure would arise when using the expected loss rating from Moody's. But it should be noted that tranches with the same rating have nearly the same expected losses in our benchmark case.

¹⁸Our reference index is the 'purchase only' index.

Table 2. Definition of the crisis scenario

Region		Year 1	Year 2	HPI_1	HPI_2	μ_3	$E(\Delta HPI_3)$	$E[HPI_3]$
US-Average	M	-0.19	-2.80	1.03	0.96	-1.4	-1.2%	0.95
Pacific	B_1	0.13	-2.60	1.03	0.85	-1.3	-7.2%	0.79
New England	B_2	-0.76	0.18	0.99	0.94	0.09	-1.0%	0.93
N. Central	B_3	-0.09	0.43	1.02	0.98	0.22	-0.5%	0.98
Atlantic	B_4	0.13	0.44	1.03	0.99	0.22	-0.4%	0.99
S. Central	B_5	0.58	1.52	1.05	1.06	0.76	2.0%	1.08

Columns 3 and 4 depict the assumed nationwide and regional factor realisations in year 1 and 2. Columns 5 and 6 give the corresponding regional HPI after one and two years. The last three columns show the mean of the distribution for the third year, the corresponding expected change in regional house prices and the corresponding expected HPI after three years according to our modeling assumptions.

**Figure 2. Distribution of discounted cashflows**

This figure shows the distributions of discounted cash flows (in percent of initial portfolio volume) for all three portfolios ('Pacific' Subprime Portfolio, Subprime Portfolio, US mortgage market Portfolio) and four different simulation scenarios.

the expected average house price development over seven years that is implied by the realisations of the first two years and our modelling assumptions.

Given these macro-factor values in the first two years, we again simulate portfolio cash flows and tranche payments. Due to the positive autocorrelation, the negative trend (as well as the positive trend) in regional house price indices affect the realisations of the latent variable in the following years. For illustration the mean of the macro factors for the third year as well as the corresponding expected cumulative HPI up to year 3 are shown in Table 2.

For the three portfolio settings the resulting portfolio and tranche characteristics given this crisis scenario are depicted in Tables 4 to 6. The crisis leads to a sharp drop in the expected level of the national house price index

after seven years from 1.23 to 1.03 (appr. 16%) which translates into significantly lower cash flows. In fact, our simulation results show that the house price index and the portfolio cash flows are positively correlated with 0.8. Whereas the expected discounted cash flow of the diversified subprime portfolios stays above the nominal issuance volume, the expected discounted cash flow of the US mortgage market portfolio drops roughly to \$ 100 million indicating that there is no premium left for originator. The 'Pacific' subprime portfolio concentrated in the Pacific, New England and North Central region shows a drop to less than \$ 96 million, a severe loss. Obviously, the crisis causes a severe first order stochastic dominance deterioration in the distributions of discounted cash flows of all three portfolios as depicted in Figure 2.

The shift in the distribution of discounted cash flows causes all tranches to exhibit much higher default probabilities and expected losses such that it would be necessary to downgrade them several rating notches. For example the AAA tranche of the diversified subprime portfolio would now receive an (A-) rating and the most junior tranches would only get a (B-) rating. The effect of the crisis on the tranches' risk characteristics is even slightly stronger for the US mortgage market portfolio. Here the default probabilities and expected losses are roughly ten times higher than before whereas for the diversified subprime portfolio the numbers only increase by a factor of about eight. Of course, the effect is largest for the Pacific subprime portfolio with default probabilities increasing by 30 times and the most junior tranche being certain to default.

The main part of the decrease in expected payments is allocated to the equity tranche. Looking at the diversified subprime portfolio, the expected present value of equity tranche payments decreases by \$ 9.2 million, 92% of the total portfolio decrease of \$ 10 million. For the US portfolio the situation is similar. The expected discounted cashflow to the equity tranche decreases by \$ 4.2 million - about 80% of the total portfolio decrease. Nevertheless the decline in expected discounted portfolio cashflows is rather moderate, only 10% for the subprime portfolio and 5% for the US mortgage market portfolio. This is due to the fact that both portfolios are assumed to be well diversified concerning the regional allocation with some regions still displaying a positive house price trend. In contrast, the Pacific subprime portfolio concentrated in regions performing poorly decreases by nearly 18% in expected discounted cash flow. Here, the equity tranche bears only 71% of this decline as the rated tranches are hit more heavily. Curiously, the equity tranche still has a positive expected cash flow of \$ 1.3 million, even though the lowest rated tranche is always hit by losses. This is due to excess spread collected in later years. Fewer excess spread would accrue in case of an interest rate freeze.

4.2 The Impact of an Interest Rate Freeze

Starting from the crisis scenario described in the previous subsection we now analyse the effects of an interest rate freeze on the sample RMBS. In particular, we assume that the interest step-up after two years is cancelled such that all mortgage loans continue to pay the low initial rates. The direct effect of this freeze will be twofold. On the one hand, lower interest rates reduce the portfolio payment claims and, thus, negatively affect payments to the issued tranches. On the other hand, an interest rate freeze takes pressure from borrowers such that there will be less foreclosures which in turn lowers the foreclosure costs. We study this trade-off of direct effects first.

In the second part of this section we investigate dif-

ferent scenarios of house price reactions following the freeze. In fact, the lower number of foreclosures may have a positive feedback effect on house prices. We find that a relatively moderate stabilization of house prices renders the net effect on most tranches positive.

4.2.1 Pure Interest Rate Freeze

As noted before, the interest rate freeze does not only lead to less interest payments from the portfolio, but has also a positive effect on the portfolio default rate. In particular, there are less downward migrations and also less defaults in year three because the stress component of all Alt-A and subprime debtors disappears (see Equation 2) due to unchanged payment obligations. In effect, by avoiding downgrades the interest rate freeze does not only decrease default rates after three years but also results in lower cumulative default probabilities in subsequent years.

We simulate portfolio repayments and tranche characteristics for this scenario. The results are shown in Figure 2 and Tables 4 to 6. Although the interest rate freeze lowers the default rate of the underlying portfolio, this does not compensate for the decline in interest payments from years three to seven. Thus, the freeze leads to a deterioration in the distributions of discounted cashflows. For the US mortgage market portfolio we see a first order stochastic dominance deterioration with the expected discounted portfolio cashflow being further reduced by \$ 1 million. Also all RMBS tranches deteriorate as compared to the crisis scenario. The former AAA tranche which would have to be downgraded to BBB+ due to the crisis would now only receive a BBB rating. Again a substantial share of the additional loss is allocated to the equity tranche (appr. 87%).

For the diversified subprime portfolio we see an additional loss of \$ 2.2 million due to the interest rate freeze and a second-order stochastic dominance deterioration in the distribution. In fact, lower quantiles slightly improve as compared to the crisis scenario. Consequently, the senior tranche benefits from the interest rate freeze whereas all other RMBS tranches suffer additional losses. Here the equity tranche takes 69% of the additional expected loss. Concerning the 'Pacific' subprime portfolio the additional expected loss is only \$ 0.8 million. As the regions represented in this portfolio saw the steepest downturn in house prices bringing many debtors close to default, waiving interest claims will avoid most foreclosures such that the positive effect of a rate-freeze is strongest in this case as compare to the other portfolios. Again we see a second-order stochastic dominance shift in the expected discounted cashflow distribution leaving the senior tranche better off at the expense of lower rated tranches and the equity piece.

Table 4. 'Pacific' subprime portfolio - simulation results

	Benchmark	Crisis	Crisis & Freeze	Crisis & Freeze Positive Feedback	Crisis & Freeze Robustness 1	Crisis & Freeze Robustness 2
Portfolio Characteristics						
Exp. Discounted CF (in \$)	113,417,751	95,678,380	94,870,489	98,466,841	96,323,606	95,954,193
% of initial Volume	113.42%	95.68%	94.87%	98.47%	96.32%	95.95%
Standard Deviation	5.60%	5.32%	3.94%	2.58%	4.01%	3.77%
1%-Quantil	94.27%	80.79%	82.75%	89.58%	84.05%	83.89%
Tranche Characteristics						
	Size	<i>Default Probabilities</i>				
Rating						
AAA	84.50%	0.29%	8.07%	0.80%	4.97%	5.34%
AA	4.70%	0.71%	24.69%	3.53%	16.15%	17.37%
A	3.90%	1.37%	53.09%	13.26%	37.72%	40.10%
BBB	4.40%	99.99%	99.99%	99.97%	99.98%	99.99%
	<i>Expected Loss</i>					
AAA	0.01%	0.37%	0.26%	0.02%	0.15%	0.16%
AA	0.36%	12.86%	11.61%	1.42%	7.45%	7.94%
A	0.81%	27.87%	28.80%	5.66%	20.10%	20.91%
BBB	2.03%	66.48%	67.03%	48.39%	60.72%	61.46%
Equity Tranche	2.50%					
Exp. Discounted CF (in \$)	13,829,942	1,200,436	297,281	1,358,743	800,465	536,498

Each column of this table shows the simulation results for a different simulation scenario: The benchmark scenario (without any crisis) the crisis scenario, the crisis scenario together with an interest rate freeze as well as the crisis scenario with an interest rate freeze and an additional positive feedback effect which resets house price trends after 2 years to zero, or lowers positive autocorrelation by a half (robustness 1) and one fourth (robustness 2), respectively. The first lines show some portfolio characteristics - the expected discounted cash flow net of transaction costs, the standard deviation and the 99% -quantile of the discounted cashflows distribution. The following lines show the default probabilities and expected losses of the rated tranches. Finally the expected discounted cash flow to the equity tranche is depicted.

Table 5. Subprime portfolio - simulation results

	Benchmark	Crisis	Crisis & Freeze	Crisis & Freeze Positive Feedback	Crisis & Freeze Robustness 1	Crisis & Freeze Robustness 2
Portfolio Characteristics						
Exp. Discounted CF (in \$)	113,409,975	103,433,389	101,204,384	103,870,437	102,743,287	101,984,251
% of initial Volume	113.41%	103.43%	101.20%	103.87%	102.74%	101.98%
Standard Deviation	4.97%	4.43%	3.15%	2.15%	2.70%	3.03%
1%-Quantil	96.94%	90.74%	91.53%	96.49%	94.03%	92.39%
Tranche Characteristics						
Rating	Size	<i>Default Probabilities</i>				
AAA	88.10%	1.58%	1.41%	0.10%	0.51%	0.93%
AA	4.60%	5.61%	6.67%	0.78%	2.66%	4.71%
A	2.80%	12.31%	15.72%	2.40%	6.59%	11.15%
BBB	2.90%	36.99%	39.99%	16.95%	25.26%	32.61%
<i>Expected Loss</i>						
AAA	0.01%	0.04%	0.03%	0.002%	0.01%	0.02%
AA	0.37%	2.49%	2.72%	0.25%	0.97%	1.85%
A	0.81%	6.82%	8.12%	1.06%	3.29%	5.64%
BBB	2.01%	17.05%	18.20%	3.99%	8.88%	13.45%
Equity Tranche	1.60%					
Exp. Discounted CF (in \$)	13,003,043	3,804,631	1,656,349	3,521,528	2,655,818	2,163,384

This table present the same simulation results as in Table 4 but now for a regionally well diversified Subprime Portfolio.

Table 6. US mortgage market portfolio - simulation results

	Benchmark	Crisis	Crisis & Freeze	Crisis & Freeze Positive Feedback	Crisis & Freeze Robustness 1	Crisis & Freeze Robustness 2
Portfolio Characteristics						
Exp. Discounted CF (in \$)	105,506,354	100,249,781	99,273,594	101,411,345	100,540,460	99,921,438
% of initial Volume	105.51%	100.25%	99.27%	101.41%	100.54%	99.92%
Standard Deviation	2.41%	2.79%	2.50%	1.54%	2.05%	2.36%
1%-Quantil	96.50%	91.50%	91.26%	95.95%	93.55%	92.08%
Tranche Characteristics						
Rating	Size	<i>Default Probabilities</i>				
AAA	89.40%	0.28%	2.24%	2.90%	0.84%	1.90%
AA	3.10%	0.73%	6.86%	9.19%	3.44%	6.10%
A	1.90%	1.35%	13.53%	18.32%	7.49%	12.53%
BBB	3.90%	4.50%	45.55%	62.42%	36.92%	49.38%
	<i>Expected Loss</i>					
AAA	0.01%	0.05%	0.06%	0.004%	0.02%	0.04%
AA	0.37%	3.20%	4.16%	0.37%	1.48%	2.81%
A	0.77%	7.61%	10.10%	1.23%	3.92%	6.81%
BBB	2.00%	20.53%	27.54%	5.95%	13.72%	20.31%
Equity Tranche	1.70%					
Exp. Discounted CF (in \$)	5,178,193	972,104	118,113	1,258,858	818,838	598,038

This table present the same simulation results as in Table 4 but now for the US Mortgage Market portfolio.

4.2.2 Interest Rate Freeze and Positive Feedback Effect

As shown in the previous subsection, the first round effects of an interest rate freeze are not sufficient to attenuate the crisis. Yet a decrease in foreclosure rates may take pressure off the housing market such that the negative trend in the regional house prices is mitigated¹⁹. This in turn will lead to a positive effect on subsequent foreclosure rates.

In the previous scenarios, persistent trends in the house price index are implemented by positive autocorrelation in the house price index. Therefore the downturn in years one and two leads to an expected downturn in year three, i.e. the conditional mean of the variable describing changes in the house price index is negative. Combined with the regional components, this yields expected house price changes of between -7.2% and 2.0% for the respective regions, nationwide -1.2% (see Table 2) which is substantially below the long-term mean of 3%. We now assume these negative trends to be stopped by the interest intervention. In effect, expected house price appreciations rise to the long-term trend of 3% for region 1 (Pacific) and up to 6.4% for region 5 (South Central) in year 3. We implement this by excluding the autocorrelation effects from year two to year three in the nationwide factor as well as in all regions with negative factor realization in year 2. In total, the average HPI stabilizes by four percent in year three. Feedback effects in subsequent years result in an average HPI that by the end of year seven is approx. 11 percent higher than in the crisis scenario, as shown in Figure 1. The results for the portfolios are displayed in Figure 2 and Tables 4 to 6.

For both subprime portfolios the expected discounted cashflow exceeds the value in the crisis scenario without an interest rate freeze. Comparing the cumulative expected discounted cash flow distributions this positive feedback scenario second-order stochastically dominates the crisis scenario with all lower quantiles being substantially improved. Consequently, all rated RMBS tranches benefit concerning their default probabilities and also their expected losses. Given the diversified subprime portfolio the AAA and the AA tranche perform as well as in the benchmark scenario without crisis meaning that no downgrade would be necessary. The costs of the interest rate freeze are completely borne by the owner of the equity tranche.

The US mortgage market portfolio loses less interest payments due to the interest rate freeze as prime mortgage loans representing 60% of the portfolio do not incorporate an interest step-up. Here, the positive effect of stopping the house price decline overcompensates the foregone interest resulting in higher expected discounted

cash flow compared to the crisis scenario. All tranches including the equity tranche profit with higher tranches benefiting most. This is due to a more narrow distribution of losses. Compared to the crisis we again observe a second order stochastic dominance shift in cumulative repayments.

Summarizing, our results indicate that an interest rate freeze may help to alleviate the current crisis. Even though RMBS tranche investors lose a significant portion of their loss protection, this deterioration may be overcompensated by improvements in mortgage payments due to lower foreclosure rates and a positive feedback effect in the housing market. For all three portfolios we derive positive net effects on all rated RMBS tranches as compared to the crisis scenario. The higher the tranche, the more it improves. Especially, the AAA tranche benefits from the rate freeze. Thus, the RMBS market will benefit from an interest rate freeze which can induce positive spill over effects on other markets. In particular, markets for other structured instruments containing RMBS tranches may stabilize. Especially, special investment vehicles backing their commercial paper funding with senior RMBS tranches may recover.

4.3 Robustness Checks

1) Assumptions concerning House Price Developments

Our previous results depend on several assumptions concerning house price developments which are motivated by empirical findings. We set the *crisis scenario* to match house price developments in the main US regions during the last two years. When discussing the positive feedback effect of an interest rate freeze we had to make a specific assumption concerning house price stabilization. Naturally, other house price reactions are also possible.

As robustness checks we derive portfolio and tranche repayments for less favorable assumptions concerning house price stabilization. In particular we assume the negative house price trend only to be partially offset by the interest rate freeze. Instead of zero autocorrelation in year three increasing the average house price index by four percent compared to the crisis situation, we now assume that only half (one fourth) of this effect is realised. Figure 1 shows the expected average house price development for these two scenarios. Tables 4 to 6 display the tranche and portfolio characteristics for these additional scenarios.

As can be seen, a more moderate stabilization of two percent in year three (translating into 5.5 percentage points until year seven) is sufficient to substantially improve all rated RMBS tranches (see *Robustness 1*). Even a stabilization of only one percent in year three (increasing to 2.7 percent in year seven) leaves the rated tranches slightly better off than in the crisis scenario (see *Robustness 2*).

¹⁹Cagan [3] finds significant additional foreclosure discounts in regions with high foreclosure rates. This indicates limited buyer capacities unable to absorb the excess supply without additional discounts.

Given these results we conclude that the qualitative results are quite stable towards changes in the assumption of house price stabilization: Due to lower excess spread, the payments to the equity tranche will be reduced the most and due to lower probabilities of high losses the highest rated tranche will profit most from an interest rate freeze. Even for modest house price reactions the net effect of the freeze is positive.

2) *Assumptions concerning RMBS-Structure*

A further assumption which needs to be critically reviewed is our assumption concerning the payment waterfall for our RMBS tranches. In the previous simulations we always assumed the existence of an unlimited reserve account, which means that the holder of the equity tranche only receives payments at final maturity and that at each annual payment dates all excess cash flows are placed in an extra account which can be used to cover future losses. In fact other reserve account specifications are possible, e.g. a capped reserve account, where all excess cash flows above this cap are paid out to the holder of the equity piece periodically, or even a structure without any reserve account, in which the holder of the equity tranche receives all excess cash flow at each payment date.

This assumption mainly influences the calibration of tranche sizes in the benchmark case. In particular, a structure without a reserve account will lead to a much smaller AAA tranche a bigger equity tranche. In this case the effect of the interest rate freeze is less pronounced since the tranche sizes are already calibrated to provide a better protection against interest losses. Nevertheless, the qualitative effects stay the same with the difference that now an even more moderate house price stabilization is sufficient to make all rated tranches better off than in the pure crisis scenario.

4.4 Other Policy Options

1) Interest rate cuts

Throughout the paper we assumed that the risk-free interest rate on top of which credit spreads are paid stays the same over seven years. In fact the crisis might lead to a cut in this reference rate. Looking at the repayments of the mortgage loans analysed in this paper lowering the reference rate would have a positive effect. Thus, interest rate cuts are an additional policy option worth examining. However, discussing the macroeconomic consequences of interest rate cuts is beyond the scope of this paper.

2) Housing Bill

For mortgage debtors the effects of the proposed interest rate freeze are comparable to the sought impact of the Housing Bill of July 2008. Here, state guarantees help troubled borrowers to refinance at lower rates. The key difference between the two policy options is on the lender side. With the interest rate freeze, mortgage banks and equity tranche holders bear the potential costs.

Looking at a portfolio of mortgage loans the state guarantee included in the Housing Bill adds a large state owned first loss position, irrespective of the portfolio being securitized. Compared to our simulation results above, lenders, all rated tranches and especially the equity tranche would profit at the tax payers expense.

5. Conclusions

The discussed interest rate moratorium for subprime mortgages is one option to tackle the current crisis. It is an agreement between two parties - the U.S. government and the originating banks - that affects two different third parties: the mortgage debtors and investors in RMBS tranches. The first group will unambiguously profit from an interest rate freeze. Some of their payment obligations are waived, thus they might avoid default. Additionally, they benefit from a stabilizing housing market.

The effect on RMBS-tranches is more ambiguous. First, we show that the pure interest rate freeze decreases the portfolio payment stream's expected value by one to 2.2 percent, depending on portfolio composition. The vast majority of this decrease is borne by the equity tranche. Default probability and expected loss of rated tranches only slightly deteriorate as compared to the crisis scenario with senior tranches even being better off. Second, we take into account that the interest rate freeze may have a positive second round effect as a reduction in foreclosures takes pressure off the housing market. In this case we find that already a very moderate mitigation of the house price downturn yields a positive net effect on all rated tranches. A stabilization of one percent in year three leaves all tranches in each of our three sample transactions better off (compared to the crisis scenario without policy reaction). For the holder of the equity tranche, the situation is different. Looking at the average of the three sample RMBS a four percent stabilization in house prices is needed for him to slightly benefit from the interest rate freeze. Thus, should the US government and loan and savings associations decide on an interest moratorium on adjustable rate subprime mortgages, this would probably not come at the expense of the RMBS investors as a third party. If additional losses occur they are borne by originators and equity tranche investors. As these parties are the ones in charge of the criticized lending standards we argue that reconsidering this policy option is worth while.

REFERENCES

- [1] International Monetary Fund, Global Financial Stability Report, Washington DC, April 2008.
- [2] Keys, J. Benjamin, T. Mukherjee, A. Seru, and V. Vig, "Did securitization lead to lax screening? Evidence from Subprime Loans," Working Paper, 2008.
- [3] Cagan, L. Christopher, "Mortgage payment reset: The issue

- and impact,” First American CoreLogic, March 19, 2007.
- [4] Ashcraft, B. Adam, and T. Schuermann, “Understanding the securitization of subprime mortgage credit,” Federal Reserve Bank of New York Staff Reports, No. 318, 2008.
 - [5] Chomsisengphet, Souphala and A. Pennington-Cross, “The evolution of the subprime mortgage market,” Federal Reserve Bank of St. Louis Review, Vol. 88, No. 1, pp. 31–56, 2006.
 - [6] Kiff, John and P. Mills, “Money for nothing and checks for free: Recent developments in U.S. subprime mortgage markets,” Working Paper, International Monetary Fund, 07/188, 2007.
 - [7] G. Dell’Ariccia, D. Igan, and L. Laeven, “Credit booms and lending standards: Evidence from the subprime mortgage market,” Working Paper, 2008.
 - [8] Demyanyk, Yuliya and Otto Van Hemert, “Understanding the subprime mortgage crisis,” Working Paper FRB St. Louis and NYU Stern, 2007.
 - [9] Gerardi, Kristofer, A. H. Shapiro and P. S. Willen: “Subprime outcomes: Risky mortgages, homeownership experiences, and foreclosures,” Working Paper Federal Reserve Bank of Boston, No. 07–15, 2007.
 - [10] Pennington-Cross, Anthony “Credit history and the performance of prime and nonprime mortgages,” *Journal of Real Estate Research*, Vol. 27, No. 3, pp. 279–301, 2003..
 - [11] Englund, Peter and Y. M. Ioannidis, “House Price Dynamics: An International Empirical Perspective,” *Journal of Housing Economics*, Vol. 6, pp. 119–136, 1997.
 - [12] Franke, Günter and Krahnen, Jan Pieter: “Default risk sharing between banks and markets: The contribution of collateralized loan obligations,” in: *The Risks of Financial Institutions*, NBER book edited by Mark Carey and Rene Stulz, University of Chicago Press, pp. 603–631, 2006.
 - [13] Hull, John, and A. White: “Valuation of a CDO and an nth to Default CDS without monte carlo simulation,” *Journal of Derivatives*, Vol. 12, No. 2, pp. 8–23, 2004.
 - [14] Duffie, Darrell, and N. Garleanu: “Risk and valuation of collateralized debt obligations,” *Financial Analysts Journal* (January/February), pp. 41–59, 2001.
 - [15] Longstaff, A. Francis, and A. Rajan, “An empirical analysis of the pricing of collateralized debt obligations,” *Journal of Finance*, Vol. 63, No. 2, pp. 529–563, 2008.
 - [16] Pennington-Cross, Anthony, “The value of foreclosed property,” *Journal of Real Estate Research*, Vol. 28, No. 2, pp. 193–214, 2004.

Flight Attendants' Emotional Labor and Exhaustion in the Taiwanese Airline Industry

Cheng-Ping CHANG, Ju-Mei CHIU

Nan-Tai Street, Yung-Kung City, Taiwan, China
Email: justin@mail.stut.edu.tw

Received May 2, 2009; revised July 29, 2009; accepted August 4, 2009.

ABSTRACT

Few research studies have discussed the two variables of emotional labor and emotional exhaustion, and even fewer have examined flight attendants as the research subject. The current study employed a questionnaire method to examine 353 Taiwanese flight attendants' feelings about emotional labor, the status of their emotional exhaustion, and the relationship between emotional labor and emotional exhaustion. The research results indicate that: 1) while the degree of emotional labor operating on female flight attendants is on the medium to high side, the attendants' perception of emotional exhaustion is only medium; 2) female flight attendants' emotional labor has a significant positive correlation with their emotional exhaustion; and 3) among the perspectives of emotional labor, the qualities of "deep emotional masking" and "multiformity" have a significant predictive effect on emotional exhaustion.

Keywords: Flight Attendant, Emotional Labor, Emotional Exhaustion

1. Introduction

In recent years, the industrial structure in Taiwan has changed such that the service industry has gradually replaced traditional manufacturing as the leading industry. Statistical data reported in the *Quarterly National Economic Trends* issued by the Directorate-General of Budget, Accounting and Statistics, Executive Yuan, indicated that in the third quarter of 2006 the service industry accounted for 71.98% of the country's industry.

The airline industry has been part of the service industry since the government implemented the open sky policy of airline deregulation in 1987. As a result, many private airlines entered the market (Liou, 2006). High speed rail opened completely to traffic in 2007, taking away many passengers from airlines in western Taiwan. Understanding that a low price strategy cannot maintain passenger loyalty for any length of time, the airlines have changed their strategy to emphasize improved quality of service and maintenance of flight safety. These strategies encourage passengers to maintain a positive opinion about the airlines and to appreciate the quality of service they provide.

Flight attendants constitute the majority of customer-service employees in the airline industry. Compared to their colleagues in other departments, flight attendants have more contact with passengers, and for longer periods of time. Within the airline industry, flight attendants

are referred to as the first-line service attendants [1]. Passengers' images of the airlines are heavily influenced by the manners and emotional attitudes of flight attendants. The working environment of flight attendants is noteworthy in that, over time, it will have a negative impact on flight attendants' psychological health. During international service, flight attendants face numerous stresses. They must provide service over a long period of time; the pressure in the aircraft cabin is high and the space is hermetic; the types and temperaments of passengers are complex; and the environment may foster various diseases. Furthermore, the work hours of flight attendants are uncertain, and they often deal with night-shift assignments and time-zone changes. Such anomalous shifts over a long period of time constitute the main influence on the health of flight attendants. Beh & McLaughlin [2] suggested that stress experienced during long flights affects some aspects of mental performance. In addition, the uncertainty of flight attendants' work schedules limits their private time, potentially causing conflicts between work and family. Additionally, an overly-heavy workload may induce burnout among flight attendants. Such burnout, in turn, increases flight attendants' alienation from work and decrease their sense of connection to the company. Liang & Hsieh [3] suggested a relationship between burnout and workplace deviance, identified as a component of job performance. Too much

physiological and psychological stress over a long period of time causes lower service quality and a higher turnover rate.

Flight attendants have to control their overt behavior and private emotions in order to maintain positive interactions with colleagues and passengers. This kind of emotional control is dictated by the job performance rules of the company, and attendants are required to adjust their emotions to the requirements of the job. This is precisely how Hochschild [4] defined "emotional labor." Engaging in emotional labor over an extended period may cause emotional labor overload and make adjustment to work demands difficult. This situation, in turn, may lead to emotional dissonance, that is, a conflict between the attendant's internal emotions and the organization's rules regarding emotional expression. Over an extended period, this may have a negative effect on the employee's physiology and psychology. Long-term emotional stress and relatively intense emotional labor result in emotional exhaustion. (see Figure 1)

There have been few research studies on emotional labor among Taiwanese flight attendants. This research focuses on flight attendants in Taiwan, addressing their experiences of emotional labor and emotional exhaustion. The goal of this study was to develop suggestions for additional research, provide practical advice for airlines and related management organizations about the management of flight attendants' emotional labor, and help those in the flight attendant industry better understand the work characteristics of that role. The research objectives are as follows:

- 1) To understand the status of flight attendants' emotional labor and emotional exhaustion.
- 2) To examine differences in emotional labor and emotional exhaustion among flight attendants from different backgrounds.
- 3) To explore the relationship between emotional labor and emotional exhaustion among flight attendants.
- 4) To understand whether flight attendants' emotional labor predicts emotional exhaustion.

1.1 Emotional Labor

The concept of emotional labor was first proposed by Hochschild [4]. She described emotional labor as follows: "Attendants who are in contact with clients over an extended period are required to control their emotions

during work time and to display appropriate facial expressions and behaviors. At the same time, the organizations' regulations and salary structure require that these attendants control their emotions to create the work climate that the organization wants." In order to meet the organization's expectations for client service, most employees must frequently control their own emotions in such a way as to convey the emotion expected by the organization.

Referring to the Hochschild's [4] model of emotional labor, Wharton [5] proposed three work characteristics or perspectives:

- 1) The worker has ongoing contact with the public in both face to face and conversational exchanges.
- 2) When working, the employee is expected to express certain emotions that will have the desired influence on others.
- 3) The range of acceptable emotional expressions is limited by employer-imposed regulations.

Goffman [6] argued that the form of emotional expression imposed by emotional labor amounted to surface acting. Goffman suggested that emotional labor amounts to a dramatic enactment, one similar to that of an actor who is expected to express emotion in a screenplay. Goffman [6] also posited that, in addition to surface acting, workers may choose to modify how they think about their own feelings or to resist conforming to the emotional expression as dictated by the external organization. Goffman's model of emotional labor as drama suggests that employees may also adopt "deep acting".

Earlier researchers discussed emotional labor from the perspectives of high- and low-level emotional labor as proposed by Hochschild [4]. Discussion about these two levels, however, does not completely address the emotional labor load and its component factors. The four perspectives proposed by Morris and Feldman [7] lacked the support of empirical research. A test of Morris and Feldman's [8] approach discovered that, although their model held promise for explaining the capacity for emotional labor, it was unable to differentiate causality from correlation in explaining emotional labor and its consequences.

Long-term emotional stress or an environment laden with too much emotional labor may cause emotional exhaustion, another consequence of emotional labor. When employees have contact with clients for a long time, the heavy emotional load may induce energy exhaustion [9]. This may cause employees to withdraw from their duties and may lead to illness [10,11].

1.2 Emotional Exhaustion

According to the literature on work burnout, the first phase of work burnout is emotional exhaustion [12]. Emotional exhaustion occurs when individuals lack energy and have the sense that they are becoming exhausted

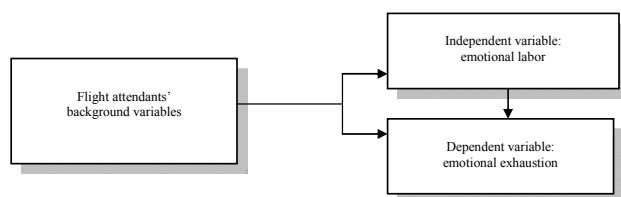


Figure 1. Research framework

[10]. Saxton, Phillips, and Blakeney [9] also pointed out that emotional exhaustion is related to excessive emotional demand during interpersonal interaction; this causes energy exhaustion and increased withdrawal. These conditions decrease productivity and increase certain stress reactions. In the same vein, Maslach and Jackson [13] found that workers who had to engage in face to face contact with clients as well as manage emotional expression for extended periods were most susceptible to emotional exhaustion.

Based on these findings, it seems probable that flight attendants develop emotional exhaustion as a result of frequent and extended interactions with customers. Those who experience excessive interaction with the clients or who obey the organization's expectations for emotional control [14] show increasingly frequent emotional exhaustion.

Emotional exhaustion harms an individual's body and mind as well as the organization. On the psychological level, emotional exhaustion leads to lower self-esteem, depression, nervousness, and irritability. On the physiological level, individuals might experience headaches, insomnia, stomach upset, and so on [15].

2. Method

2.1 Sample and Procedure

The current research focused on first-line employees, i.e., flight attendants, employed by domestic and international airlines in Taiwan (Far Eastern Air Transport Corp., Mandarin Airlines, UNI Ai, TransAsia Airways, EVA Air, and China Airlines). We distributed 500 questionnaires and received 380 responses. After eliminating defective questionnaires, we were left with 353 questionnaires, for an effective response rate of 70.16%. Because of time and resource limitations as well as the work characteristics of flight attendants and their variable schedules, we delivered the questionnaires to selected senior flight attendants and asked them to transmit the questionnaires to other attendants working for national and international airlines through their own company and other companies. We also asked these flight attendants to complete the questionnaire when they were on duty. Because the work time and airline schedules of flight attendants are not fixed, and because the shift schedules are arranged by computer, flight attendants are subject to arbitrary posting by the airline that trained them. Therefore, each flight attendant may work with a different flight team at any given time and may work with the same flight team over the short term only rarely. Thus, we provided the questionnaires without knowledge of which flight attendants would receive them. Therefore, this sampling method amounts to simple random sampling.

2.2 Hypotheses

H1: Flight attendants with different backgrounds have

significantly different feelings about emotional labor.

H2: Flight attendants with different backgrounds have significantly different feelings about emotional exhaustion.

H3: Flight attendants' emotional labor is positively correlated with emotional exhaustion.

H4: Flight attendants' emotional labor is predictive of emotional exhaustion.

2.3 Measures

2.3.1 Emotional Labor

Scale development: The first part of the emotional labor scale, cited as the "Organization Emotional Labor Scale" developed by Lin [16] especially for Taiwan, measures emotional labor. It adopts a 5-point Likert scale to score the items from 1 to 5 as absolute disagreement, disagreement, neutral, agreement, and absolute agreement, respectively. The higher the total score, the greater the employee's level of emotional labor. The original emotional labor scale had 24 questions divided into five themes: basic emotional expression, surface acting, deep acting, multiformity, and interpersonal interaction. Taking into account the test of professional validation as well as suggestions by experts, the descriptions of two questions in the original scale were modified. Finally, the authors designed the pre-questionnaire. Using SPSS, the authors conducted an item analysis on responses to the pre-questionnaire designed to test the scale questions. This research will adopt, as indices of internal consistency, the criterion method of discrimination analysis and correlative analysis on questions and the total score on a test of isomorphism type to analyze the pre-questionnaire.

Scale reliability and validation analysis: Analysis of the internal consistency of question 16 showed that the correlation coefficient of items and total scores was lower than 0.3 and that the discrimination was unacceptably low. Therefore, we deleted this question. The discrimination values of Question 15 and Question 17 were low, and they did achieve the level of significance. Therefore, the descriptions of these two questions were modified. Based on the item analysis, the emotional labor scale retained 23 questions. Cronbach's α calculated for the entire emotional labor scale was 0.883, and the reliability the entire scale achieves was over 0.7. The reliability value of the themes within the scale was calculated at 0.6, which is within the acceptable range. These results demonstrated the reliability of this emotional labor scale.

2.3.2 Emotional Exhaustion

Development of scale: The second part of the emotional exhaustion scale cites the third edition of the Maslach Burnout Inventory-General Survey (MBI-GS) developed by Maslach and Leiter [17]. There are five questions in this scale. This research adopts a 5-point Likert scale to score the items from 1 to 5 as absolute disagreement,

disagreement, neutrality, agreement, and absolute agreement, respectively. The higher the total score, the higher the degree of emotional exhaustion an employee feels. The original scale was tested by expert validation and we retained all six terms to develop the pre-questionnaire by referring to the experts' suggestions.

Scale reliability and validation analysis: Referring to the item analysis results, the correlations of all questions and the subscale of emotional exhaustion scale were above 0.3, achieving significance. In addition, the critical ratios were higher than 3 and achieved significance. Therefore, the five questions of the emotional exhaustion scale are all retained, in keeping with results of the item analysis. The Cronbach's α for the entire emotional exhaustion scale was 0.755, and the reliability of the entire scale was more than 0.7. Both are in the acceptable range, indicating that this emotional exhaustion scale is reliable.

3. Results

The effective sample of flight attendants was 353 respondents. The demographic data were as follows: Age: One hundred twelve flight attendants were 31-35 years old, accounting for 31.7% of the sample. Marital status: Two hundred attendants were single, accounting for 56.7% of the sample. Educational background: One hundred seventy-one attendants had graduated from junior college, accounting for 48.4% of the sample. Service year: One hundred eighty-four attendants had served for three to seven years, accounting for 51.1% of the sample. Average salary: One hundred sixty-two flight attendants reported salaries ranging from NTD60,000 to NTD70,000, accounting for 45.9% of the sample. The data regarding flight attendants' emotional labor and emotional exhaustion are displayed in Table 1.

There were 23 questions on the emotional labor scale. Results showed an average emotional labor score of 4.15 ($SD=0.392$) on the 5-point Likert scale; the average score

is thus moderately high. This indicates that the flight attendants believe that they must obey the company's regulations regarding their emotional expression while working. Results indicate that these flight attendants engaged in a high level of emotional labor. Analyses of each element of emotional labor indicated that interpersonal interaction scored highest on the emotional labor scale ($M=4.39$, $SD=0.569$); next highest was basic emotional expression ($M=4.28$, $SD=0.482$); multiformity achieved the lowest score ($M=3.66$, $SD=0.642$). There were five questions on the emotional exhaustion scale. The average score over all questions was 3.81 ($SD=.528$). This result indicates that the feeling of exhaustion experienced by flight attendants in response to emotional labor is moderately high.

The current study examined whether different backgrounds cause differences in feelings about emotional labor. Background variables included age, marital status, educational background, service year, and average salary. Results are displayed in Table 2:

Age: Flight attendants of different ages showed significant differences in their perspectives on emotional labor. According to the post hoc analysis using Scheffe's test, feelings regarding surface acting were experienced more strongly by flight attendants 20-25 years of age, and those older than 41, than by those 26-30 years old. Feelings about deep acting were stronger among attendants 20-25 years old, 31-35 years old, and older than 41 years old than among those 26-30 years old. Feelings of multiformity were more profound among flight attendants 31-35 years old than among 36-40 year-old attendants. The degree of emotional exhaustion of flight attendants 36-40 years old was more pronounced than that among those 20-30 years old. In addition, feelings of emotional exhaustion were stronger among flight attendants 31-35 years old than among attendants 20-25 years old. And, finally, flight attendants older than 41 years had the strongest feelings about emotional exhaustion.

Table 1. Abstraction of the analysis on domestic flight attendants' emotional labor and emotional exhaustion measured on a 5-point Likert scale (N=353)

Name	Number of Questions	Minimal Value	Maximal Value	Average of Question	Standard Deviation of Question
Entire Emotional Labor	23	2.96	5.00	4.15	0.392
Basic Emotional Expression	7	2.86	5.00	4.28	0.482
Surface Acting	3	2.33	5.00	4.13	0.598
Deep Acting	7	2.71	5.00	4.15	0.531
Multiformity	3	1.67	5.00	3.66	0.642
Interaction	3	3.00	5.00	4.39	0.569
Entire Emotional Exhaustion	5	2.80	5.00	3.81	0.528

Table 2. Analysis of the difference between the basic attribute of research object and research variable (n=353)

Item	Age (F Value)	Marital Status (F Value)	Educational Background (F Value)	Service Year (F Value)	Average Salary (F Value)
Entire Emotional Labor	4.004*	2.445	0.531	2.248	2.928*
Basic Emotional Expression	0.771	1.621	0.527	1.729	3.776**
Surface Acting	4.868**	0.462	1.011	2.780*	4.745*
Deep Acting	9.083***	4.928**	0.204	3.071*	2.636*
Multiformity	3.119*	0.290	2.620	1.552	0.798
Interaction	2.120	3.180*	0.454	3.441*	1.940
Entire Emotional Exhaustion	16.290***	7.798**	6.355***	10.803***	13.340***

*p < 0.05 significant level; **p < 0.01 extraordinary significant level; *** p < 0.001 absolutely significant level

Table 3. Abstract of the analysis on the correlation between emotional labor and the spectrum of the aspects of emotional exhaustion (N=353)

	1.	2.	3.	4.	5.	6.	7.
1. Spectrum of Emotional Labor							
2. Basic Emotional Expression	0.839 **						
3. Surface Acting	0.751 **	0.625 **					
4. Deep Acting	0.765 **	0.501 **	0.528 **				
5. Multiformity	0.338 **	0.032	0.110*	0.114 *			
6. Interaction	0.729 **	0.624 **	0.416 **	0.412 **	0.222 *		
7. Entire Level of Emotional Exhaustion	0.256 *	0.138**	0.138**	0.371 *	-0.055	0.211**	

Data source: Integrated by this research; * P<0.05; ** P<0.01 (Single tail)

Marriage: The deep acting of single flight attendants was more pronounced than that of married flight attendants. Additionally, the emotional exhaustion of flight attendants who selected the term “other” to describe their marital status was stronger than that reported by married flight attendants.

Educational background: The feeling of emotional exhaustion among flight attendants who had masters degrees was greater than among those who had graduated from high school, junior college, or college.

Service years: Flight attendants who had served for three to seven years reported stronger feelings of emotional exhaustion than did those who had served fewer than three years. Correlatively, the feeling of emotional exhaustion reported by flight attendants who had served more than 15 years was more pronounced than that reported by flight attendants who had served for three to 11 years.

Average salary: Flight attendants with an average salary of NTD90-100,000 expressed more marked feelings regarding the overall level of emotional labor than did flight attendants with an average salary of NTD60-90,000. Feelings regarding the basic emotional expression expected of flight attendants with an average salary of NTD90-100,000, in turn, were more marked than those

expressed by those with an average salary of NTD80-90,000. Feelings of engaging in deep acting were stronger among flight attendants with an average salary of NTD90-100 than among those whose average salary was less than NTD60,000. Finally, feelings about emotional exhaustion expressed by flight attendants whose average salary was NTD90-100,000 were stronger than those reported by flight attendants with an average salary of less than NTD60,000, and those with an average salary of NTD60-90,000.

In combination, these results indicate that flight attendants from different backgrounds have significantly different experiences regarding emotional labor and emotional exhaustion. Thus, Hypothesis 1 and Hypothesis 2, which posited that existence of such differences in experience based on background variables, are both partially supported.

Referring to Table 3, the measure of overall emotional labor showed a significant positive correlation with each element ($P < 0.01$) within the emotional labor measure, and the score for overall emotional labor showed a significant positive correlation with emotional exhaustion ($p < 0.05$). Among the elements of emotional labor, only deep acting showed a significant positive correlation with emotional exhaustion ($p < 0.05$). The other elements did

Table 4. Abstract of the stepwise multiple regression analysis on each aspect of emotional labor for emotional exhaustion (N=353)

Selective Sequence of Variables	Multiple Correlation Coefficient (R)	Determination Coefficient (R ²)	R ² after Regulation	F Value	Original Regression Coefficient (β)	Standardized Regression Coefficient
Intercept (Constant Term)					2.526	
Deep Acting	0.371	0.137	0.135	55.871***	0.380	0.382***
Multiformity	0.383	0.147	0.142	30.141***	-0.081	-0.099*

*p < 0.05 significant level; *** p < 0.001 absolutely significant level

not reach significance. The correlations among these perspectives are described as follows:

The score for overall emotional labor showed a significant positive correlation with each element of emotional labor ($P < 0.01$). Furthermore, overall emotional labor showed a significant positive correlation with overall emotional exhaustion ($p < 0.01$). Among the elements of emotional labor, only multiformity failed to show a significant correlation with emotional exhaustion ($r = -0.055$, $p = 0.15 > 0.05$); the other elements of emotional labor all showed significant correlations with emotional exhaustion. Finally, the overall measure of emotional labor showed a significant positive correlation with the overall measure of emotional exhaustion. Thus, Hypothesis 3, which predicted just such a relationship, was also supported.

Referring to the stepwise multiple regression analysis shown in Table 4, the two perspectives labeled deep acting and multiformity (from among the five elements of emotional labor) predicted criterion variables in the regression model. The multiple correlation coefficient, R , was 0.383, and its united explanation variance was 0.147, indicating that these two variables explain 14.7% of the entire emotional exhaustion measure. Referring to the results above, deep acting and multiformity have significant positive and negative predictive capability for emotional exhaustion, respectively. Therefore, Hypothesis 4, which posited that flight attendants' emotional labor would have a significant predictive effect on emotional exhaustion, was also partly supported.

4. Conclusions

Referring to Table 1, it appears that domestic flight attendants' feelings about emotional labor can be described using a 5-point Likert scale, with the average score being 4.15. Thus, flight attendants' average feeling about the role of emotional labor in their lives is moderately strong. This result indicates that flight attendants can be classified as workers who engage in a high level of emotional labor. Regarding feelings about various forms of emotional labor, the element of emotional labor having to do with interaction showed the strongest effect in the present findings. Compared to colleagues in other departments,

flight attendants had face to face contact with clients, and had conversations for longer periods, and with higher frequency. Therefore, flight attendants are also first-line service people in the airline industry [1]. Adelman's research [18] mentioned that worker performance and satisfaction were lower among those engaging in high levels of emotional labor than among workers with less emotional labor. High levels of emotional labor can easily generate feelings of unhappiness, lack of self-respect, and depression. Therefore, aviation management organizations must consider whether long-term emotional labor will have negative emotional effects on flight attendants.

Flight attendants' feelings about emotional exhaustion can also be described using a 5-point Likert scale, with an average score of 3.81. This indicates that, on average, flight attendants in this study experienced a moderate level of emotional exhaustion. More specifically, examination of the detailed questions regarding emotional exhaustion revealed that the highest average score was given to the statement, "The whole work day makes me feel tired." this was followed in importance by the statement, "My work makes me feel tired emotionally." Thus, most flight attendants reported that they felt tired when working for a long period of time. This research highlights the consequences of the tiring nature of flight attendants' working environment. Take offs and landings are frequent for domestic airlines, which means that flight attendants must repeatedly provide customer service. Thus, it is easy for flight attendants to feel that their work is dull and repetitive. Flight attendants serving international airline routes have to face a variety of passengers with distinctive needs; thus, the company is stricter about the entire range of services. All of these factors contribute to flight attendants' feelings of exhaustion when working for a long period of time.

Referring to Table 2, findings addressing differences among flight attendants with different background variables and the relationships of these variables to emotional labor and emotional exhaustion indicate that all background variables play a role. Relationships between emotional labor and variables of age, marital status, service years, and average salary all reached significance. Flight attendants with longer service years have deeper feelings

about emotional labor. This may be because service year is directly related to a particular level of customer service. New attendants may mainly serve in economy class, those with intermediate service years may serve business class, and those with the most service years may work in first class. Because the passengers in first class have high social status, the flight attendants must be especially attentive; they have to concentrate on their movements, emotional expressions, and so on. Therefore, we suggest that this is the reason that flight attendants with longer service years have deeper levels of emotional labor.

For feelings of emotional exhaustion among flight attendants with different population backgrounds, relationships between emotional exhaustion and the population variables of age, marital status, educational background, service years, and average salary all achieved significance. As regards educational background, our findings indicate that the higher the educational background of the flight attendant, the higher are the requirements for their work performance. Cordes and Dougherty [10] pointed out that high work achievement motivation and serious attention to work can result in high stress and can easily generate emotional exhaustion.

For the relationship between domestic flight attendants' emotional labor and emotional exhaustion, overall emotional labor scores showed a significant positive correlation with emotional exhaustion. This means that the stronger the feelings about emotional labor, the higher the emotional exhaustion. As regards the predictive effect on emotional exhaustion of the various elements of emotional labor, deep acting had the strongest value for predicting emotional exhaustion. This means that the stronger the feeling about emotional labor, the higher the level of emotional exhaustion. In particular, when flight attendants realize that the company always responds to customer complaints by supporting the client rather than the employee, the degree of deep acting is likely to increase. Therefore, airlines should consider carefully how to decrease their employees' feeling that they must engage in deep acting; they should also consider more carefully their treatment of employees, and whether they should be more understanding and tolerant of employees.

REFERENCES

- [1] S. J. Yuan, "The steward obeys the positive research of hardworking ability and passenger's satisfaction dependence," Department of Marketing and Distribution Management National Kaohsiung First University of Science and Technology Master's thesis, 2005.
- [2] H. C. Beh and P. McLaughlin, "Effect of long flights on the cognitive performance of air crew," *Perceptual and Motor Skills*, Vol. 84, pp. 319–322, 1997.
- [3] S. C. Liang and A. T. Hsieh, "Burnout and workplace deviance among flight attendants in Taiwan," *Psychological Reports*, Vol. 101, pp. 457–468, 2007.
- [4] A. R. Hochschild, "The managed heart: Commercialization of human feeling," Berkley, University of California Press, 1983.
- [5] A. S. Wharton, "The affective consequences of service work: Managing emotions on the job," *Work and Occupations*, Vol. 20, No. 2, pp. 205–232, 1993.
- [6] E. Goffman, "The Presentation of self in everyday life, Doubleday and Co," New York, NY, 1959.
- [7] J. A. Morris and D. C. Feldman, "The dimensions, antecedents, and consequences of emotional labor," *Academy of Management Review*, Vol. 21, No. 4, pp. 986–1010, 1996.
- [8] M. J. Morris and D. C. Feldman, "Managing emotions in the workplace," *Journal of Managerial Issues*, Vol. 9, No. 1, pp. 257–274, 1997.
- [9] M. J. Saxton, J. S. Phillips, and R. N. Blakeney, "The changing nature of work," San Francisco: Jossey-Bass, pp. 365–410, 1991.
- [10] C. L. Cordes and T. W. Dougherty, "A review and integration of research on job burnout," *Academy of Management Review*, Vol. 18, No. 4, pp. 621–656, 1993.
- [11] S. E. Jackson, R. Schwab, and R. S. Schuler, "Toward an understanding of the burnout phenomenon," *Journal of Applied Psychology*, Vol. 71, No. 1, pp. 630–640, 1986.
- [12] C. Maslach, "Burnout: The cost of caring," Englewood Cliffs, NJ: Prentice-Hall, 1982.
- [13] C. Maslach and S. E. Jackson, "Maslach burnout," San Francisco, CA: Jossey-Bass, 1986.
- [14] A. Rafaeli and R. I. Sutton, "The expression of emotion in organizational life," *Research in Organizational Behavior*, Vol. 11, No. 1, pp. 1–42, 1989.
- [15] L. C. Wang, "Role's pressure influences the course discussion that mood exhausts," Department of Business Administration, Institute of Behavior Medicine, National Cheng Kung University, Master's thesis, 1999.
- [16] S. P. Lin, "Organize the development of the form of amount of load of labor service of mood," *Sun Yat-Sen Management Review*, Vol. 8, No. 3, pp. 427–447, 2000.
- [17] C. Maslach and M. P. Leiter, "The truth about burnout," San Francisco, CA: Jossey-Bass, 1997.
- [18] P. K. Adelman, "Emotional labor and employee well-being," Doctoral Dissertation, The University of Michigan, 1989.

Towards a More Business-Oriented Definition of Corporate Social Responsibility: Discussing the Core Controversies of a Well-Established Concept

Matthias S. Fifka

School of Business and Economics, Friedrich-Alexander-University Erlangen-Nuernberg, Erlangen, Germany.
Email: fifka@wiso.uni-erlangen.de

Received September 6, 2009; revised October 19, 2009; accepted November 23, 2009.

ABSTRACT

The concept of Corporate Social Responsibility (CSR) has been controversially discussed for over 50 years. Consequently, a wide variety of definitions and understandings of CSR have been developed throughout the decades. That has made it increasingly hard, or not to say impossible, to agree on a common perception of CSR. Concerning the various notions of CSR, four core controversies can be identified which revolve around certain elements of CSR: First of all, there is the underlying question if CSR is the business of business or if it is none of its business as Friedman has famously argued. Second, should CSR contain legal obligations or is it a purely voluntary concept and, thus, ethical in nature? Strongly connected to that is the third controversy on whether CSR should be self-serving or if it has to be purely altruistic. Finally, there is widespread disagreement on the scope of CSR. Does it have a local, community-oriented focus or should it address concerns of a wider geographical scope? These controversies are analyzed and discussed here with the aim of developing a definition of CSR that does not remain confined to the academic world.

Keywords: Corporate Social Responsibility, Corporate Citizenship, Corporate Social Performance, Corporate Community Involvement

1. Introduction

While the idea that businesses should voluntarily contribute to the well-being of the communities where they operate dates back far into the 19th century, the scientific discussion of that phenomenon began in the 1930s. E.M. Dodd [1] and Chester Barnard [2] asked whether the executives of large corporations had an obligation to society extending beyond daily business. Theodore Kreps [3] even developed a framework for measuring the social involvement of businesses in his widely regarded book *Measurement of the Social Performance of Business*. The key term “social responsibility” was finally introduced in 1953 by Howard Bowen in his path-breaking work *Social Responsibilities of the Businessman* [4]. Aside from being the first to use the term, Bowen extensively elaborated on the subject and thus can be called “the father of corporate social responsibility” [5].

Today, more than 50 years after its introduction by

Bowen¹, “corporate social responsibility” (CSR) has become a highly popular term among scholars and practitioners alike. However, with the increasing usage and an ever growing number of publications, the meaning of the term has become increasingly blurred. In this context, it deserves mentioning that already in 1975 the *Handbook of Corporate Social Responsibility* remarked: “It seems that everyone has his/her own definitions for terms like [...] Corporate Social Responsibility, Public Affairs, Community Relations, Urban Affairs and Corporate Responsibility” [6]. The intensifying discussion among academics, especially in the 1990s, has not lead to any more clarity of the meaning of “corporate social responsibility.” Consequently, the World Business Council for Sustainable Development – a CEO-led, global association of more than 200 multinational companies – observed that “no universally acceptable definition of CSR exists” and that the “lack of an all-embracing definition” [7] will most likely persist.

This lack of consistency can mainly be attributed to three reasons. First, scholars from very different fields – business, economics, law, sociology, philosophy, and

¹It must be noted that Bowen only used the term “social responsibility,” but exclusively applied it to corporations. Therefore, the introduction of the concept of “corporate social responsibility” can be attributed to him.

even theology – have become increasingly interested in CSR. Inevitably, a coherent scientific discussion has become more difficult since different understandings, aims and research methods frequently collide. Second, many related terms have been created over the decades which sometimes serve as synonyms and at other times they describe different concepts. Among the most widely used are: corporate societal responsibility, corporate citizenship, corporate social responsiveness, business ethics, sustainable development, triple bottom-line, and corporate community involvement. Although many more could be mentioned, it should be remarked that next to CSR, corporate citizenship (CC) without question has become the most popular among scientists and practitioners alike. Third and most important, the understanding of CSR is inevitably based on an underlying view of what role businesses play or should play in society and how they can fulfill that role. With diverging core conceptions of the business-society relationship, the understanding of CSR will automatically vary, too.

In order to understand the different notions and definitions of corporate social responsibility it is necessary to look at their development over the course of time in a first step. Here, related concepts, especially CC, will be taken into account as well, because they reflect different approaches taken by scholars and practitioners. In a second step, the central controversies on the meaning of CSR will be isolated and discussed with regard to what they imply for businesses. This serves as a basis for the development of a definition of CSR that is applicable for businesses. As long as definitions and concepts of CSR remain mostly abstract and thus confined to the academic discussion, they are hardly useful in the long run and do not contribute to what all of them actually demand: the obligation of corporations to take on social responsibilities.

2. The Evolution of CSR Rhetoric Since 1953

When Howard Bowen provided a first vague definition of CSR in 1953, he stated that it “refers to the obligations of businessmen to pursue those policies, to make those decisions, or to follow those lines of action which are desirable in terms of the objectives and values of our society” [4]. Although Bowen’s definition is abstract in the sense that it is hardly possible to determine the objectives and values of any society, he leaves no doubt that businessmen have an obligation to consider social objectives and values when they make decisions. Ten years later, Joseph McGuire argued in the same vein, although he did not refer to individual “businessmen” any longer, but regarded the corporation as a whole: “The idea of social responsibilities supposes that the corporation has not only economic and legal obligations but also certain

responsibilities to society which extend beyond these obligations” [8]. While McGuire is more precise than Bowen in describing social responsibilities as extending beyond economic and legal obligations, he, too, assumes that a corporation has these obligations *per se*. This assumption can also be seen in a definition by Clarence Walton, which takes managers as individuals into account as well as the corporation as a whole: “The new concept of social responsibility recognizes the intimacy of the relationships between the corporation and society and realizes that such relationships *must* [my emphasis] be kept in mind by top managers as the corporation and the related groups pursue their respective goals” [9].

It is remarkable that all the early thinkers on CSR took for granted that corporations have certain social responsibilities and none of them spent considerable time to discuss that assumption. The first one to do so was Milton Friedman in 1970 in his now famous essay “The Social Responsibility of Business is to Increase Its Profits,” in which he vigorously denied any such responsibilities. Friedman argued that corporations are “artificial persons” [10] and thus cannot assume moral responsibility for what they do. Moreover, he stated that it is the sole obligation of managers to increase the wealth of the shareholders as it is their money they administer. Finally, he saw the social involvement of corporations as a violation of the government’s exclusive jurisdiction for social affairs.

Nevertheless, it must be noted that Friedman did not condemn social activities by a corporation as long as they were aimed at improving the corporation’s image in order to increase its sales and profits. This, however, would have to be counted as “window dressing” [10] and could not be classified as CSR according to Friedman, because he saw CSR based upon an altruistic and ethical foundation. This perception of CSR was by no means unique and reflects the dominant view on CSR in the late 60s and 70s. It was no coincidence that Friedman wrote his essay at a time when ideas of the “New Left” enjoyed widespread popularity and the American public became increasingly sceptic about large corporations [11]. He was very aware of that notion and observed a “present climate of opinion, with a wide spread aversion to ‘capitalism,’ ‘profits,’ ‘the soulless corporation’ and so on” [10]. This aversion would eventually lead to the public demand that corporations should contribute more to the social well-being because of a moral obligation to do so. In 1975, the *Handbook of Corporate Social Responsibility* also illustrated this view of CSR: “‘Corporate Social Responsibility’ seems to be the most favored term among reformers who use it as a moralistic rallying post. Anti-reform conservatives feel this generic icon smacks of do-goodism, and thus has no relationship to the hard-nosed profit motive of business” [6].

The notion that CSR includes ethical responsibilities was cemented by Carroll's "Four-Part Model of Corporate Social Responsibility" from 1979, which probably has become "the most established and accepted model of CSR" [12]. He stated that the social responsibility of business "encompasses the economic, legal, ethical, and discretionary expectations that society has of organizations at a given point in time" [13]. Carroll later revised this definition slightly and replaced "discretionary expectations" with the term "philanthropic" [14]. Finally, intending to use a more pragmatic language, he once again revised his definition: "The CSR firm should strive to make a profit, obey the law, be ethical, and be a good corporate citizen" [14].

That rephrased definition is interesting in the respect that "being a good corporate citizen" becomes equal with "philanthropy," replacing it in the definition. Therefore, Carroll's concept of CC is a very narrow one, basically reducing it to "corporate giving." Thus, it is no surprise that practitioners came to prefer this limited and easy approach to CC over the "ethically and morally loaded" idea of CSR. Until today, CSR has carried the notion of being a concept which assumes an ethical obligation of businesses to "do good" without pursuing a material self-interest. Some scholars have even come to regard socially responsible behavior as the primary objective of business: "Finally, assuming that a business firm has enhanced socially responsible discretion, obeyed society's major ethical principles, and followed the law, then the firm is responsible for producing and distributing goods and services, enhancing shareholder wealth [...]" [15].

Moreover, in comparison to CSR, CC did not emphasize the responsibilities of corporations one-sidedly, but also implied rights that emerge from the status as a citizen. It also did not demand that CC had to be of an altruistic nature. Instead, it assumed that the social activities of a business would be expanded if a business profited from them, because that way a lasting and structured engagement was guaranteed. In this context, Wood and Logsdon correctly observe that CC stresses a " 'give-back' and 'tit-for-tat' grounding" which "does not require any philosophy nor deep thinking," while CSR has a strong "moral grounding" and "a 'Sixties-liberal' orientation" [16].

However, the notion that CC is of a purely pragmatic nature without an ethical foundation has increasingly been challenged since the turn of the century. In 2001, David Birch noted that "corporate citizenship is not simply about philanthropy, corporate generosity, business community partnerships, executive leasing to community organizations, cause-related marketing, good causes and so on (though these may well be some of the concrete realizations of corporate social responsibility) – it is

about a changing business ethos" [17]. Definitions such as these [18,19], stressing a moral component of CC, have made it increasingly hard to differentiate between CSR and CC. Although CC was originally invented to express a distinction from CSR, a trend has become observable during the last decade that emphasizes a moral-ethical component of CC. Consequentially, CC and CSR are increasingly used interchangeably. Maignan, Ferrell and Hult, for example, recently defined CC as "the extent to which businesses meet the economic, legal, ethical, and discretionary responsibilities placed on them by their various stakeholders" [20], which is almost identical to how Carroll had defined CSR thirty years ago. For good reason, Matten and Crane [21] thus speak of an "equivalent view" of CSR and CC that today can frequently be found among academics and practitioners.

The more traditional view of seeing CC as corporate giving or philanthropy, which Matten and Crane refer to as "limited view of CC," has become less popular among academics and practitioners alike. Although mere corporate giving is easy to handle and does not require an elaborate strategy or long term involvement, executives have become less favorable towards it, because it does not contain concepts of how corporations can actually benefit from this form of social responsibility. Instead, more and more businesses are trying to implement diverse forms of social activities into their business strategy. This is why the Center for Corporate Citizenship at Boston College has defined CC as "the business strategy that shapes the values underpinning a company's mission and the choices made each day by its executives, managers and employees as they engage with society" [22]. This clearly emphasizes the notion that CC should self-serve businesses and does not have to be altruistic by any means. The idea that businesses and society can both benefit from social activities at the same time has been expressed most clearly by Windsor: "Although corporate citizenship embeds older traditions of corporate social responsibility and responsiveness, fundamentally it crafts an instrumental, self-serving view of the relationship between business and society" [23].

Although one might assume that executives today would thus exclusively prefer the term "corporate citizenship" over the moralistic "corporate social responsibility", this is not the case. While it is true that the term CC was mainly coined by practitioners to express a distinction from CSR, the latter term is still being used prominently by many corporations to describe their social activities. It comes as no surprise that empirical research has shown that executives remain uncertain about what the terms actually mean and that they find it hard to differentiate between them [24]. Although there is a vivid academic discussion about the different potential meanings of CC and CSR [16,21,25], in the executive world

both terms are being used to describe the same ideas and forms of activity.

However, regardless of the term that is being applied, the core controversies that surround CSR or CC are identical and shall be addressed in the next chapter. Paying tribute to the facts that there is no established differentiation between CSR and CC and that both are used synonymously by many scholars and businesses to describe the same phenomena, an equivalent view is taken here. Thus, for reasons of simplicity, only the expression CSR will be applied in the following.

3. Core Controversies on CSR

Beginning with Friedman's essay in 1970, a vivid discussion on the characteristics of CSR has erupted. Four central controversies can be determined in this context:

- Should businesses take over social responsibilities or is that none of their "business"?
- Assuming that they have social responsibilities, should there be legal obligations to enforce CSR or is it a concept that is of voluntary nature and may be built on a moral foundation? Does CSR also contain an economic responsibility towards society?
- Strongly connected to the dispute on its voluntary character is the question, if CSR should be free of self-interest or if it may serve the aims of businesses as well?
- Finally, there is dispute on the scope of CSR? Should CSR activities be targeted only at the local environment of a business or should they be of a broader scope and address more general problems?

The debate on these central controversies among scholars as well as practitioners is addressed in the following chapters and the implications that the different views have for businesses are discussed.

3.1 Is CSR the Business of Business?

As mentioned above, in the 50s and 60s the notion became popular, at least among scholars, that businesses and executives should take over social responsibilities. In the 70s, Friedman was the first to challenge that perception. It is interesting to note that today, almost 40 years after Friedman's famous essay was published, the perception that businesses should act socially responsible remains once again largely unquestioned in the academic world. Not only that; most scholars take a certain social responsibility for granted and do not spend any time or word in their publications to address the important question if businesses can really be assumed to carry that kind of responsibility. The opponents of CSR, who are very few in number today, still mostly refer to Friedman's arguments that businesses should act not socially responsible because of three reasons. First, they are artificial

persons who cannot be held morally responsible. Second, the managers as agents have to exclusively serve the shareholders as principals and increase their wealth. CSR is adverse to that mission as it solely causes costs which reduce profits. Third, the maintenance and regulation of the social system is solely the job of the government.

Although Friedman certainly makes important points, there are several arguments that can be held against him. First of all, the assumption that businesses as artificial actors cannot be held morally responsible because all decisions connected to them are made by individuals is an oversimplification. One could argue that although decisions in a business are being made by individuals, it is still the business that is held legally responsible for the outcome of the decisions most of the time. Hence, if businesses can take legal responsibility for actions initiated by individuals, why should they not be able to take moral responsibility as well? Moreover, all businesses have a certain structure and culture as well as superordinate goals that transcend individual decision-making. They provide a framework to which all individuals within the business are bound and establish "an explicit or implicit purpose" [12] for their decisions. Therefore, businesses also carry a moral responsibility for the decisions that are made by individuals within this framework.

The second argument and underlying perception that CSR inevitably increases costs and reduces profits dominated the CSR discussion for a long time. In 1972, Manne and Wallich even argued that business expenditures could only be counted as socially responsible, if they produced lower profits than alternative investments which paid no attention to CSR [26]. One year earlier, Johnson had argued that only businesses which had already met their economic goals would set aside money for CSR [27]. This so called "lexicographic" view of CSR also implied that CSR was a financial burden. This notion slowly began to change with Drucker's argument that profitability could not only be maintained despite CSR activities, but that it could actually be increased, because CSR creates new opportunities for businesses. According to him, social problems should be seen as chances for new fields of activity and thus be turned into profits [28]. Drawing on that idea, scholars set out to develop theories on how to integrate CSR activities into the business strategy [29,30].

Over the decades, numerous studies have also been conducted to examine the relation between corporate financial performance (CFP) and corporate social performance (CSP). The latter concept had already been developed in the 1970s [13,31] and was refined in the 1980s and 1990s. It tried to offer a framework for executives how to first implement and then measure the outcomes of CSR. In 1994, Wood defined CSP accordingly as "a business organization's configuration of principles

of social responsibility, processes of social responsiveness, and policies, programs, and observable outcomes as they relate to the firm's societal relationships" [15]. The concept of CSP – just as CSR – has also suffered from inconsistencies and very diverging understandings of it. Therefore, it comes as no surprise that studies which examine the relation between CSP and CFP come to highly different conclusions. In 1999, Roman et al. conducted a meta-survey of 52 studies and found that 33 suggested a positive relationship between CSP and CFP, 14 found no effect and five found that CSP negatively affects CFP [32]. This rather positive overall conclusion could not be fully affirmed by a meta-analysis of Mahon and Griffin [33]. Of the 62 studies they evaluated, 33 found a positive correlation, nine no correlation at all, and 20 a negative one. It is interesting to note that all authors observed a highly heterogeneous understanding of CSR and CSP throughout the studies they examined, which makes valid comparisons nearly impossible.

Nevertheless, despite the varying understanding displayed in the individual studies, more than half in both meta-analyses found CSR to increase financial performance. Therefore, Friedman's assumption that CSR measures would inevitably decrease profitability cannot be maintained, which makes his second argument obsolete. Moreover, it cannot only be argued that CSR activities are likely to increase profitability, it can also be brought forward that not to act socially responsible might decrease profitability. Consumers have come to expect that businesses "behave" accordingly and retaliate with protests and boycotts if they do not: Exxon, e.g., had to face extensive consumer boycotts when it refused to sign the Kyoto Protocol; Nike experienced declining sales when its employment of childhood labor in Asia became public; and consumers avoided Shell gas stations when the company intended to sink its swimming oil tank "Brent Spar" into the North Sea instead of having it dismantled on shore. The latter case is especially interesting since Shell had acquired all necessary permission for deep sea disposal and thus would not have acted illegally. However, alarmed by Greenpeace activists, the public did not consider mere compliance with legal norms to be sufficient and eventually forced Shell, which at the same time experienced a sharp drop in sales and share prices [34], to abandon its plans.

The public has come to expect businesses to do more than to simply act accordingly to the laws. While legal compliance is *sine qua non*, it alone is not sufficient any more in times when consumers believe that businesses are more than mere providers of goods and jobs. This leads us to Friedman's third argument against CSR, which was based on the assumption that it is the government's exclusive duty to provide and regulate the social system. Already at the time of publication, Fried-

man's position did not reflect a dominant understanding that saw social support solely as any longer. The public, especially in the US, had already begun to articulate mistrust of businesses because of perceived "corporate greed" and demanded more social involvement [11]. Today, almost 40 years later, that notion has become even stronger as governments are increasingly unable to solve social or ecological problems. With rising government debt, a smaller and smaller portion of the budget remains available for social spending, the rest flows into predetermined or mandatory spending, especially debt service [35]. Therefore, governments have to rely more on more upon private actors to take over social responsibilities in order to provide certain services and support for those in need. Friedman's argument that the social system should be the responsibility of the government only thus has become obsolete, as governments in most industrialized nations are no longer able to carry that responsibility alone.

Overall, Friedman's core argument that the "social responsibility of business is to increase its profits" cannot be maintained. Businesses can very well be considered to be moral actors that can assume social responsibility. That in turn does not automatically decrease their profits; quite the contrary is the case: the negligence to act socially responsible is often sanctioned by the public through protests and boycotts leading to reduced revenues. The public has come to expect businesses to be socially engaged as governments cannot shoulder social responsibility on their own any longer.

The question which inevitably arises in the governmental context is whether CSR should be enforced upon businesses through legal provisions or if it is a concept that is built upon voluntary participation.

3.2 CSR – Economic and Legal Obligation or Voluntary Involvement?

The term "CSR" by itself does not allow any conclusions if the corresponding action has to be voluntary in nature or if it has to be enforced legally. Social responsibility can either be assumed voluntarily or because of legal provisions.

Until today, many authors have followed Bowen's original idea that CSR was primarily a voluntary concept. That perception was most clearly expressed by McGuire in 1963: "The idea of social responsibilities supposes that the corporation has not only economic and legal obligations but also certain responsibilities to society which extend beyond these obligations" [8]. McGuire assumes corporations to have certain economic and legal responsibilities, he emphasizes the "responsibilities to society which extend beyond these obligations." That would lead

to the conclusion that economic and legal obligations are a part of CSR and can be regarded as its foundation, with a voluntary or discretionary component resulting from a moral or ethical obligation on top. In 1991, Carroll expressed this perception in his “Pyramid of CSR,” which was based on his definition of 1979: “The social responsibility of business encompasses the economic, legal, ethical, and discretionary expectations that society has of organizations at a given point in time” [14].

At times, the necessity to include economic and legal obligations in a definition of CSR is questioned because these obligations could be taken for granted. The economic responsibility to be profitable can indeed be assumed, otherwise, a business will not be able to exist in a free market economy in the long run – unless it was heavily subsidized. The legal responsibility to obey the law is another story since businesses may very well benefit from not acting in accordance with the laws. One must only think of the corporate scandals in the first half of this decade, when companies profited from fraudulent accounting and false balance sheets. In reaction, tighter legal standards were enforced upon the corporations since the voluntary commitments made by industry groups obviously fell short.

Therefore, law abidance must be considered a vital element of CSR. However, this must not lead to the dangerous notion that merely obeying rules and regulations is sufficient. Without a voluntary component, the concept of CSR becomes obsolete. While certain legal standards are inevitable and necessary, the “added value” of CSR results from the voluntary involvement of businesses. They can decide best how and where to become active in the most efficient manner. Let us look at an example where a group of businesses decides to support the construction of a new playground in their neighborhood. One might argue that this construction could simply be financed by raising taxes or other levies and thus resolve social problems through governmental action. However, higher taxes might drive companies out of business that are not highly profitable or force them to lay off workers. Other might consider relocating their facilities. Moreover, politicians might be less inclined to raise taxes out of fear to lose the next election, and thus the necessary governmental funds are missing. Instead, a voluntary and discretionary approach can provide businesses with a variety of ways to contribute: A real-estate firm could set aside a lot which it owns; a construction company could provide workers or machinery which might be less burdensome for it than making a cash donation, while the contribution to the success of the project might be equal; and a law firm, for example, can give one of its lawyers some time off to take care of the necessary administrative procedures.

Furthermore, we must keep in mind that not all problems, of whatever kind they may be, can be solved through legal provisions. This would lead to an excessive legislative process creating a bloated and non-transparent body of regulation. Automatically, such a regulatory approach would create enormous costs for the development of the norms by governmental institutions and later for the businesses which have to comply with them.

Although it is evidently impossible to tackle all social problems through legal provisions, the effectiveness of voluntary approaches was doubted for a long time. Especially in the 70s and 80s, it was argued that reliance on the voluntary involvement of businesses as a part of CSR was naive because in the hard-nosed world of business, companies would and should not do anything which did not serve their interests unless it was required by the law [10]. This belief was strongly linked to the perception of CSR as – aside from its obligatory component – a concept that was altruistic and not self-serving in nature. But can businesses really be assumed to make meaningful social contributions if they do not receive anything in return?

3.3 Altruism or Self-Interest?

The early concepts of CSR developed by scholars like Bowen, Frederick or McGuire were normative in nature. From corporations they demanded social involvement which was motivated by the realization of an ethical responsibility and not by the pursuit of self-interests. This in turn did not mean the violation of economic principles to them, but the guarantee to reach a higher level of economic welfare for society as a whole. Frederick underlined this approach in 1960:

“Businessmen should oversee the operations of an economic system that fulfils the expectations of the public. And this means in turn that the economy’s means of production should be employed in such a way that production and distribution should enhance total socio-economic welfare.

Social responsibility in the final analysis implies a public posture toward society’s economic and human resources and a willingness to see that those resources are used for broad social ends and not simply for the narrowly circumscribed interests of private persons and firms” [36].

Ten years later, as already indicated, Friedman was the first to prominently challenge this position by stating that businesses had no social responsibilities as demanded by Bowen and his companions. To him, their concepts of CSR were pure socialism:

“The businessmen believe they are defending free enterprise when they declaim that business is not concerned ‘merely’ with profit but also with promoting desirable ‘social’ ends; that business has a ‘social conscience’ and

takes seriously its responsibilities for providing employment, eliminating discrimination, avoiding pollution and whatever else may be the catchwords of the contemporary crop of reformers. In fact they are – or would be if anyone took them seriously – preaching pure and unadulterated socialism. Businessmen who talk this way are unwilling puppets of the intellectual forces that have been undermining the basis of a free society these past decades [10].”

Nevertheless, it is interesting to note that Friedman did not object against the social involvement of businesses as long as they profited from it: “This is one way for a corporation to generate goodwill as a by-product of expenditures that are entirely justified in its own self-interest” [10]. Such measures, however, could not be counted as CSR in Friedman’s eyes, which to him only became a pretence for activities that were undertaken to foster the public perception of a company: “Of course, in practice the doctrine of social responsibility is frequently a cloak for actions that are justified on other grounds rather than a reason for those actions” [10].

With regard to terminology, this leads to the paradox situation that social involvement – in the eyes of Friedman – is desirable if it benefits business, but when it does, it cannot be classified as CSR, because CSR demands an altruistic motivation. Now, one might certainly ask if this terminological skirmish is of any importance to those who profit from social activities by businesses. Does it matter to the hungry whether a company provides food out of selflessness or because it wants to improve its image?

It may very well be pointed out that unselfish involvement is more efficient because it is focused on the needs of the recipients and does not have to be subordinate to a company’s goals. But it can be countered that it is far more likely that companies get engaged on a larger scale if they benefit from their social involvement as well. Only this way it can be guaranteed that social activities are conducted permanently and in a structured, well developed manner, and not only punctually if the current financial situation permits it.

While some business ethicists decline social activities driven by a self-serving intention as utilitarian exploitation [15,37], from an economic point of view it makes no sense to demand that CSR has to be altruistic. Nevertheless, one should not automatically assume the pursuit of mere self-interest behind every social activity of a company, although it is certainly possible to make such an allegation in any case. If a company supports schools and universities, one could allege that it solely tries to ensure the availability of qualified workers in the future. If it builds recreational facilities for its employees, it could be argued that it only tries to maintain a healthy workforce. Finally, the intention to polish up one’s own reputation

could always be insinuated.

From an empirical point of view, corporate motives for social involvement are difficult or even impossible to examine [21]. While the true conviction to help people in need may indeed be the grounding for social activities in some cases, this motivation would be mentioned more often than self-interest if one did a survey among companies, because everybody likes to presume to have acted unselfishly.

Overall, the conclusion is that individual corporate and overall social benefits can go hand in hand and are not necessarily diametrical to each other. It is exactly this “give and take” which serves as a viable driving force for social activities by businesses. It makes sense to assume that this “tit-for-tat,” as Wood and Logsdon call it [16], should be one of the bases of CSR, but one must be aware that the understanding of CSR in many publications still is – resulting from the term’s history – a different one. There CSR is often perceived as an altruistic concept. This narrow perception, however, limits the advantages resulting from a broader understanding, which allows self-interest behind CSR: A business is getting involved and tackles certain social grievances, and it is rewarded by improving the environment where it does business and its public image.

The final controversy that is to be addressed here is the scope and constitution of the business environment. Should CSR activities be narrow in scope and be targeted at the immediate environment of a business or should they be broader in nature and address more general problems?

3.4 CSR – Local Community Involvement or Global Activism?

Early definitions of CSR had left open the form and scope of engagement and remained rather vague as we have seen. Even later concepts that were specifically targeted at businesses often were of an abstract nature, like the one provided by the Committee on Economic Development (CED) from 1971 demonstrates. The CED used three concentric circles to explain CSR. The inner circle referred to economic tasks like providing jobs and goods and operating profitable. The second circle suggested that these economic aims must be pursued with sensitivity to changing social values and the social contract that exists between society and businesses. The outer circle emphasized the need to actively become involved in improving the social environment [38]. Carroll, for good reason, referred to that latter part of the concept as “amorphous” [14], because it still remained vague.

The lacking possibility to apply CSR and the abstract character of the respective definitions were major reasons for why the term “corporate citizenship” gained more prominence among practitioners as it emphasized prag-

matic involvement in the local community. Wood and Logsdon thus have correctly observed that CSR articulates a “broad concern for many issues and stakeholders, and for society at large,” while CC stresses a “narrow focus on local community and charity” [16]. Although definitions of CC have been increasingly widened [17,21,39] the term to many still has a more operable or “hands on” character and a local focus.

Certainly, the problem resulting from the abstract and all-embracing character of CSR definitions is that it is “too broad in its scope to be relevant for organizations” [40]. Henderson argues in the same vein and finds that there is “no solid and well-developed consensus which provides a basis for action” [41]. Therefore, CSR might mean nothing to businesses because they simply do not know how to make it operable.

Nevertheless, as tempting as a precise definition or concept may seem, there is a high risk of creating a framework that is far too rigid. First of all, as was shown above, the great strength of CSR lies in the fact that it gives businesses the possibility to become socially involved in many different ways. Second, the scope of CSR activities can hardly be predetermined because they are different from business to business. A small manufacturing company that relies on employees, suppliers and buyers from a specific region will take a different approach than a multinational corporation, because it will not be able to afford meaningful CSR on an international scale and also would not benefit from it.

With increasing globalization, however, it becomes harder and harder to precisely determine the business environment of a company. For many businesses, shareholders as well as suppliers, buyers and other stakeholders are distributed all over the globe, which would demand a concept of global CSR with specific national approaches as, e.g., legal obligations which have to be adhered to vary from country to country.

This inevitably leads to a definitional dilemma. On the one hand, definitions regarding the form and scope of CSR should not be too abstract, but at the same time broad enough not to rule out potential forms of activity and also pay tribute to the individual possibilities of businesses and their varying business environments. The solution lies in considering the possibilities of businesses to become socially involved, which vary considerably with regard to form and scope. The business environment – as the area where the stakeholders of a business are located – thus seems appropriate for determining the geographic scope of CSR. Although it will not be possible for a business to identify all of its stakeholders [42], the identification of the most important stakeholders provides businesses with a guideline on where to become active.

4. Conclusion A Business-Oriented Definition of CSR

From the examination of the central controversies on CSR, we can reach the following conclusions. First, CSR clearly is the business of business. Even when one assumes that businesses cannot or should not take on morally grounded obligations, CSR still remains the business of business because of a self-serving, elementary necessity. Not to act socially responsible is negatively rewarded today and no company can afford to incur the risk of being labeled an irresponsible member of society. Nevertheless, it can be questioned if the term CSR still applies to the behavior of a company that only acts socially responsible because of self-interest. The term itself undoubtedly carries a moral or ethical component, which, however, is of only limited importance to the practical outcome of CSR. It does not matter whether a business supports a social cause because of self-interest or because of a felt moral obligation as long as the support is of equal size and scope. There is only one solid argument for assuming an ethical foundation of CSR. Such an ethical basis will most likely lead a business to consider the social interests of its environment to a larger extend in relation to its own interests when planning CSR-activities. Overall, independent from an ethical or a purely pragmatic perspective, CSR in any case is the business of business.

Concerning the question whether CSR encompasses economic and legal obligations, it can be concluded from the deliberations above that both are fundamental for CSR. At first sight, the obligation to make responsible economic decisions could be regarded as redundant in a definition of CSR because it is a prerequisite for the permanent existence of a business in a free-market economy and thus could be taken for granted. However, businesses as members of a society also have the economic obligation to provide goods and employment which constitutes a social responsibility at the same time. While the negligence of economic responsibilities is unlikely due to the negative consequences for a business itself, violations of legal obligations such as tax fraud or violations of environmental standards might consciously be made, especially when they are unlikely to be detected or can be covered up by bribery and corruption. In these cases, a business can profit, but the overall effects on society in sum will be negative. Therefore, legal obligations also must form part of the fundament of CSR.

Despite its legal basis, CSR has to be understood as a mostly voluntary concept in order to provide an “added social value”. Mere compliance with legal obligations is not enough to tackle social problems which cannot be solved through regulatory measures alone. It is the strength of CSR that it gives businesses the possibility to

address such problems in their environment. In contrast to legal provisions, CSR can be especially effective due to its voluntary and discretionary component, because businesses can apply their unique resources to those problems where they can contribute most effectively. Therefore, a definition of CSR also has to consider that CSR works most effectively when businesses are given the possibility to take into account what kind of social involvement their own resources allow best.

This also applies to the geographic scope of their social involvement. Businesses should be expected to practice social responsibility in the markets where they operate. This will match the resources that they have available and also guarantees that they become socially involved in environments which they already know to a certain degree through their business activities.

In return, businesses can also expect to benefit the most from CSR-activities in environments where their suppliers, customers, employees and other stakeholders are located. It makes no sense to demand that CSR has to be of altruistic or selfless nature. Only if businesses benefit themselves from CSR in return, their involvement will be permanent and meaningful.

Taking these considerations into account, the following business-oriented definition of CSR can be developed:

“Corporate Social Responsibility encompasses the adherence to fundamental economic and legal obligations which a business encounters in the environments where it operates as well as the responsibility to voluntarily contribute to the social development of these environments in an adequate and structured manner that is in accordance with the resources available to each business and the underlying business strategy.”

This less abstract, geographically more focused and resource-oriented approach which gives room to a business's individual possibilities and goals is aimed at improving the understanding, acceptance and applicability of a concept that often has remained too abstract and too “morally loaded” in theoretical disquisitions. As a consequence, businesses have refrained from considering the implementation of comprehensive CSR-programs. However, the willingness to do so is a necessary and decisive factor in maintaining and improving the social welfare of nations in times when governments are less and less able to sustain that welfare on their own.

REFERENCES

- [1] E. M. Dodd, “For whom are corporate managers trustees?” *Harvard Law Review*, Vol. 45, No. 8, pp. 1365–1372, 1932.
- [2] C. Barnard, “The Functions of the Executive,” 30th Anniversary Edition, Cambridge, Harvard University Press, 1968.
- [3] T. Kreps, “Measurement of the social performance of business,” U.S. Government Printing Office, Washington, D.C., 1940.
- [4] H. R. Bowen, “Social responsibilities of the businessman,” Harper & Brothers, New York, 1953.
- [5] A. B. Carroll, “Corporate social responsibility – Evolution of a definitional construct,” *Business & Society*, Vol. 38, No. 3, pp. 268–295, 1999.
- [6] Human Resources Network, “The handbook of corporate social responsibility,” Chilton, Book Company Radnor, 1975.
- [7] World Business Council, “Meeting changing expectations,” 2000, <http://www.wbcsd.org/DocRoot/hbdf19Txhmk3kDxBQDWW/CSRmeeting.pdf>.
- [8] J. McGuire, “Business and society,” McGraw Hill, New York, 1963.
- [9] C. Walton, “Corporate social responsibilities,” Wadsworth, Belmont, 1967.
- [10] M. Friedman, “The social responsibility of business is to increase its profits,” *New York Times Magazine*, pp. 32, 13 September 1970.
- [11] S. M. Lipset and W. Schneider, “The confidence gap – Business, labor, and government in the public mind,” The Free Press, New York, 1983.
- [12] A. Crane and D. Matten, “Business ethics – A European perspective,” Oxford University Press, Oxford, 2004.
- [13] B. Carroll, “A three dimensional conceptual model of corporate performance,” *Academy of Management Review*, Vol. 28, No. 2, pp. 446–505, 1979.
- [14] B. Carroll, “The pyramid of corporate social responsibility: Toward the moral management of organizational stakeholders,” *Business Horizons*, Vol. 34, No. 4, pp. 39–48, 1991.
- [15] D. J. Wood, “Business and society,” Harper Collins, New York, 1994.
- [16] D. J. Wood and J. M. Logsdon, “Theorizing business citizenship,” In J. Andriof and M. McIntosh (Eds.), *Perspectives on Corporate Citizenship*, Greenleaf, Sheffield, pp. 83–103, 2001.
- [17] D. Birch, “Corporate citizenship – Rethinking business beyond corporate social responsibility,” In J. Andriof and M. McIntosh (Eds.), *Perspectives on Corporate Citizenship*, Greenleaf, Sheffield, pp. 53–65, 2001.
- [18] K. Davenport, “Corporate citizenship: A stakeholder approach for defining corporate social performance and identifying measures for assessing it,” *Business and Society*, Vol. 39, pp. 210–219, 2000.
- [19] J. Andriof and M. McIntosh (Eds.), “Perspectives on corporate citizenship,” Greenleaf, Sheffield, 2001.
- [20] Maignan, O. C. Ferrell, and G. T. Hult, “Corporate citizenship: Cultural antecedents and business benefits,” *Journal of the Academy of Marketing Science*, Vol. 27, No. 4, pp. 455–469, 1999.
- [21] D. Matten and A. Crane, “Corporate citizenship: Toward

- an extended theoretical conceptualization," *Academy of Management Review*, Vol. 30, No. 1, pp. 166–179, 2003.
- [22] The Center for Corporate Citizenship, "Corporate social responsibility statements," 2009, <http://www.bccccc.net/index.cfm?fuseaction=Page.viewPage&pageId=596&parentID=473>.
- [23] D. Windsor, "Corporate citizenship – evolution and interpretation," In J. Andriof and M. McIntosh (Eds.), *Perspectives on Corporate Citizenship*, Sheffield: Greenleaf, pp. 39–52, 2001.
- [24] M. Abshire, "State of corporate citizenship: Survey," *Corporate Philanthropy Report*, Vol. 18, No. 8, pp. 1–3, 2003.
- [25] Valor, "Corporate social responsibility and corporate citizenship: Towards corporate accountability," *Business and Society Review*, Vol. 110, No. 2, pp. 191–212, 2005.
- [26] H. G. Manne and H. C. Wallich, "The modern corporation and social responsibility," *American Enterprise Institute for Public Policy Research*, Washington D.C., 1972.
- [27] H. L. Johnson, "Business in contemporary society: Framework and issues," Wadsworth, Belmont, CA, 1971.
- [28] P. F. Drucker, "A new look at corporate social responsibility," *McKinsey Quarterly*, No. 4, pp. 17–28, 1984.
- [29] M. E. Porter and C. van der Linde, "Green and competitive: Ending the stalemate," *Harvard Business Review*, Vol. 73, pp. 120–134, September/October 1995.
- [30] S. L. Hart, "A natural-resource-based view of the firm," *Academy of Management Review*, Vol. 20, No. 4, pp. 986–1014, 1995.
- [31] S. P. Sethi, "Dimensions of corporate social performance: An analytic framework," *California Management Review*, Vol. 17, No. 1, pp. 58–64, 1975.
- [32] R. M. Roman, S. Hayibor, and B. Agle, "The relationship between social and financial performance – Repainting a portrait," *Business and Society*, Vol. 38, No. 1, pp. 109–125, 1999.
- [33] F. Mahon and J. J. Griffin, "Painting a portrait – A reply," *Business and Society*, Vol. 38, No. 1, pp. 126–133, 1999.
- [34] P. Owen and T. Rice, "Decommissioning the Brent Spar," London: Spon Press, 1999.
- [35] U. Schrader, "Corporate citizenship – Die Unternehmung als guter Bürger?" Logos, Berlin, 2003.
- [36] W. C. Frederick, "The growing concern over business responsibility," *California Management Review*, Vol. 2, pp. 54–61, 1960.
- [37] P. Ulrich, "Integrative Wirtschaftsethik: Grundlagen einer lebensdienlichen Ökonomie," Haupt, Bern, 2008.
- [38] Committee for Economic Development, "Social responsibilities of business corporations," New York, NY, A statement on national policy by the Research and Policy Committee for Economic Development, June 1971.
- [39] Seitz, "Corporate citizenship – Rechte und Pflichten der Unternehmung im Zeitalter der Globalität," Deutscher Universitäts-Verlag, Wiesbaden, 2002.
- [40] Banerjee, "Corporate citizenship and indigenous stakeholders – exploring a new dynamic of organisational-stakeholder relationships," *The Journal of Corporate Citizenship*, No. 1, pp. 39–55, 2001.
- [41] Henderson, "Misguided virtue," *False Notions of Corporate Social Responsibility*, New Zealand Business RoundTable, Wellington, 2001.
- [42] R. E. Freeman, "Strategic management: A stakeholder approach," Pitman, Boston, 1984.

Combining Personal Ontology and Collaborative Filtering to Design a Document Recommendation System

Deng-Neng CHEN, Yao-Chun CHIANG

Department of Management Information Systems, National Pingtung University of Science and Technology, Taiwan, China.
Email: dnchen@mail.npust.edu.tw

Received July 31, 2009; revised September 15, 2009; accepted October 24, 2009.

ABSTRACT

With the advance of information technology, people could retrieve and manage their information more easily. However, the information users are still confused of information overloading problem. The recommendation system is designed based on personal preferences. It can recommend the fittest information to users, and it would help users to obtain information more conveniently and quickly. In our research, we design a recommendation system based on personal ontology and collaborative filtering technologies. Personal ontology is constructed by Formal Concept Analysis (FCA) algorithm and the collaborative filtering is design based on ontology similarity comparison among users. In order to evaluate the performance of our recommendation system, we have conducted an experiment to estimate the users' satisfaction of our experiment system. The results show that, combining collaborative filtering technology with FCA in a recommendation system can get better users' satisfaction.

Keywords: Document Recommendation System, Personal Ontology, Formal Concept Analysis (FCA), Collaborative Filtering

1. Introduction

With the internet technology has been widely used in human life, huge amounts of websites have been built and updated every day. This phenomenon usually makes the internet users at a loss in such a huge amount of information, and this problem is known as "information overloading". Furthermore, the information that hides in the databases is beyond the search engines' reach. In this case, although many internet search engines are available, it is still useless to information users to find what they want. Therefore, many websites, such as Yahoo! news and Amazon online bookstores, launch their own recommendation service on their platforms. They hope their systems could recommend products or information to users automatically and help users to find what they are searching for more quickly. In advance, the recommendation systems could even assist in answering to the potential information in which the users are interested.

Collaborative filtering technology is considered to be an effective way to solve the information overloading problem [1]. This technology mainly emphasizes on the cooperation between people. The system first collects the information of the users and then calculates the similari-

ties among the users. Through this way, the system could learn the preferences of every user and those preferences in common which could be recommended to the users. It will not only present the information that the users are interested in, but also some potential information that may surprise the users. Currently, some famous websites such as Amazon have adopted this technology. This shows that among these recommendation systems, collaborative filtering technology is relatively successful and most commonly used, as well as an excellent system used in electronic commerce [2-4].

Apart from helping finding the demanded information, the recommendation system aims to help the users to search with a faster speed and accuracy by constructing the shared documents and common preferences. It also makes the resources and services on the internet easier to access and share [5]. In this research, we integrate ontology and collaborative filtering to design a system to provide information recommendation service. We adopt Formal Concept Analysis (FCA) to construct a personal ontology to show the conceptual structure of personal preferences. FCA technology has been proved to be helpful in the development of ontology [6-10].

This research is not only engaged in constructing a recommendation system which combines ontology with the collaborative filtering technology, but also compare the users' satisfaction with the system without that technology. We have developed a prototype system and conducted a laboratory experiment to evaluate the users' satisfaction on different recommendation mechanism. The remainder of the paper is organized as the following. Section 2 reviews major literature concerning recommendation systems, ontology and FCA. System architecture and experiment design are shown in Section 3, and data analysis results are discussed in Section 4. Finally, implications and conclusions are described in Section 5.

2. Literature Review

2.1 Recommendation Systems

At the present time, recommendation systems hold more extensive definitions. It can be used to describe personal recommendations from any system or direct the users to find interested or useful targets from multiple possible choices. In this information overloading era, the design and development of recommendation systems is virtually more attractive than search for information depending on individuals, because it could help people make decisions from the complicated information. Currently, recommendation systems are already included in some electronic commerce websites such as Amazon [3].

The earliest recommendation system, developed by Goldberg *et al.* [11] is called Tapestry. It filters the useful information by collaborative filtering system. Collaborative recommendation system is the most famous and commonly used one. The system analyses the behaviors or preferences from the set of users within the system. It finds out the set of users with similar characteristics and takes this relevance as an evidence to induct the potential preferences of the users. Therefore, besides recommend the interested information to the users, this research is expected to recommend the information that may arouse the users' potential demands. In our recommendation system, it will first collect the users' information and calculate similarities of every user. From this way, the system could learn the preferences and the ones in common and find out the users who hold the similar preferences.

2.2 Ontology

Ontology could be defined from many aspects. Schreiber *et al.* [12] defined it as ontology provides a clearly description and conceptualization to express the knowledge in knowledge base from the aspect of knowledge base construction. In addition, Bernaras *et al.* [13] agreed that ontology provides a clear description to conceptualize knowledge in knowledge base. William and Austin [14] also proposes that ontology is a set describing or ex-

pressing concepts or terms of a certain field and can be used to organize the higher level of conceptual knowledge in knowledge base or describe the knowledge of a certain field. The process of its development leads to different definitions of ontology, but one point in common is ontology could help describe knowledge and the conceptual structure. In addition, the importance of ontology is that it matters the expression of knowledge structure and the analysis through ontology so as to present a clear knowledge structure. In one certain field, ontology is the core of expressing knowledge system and would help effectively express through analysis of ontology.

Therefore, the utmost task is to develop terms and relations that could effectively express knowledge so that the certain field or category would be analyzed efficiently. Moreover, the development of ontology would help share the knowledge. Knowledge base could be constructed according to different circumstances due to the share of ontology. For example, different manufacturers could use common terms and grammars to construct and describe the catalog indexes of some product, and then they share and use these indexes in automatic data exchanging systems. This kind of sharing could greatly increase the chances of knowledge reusing [15]. Now that ontology could familiarize the users with knowledge in specific field, users could utilize the conceptual correspondence of ontology to avoid the confusion of conceptions and rapidly find conceptual category in individual ontology. This could make browsing websites and searching information more efficient and convenient [5,16].

2.3 Formal Concept Analysis and Ontology

Formal Concept Analysis (FCA), proposed by Rudolf Wille in 1982, is a data analysis theory to disclose conceptual structures from data set [17]. The characteristic is that structures of data set could produce the graphical visualization, especially the quantitative analysis that the social sciences cannot be fully captured. Ganter and Wille [18] considered that FCA could mainly be used on data analysis such as investigate and process definite data. This data is based on Formal Abstractions of Concepts which is prominent and understandable. Wille [17] combined the target, property and relevance (each target possesses a property) together to present these relations by mathematical definitions of Formal Context and define Formal Concept [19].

The goal of both ontology and FCA is to build conceptual models of knowledge domain. FCA can be viewed as a technology of ontology construction to obtain structured data by concept lattices; it can be used as foundation of developing ontology manually and automatically by extracting concepts from the data set; it can also be used to present the visualization of ontology and

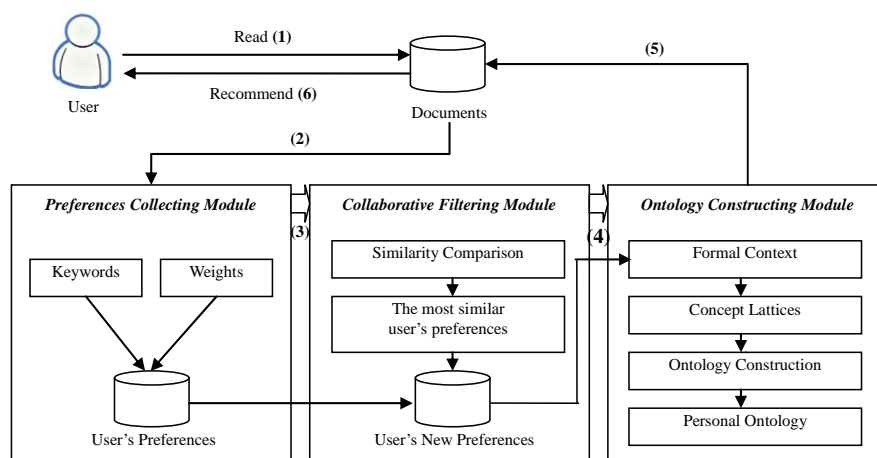


Figure 1. System architecture

help browse and analyze tasks. Among the theories combining FCA and ontology, the most prominent application is to identify the concept of ontology through formal concept [20]. Moreover, Hsu [7] proposes to automatically construct ontology based on FCA theory. It firstly extracts terms that stands for document concepts from term extraction system. Then integrate the binary matrix of document and terms to express independent, interlaced and inherited relations among different concepts and form the diagram of relations of concepts of ontology. The above documents all consider the property of FCA as the concept of ontology and the other relevant concepts as properties. Based on this view ontology is constructed or combined. The researches mentioned above prove that FCA and the concept of ontology could effectively help construct ontology. This research will use the ontology construction by FCA in recommendation systems.

3. System Architecture and Experiment Design

In this research, we aims to develop a recommendation system based on the combination of collaborative filtering technology and ontology. It will not only construct personal ontology with the FCA, but also calculate the users' familiarity to the keywords of all the documents. The users will give scores on those they read and are interested in while browsing them. These scores could show the users' preferences and work as a weights standard in the construction of ontology.

3.1 System Architecture

Figure 1 shows the system architecture of our recommendation system. In the step 1, the users enter the system, and the system assigns 20 documents randomly to users. The users browse and choose the top five docu-

ments they prefer to and give scores from 1 to 5 on the familiarity of the keywords of the 5 documents. In the step 2, the system analyze the collection of keywords and scores in the preference documents and make weights computing in users' preference collection module to prepare for the ontology construction and similarity comparison. In the step 3, with the weights computed in the previous module, the collaborative filtering module will compare the keywords and weights of preference with others. For the sake of time and efficiency, the system will only compare the first 100 users in the database and find the users with the highest similarity. The preference keywords and weights of these couple users will be sent to users' ontology construction module to prepare for the ontology construction. In the step 4, the system intermix the keywords and weights of the user with the highest similar one's. The sum of the keywords and weights will be used to construct the users' preference ontology by ontology construction module based on FCA technology. In the step 5, the system will send the new personal preferences back, and then the system will calculate the weights of each document. Finally, in the step 6, after calculating weighs of each document in the database, the system recommend the top five documents with the highest weights, and measure the user's satisfaction by online questionnaires.

The major modules in the system architecture are shown as follows.

1) Document database

The experimental system recommends documents to the users to read. In the document database, there are 210 mater dissertations focus on electronic commerce selected from Electronic Theses and Dissertations System in Taiwan¹. The data schema of documents database is composed of eleven fields, including serial number, author's name, year, paper's title, affiliation, abstract, and

five keywords.

2) Preferences collection module

For constructing personal preferences ontology by FCA, we need to collect user's preferences of keywords of documents. We believe that choosing their preference documents of the users cannot fully reflect the degree of their preferences. Therefore, we propose the scoring mechanism of the keywords to modify the weights between the concepts in the process of constructing ontology. In this module, user should select 5 preferred documents and score from 1 to 5 for each keyword in the documents to show their preference degree.

3) Collaborative filtering module

For the collaborative filtering mechanism, our system should have some users' preferences first. Therefore, when a user enters our system, the system can select the fittest user from the database and finish the collaborative filtering. In our experiment, we collect 105 participants' preferences in the database before collaborative filtering mechanism is running.

To find the fittest user from the database, we need a function to calculate the similarities between the users. We define Sims as the degree of the similarities of two users' preferences, and its function is shown as follows.

$$Sims = \frac{\sum_{t=1}^r K_t W_t}{\sum_{i=1}^m K_i W_i + \sum_{j=1}^n K_j W_j}$$

$\sum_{i=1}^m K_i W_i$: the sum of weights of user's preferred concepts

$\sum_{j=1}^n K_j W_j$: the sum of weights of the other user's preferred concepts

$\sum_{t=1}^r K_t W_t$: the sum of weights of the two users' conjunctive preferred concepts

4) Ontology construction module

This module mainly focuses on the weights of keywords collection and constructs the personal ontology. We adopt FCA [17] construct ontology. The steps are as follows.

Step 1: produce the formal contexts of the documents and keywords.

We first extract the collection of the keywords of the chosen documents from the document database. Then we match all the documents with the keywords collection. If the document includes certain keyword it will be marked

as "1". In this way form the formal contexts of the documents and keywords. Because of the scoring mechanism in this research, the keywords collection will be sequenced according to the weights of the users. In the later part the preference discussion will be transformed into the section of tree framework and make the concepts with high weights as higher hierarchy. According to the definition of FCA, this research defines the definition of formal contexts as K , the document collection on e-commerce as E , the keywords collection as T and the binary of the document collection and keywords collection as R . Then their relation can be put into $K := \langle E, T, R \rangle$.

Step 2: Produce all the concepts C

Define A as the subset of E and B as the subset of T , that is, $A \subseteq E$, $B \subseteq T$. If a certain concept is $A \subseteq B$, then it is marked as concept $c(A, B)$. For a concept $c(A, B)$, if all the relations R between A and B can form a biggest matrix, then all the collection of concept c is marked as C .

Step 3: produce the concept lattices between all the concepts

If the collection of all the documents with the keyword B_1 is included in the collection of all the documents with the keyword B_2 , the keyword B_1 is marked as the sub-concept of the keyword B_2 . That is, for all the concepts C , if $B_1 \subseteq B_2$, then $c_1(A_1, B_1)$ is the sub-concept of $c_2(A_2, B_2)$ and expressed as $(A_1, B_1) \leq (A_2, B_2)$. The $\text{sign} \leq$ stands for hierarchy of concepts.

Step 4: transform into tree diagram of ontology

While transforming the concept lattices diagram into tree framework of ontology by using breadth-first search, the relations of nodes may be fairly complicated and make the system spend too much time computing. This would lead to the inefficiency of recommendation and failure to promptly recommend documents to users. In order to avoid this, while constructing concept lattices, this research does not take the interlaced relations into account and make the concepts with high weights higher hierarchy. Then the relation contains only the concepts of higher hierarchy and the lower hierarchy. Then by breath-first search transform the relevance of formal contexts into tree framework which is the users' preference ontology.

3.2 Experiment Design

This experiment aims to recommend the users documents through two different recommendation systems and test their satisfaction. First, to be the experiment group, this system constructs ontology with the FCA theory, the scoring system and collaborative filtering technology. The other one, to be the control group, this system constructs ontology with the FCA theory and the scoring system without collaborative filtering. We will introduce

¹<http://etds.ncl.edu.tw/theabs/index.html>

Table 1. User's satisfaction measurement

-
1. Do the recommendation documents meet your demands?
 2. Are the recommendation methods accurate?
 3. Are the recommendation methods satisfying?
 4. Do you understand the recommendation methods?
 5. Do you think the recommendation methods practical?
 6. Do you think the recommendation methods reliable?
 7. Do you think the recommendation methods clear?
 8. Do you think the way of recommending understandable?
-

the recommendation steps of the first system as follows:

Step 1: Enter into the system: the users first read the introduction of the first page to learn the purpose and contents of the experiment.

Step 2: Assign documents randomly: the system extracts 20 documents randomly from the 210 ones for the users to read.

Step 3: Choose the documents the users prefer to and give scores: the users click the 20 ones to further read the contents and give scores on five interested ones. The system will store the keywords collection and preference scores of the five documents to prepare for the computing or collaborative filtering of the preferences.

Step 4: Ontology constructing for the users and recommends 5 documents to users based on ontology.

Step 5: After reading the recommendation documents, the users could fill in the questionnaires. The satisfaction refers to the satisfaction with information quality. The users should answer eight questions with Likert's five point scale from very dissatisfying to very satisfying. The experiment finishes after the users answer these questions.

4. Data Analysis

To evaluate the user's satisfaction on our experiment system, we have conducted a laboratory experiment research. The system combining personal ontology and collaborative filtering is served as the experiment group, and the system that has only personal ontology recommendation without collaborative filtering is served as the control group. There are totally 250 qualified participants have been invited to the experiment. By randomly dispatched by the system, 145 samples are assigned for experiment group and 105 for control group. User's satisfaction is measured by questionnaires online. The questionnaire is designed based on DeLone and Mclean's IS (information systems) success model [21,22]. This model proposes a comprehensive perspective to measure the success of an information system and has been widely used to appraise the quality of information systems. In a nutshell, a suc-

cessful information system should have qualified information quality and system quality to satisfy the users. In our research, due to both the experiment group and control group are conducted in the same platform, the system quality are the same in certain. We only adopt the measurements for information quality in our questionnaires. Table 1 shows the user's information quality satisfaction measurements and Likert's five point scale, from very disagree to very agree, is applied.

Factor analysis is applied to evaluate the validity of our measurements. The KMO value of this construct is 0.856. It shows that these measurements are feasible to factor analysis. Extract the dimensions whose eigenvalue is larger than 1 by using principal component analysis and orthogonal rotation through VARIMAX. After factor analysis, we divide the eight questions into two factor components. Question 2, 1, 6, 3 make up the first factor component, and this construct is named as *satisfaction with recommendation results*. Question 8, 4 and 7 make up the second one, and is named as *satisfaction with recommendation process*. Question 5 has the similar factor loading in both the two components (both are more than 0.5). We would delete question 5 after factor analysis.

Table 2 shows the descriptive statistics results of our experiment. The experiment group always gets higher satisfaction both on recommendation results and process.

To verify the experiment group really gets higher users' satisfaction than the control one in statistics, the independent-samples T test is applied. The results are shown in Table 3. No matter on recommendation results or process, users get higher satisfaction significantly. That is to say, the recommendation system based on the combination of ontology and collaborative filtering system is more satisfying than the one based only on personal ontology.

The higher satisfaction of experiment group might cause by the extension capability of combining collaborative filtering results with personal preferences. Due to the original personal ontology is built based on only five

Table 2. Descriptive statistics results of research constructs

Users' Satisfaction	Group	N	Mean	Std. Dev.
Recommendation Results	Experiment	145	3.641	0.591
	Control	105	3.441	0.580
	Total	250	3.557	0.594
Recommendation Process	Experiment	145	3.740	0.658
	Control	105	3.578	0.526
	Total	250	3.672	0.610

Table 3. Independent-samples t test results

Users' Satisfaction	Levene's Test for Equality of Variances		T-test for Equality of Means	
	F	Sig.	t	Sig.
Recommendation Results	0.011	0.916	2.673	0.008**
Recommendation Process	1.841	0.176	2.092	0.037*

interested documents, it will not include all the user's preferences certainly. Collaborative filtering would help to capture user's other preferences that have not been defined in the original personal ontology. In the other words, the expanded personal ontology, combining collaborative filtering results, might cover the potential preferences that have not been discovered. Therefore, the system recommends documents to users based on the expanded personal ontology would cause higher satisfaction.

5. Conclusions

This research is expected to take advantages of collaborative filtering and personal ontology to design an effective recommendation system. Therefore, we have first discussed how to construct personal ontology based on one self's and others' preferences. The personal ontology is built up by FCA method, in advance, we used scoring mechanism to intensify the weights of users' preferences. Then, we elaborated on how to make use of this method to provide personal recommendation service in an electronic documents repository system. We have implemented a prototype system and conducted a laboratory experiment to evaluate the system's performance. The research results show that the users have higher satisfaction with the recommendation system that combined collaborative filtering and ontology technology.

In practice, this research applies collaborative filtering and ontology to provide personal recommendation service on an electronic documents website. This personal recommendation method can be used widely in different online websites such as electronic news website, or e-retail website to recommend news/products to

customers.

However, in our experiment, the recommendation service is built based only 210 master theses. Due to the FCA method should calculate the relations among each document, it might cause performance problem when it were used in real repository system that usually has more than ten or hundred thousands of documents. The FCA method should be improved in calculation efficiency when it is used in the larger scale system. In the other, how to extract the proper keywords from documents would be another important and interesting issue. In our system, the recommendation documents database is composed of master theses. As usual, the master theses have accurate keywords that are defined by the author. However, in some other documents repository system, such as news website, there is no well-defined keyword in the system. How to extract proper keywords from this kind of system would be another critical problem when our recommendation system is implemented.

REFERENCES

- [1] B. Sarwar, J. Konstan, A. Borchers, J. Herlocker, B. Miller, and J. Riedl, "Using Filtering Agents to Improve Prediction Quality in the GroupLens Research Collaborative Filtering System," Proceedings of the 1998 Conference on Computer Supported Cooperative Work, pp. 345–354, 1998.
- [2] M. Balabanovic and Y. Shoham, "Fab: Content-Based, Collaborative Recommendation," Communications of the ACM, Vol. 40, pp. 66–72, 1997.
- [3] J. B. Schafer, J. Konstan, and J. Riedl, "Recommender System in E-Commerce," Proceedings of the first ACM

- Conference on Electronic Commerce, pp. 158–166, 1999.
- [4] S. Staab, and H. Werthner, “Intelligent Systems for Tourism,” *IEEE Intelligent Systems*, Vol. 17, pp. 53–66, 2002.
 - [5] L. Khan, D. McLeod, and E. Hovy, “Retrieval Effectiveness of an Ontology-based Model for Information Selection,” *Very Large Data Bases*, Vol. 13, pp. 71–85, 2004.
 - [6] H. M. Haav, “A semi-automatic method to ontology design by using FCA,” *Proceedings of the 2nd International CLA Workshop*, Ostrava, pp. 13–25, 2004.
 - [7] C. H. Hsu, “Automatically Constructing Ontology on Semantic Web,” M.S. thesis, Fu Jen Catholic University, Taiwan, 2004.
 - [8] M. Obitko, V. Snasel, J. Smid, and V. Snasel, “Ontology Design with Formal Concept Analysis,” In V. Snasel, and R. Belohlavek, (eds.) *Concept Lattices and their Applications*, Ostrava: Czech Republic, pp. 111–119, 2004.
 - [9] G. Stumme, and A. Maedche, “FCA-MERGE: bottom-up merging of ontologies,” *Proceedings of the Seventeenth International Conference on Artificial Intelligence*, pp. 225–234, 2001.
 - [10] Y. Zhao, X. Wang, and W. Halang, “Ontology Mapping based on Rough Formal Concept Analysis,” *Proceedings of the Advanced International Conference on Telecommunications and International Conference on Internet and Web Applications and Services*, pp. 180, 2006.
 - [11] B. Ganter, and R. Wille, “Applied lattice theory: Formal Concept Analysis,” In G. Grater, (edition), *General Lattice Theory*, Springer, Berlin, 1997.
 - [12] G. Schreiber, B. Wielinga, and W. Jansweijer, “The kactus view on the “o” word,” *Workshop on basic ontological issues in knowledge sharing: international joint conference on artificial intelligence*, pp. 159–168, 1995.
 - [13] A. Bernaras, I. Laresgoiti, and J. Corera, “Building and reusing ontologies for electrical network applications,” *Proceedings of European Conference on Artificial Intelligence*, pp. 298–302, 1996.
 - [14] S. William and T. Austin, “Ontologies,” *IEEE Intelligent systems*, Vol. 14, pp. 18–19, 1999.
 - [15] B. Chandrasekaran, J. R. Josephson, and V. R. Benjamins, “What are ontologies, and why do we need them?” *IEEE Intelligent systems*, Vol. 14, pp. 20–26, 1999.
 - [16] J. Chaffee and S. Gauch, “Personal Ontologies for Web Navigation,” *Proceedings of Conference on Information Knowledge Management*, pp. 227–234, 2000.
 - [17] R. Wille, “Restructuring lattice theory: An approach based on hierarchies of concepts,” In I. Rival, (edition) *Ordered Sets*, Reidel, Boston-Dordrecht, pp. 445–470, 1982.
 - [18] A. Formica, “Ontology-based concept similarity in Formal Concept Analysis,” *Information Sciences*, Vol. 176, pp. 2624–2641, 2006.
 - [19] K. E. Wolff, “A first course in Formal Concept Analysis - how to understand line diagrams,” *Advances in Statistical Software*, Vol. 4, pp. 429–438, 1994.
 - [20] P. Cimiano, A. Hotho, G. Stumme, and J. Tane, “Concept Knowledge Processing with Formal Concept Analysis and Ontologies,” *Proceedings of Second International Conference on Formal Concept Analysis*, pp. 189–207, 2004.
 - [21] W. H. Delone and E. R. Mclean, “Information System Success: The Quest for the Dependent Variable,” *Information Systems Research*, Vol. 3, pp. 60–95, 1992.
 - [22] H. D. William and R. M. Ephraim, “Information Systems Success: The quest for the dependent variable,” *Information Systems Research*, Vol. 3, pp. 60–95, 1992.

The Dynamic Multi-Task Supply Chain Principal-Agent Analysis

Shanliang LI^{1,2}, Chunhua WANG³, Daoli ZHU¹

¹Management School, Fudan University, Shanghai, China; ²Management School, Soochow University, Suzhou, China; ³Information School, Shanghai Ocean University, Shanghai, China.
Email: Lisl@fudan.edu.cn

Received August 18, 2009; revised September 25, 2009; accepted November 5, 2009.

ABSTRACT

In the supply chain by the composition of the supplier and the retailer, the supplier offers products to the retailer for sales while the retailer affects the sales outcome by his effort which is divided into two dimensions. One is for the short-term sales task and the other is for the long-term sales task. For the long-term development of the enterprise, the supplier wants to inspire the retailer to make more effort for the long-term task. However, due to the asymmetric information, the supplier can't observe the retailer's action and the moral hazard will come into being. To deal with this problem, we construct the dynamic multi-task supply chain principal-agent model, by which we analyze the impact of the information asymmetry to the supply chain contract. Furthermore, by comparing the contracts between the single-term multi-task and two-term multi-task, we have analyzed their different effect on the commission rate.

Keywords: Supply Chain Management, Multi-task Principal-agent, Dynamic Incentive, Moral Hazard

1. Introduction

In the supply chain system, there exists the conflict between the local interests of the supply chain members and the overall performance of the supply chain, which leads to the system inefficiency. At present, one of the most important research areas of supply chain is to design the suitable coordination mechanism in order to obtain the global optimization of the supply chain performance. In this process, the information plays a very important position. As the supply chain members tend to hide their private information to maintain information superiority, this will lead to "Adverse Select" and "Moral Hazard" in various fields [1].

In the recent decade, scholars have studied on the issue of the supply chain coordination from various angles. These studies can be roughly divided into two categories. One is to resolve the "double marginalization" problem by contract design in the situation of the full information [2–4]. Such contracts do not involve the information incentive. The other is to study the supply chain incentive problem in the situation of the asymmetric information. Corbett etc. studied that the optimal quantity discounts incentive contract between the supplier and the retailer [5]. Basu etc. studied the incentive issues of the sales force under asymmetric information based on agency theory [6]. Lal etc. [7,8] and Chen [9] extended the above

studies. Many Chinese academics are also carried out research on this issue [10–14]. For the supply chain coordination, the author's research team had a systemic research on the issue earlier. Some relevant research results have been published [15–22]. This paper is the important one of the systemic study. In the simple principal-agent model, the agent is engaged in one job and the agent's effort is one-dimensional. However in many cases of the real life, agents are engaged in the job of more than one. Or, even if there is only one job, it involves more than one dimension. Furthermore, it exists conflict in the distribution of the same agent's energy between the different jobs. To deal with this problem, we construct the dynamic multi-task supply chain principal-agent model, by which we analyze the impact of the information asymmetry to the supply chain contract. Furthermore, by comparing the contracts between the single-term multi-task and two-term multi-task, we analyzed their different effect on the commission rate.

2. Assumptions and Parameters Set

We make the following assumptions for the tractable analysis. Considering a Stackelberg model between a supplier *S* who is the principal and a retailer *R* who is the agent, the supplier offers the retailer products to sale and pays the retailer according to sales outcome which is

affected by the retailer's effort and the random factors. Set $B_1(a_1, a_2)$ is the retailer's expected profit whose ownership belongs to the supplier. a_1 denotes the retailer's effort for the short sales goal. a_2 denotes the retailer's effort for the long sales goal. $C(a_1, a_2)$ denotes the cost of the retailer's effort, satisfying $\frac{\partial C}{\partial a_{1,2}} \geq 0$, $\frac{\partial^2 C}{\partial a_{1,2}^2} \geq 0$,

i.e. the cost of the effort increases and the marginal cost increases. For the simplicity, Assume $C(a_1, a_2) = \frac{a_1^2}{2} + \frac{a_2^2}{2}$.

The supplier can't observe a_1 and a_2 , but can observe and verify the sales outcome x , which is affected by the retailer's effort together with the random variables, denoted by $x = \mu(a_1, a_2) + \theta$, where $\mu(a_1, a_2)$ is the output function of the effort, satisfying $\frac{\partial \mu}{\partial a_{1,2}} \geq 0$, which

means the marginal sales outcome of the effort is positive. i.e. more efforts mean more sales; $\frac{\partial^2 \mu}{\partial a_{1,2}^2} \leq 0$, which means

marginal sales outcome decrease (When the equal sign is set up, marginal sales unchanged). Set $\theta^T = (\theta_1, \theta_2)$, which is the random variable of Normal distribution and satisfy $N \sim (0, \sigma_1^2; 0, \sigma_2^2; r)$; Set $x^T = (x_1, x_2)$, For the sake of simplifying the calculating, assume that $x_1 = \mu(a_1) + \theta_1 = a_1 + \theta_1$; $x_2 = \mu(a_2) + \theta_2 = a_2 + \theta_2$. i.e. different efforts result in different information (However, different information may be relevant if θ_1 and θ_2 are relevant.): x_1 reflects a_1 , x_2 reflects a_2 . The ownership of the sales profits belongs to the supplier, and the supplier offers the linear salary to pay the retailer.

$$s(x) = \alpha + \beta^T x = \alpha + \beta_1 x_1 + \beta_2 x_2 \quad (1)$$

where $s(x)$ is the wage paid to the retailer if the total sales outcome is x . α denotes the one-off wealth transfer which doesn't affect the incentive intension (Called Salary); $\beta^T = (\beta_1, \beta_2)$ which denotes the incentive intension (Called Commission Rate), that means if x increase by one unit, the wage of the retailer increased by β unit.

3. The Single-Stage Multi-Task Model

In the single-stage model, the supplier offers a one-time wage schedule, $s(x)$, according to which the retailer is rewarded. Assume the supplier is risk-neutral, the expected utility function is as follows:

$$\begin{aligned} EU_s &= B(a_1, a_2) - \alpha - E\beta^T \mu(a_1, a_2) \\ &= B(a_1, a_2) - \alpha - \beta_1 a_1 - \beta_2 a_2 \end{aligned} \quad (2)$$

Assume the retailer is risk-averse, and the utility is that $V(x) = -e^{-\rho x}$, where ρ is the risk aversion coefficient. When $\rho = 0$, the retailer is risk-neutral. When $\rho > 0$, the retailer is risk-averse. When $\rho < 0$, the retailer is risk preference. The retailer's expected utility is as follows:

$$EU_R = EV(s(x) - C(a_1, a_2)) \quad (3)$$

To make the analysis simple, we use the certainty equivalent (CE) instead of the expected utility of the retailer [18].

$$CE_R = \alpha + \beta^T \mu(a_1, a_2) - \frac{1}{2} \rho \beta^T \Sigma \beta - C(a_1, a_2) \quad (4)$$

where $\alpha + \beta^T \mu(a_1, a_2)$ is the expected wage, ρ is risk aversion coefficient, $\beta^T \Sigma \beta$ is the income variance, $\rho \beta^T \Sigma \beta / 2$ is the risk cost. Σ is the covariance matrix of θ_1 and θ_2 , denoted by $\Sigma = \begin{pmatrix} \delta_1^2 & r\delta_1\delta_2 \\ r\delta_1\delta_2 & \delta_2^2 \end{pmatrix}$.

The supplier is the leader in the Stackelberg model, who has first-step advantage in the game. However, when he/she pursues the profit maximization, he/she must consider the incentive compatibility constraint and participation constraint. Thus, the principal-agent model between the supplier and the retailer can be rewritten as the following optimization programming.

$$(P) \quad \underset{\alpha, \beta, a}{\text{Max}} EU_s = B(a_1, a_2) - \alpha - \beta_1 a_1 - \beta_2 a_2 \quad (5)$$

$$\begin{aligned} \text{s.t. (IR)} \quad CE_R &= \alpha + \beta_1 a_1 + \beta_2 a_2 \\ -\frac{1}{2} \rho (\beta_1, \beta_2) &\begin{pmatrix} \delta_1^2 & r\delta_1\delta_2 \\ r\delta_1\delta_2 & \delta_2^2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} - \frac{a_1^2}{2} - \frac{a_2^2}{2} \geq 0 \end{aligned} \quad (6)$$

$$\text{(IC)} \quad \arg \max_{a_1, a_2} CE_R \quad (7)$$

where (6) is participation constraint (IR), and (7) is incentive compatibility constraint (IC).

3.1 The Full Information Benchmark

In this section, let's begin with the full information case where the retailer's effort is observable and verifiable. Then the supplier can assign an effort level to the retailer by designing a forcing contract. Under this condition, the incentive compatibility (7) is invalid and we only consider the participation constraint (6), which is binding. Namely, (P) can be rewritten as follows:

$$\begin{aligned} (P') \quad \underset{\alpha, \beta, a}{\text{Max}} U_s &= B(a_1, a_2) - \alpha - \beta_1 a_1 - \beta_2 a_2 \\ \text{s.t.} \quad CE_R &= 0 \end{aligned} \quad (8)$$

Solving (P'), we can obtain that:

$$\begin{aligned}\frac{\partial B(a_1, a_2)}{\partial a_1} &= \frac{\partial C(a_1, a_2)}{\partial a_1} = a_1 \\ \frac{\partial B(a_1, a_2)}{\partial a_2} &= \frac{\partial C(a_1, a_2)}{\partial a_2} = a_2\end{aligned}\quad (9)$$

The Equation (9) is the class condition of the Pareto optimality: the expected marginal profit of the effort is equal to the expected marginal cost. That is similar to the single-task case. We have the following conclusion.

Proposition 1: Under the condition of full information, by designing the linear incentive contract, the game between the supplier and the retailer can achieve the Pareto optimality when the retailer has multi-dimensional effort.

$$Max_{a, \beta} B(\beta_1, \beta_2) - \left(\frac{1}{2} \rho (\beta_1^2 \delta_1^2 + 2r\beta_1\beta_2\delta_1\delta_2 + \beta_2^2 \delta_2^2) + \frac{\beta_1^2}{2} + \frac{\beta_2^2}{2} \right) \quad (11)$$

Solve the first order derivative of (13), (14) about β_1 , obtain

$$\frac{\partial B}{\partial \beta_1} - \left(\frac{1}{2} \rho (2\beta_1\delta_1^2 + 2r\beta_2\delta_1\delta_2) + \beta_1 \right) = 0 \quad (12)$$

Solving the above equation, we obtain that

$$\beta_1 = \frac{\frac{\partial B}{\partial \beta_1} - \rho r \beta_2 \delta_1 \delta_2}{1 + \rho \delta_1^2} \quad (13)$$

Similarly, we get

$$\beta_2 = \frac{\frac{\partial B}{\partial \beta_2} - \rho r \beta_1 \delta_1 \delta_2}{1 + \rho \delta_2^2} \quad (A) \quad (14)$$

By (13), (14), we get the following conclusion.

Proposition 2: When $\rho = 0$, the retailer is risk-neutral, then $\beta_i = \frac{\partial B}{\partial \beta_i} = \frac{\partial C}{\partial a_i} = a_i$ ($i=1,2$), which means the game can get the Pareto optimization just as the full information case. When $\rho > 0$, β_i ($i=1,2$) is in inverse ratio with ρ , the risk aversion coefficient will reduce the incentive intensity β_i ; β_i ($i=1,2$) is in inverse ratio with variance δ_i^2 ($i=1,2$); in inverse ratio with the Covariance $r\delta_1\delta_2$, i.e. the random factors also reduce the incentive intensity of β_i . β_1 is also in inverse ratio with β_2 . More β_2 means less β_1 , and vice versa.

4. Two-Stage Multi-Task Game Model

In the two-stage multi-task model, suppose the retailer's effort a_2 for the long task in the first stage will affect the profit in the second stage of the supply chain. Set $B_1(a_1, a_2)$ denotes the expected effort profit of the first

3.2 The Asymmetric Information Case

Generally, the supplier can't observe the retailer's action a , and only can observe outcome x . In this case, the incentive compatibility constraint (7) is valid. Substituting (7) by the first-order condition, we can obtain the equivalent programming. i.e. (7) is equal to that

$$\begin{aligned}a_1 &= \beta_1 \\ a_2 &= \beta_2\end{aligned}\quad (10)$$

Solving the model (P)

$$Max_{a, \beta} B(a_1, a_2) - \left[\frac{1}{2} \rho (\beta_1, \beta_2) \left(\delta_1^2, r\delta_1\delta_2 \right) \left(\beta_1 \right) + \frac{a_1^2}{2} + \frac{a_2^2}{2} \right]$$

Substituting by (10), we get

stage of the retailer, $B_2(a_{21}, a_{22}, a_2)$ denotes the expected profit of the second stage. Where a_{21} and a_{22} denote the effort for the short task and long task respectively. Because the effort a_2 in the first stage will affect the profit in the second stage, it will be the variable of the output function of the second stage. The ownership of $B_1(a_1, a_2)$ and $B_2(a_{21}, a_{22}, a_2)$ belongs to the supplier, the supplier offers the wage schedule according to the two-stage outcome. Similarly to the assumption of one-stage, the observed outcome in the second stage is that

$$x2^T = (x_1, x_2, x_3, x_4, x_5) \quad (15)$$

where $x_1 = a_1 + \theta_1$, $x_2 = a_2 + \theta_2$, $x_3 = a_{21} + \theta_3$, $x_4 = a_{22} + \theta_4$, $x_5 = a_2 + \theta_5$. Assume $\theta'^T = (\theta_3, \theta_4, \theta_5)$, which is the random variable in the second stage, Independent with $\theta^T = (\theta_1, \theta_2)$, the random variable in the first stage. The supplier offers the two-stage payoff contract according the observed outcome as follows.

$$\begin{aligned}s_1(x) &= a_1 + \beta_1 x_1 + \beta_2 x_2 \\ s_2(x) &= a_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5\end{aligned}\quad (16)$$

The supplier's expected utility is that:

$$EU_S = B_1(a_1, a_2) - Es_1(x) + B_2(a_{21}, a_{22}, a_2) - Es_2(x) \quad (17)$$

Now, the certainty equivalent of the retailer in the first stage and the second stage is that

$$CE_R^1 = a_1 + \beta_1 a_1 + \beta_2 a_2 - \frac{1}{2} \rho \beta^T \Sigma \beta - \frac{a_1^2}{2} - \frac{a_2^2}{2} \quad (18)$$

$$CE_R^2 = a_2 + \beta_3 a_{21} + \beta_4 a_{22} + \beta_5 a_2 - \frac{1}{2} \rho \beta^T \Sigma_1 \beta - \frac{a_{21}^2}{2} - \frac{a_{22}^2}{2} \quad (19)$$

where Σ_1 is covariance matrix θ_3 , θ_4 and θ_5 . Denoted by

$$\Sigma_1 = \begin{pmatrix} \sigma_1^2, r_1\sigma_1\sigma_2, r_2\sigma_1\sigma_3 \\ r_1\sigma_1\sigma_2, \sigma_2^2, r_3\sigma_2\sigma_3 \\ r_2\sigma_1\sigma_3, r_3\sigma_2\sigma_3, \sigma_3^2 \end{pmatrix} \quad (20)$$

For obtaining the retailer's optimal effort of the second stage, solve $\arg \max_{a_{21}, a_{22}} CE_R^2$, and get

$$\begin{aligned} a_{21} &= \beta_3 \\ a_{22} &= \beta_4 \end{aligned} \quad (21)$$

Considering the participation constraint and incentive compatibility constraint, the supplier need solve the following programming:

$$\begin{aligned} (P'') \quad & \text{Max}_{\alpha, \beta} E U_p = B_1(a_1, a_2) - E_{s_1} + B_2(a_{21}, a_{22}, a_2) - E_{s_2} \\ \text{s.t.} \quad & \text{(IC)} \quad \arg \max_a CE^1 + CE^2 \\ & \text{(IR)} \quad CE^1 + CE^2 \geq 0 \end{aligned} \quad (22)$$

Instead (IC) in (22) by the first-order condition and substitute by (10), (21), Solve the programming P'' and get

$$\begin{aligned} B_1(a_1, a_2) & - \left[\frac{1}{2} \rho (\beta_1^2 \delta_1^2 + 2r\beta_1\beta_2\delta_1\delta_2 + \beta_2^2 \delta_2^2) + \frac{\beta_1^2}{2} + \frac{\beta_2^2}{2} \right] \\ & + B_2(a_{21}, a_{22}, a_2) - \frac{1}{2} \sigma_1^2 + r_1\beta_3\beta_4\sigma_1 \left[\frac{1}{2} (\beta\sigma_2 + r_2\beta_3\beta_5\sigma_1\sigma_3 \right. \\ & \quad \left. + r_1\beta_3\beta_4\sigma_1\sigma_2 + \beta_4^2\sigma_2^2 + r_3\beta_4\beta_5\sigma_2\sigma_3 \right. \\ & \quad \left. + r_2\beta_3\beta_5\sigma_1\sigma_3 + r_3\beta_4\beta_5\sigma_2\sigma_3 + \beta_5^2\sigma_3^2) + \frac{\beta_3^2}{2} + \frac{\beta_4^2}{2} \right] \end{aligned} \quad (23)$$

Solve β_1 by derivate (23) about β_1 and get

$$\frac{\partial B_1}{\partial \beta_1} + \frac{\partial B_2}{\partial \beta_1} - \beta_1(1 + \rho\delta_1^2) - \rho r\beta_2\delta_1\delta_2 = 0 \quad (24)$$

$$a_1 = \beta_1 = \frac{\frac{\partial B_1}{\partial \beta_1} + \frac{\partial B_2}{\partial \beta_1} - \rho r\beta_2\delta_1\delta_2}{1 + \rho\delta_1^2} \quad (25)$$

Solve β_2 by derivate (23) about β_2 and get

$$\frac{\partial B_1}{\partial \beta_2} + \frac{\partial B_2}{\partial \beta_2} - \beta_2(1 + \rho\delta_2^2) - \rho r\beta_1\delta_1\delta_2 = 0 \quad (26)$$

$$a_2 = \beta_2 = \frac{\frac{\partial B_1}{\partial \beta_2} + \frac{\partial B_2}{\partial \beta_2} - \rho r\beta_1\delta_1\delta_2}{1 + \rho\delta_2^2} \quad (B) \quad (27)$$

Because $B_2(a_{21}, a_{22}, a_2)$ doesn't involve the variable a_1 , $\frac{\partial B_2}{\partial \beta_1} = 0$. The Equation (13) is the same to the Equation (25). Comparing (14) with (27), because $\frac{\partial B_2}{\partial \beta} > 0$, it is evident $B > A$. Thus, we have the following conclusion.

Proposition 3: By designing dynamic multi-task contract, the supplier can inspire the retailer to pay more effort for the long-term goal without the premise of changing the retailer's effort for the short goal. It shows that the dynamic contract design is conducive to maintain the long-term supply chain partnership.

5. Conclusions

The supply chain contract design is the important means of the supply chain coordination. For different environment, it will greatly improve the level of supply chain collaboration by the design of appropriate contract. In this paper, we have studied the incentive contract between the supplier and the retailer. Because of asymmetrical information, the supplier can't observe the effort level of the retailer. Therefore, the supplier can only inspire the retailer's different effort level by the incentive mechanisms design. The major study of the paper is on how to design the dynamic incentive contract to stimulate retailers to pay more efforts for the long-term under asymmetric information and multi-task environment, which has the guiding role for establishing the supply chain dynamic alliance. At the same time, our study extends the existing research results of the principal-agent. In our research work, for the sake of simplifying the technical analysis and the calculating, we focused on the second-term multi-task game. In the future, we will extend our research to multi-term multi-task model, which would be challenging and meaningful.

6. Acknowledgements

The authors would like to thank the referees for their helpful suggestions. The work is supported by China Postdoctoral Science Foundation under Grant No. 20080430075 and Special Grade of Financial Support from China Postdoctoral Science Foundation under Grant No. 200902199 and National Natural Science Foundation under Grant No. 70171010.

REFERENCES

- [1] W. Y. Zhang, "Game Theory and Information economics [M]," Shanghai People's Publishing House, 1997.
- [2] G. Cachon, "Supply chain coordination with contracts," Handbooks in operations research and management science: Supply Chain Management, North Holland, 2003.
- [3] G. Cachon and M. Lariviere, "Contracting to assure supply:

- How to share demand forecasts in a supply chain [J],” *Management Science*, Vol. 47, No. 5, pp. 629–646, 2001.
- [4] B. Pasternack, “Optimal pricing and return policies for perishable commodities [J],” *Marketing science*, Vol. 4, No. 4, pp. 166–176, 1985.
- [5] C. Corbett, D. Zhou, and C. Tang, “Designing supply contracts: Contract type and information asymmetry [J],” *Management Science*, Vol. 50, No. 4, pp. 550–559, 2004.
- [6] A. Basu, R. Lal., V. Srinivasan, and R. Staelin, “Sales-force compensation plans: An agency theoretic perspective [J],” *Marketing science*, Vol. 4, pp. 267–291, 1985.
- [7] R. Lal and V. Srinivasan, “Compensation plans for single-and multi-product salesforces: An application of the Holmstrom-Milgrom model [J],” *Management Science*, Vol. 39, No. 7, pp. 777–793, 1993.
- [8] R. Lal and R. Staelin, “Salesforce compensation plans in environments with asymmetric information [J],” *Marketing science*, Vol. 5, pp. 179–198, 1986.
- [9] F. Chen, “Sales-force incentives and inventory management [J],” *Manufacturing & Service Operations Management*, Vol. 2, No. 2, pp. 186–202, 2000.
- [10] S. Z. Bai and X. Y. Zhu, “Studyon Incentive- Mechanism of S. C MoralRisk Based on Principal-agent [J],” *Logistics Technology*, Vol. 10, 2007.
- [11] X. Y. Bao and Y. Pu, “Incentive contract in reverse supply chain with asymmetric information [J],” *Computer Integrated Manufacturing Systems*, Vol. 9, 2008.
- [12] S. P. Wu and X. Y. Xu, “Incentive mechanism for two-echelon supply chain under asymmetric information,” *Computer Integrated Manufacturing Systems*, Vol. 14, No. 3, 2008.
- [13] C. Q. Yi and J. P. Wan, “Dynamic Principal-Agent Model with Two stage [J],” *Chinese Journal of Mangement*, S2, 2005.
- [14] H. S. Yu, L. D. Zhao, and Y. H. Long, “An Evolutionary Game Model for Supply Chain Partnerships System Based Reputation Incentive,” *Mathematics in Practice and Theory*, Vol. 38, No. 4, 2008.
- [15] S. L. Li and C. H. Wang, “The Analysis of the Supply Chain Incentive Contract under Asymmetric Information [C],” *The 4th International Conference on WICOM: EMS2008*, 2008.
- [16] S. L. Li and C. H. Wang, “Linear Incentive Contract for Principal-agent Problem with Asymmetric Information and Moral Hazard [C],” *2006 IEEE Asia Pacific Conference on Circuits and Systems*, 2006.
- [17] S. L. Li and C. H. Wang, “The Supply Chain Incentive Contract under Double Moral-Hazard [C],” *2007 International Conference of Systems Science, Management Science and System Dynamics*, 2007.
- [18] S. L. Li and D. L. Zhu, “Principal-agent analysis of supply chain incentive contract with asymmetric information,” *Computer Integrated Manufacturing Systems*, Vol. 12, 2005.
- [19] S. L. Li and D. L. Zhu, “Supply Chain Linear Incentive Contract with Asymmetric Information and Moral Hazard [J],” *OR Transactions*, Vol. 2, 2005.
- [20] S. L. Li, D. L. Zhu, and B. Wang, “Study on the Principal-agent Problem in Supply Chain,” *Logistics Technology*, Vol. 11, 2004.
- [21] S. L. Li, M. Zuo, and D. L. Zhu, “The Analysis on the Pricipal-agent Model of a Product Line Design,” *Chinese Journal of Management Science*, Vol. 13, No. 1, 2005.
- [22] Q. Xu, D. L. Zhu, and S. L. Li, “The Supply Chain Optimal Contract Design under Asymmetrical Information,” *System Engineering-Theory & Practice*, Vol. 4, 2007.

Narrating National Geo Information Infrastructures: Balancing Infrastructure and Innovation

Henk Koerten¹, Marcel Veenswijk²

¹Department of Delft University of Technology, Amsterdam, Netherlands; ²Department of Culture, Organization and Management, VU University Amsterdam, Amsterdam, Netherlands.

Email: h.koerten@tudelft.nl, m.veenswijk@fsw.vu.nl

Received September 10, 2009; revised October 16, 2009; accepted November 18, 2009.

ABSTRACT

This paper examines narratives relating to the development of National Geo Information Infrastructures (NGII) in ethnographic research on a Dutch NGII project which was monitored throughout its course. We used an approach which focuses on narratives concerning the environment, groups and practice to elicit sense-making processes. We assert that narratives are relatively fixed and that they only change under specific circumstances. Moreover, the fixing of or change in narratives takes place in practice, so our research approach aimed at analysing narratives of practice, which we label 'storyboards'. For this purpose, project meetings and conferences were observed, key persons both within and outside the project environment were interviewed, and an analysis of relevant documents and video footage was undertaken. Storyboards are created by actors as a result of day-to-day challenges related to project goals, technology and infrastructure. In our research we found that these storyboards occur as vicious circles from which actors cannot escape. In the specific case analysed, our interpretation of the narrative storyboards suggested that these vicious circles are caused by the inability of project participants to distinguish between infrastructure and innovation requirements in their daily work.

Keywords: Narrative Analysis, Narrative Approach, Innovation, Organisation, NGII, Infrastructure

1. Introduction

There is a worldwide tendency to create facilities on a national scale to collect and disseminate location-based information, usually called geo information [1]. Car navigation systems are a form of geo information used by the general public, and geo information is also applied in government and business organisations to make them more effective. Within organisations, this information is often managed by a Geographical Information System (GIS), and between organisations, through National Geo Information Infrastructures (NGIIs) with national governments playing a key role in their dissemination [2–5].

When setting up a program aimed at establishing NGIIs, policy advisors take organisational aspects seriously, but do not treat them as manageable phenomena [6,7]. Technical aspects are regarded as crucial [3], and those involved in implementing the programs generally seem to overlook the organisational consequences, denying the relationship between organisational change and NGII implementation [8]. Therefore, organisational structures, modes of cooperation and work relationships have not

been topics of interest in the context of research into NGII implementation [9].

However, while technological developments are still regarded as crucial, those involved in implementation are now more inclined to take the organisational aspects of NGII development into account, culminating in design rules borrowed from political science, economics and management science [10–12]. Practitioners still point to difficulties with infrastructure development—mostly in the context of specific projects—of which we still have little knowledge of the lived experience of the project members [13,14].

Our aim is to find out why people who have problems in their daily work nevertheless maintain their current practices and refrain from looking for alternative methods. In relation to NGII development, there is a tendency to continue developing design rules while rarely taking the implementation processes into account. Our focus is practical: on how NGIIs are discussed in meetings, interviews and policy documents, where these discussions culminate in the creation of narratives. The research question guiding this paper is: How can we understand NGII

implementation using *narrative analysis*? Our secondary questions are: How do technological and organisational aspects interact with each other? How are goals and results perceived? And do these goals and results change over the course of the project?

Using a narrative approach, this paper provides an in-depth ethnographic case study of a Dutch NGII implementation project called Geo Portals. The project was intended to realise a part of the Dutch NGII by disclosing governmental geo information in a thematically organised way. Our research findings demonstrate that the initial project goal of building an infrastructure gradually changed over the course of the project, moving towards knowledge creation to facilitate innovation aimed at the further development of the NGII.

We will start with a description of the theoretical assumptions underlying the narrative analysis approach to research, followed by an account of the research methodology. An analysis of the project in terms of the theory will follow, and finally, we will provide a summary and some concluding remarks.

2. The Narrative Analysis Approach to Research

Symbolic interactionism introduced the idea that human thought is shaped by social interaction, and treated the modification of meanings and symbols as a process [15]. Goffman expanded this notion by adding the ability of human beings to look at themselves from another point of view [16], framing the notion using the theatrical terms of a 'front-stage' and a 'back-stage' [17]. Over his career, Goffman became aware of the ritualistic and institutionalising aspects of social interaction, but failed to specify how and why these frames or structures emerge [18–20].

Sociologists have attempted to understand society by gaining insight into how the structures involved in the process of modernisation affect our lives [21–24]. Some have made efforts to integrate micro and macro approaches [25–27]. For example, Bourdieu implicitly rejected the assumption of an objective truth, implying that structures are socially constructed, and he attempted to take a middle position which he labelled both 'constructivist structuralism' and 'structuralist constructivism' [27,28]. Bourdieu conceptualised *habitus* as the cognitive structures through which people deal with the social world, being both individual and collective, dialectically developed and internalised, a process which he labelled 'practice'. A 'field' was conceptualised as a network of relations among objective positions and not as a network of interactions or inter-subjective ties among individuals. These relationships, regarded as existing externally with respect to individuals, determined the position of individual agents through their acquisition of various kinds of capital: economic, cultural (knowledge), social (relationships) and symbolic (prestige). In this process, field and *habitus* define each other in a dialectical relationship.

Bourdieu and Goffman may have different points of departure, but there are similarities in their conceptualisations. Goffman's dramaturgical perspective may, to a considerable extent, be comparable to Bourdieu's *habitus*, while Goffman's notion of frames resembles Bourdieu's field concept and practice is more or less interchangeable with Goffman's concept of the 'front-stage'. While this comparison may appear to be a broad generalisation, these observations will prove useful in blending the two approaches together into one theoretical concept for research. Nevertheless, while an intersectional framework such as this might provide useful notions about the life world affecting individual, group and inter-group behaviour, the very aspect of meaning creation remains unaddressed. It remains unclear how images come to life and develop over time, as this framework assumes univocality, iniquitousness and fully informed actors and as such the ambivalence, ambiguity and incompleteness of world-views is overlooked.

Thus, the theoretical notions presented above provide useful hints for a theoretical approach but do not address the process of sense-making needed to answer the research question. Therefore, we will focus on the interpretation of lived experience as a guide for action and extend this towards a narrative approach using linguistic, anthropological and social psychological insights [29–31]. Interpretation, meaning creation and sense-making have become guiding concepts in the development of less positivistic methods [32,33]. Two sources that have inspired narrative theory may be distinguished: a 'linguistic turn', inspired by Saussure, Wittgenstein, Chomsky and Derrida, and a 'narrative turn', with more emphasis on stories and meaning, represented by authors such as Barthes, Bakhtin, Boje, and Gabriel [34].

In itself, language has no relationship to time or the originator of an utterance [35]. The concept of discourse, however, is treated as a combination of spoken word and written text, linked to time and space and used to make sense of the world, without drawing a distinction between the two [36]. In relation to the concept of discourse, the process of enactment is conceived of as communication through written and spoken symbols, usually linguistic. For example, to complete a management task, people write, read, speak, listen and discuss, using messages which convey myths, sagas, results, setbacks, challenges or strategies [37].

While language has been recognised as the dominant vehicle for the development of meaning in the discursive approach, the dynamic character of organisational practice has invoked interest in linguistic aspects other than text alone, such as metaphor, stories, novels, rituals, rhetoric, language games, drama, conversations, emotions and sense-making [38]. Grounded in literary criticism, new methods of analysis have emerged and been labelled as the narrative turn, which is aimed at delineating stories and storylines rather than texts [39–42].

Meaning is created, maintained, altered and destroyed and may be used to contemplate, to manipulate, be purposeful and invoke change [43–47]. The narrative turn has been considered fundamental in interpretive organisational research for conceptualising the notion of organisation in a more dynamic way (Hatch and Yanow, 2003). These dynamics have been envisioned in terms of people using and producing frames of reference in a cyclical process of enactment-selection-retention [48], as a dialectical process moving towards objectification [28], or as a narrative that is edited under particular circumstances [31,49]. The concept of narrative is broad, in the sense that it can be regarded as structuring human memory, and therefore should be conceptualised as both medium and process [50]. The concept of discourse, however, is more defined, referring to meaning produced in the exchange of signs and symbols, and in this respect more closely related to symbolic interactionism [50,51].

Narrative has been regarded as story [42], as telling a story [38] and as the art of telling a story [52], while there are also other approaches concerned with linking stories and narratives [49,53–55]. Living in a world composed of stories, we use narratives for communication and to give meaning to experience [42]. In providing an account of events, narratives allow us to create an interpretation of these events, relating the story in a favourable manner. Some stories are created for single use, while others are retold and altered and in the process gain a meaning they would never have had if they had been told only once. In this way they become a frame of reference for future stories and actions [56]. Once stories begin to have a life of their own, they grow further to become narratives which might be only loosely or even poorly connected to the original [55]. They become universal images, constituting all aspects of society, referring to the culture of all kinds of people, culminating in identity-creation using social categories [57]. From a manager to a company car, human and non-human identities are created by storytelling, leading to narratives that are continuously constructed and therefore subject to change. Having a plot does not imply that narratives are always visible and recognisable, they can be prominent or latent, and can also sometimes be unconsciously present to actors. They are an interpretation of assembled, either real or imagined stories, which Boje, after Clair, called ‘narratives dressed as theories’ [55].

The hermeneutic approach implies that a specific narrative can only be understood when it is interpreted in relation to other narratives, for example if we conceptualise narration as a ‘grand narrative’ grounded in many ‘micro stories’ which are mutually dependent [49,55]. This notion is reminiscent of the sociological micro-macro debate which links Bourdieu to the insights presented above. To avoid being confined to a type of hierarchically layered concept, one can focus on the morphology of narra-

tives over time, conceptualising how such narratives are edited by the actors involved so as to invoke the narrative, as well as sustain or to change it [31,49]. However, because the editing process is associated with the particular editors, there is a danger of overemphasising the role of individuals and in so doing implicitly sustain the idea of ‘culture creation’ or ‘cultural intervention’, which we have seen before [43,58].

The notion of narrative has also been distinguished by declaring everything before narrative to be ‘ante-narrative’. Verduijn refers to ante-narrative as ‘lived experience’, which she finds to be speculative, multifaceted and ambiguous, and while it always tends towards a coherent story, it is always prior to its reification into a sensible narrative [30,34,55]. However, while this distinction may be tempting, it is difficult to sustain the division between narrative and ante-narrative. This approach also presupposes that all the storylines—the ‘Tamara of stories’—can be known by the researcher [30]. However, it is impossible to grasp the full picture, just as it is impossible to simultaneously be in all places at all times. Nonetheless, people still look for a clear, overall picture to give sense to their experiences, and therefore missing elements are filled in and the incomplete picture is supplemented with fantasies that function as experiences and thus construct the full picture [35,50]. Thus, the development of meaning requires an overall understanding of the relevant situation, in terms of both ante-narrative and narrative.

As humans, we can only understand change with great difficulty, we notice when something has changed only after a certain period of time has elapsed and we perceive this as an interval [59]. As a result, change is reduced to a series of instances: the difference between one state of affairs and another gives us clues about change, determining our thinking about time in a profound way [60,61]. Due to modernity dictating a linear concept of time, we tend to experience that as ‘concrete lived time’ [62], and while change is basic to life, it is difficult for humanity to grasp it. In this sense, we are ‘becoming’ instead of ‘being’ [60,63,64]. The concept of becoming elicits the sense we make of change. Sense-making, or meaning creation, can be envisioned as a human attempt to comprehend change, in a process in which we attempt to convert an influx of stimuli into adequate concepts [62]. However, striving for fixed concepts in the process of sense-making means that intentional shifts in meaning rarely occur because of the tendency to maintain familiar concepts. Despite this tendency, meaning does change—usually without the awareness of the meaning creator—due to the changing environment. The propensity to ignore change by creating stable narratives is prevented by these changing circumstances, giving change the quality of ‘basic assumption’ or a ‘deep structure’ [65], or of basic, dichotomous, generally subconscious human preferences [66]. For Schein, the more superficial cultural notions are, the more they are subject to change, in which case perhaps it would be better to

describe both superficial and deep structures as changing, but with the latter not being narrated as such.

2.1 A Narrative Framework for Research

As the outcome of sense-making processes, narratives are dynamic. How narratives come into being and how existing narratives enhance or constrain new narratives, creating relative stability or a momentum towards change, will be conceptualised within this framework [67]. We will discern narrative conceptualisations about scene, actors and actions, in terms of narrative setting, narrative space and narrative storyboard respectively (see Figure 1) [40,68]. A narrative setting concerns notions about the narrated environment. Narrative spaces refer to configurations of actors and how they interact with each other and narrate their world, individually and collectively. Narrative storyboards arise from reflection on practices and are transposed into relatively fixed patterns, which can be regarded as the outcome of the propensity of human beings to consider sense-making itself in terms of fixed concepts [62].

The *narrative setting* conceptualises narratives about the environment, time and space. The notions of local and global, presence and absence and home and abroad are combined in the narrative setting, and images of the technological environment are also found here. Notions about change, stability and institutionalisation come together in an enacted location, which is to say, a locus where narratives combine [66,69,70]. People act in different ways, within different groups, within a specific narrated setting, acting in a local or global manner and in an explicit or implicit way. They create narratives about that specific location at that specific moment using images from the past, present and future, from the local community to the global environment. It is their lived experience of that location which is narrated.

The narrative setting also refers to the physical environment, which includes buildings [71,72]. Gastelaars analysed a building as a site, space or skin, and as a place to be [73]. She refers to the theatre, using the metaphor developed by Burke and Goffman in their notions of a

front-stage, an offstage or backstage, and the wider environment [17,40], making us aware that physical locations may have different functions in different contexts: in one situation the location may be considered to be front-stage, while in another it may be backstage or reflect an even wider environment. The presence of props and the 'personal front' of people, realised through physical objects, also needs to be mentioned in this context.

Thus, the narrative setting has tangible and non-tangible aspects [48,69,74]. An intangible software program used through a tangible computer is an example of a complex relationship which has a fundamental impact on how things are done [75]. Bijker has suggested that technology is shaped through images of how it will be used, being conceived in subjective, partial and distorted images [76,77]. The narrative setting enacts how technology in the lived environment is linked to time and space [60]. It is a relatively stable image of the environment, however vulnerable to redefinitions. Only when they become untenable, will narratives about the setting explicitly or tacitly change, influenced by narratives about the past, present and future.

One or more *narrative spaces* may be discerned within a narrative setting. They represent groups of people and are therefore the link to human existence. They might enact a department, organisation, profession, religion or subgroup. The interplay of narrative spaces might invoke action or conversely create a deadlock or cease-fire. Narrative spaces are ever-changing kinds of 'zoning plans' for enacted human groups, determining their nature and limits, and they may form quite complex combinations, as human thought is very capable of generating and handling these complexities. They do not necessarily have links to or comply with organisational or societal structures [45,66,78]. Governed by a search for predictability, narrative spaces appear to represent stability, enacting cultural entities to create a stable environment. However, narrative spaces are also vulnerable to change, as they must adjust to new developments, which are usually conceptualised as changeable, moving from one form of stability to another [79–81]. Thus, a narrative space is enacted as stable, offering a comfort zone, an image which invokes predictability, but also a path to follow, towards an enacted, desired state of affairs. Narrative spaces allow people to know what to do, who to trust and where to go. They make clear what is important and what is not, what is consonant or dissonant, and ultimately they indicate how to progress to another stage. While experienced as stable, narrative spaces are consciously and unconsciously changing.

Narrative storyboards are the bedrock of human actions, providing predefined scripts. In a world that is made up of a constant flow of events we enact that world as stable and predictable, while also requiring fixed recipes for action. Heavily anchored in narratives on the envi-

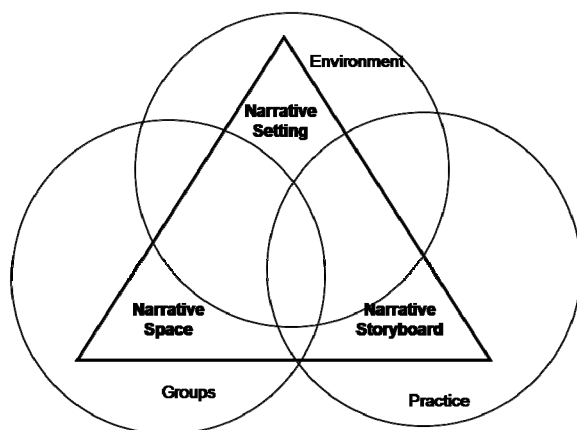


Figure 1. Theoretical focus

ronment and social groups, they are also based on past and future actions [28,48,59]. People adhere to certain unwritten rules in daily life, allowing them to present themselves as good citizens, and thus feel uncomfortable when the rules are not appropriately applied [82]. Storyboards provide us with a narrative of how to move from an initial state of affairs towards a new state within a particular context. They may relate to action that still needs to take place, that which is being undergone, or that which has already taken place, linking the action in question to time and space and thereby delimiting the storyboard's explanatory power. In this way a plot of the action is provided and related to the circumstances conceptualised in narrative settings and narrative spaces [17].

Storyboards emerge in relation to groups of people, who can be considered as apprentices becoming accustomed to a general way of doing something [83]. The people within such a group may feel confined in relation to a specific array of actions which have been proposed as a means to move from chaos to order [84]. Predictions concerning actions and outcomes are made because these allow people to know what to expect and to determine which stories are dominant and how they form a logical sequence [53]. The narrative storyboard makes us aware of the limited ways of creating a plot. It reveals how a specific storyboard connects to the setting and spaces of its constitutive narrative and what aspects of the narratives are specific to that storyboard. Their predictable features make them triggers for change. In this way, while the exact prediction of narrative progression is impossible, the narrative provides building blocks for the analysis of change, shedding light on how narrative change can be mapped [30].

3. Methods

This section will provide some information on the context of the Geo Portals project, as well as an explanation of the ethnographic research design.

3.1 Context

In early 2005, the National Initiative for Innovation Stimulation (BSIK) began the Space for Geo Information program (SGI) (*Ruimte voor Geoinformatie*). The project ran until the end of 2008, with a budget of 58 million euros. The SGI program was set up to provide grants to projects dealing with geo information and thereby stimulate innovation and promote the realisation of the Dutch National Geo Information Infrastructure (NGII). The Geo Portals project emerged from the initial discussions on the content and design of the SGI program, bringing together thirteen governmental and non-governmental organisations in the field of geo information who proposed the establishment of a network of geo portals for the disclosure of geo data. The Geo Portals project had a two million Euro budget, with 60 percent of its funding com-

ing from the SGI initiative, while the participating organisations were to supply the remaining 40 percent. Within the Geo Portals project, geo data was regarded as a crude product that should be thematically disclosed in order to obtain geo information from which society as a whole could benefit.

In relation to the multifaceted palette of the SGI program, Geo Portals was one of the larger projects, and was often described as a prestigious, key project by program officials. The projects that were set up were evaluated in terms of their ability to bring the Dutch NGII closer to completion. In this context, Geo Portals was focused on the overarching goal of the program: disclosing geo data from different sources to produce geo information.

3.2 Research Design

In the next section, we will present ethnography of the Geo Portals project, which ran from the beginning of 2005 until the end of 2008. It will become clear that narratives referring to the project changed during its progression. However, before turning to the case description we will explain the ethnographic design of our research.

One researcher monitored the project during its course. Because social scientific research on how the project was conducted was one of the project goals, the researcher was accepted as a full member of the project committee, which consisted of one representative from every participating organisation. Monthly meetings were scheduled with the intention of addressing management issues and, especially at the outset, serving as a platform for developing the scope of the project. Workshops open to and aimed at professionals within the geo information sector were also organised with the purpose of project promotion. Two brainstorming sessions were held by the project team in the first phase of the project, intending to establish a clear and univocal project approach agreed on by the project committee. These events were observed and also interviews were conducted with key persons, both during the commencement and conclusion phases of the project. Relevant documents and some video footage were also analysed. In addition, the researchers observed the presentations of the project at the geo information conferences, as well as the subsequent audience reactions.

Ethnographers have to be convincingly authentic ('been there'), plausible (relevant to the reader) and engage in critical analysis [85]. In order to do so, this research project followed writing conventions developed by Watson and extended by Duijnhoven concerning the transfer of field notes into convincing and authentic texts [86,87]. To meet these requirements, we will present excerpts from our interviews and field notes. In order to summarise the numerous discussions occurring during meetings, these have been condensed into a representation of the typical form of the discussion concerning a particular

topic. These representations of conversations are in essence fictitious; however, they are based on conversations recorded in field notes and, to a lesser extent, in interviews.

The research materials revealed the presence of narratives that developed over time. They were in continuous flux and either prominent or concealed, close or distant. The narratives within the project not only show how projects function as arenas where the narrative of change is created, contested, appropriated and diffused, but also how the quest for project narratives among members may serve both to reduce as well as to increase ambiguity. On the one hand, the project narratives seem to reduce ambiguity by providing a 'narrative of change' in terms of the use of new software applications. On the other hand, these software applications fail to offer a solution because they create a new ambiguous situation, requiring another 'narrative of change'. Coping strategies are developed through the redefinition of the initial project goals, aligning them to these narratives of change.

4. Analysis

In this section we provide a detailed description of three phases of the Geo Portals project. Each is described separately and followed by a narrative analysis that identifies the narrative setting, space and storyboard.

4.1 Getting Started

The SGI program started in 2002, with the basic idea of stimulating innovation in order to boost geo information sharing. The next step was to bring together representatives of organisations in the GI field to make goals more concrete. The result was a glossy brochure, with a program outline produced by a consortium of 10 universities, 20 research institutes, 60 companies, 40 governmental bodies and 30 geo information producers [88]. It was argued that government needed complex information about a complex society to develop convincing policies. To make the information manageable, it was to be ordered spatially as geo information, disclosed by a National Geo Information Infrastructure (NGII). The bottom line was to make geo information available in a structured manner, with it being disseminated independently by individual organisations.

To promote future projects, SGI organised 'broking and bargaining days' on which representatives of organisations from the GI sector were invited to generate project ideas. It was in this context that the concept of Geo Portals emerged. Some typical observations of those in attendance were as follows:

SGI mobilised the field. They organised broking and bargaining days in order to get rough ideas. Some 25 ideas were identified as potentially successful. In the end, these ideas were connected to organisations; it was just one big dating show. It became obvious that some central portal facility was needed and that our organisation should play a role in its development.

I remember how Geo Portals emerged. The idea behind broking and bargaining events organised by SGI was that through discussion among representatives of geo information organisations, ideas for concrete projects would pop up. During one of those meetings, the Geo Portals concept just came out of a plenary discussion. Then the moderator asked which organisations were willing to participate. Representatives of interested organisations raised their hands, as did I. So, all of a sudden I was an initiating member of an instantly formed club of enthusiastic people who wanted to disclose geo information through portals.

That the overarching concept of Geo Portals should be *liberty united*, was obvious from the outset. A central, top-down organisation was totally out of the question. The idea was a network of portals of different nature, working together with a minimum set of rules.

Those involved in the discussion saw the rudimentary concept of Geo Portals as a collective idea in need of development. The thirty organisations willing to participate were gradually reduced to thirteen, and in October 2002, representatives from these organisations presented an initial proposal which envisioned thematically categorised, colour-coded portals like red for built environment, green for nature and agriculture, and brown for sub-surface conditions [89].

After the initial submission in 2002, a rewriting process occurred, giving the project more focus. In the minutes of early project meetings, there are clear conceptions about how data should be distributed. It was stated that all the processes for disclosure, search, diffusion and payment should be web-based, while how all the different data sources were to be connected was not a matter of discussion. The first rudimentary description of the geo portal framework presented a static image: the portal would be based on proven technology and standards and also on a fixed notion of architecture [90].

While the project goals were stated clearly and unambiguously, at their regular meetings the representatives of the participating organisations expressed doubts about how to proceed. They were uncertain about the financing and procedures for reporting to SGI, but even more about the essence of the project. Now the project was about to start, the representatives felt the need for definitions about what a portal should look like, how users would be reached and what technology would be used in its setup. A typical discussion in a meeting of representatives would proceed as follows:

A: If we want to set up a proper Geo Portals, we need to be clear about standards. It is obvious that we use the most recent and commonly used standards. We are not going to use any standard that has not been accepted by the community, or that has not proved to be useful.

B: I agree on that. If nobody objects, we should proceed to the next topic, and that is user orientation. We have to be demand-driven, preventing us from making the same mistakes they made in the NCGI project. So how can we be demand-driven?

C: First and foremost we need to disclose our data in a way that it can be readily found. Furthermore, we need to present it in a format that can be read by the user. So, we need to use the proper standards.

B: I agree. We need to use proper standards, those that are widely accepted.

A: Now we agreed on how to settle the standards issue, we are discussing standards again.

The motto of Geo Portals was 'liberty united', which reflected the fact that it was a network of portals established by various organisations, each with its own autonomy, but working according to a minimal set of rules. In defending this view of the essence of Geo Portals, it was often explained as a reaction to a former national project regarding geo information, the National Clearinghouse for Geo Information (NCGI). The feeling was that NCGI had failed due to the central, top-down enforcement of detailed standards and work procedures and this had proved to Geo Portal protagonists that organisations were not inclined to comply voluntarily with strict rules. To avoid another failure, they decided to meet as a small group of motivated organisations connected through a minimal number of mutually agreed standards.

While Geo Portals was sketched out in organisational terms, discussions on how to proceed would always come down to technical matters. Standardisation was considered to be crucial, followed by the question of whether the data was accessible enough. The bottom line was that it was most important that the issue of technological standardisation should be settled properly. Technological matters dominated discussions:

A: Technology is not really a problem anymore. We can build everything we want without any limit. All the techniques needed are at our disposal.

B: That's right, the things that do matter are organisational aspects. Look at the US example of Geospatial One Stop. They just do it: American government agencies put everything they have on the web, without restrictions.

C: But its quality is doubtful at best, they don't guarantee its accuracy. I wonder if anybody actually uses it.

A: If we follow the example of Geospatial One Stop, then it will look like NCGI. We have to do better than that.

B: Just use the right standards. That is of paramount importance. The architecture we have developed is perfectly equipped to set up a network.

A: If we stick to proven technology and standards, nothing can go wrong.

B: But what is that, which standard is proven, which standard is commonly used, which one really works?

C: Here we go again!

In November 2005, the core team, made up of representatives of a few major participating organisations, attempted to tackle the problems experienced by calling the project team together for a two-day brainstorming session in

a remote countryside hotel. The technology and standardisation issues had been declared settled, but still played a role, while the intention was to produce a strategy for developing a user-driven approach. The program for the session mentions a meeting with a public relations consultant and the question of how to bring more user-drivenness into the project. In fact, user orientation was extensively discussed, eventually leading to a 'motto' of which the team was very proud: 'Able to find and allowed to use'.

The subsequent working conference, in which the project was to be presented to the GI community in December 2005, was also a pressing issue. The project team had mixed feelings about whether there was anything tangible to demonstrate and thought that if this was not the case, it would be better to cancel the presentation. After some deliberation it was agreed that a rudimentary version of the Red Portals would be demonstrated.

Thus, in December 2005 the Geo Portals project was launched before a GI audience at the conference. The core team was determined to make a convincing statement by showing that the project was user-driven and was doing the right thing in terms of technology, but also felt a little uncertain. The audience was familiar with SGI and its projects and knew of the existence of the Geo Portals project, but was unfamiliar with the details. Sheer curiosity brought about fifty GI professionals together.

In his introduction, the scientific director of SGI signified the importance of Geo Portals for SGI, proclaiming it to be a key project. The core team then gave a presentation about the demand-drivenness of the project and elucidated the 'motto'. Despite the importance with which this was regarded by the project team, it barely raised the interest of the audience. However, the demonstration of a rudimentary version of the Red Portals website using data from the built environment had an astonishing effect. What the Geo Portals team considered window-dressing was the very thing that convinced the audience of the project's importance. In subsequent discussions it became apparent that participants were convinced that the Geo Portals project was SGI's key project and that it was technically well managed and would make a difference. The Geo Portals project team celebrated the day as a success.

4.1.1 Narrative Setting, Space and Storyboard

In this case, technology is the dominating factor in the narrative setting. In the past it has been an impediment with respect to infrastructure development, but in this setting this was no longer the case, the team considering it possible to apply GI technology for the disclosure of data in a way that society as a whole would benefit. In this setting, GI technology is seen as an ever-developing and changing phenomenon that will be mastered through the application of standards and result in an infrastructure with a rather static form, divided into thematically organised compartments of data that give it a neatly arranged appearance.

In the narrative space, the project team has a direct re-

relationship with the GI community. Individual project members belong to organisations that financially support the project, but these organisations are not recognised as individual partners. As a whole, the organisations have a neutral and minor role and are all seen as equal and as supporting the common cause of sharing GI data. GI data users are recognised as a defined group through the user motto, but a clear picture of these users has still not been developed.

A storyboard emerges concerning the propensity to let technology work for the GI sector through the application of standards. The Geo Portals project is seen to be acting on behalf of the entire GI sector, detached from individual organisations and creating a stable infrastructure.

4.2 Attempting to Reduce Uncertainty

The project team continued its project meetings on a fixed day of the month in a centrally situated venue, with meetings held in a building occupied by one of the participating organisations. The morning agenda was devoted to management matters, while discussions prepared by a core team member or an external speaker took place in the afternoon. However, fundamental issues would emerge during the morning sessions and be discussed over lunch, sometimes continuing throughout the day, suggesting a certain level of insecurity. Nevertheless, a research paper written by the project members to convince European peers expressed confidence [91].

The Geo Portals project was meant to provide all possible kinds of data, to be delivered to both professional users and the general public. Professional users only needed disclosed data, while lay users could be provided with software services which had to be developed for integrating, harmonising and presenting data. Existing examples of the disclosure of geo data through websites were reviewed, the flaws convincing the project members that there were many difficulties involved in bringing together different sources. Services designed to harmonise and present data were seen as essential to Geo Portals, emphasising the user orientation of the project, which was communicated to the GI community. The core team developed the example of a beer brewer in need of geo information to assist in finding a location for a new brewing facility. In all the subsequent presentations and promotional material, including an SGI promotional film, this example—which connected different processes within different public organisations—was made prominent [92].

User orientation also generated interest in legal aspects and the issue of digital rights management. A researcher affiliated with Geo Portals translated an approach for regulating copyright on the internet into a model applicable to the field of geo information. This model, regulating legal and economic aspects of geo information, was regarded as essential for Geo Portals, although, however important it was felt to be, it was also seen as a separate entity, unlike technological issues. Technology was held to be dynamic,

while the access model was found to be static. Further development of the model was embedded in another SGI project, placing it beyond the control of the project team.

At the end of 2006, the project team began to feel uncomfortable about the lack of steering capacity at SGI. While SGI saw Geo Portals as the core project of the program, the core team thought SGI, giving voice to the management of individual organisations, should provide an overarching framework. As SGI was seen as the custodian of the National Geo Information Infrastructure, a serious discussion among project participants was devoted to this topic:

A: We are supposed to work on NGII. For SGI, Geo Portals are considered as focal, but they don't say anything about the guidelines we should follow or how to connect to other projects that are part of the NGII.

B: They are talking about a test bed for NGII, but is NGII only a test bed? Are we supposed to deliver something that actually works?

C: We are certainly working on our data viewer, but to what standards should it comply? Are there any organisations that are going to use it?

A: They say that a new GI coordinating organisation is in the making—yet another organisation that is supposed to organise something. We need guidelines and all they do is establish a new organisation. This does not sound like coordination to me!

D: I think that as a Geo Portals team we should take a stand and do what SGI refuses: take the lead!

The core team did not feel supported by SGI, which until then had been seen as the keeper of the National Geo Information Infrastructure, of which Geo Portals was a part. At the end of 2006, SGI published an article in a leading professional magazine with the provocative title: 'Where to with the Dutch Geo Information Infrastructure?' [93]. It provoked discussion and also made the core team feel that SGI had no strategy.

Geo Portals concentrated on the work to be done: new services had to be developed with new software. Choices had to be made on what technology to use and what standards to apply. The core team, representing three government-supported knowledge institutions and a software company, felt responsible for this part of the project and took up the challenge of drawing up a framework and organising software development. A participating engineering firm also did some work, but took little part in any conceptual, organisational or management activities.

During the software development process, the core team came together on a weekly basis to coordinate software development which was undertaken by software engineers from core team member's organisations. In spring 2007, these efforts resulted in a data viewer, a software device designed to be capable of consistently retrieving geo data from different sources on a computer screen. The Geo Portals core team, being enthusiastic about it, saw it as a

requirement for bringing the ultimate goal, a system of Geo Portals, one step closer.

While celebrating this achievement, project members soon felt that the newly developed data viewer was already becoming outdated because new techniques were now available. This gave software engineers the opportunity to develop even more sophisticated viewers. Thus, while having a tested product ready for implementation, the development process went on, with an enthusiastic core team managing the same team of software developers. While working with the newest technologies they gave the impression that these developments were quite normal for them—new technology had to be explored and applied.

4.2.1 Narrative Setting, Space and Storyboard

In this phase of the project, the narrative setting becomes increasingly dominated by technology. To serve lay users, services have to be developed using state-of-the-art technology. Standards are still important but they are appraised as being of lesser concern. Legal aspects are seen as a separate area that needs to be dealt with, but not necessarily by the project management team.

In the narrative space, the management of individual participating organisations is seen as collectively organised into advisory boards of the SGI program. The program itself is considered to be unsupportive, as it simply does not have a policy, and those on the boards are not seen as GI experts, but as serving the interests of individual organisations, which are not necessarily the interests of the Geo Portals project. Those involved in the Geo Portals project must recognise that in order to be successful they must plot their own course, which will be to address the newest trends in GI technology.

The storyboard at this stage is at the point of exploring the latest GI technology and incorporating this into a test website. Once the technology is ready to be used as a building block for GI infrastructure, further effort will be put into assessing newer technological improvements.

4.3 Towards Judgement Day

In 2007, the Geo Portals project was on track as far as software development was concerned, but the core team was becoming increasingly agitated, feeling that the initial goal of sharing geo information was moving out of reach. At the project team meeting in April 2007, a discussion on this point was initiated by two core team members in an attempt to engineer a breakthrough:

It is terribly sad that we cannot build on the achievements of SGI. It looks like management does not recognise what it is all about. In the Netherlands we have an abundance of geo data, distinguished scholars, high GIS penetration, a vast and schooled workforce and many knowledge exchange networks. Perfect circumstances for great ideas. But guess what? We just keep on chatting!

Nobody seemed to be in charge of developing the NGII, and the decision-makers at SGI were depicted as abstract

thinkers with no practical knowledge. It was felt that a breakthrough was needed, and the appraisal of the SGI promotional conference held in March 2007 did not display any confidence:

A: I am sad to say that real sharing of geo information is further away than ever. We have just had the SGI conference in Rotterdam. It lacked any ambition. The bottom line was: 'The NGII has to be developed, but let's move on as we did'. That's not the way to get it done.

B: It was a convention of the same people that you see all the time at such events; 'the usual suspects' were doing their ritual thing.

C: It was like being in some religious rally, people celebrating and praising something of which everybody has a different image.

B: It is a paradoxical situation. When we need a breakthrough, surprise, surprise, nobody wants to change, we keep on doing things the way we did, and nothing really changes.

C: Everybody talks about the costs of an NGII, the benefits are not mentioned.

A: An NGII will add value, that's the *raison d'être*. If we only want an NGII for incident management and fighting terrorism we're on the wrong track.

Despite the uncertainty, Geo Portals was considered to be successful because it offered technical solutions. The technology only had to be brought to a meaningful conclusion in order to establish the NGII, but failing management seemed to obstruct this. Perceptions of the role of Geo Portals started to change:

It is perfectly clear that it was unattainable to build an infrastructure. Just look at the budget we had for this project: it was clear even before we started that it was insufficient. Our job was to deliver building blocks, to innovate for the sake of an NGII.

We are good at the technological aspects. So if they ask us for such a project, we will handle technology. Without any guidance from SGI, it is impossible to develop an NGII. What we can offer for a future NGII is best practices and software tools. We form a community for NGII development.

Another working conference was organised for November 2007 with a striking theme: 'Just do it'. External experts were asked to focus on financial, legal and organisational aspects, while Geo Portals project members were keen to present the technical aspects. The message in workshops was that new software applications, as developed by Geo Portals, were fully capable of integrating geo data from different sources. This message was symbolised using Lego blocks, representing geo data building blocks which could be put together in any possible way.

Now that the finish was in sight, the project team wanted to deliver results which could be used in the future. Slowly but steadily, the project goals were redefined. The obligation to produce tangible products changed, with the Geo Portals team coming to see itself as a 'commu-

nity of practice'. The image of the project as developing building blocks for an NGII now changed, with Geo Portals being reconceived as a knowledge-creating project. The atmosphere also changed, from distress to optimism to euphoria, although one of the more sceptical project team members noted that what was occurring was 'expectation management'.

It was felt that the positive results should be disseminated to the GI community, for example in a research paper [94], and a new sector-wide policy coordinating organisation called Geonovum began to promote itself. While the Geo Portals project team had at first thought that this organisation was covering up the failings of the geo information sector, they now thought that Geonovum could secure the innovative achievements of Geo Portals for the future. The image of SGI changed accordingly, from being purely involved in funding to becoming a knowledge-boosting program that should be continued.

At the closing conference in December 2008 there was confidence about the results. The highest civil servant responsible for geo information in the Ministry of Housing, Spatial Planning and the Environment was the keynote speaker, addressing 150 people in a prestigious location. A specially produced video presented the improvement of the accessibility of geo information as an ongoing project, suggesting that there was much work still to be done. Software applications were presented as stepping stones in a continuous progression, invoking a great deal of interest in newly developed techniques. A new website with a new name (Carta Fabrica) was also launched, where the achievements of Geo Portals were to be made available. Both the core team and the audience were optimistic about the future.

In interviews held after the completion of the project, the image of technology as dominating all developments was persistent. Standards were seen as a thing of the past because technology was now seen as being capable of connecting all forms of data. The approach was referred to as 'Web 2.0', signifying that the new technology was obviously web-based. It was also noted by Geo Portals project members that Geonovum was still working on a National Geo Register aiming at the registration and standardisation of all governmental geo data but that this project was obsolete because Web 2.0 would solve all the connection problems where standardisation had failed. However, most importantly, the National Geo Register was seen as a project that hampered innovation in the geo information sector.

4.3.1 Narrative Setting, Space and Storyboard

In the narrative setting, technology is now treated as the essence of Geo Portals. Technology is seen as an unleashed phenomenon, now labelled as 'innovation', and it is ready to solve any problem, with the aim of making the world a better place. Innovation is thus seen as an enabler of dynamic geo information management, without being chained by standards. However, the solutions created by this

technology are found to be obsolete before they can be used, not because they do not function properly but because they are superseded by solutions powered by even more sophisticated technology.

In the narrative space, both diverging and converging tendencies can be observed. The GI sector management, speaking through organisations such as Geonovum and SGI with their emphasis on standards, is found by Geo Portals project members to inhibit the possibilities created by the application of technology. By providing insufficient funding they are also seen as responsible for not delivering Geo Portals as originally planned. Realising that the initial goals were untenable, the Geo Portals team redirected their aim towards creating innovation to facilitate the creation of an NGII. As the SGI was supposed to stimulate innovation in geo information sharing, the Geo Portals project team felt quite comfortable with their new goals, knowing that their project would stimulate innovation.

The storyboard that can be identified here aims at the production of new technologies which could be made available to the GI sector. It affects the reframing of goals, moving from the creation of a static infrastructure into making new technologies available. This reframing is justified through concluding that the funds originally granted by SGI were inadequate to realise the GI infrastructure considered in the initial plan.

5. Discussions

In this paper, we have used the framework of narrative setting, space and storyboard to analyse the Geo Portals project. Three phases of the project were identified, in which the narrative setting and space could be placed in a relationship with a developing storyboard. The Geo Portals project had a clear beginning and end, and there were also some preparatory activities which were considered to be important for the analysis, as well as the impact of the project on the Dutch GI sector.

Initially, the Geo Portals project proposal was to develop an infrastructure serving societal needs. These needs were converted into user profiles with different demand structures. As project participants became dissatisfied with the lack of guidelines for an overarching strategy, they started to develop software applications. Because they considered themselves to be the vanguard of ever-changing technology, the idea of building an infrastructure slowly faded. Consequently, the goal shifted towards providing a toolbox, which in turn changed into the image of the project as stimulating innovation.

The narrative setting, dominated by rapidly developing information technology, encouraged project participants to look to the future, and the Geo Portals project acted as a means to deal collectively with this task and to apply the latest technology to create newly developed software applications. Geo Portals project members, acting inde-

pendently of their respective organisations, made new technology available, while unintentionally ensuring that no individual or organisation could be blamed for failure. Because the Geo Portals project was supposed to be beneficial to the whole GI sector, the project team decided to supply state-of-the-art technology.

In the narrative space, SGI was seen as an enabling organisation, acting on behalf of the Dutch GI community, in relation to which the Geo Portals project would be beneficial to the whole sector. The Geo Portals project team saw SGI as a temporary organisation, being part of the GI community and primarily involved in sustaining the Geo Portals concept through funding with money budgeted for the GI community. This relationship made the project team cautious, responsible and somewhat self-reflective. Thus, SGI brought the GI community together around a financing source, forcing individual organisations to cooperate with each other in order to be eligible for funding.

The analysis shows a cyclical storyboard: whenever new technology was tested and approved, newer technology was already virtually available to be tested and eventually to be confirmed as a new standard. The data shows two of these cycles, with the typical pattern being depicted in Figure 2. This is the storyboard of the action occurring within the project, which can also be interpreted as a vicious circle [95].

In a world with a pressing and increasing turnover of technological innovations, reliable infrastructures might create stability. The two competing narratives of stability and change always struggle for dominance. An infrastructure is a fixed, predictable, stable, unambiguous and ubiquitous facility that users almost take for granted [96] and a focus on the development of a standardised infrastructure utilises the narrative of stability, a prominent feature in the initial Geo Portals project proposal. The difficulties involved in standardisation were already recognised in the project's subtitle: 'liberty united' and a strict regime of standardisation was also feared, as well as being considered difficult to implement. Therefore, a limited, 'light' version of standardisation was proposed.

Throughout the project, from the initial presentation of the Red Portals, which was hailed as innovative, until the conclusion, when the entire Geo Portals project was declared innovative, the emphasis was on change. Newly developed software, already obsolete on the day of its realisation, was not considered a problem. Moreover, it was seen as essential, as the average GI professional sees tomorrow's technology as the solution to problems encountered today.

The storyboard of innovation remains prominent. The core message of SGI, to be innovative, hampered the development of an infrastructure. For this reason, the project was reframed into a knowledge-generating endeavour, driven

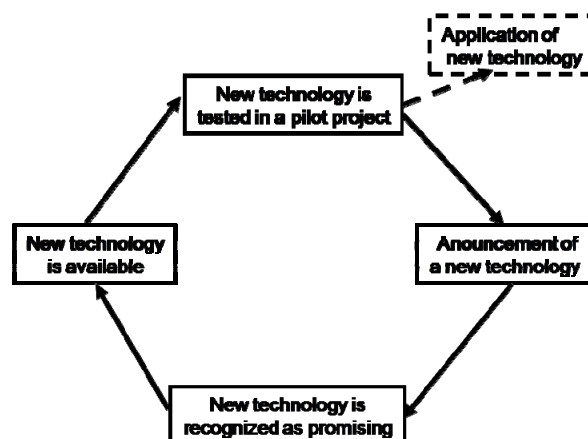


Figure 2. The storyboard of innovation

by a storyboard of innovation. Ultimately, the GI community would judge the project on its innovative qualities, presented through state-of-the-art software. While this is a tangible result of the four-year Geo Portals project, it is only temporary, with no reference to infrastructure.

6. Conclusions

Delivering infrastructure seems to involve two contradictory aspects [97]. On the one hand there is a narrative of change, expressing the urge to work with the newest technology, and on the other hand there is a narrative of stability which sees infrastructure as predictable and stable and thus useful. These two narratives seem to fight for attention.

As the Geo Portals program basically aimed at innovation, the narrative of change was dominant, and can be identified in the innovation storyboard. Infrastructure development rather than infrastructure building was paramount, and thus a stable, recognisable infrastructure was absent.

The narratives reflect the basic stability/change contradiction [66]. The confrontation of these two differing narratives is not uncommon and has been called the 'innovation paradox'. It is found in large public sector projects where a fixed infrastructure has to be delivered in an unstable environment [49].

It has been suggested that when problems with the construction of infrastructures emerge it is necessary to focus on project designs in the light of cultural settings [98]. However, here there was more at stake. A GI community, seemingly preoccupied with innovation, desperately required a useable infrastructure. While one of those involved in the project suggested that infrastructures are always in a process of innovation and should be regarded as 'moving targets', in order to be used, infrastructures also need to be stable. Thus, the sector as a whole must find equilibrium between stability and change in relation to infrastructure. Now that these driving forces have been identified, a breakthrough is within reach.

REFERENCES

- [1] J. Crompvoets, "National spatial data clearinghouses, worldwide development and impact," Wageningen Universiteit, Wageningen, 2006.
- [2] A. Rajabifard and I. Williamson, "Spatial data infrastructures: Concept, sdi hierarchy and future directions," *Geomatics* 80, Tehran, Iran, 2001.
- [3] D. Nebert, "The SDI cookbook," 2004. <http://www.gsdi.org/pubs/cookbook/>
- [4] F. de Bree and A. Rajabifard, "Involving users in the process of using and sharing geo-information within the context of SDI initiatives," *Pharaohs to Geoinformatics*, FIG Working week, 2005.
- [5] I. Masser, "GIS worlds: Creating spatial data infrastructures," ESRI Press, Redlands CA, 2005.
- [6] Y. Georgiadou, S. K. Puri, and S. Sahay, "Towards a potential research agenda to guide the implementation of Spatial Data Infrastructures: A case study from India," *International Journal of Geographical Information Science*, Vol. 19, No. 10, pp. 1113–1130, 2005.
- [7] J. Crompvoets *et al.* eds., "A multi-view framework to assess SDIs," *Space for Geo-Information (RGI)*, Wageningen University and Centre for SDIs and Land Administration, Department of Geomatics, University of Melbourne, Melbourne, Australia, 2008.
- [8] H. Koerten, "Assessing the organisational aspects of SDI: Metaphors matter," *A Multi-View Framework to Assess SDIs*, J. Crompvoets *et al.*, Editors, Melbourne, pp. 235–254, 2008.
- [9] Y. Georgiadou, O. Rodriguez-Pabón, and K. Lance, "Spatial data infrastructure (SDI) and e-governance: A quest for appropriate evaluation approaches," *URISA Journal*, Vol. 18, No. 2, pp. 43–55, 2006.
- [10] W. van den Toorn and E. de Man, "Anticipating cultural factors of GDI," in *Geospatial data infrastructure: Concepts, cases and good practice*, R. Groot and J. McLaughlin, Editors, Oxford University Press, Oxford, 2000.
- [11] E. de Man, "Cultural and institutional conditions for using geographic information: Access and participation," *URISA Journal*, Vol. 15(APA I), pp. 29–33, 2003.
- [12] H. Koerten, "Blazing the trail or follow the Yellow Brick Road? On geo-information and organizing theory in GI days," *Young researchers Forum*, 10–12 September 2007, F. Probst and C. Kessler, Editors, Institut für Geo-informatik, Universität Münster: Münster, Germany, pp. 85–104, 2007.
- [13] D. Hodgson and S. Cicmil, eds., "Making projects critical," in *Management, Work and Organisations*, Palgrave Macmillan, New York, 2006.
- [14] A. van Marrewijk and M. Veenswijk, "The culture of project management: Understanding daily life in complex megaprojects," Pearson Education Limited, Harlow, UK, 2006.
- [15] H. Blumer, "Symbolic interactionism, perspective and method," Prentice Hall, Englewood Cliffs, NY, 1969.
- [16] G. Ritzer, "Modern sociological theory," Mc Graw-Hill International Editions, NY, 1996.
- [17] E. Goffman, "The presentation of self in everyday life," Doubleday Garden City, NY, 1959.
- [18] G. Goffman, "Frame analysis," Harper & Row, NY, 1974.
- [19] G. Gonos, "'Situation' versus 'Frame': The 'Interactionist' and the 'Structuralist' analysis of everyday life," *American Sociological Review*, Vol. 42, pp. 854–867, 1977.
- [20] P. Manning, "Erving Goffman and modern sociology," Stanford University Press, Stanford CA, 1992.
- [21] J. Habermas, "The theory of communicative action," *Life-world and System*, Beacon Press, Boston MA, Vol. 2, 1987.
- [22] A. Giddens, "Modernity and self-identity: Self and modernity in the late modern age," Stanford University Press, Stanford CA, 1991.
- [23] U. Beck, "Risk society," Sage Publications, Newbury Park, California, 1992.
- [24] G. Ritzer ed., "The McDonalidization of society revised edition," Pine Forge Press, Thousand Oaks CA, 1996.
- [25] A. Giddens, "Central problems in social theory: Action, structure and contradiction in social analysis," Basingstoke, MacMillan, 1979.
- [26] M. Archer, "Culture and agency: The place of culture in social theory," Cambridge University Press, Cambridge, 1988.
- [27] P. Bourdieu and D. Pels, "Opstellen over smaak, habitus en het veldbegrip," Van Gennep, Amsterdam, 1989.
- [28] P. L. Berger and T. Luckmann, "The social construction of reality," Penguin Books, Harmondsworth, 1967.
- [29] K. Gergen, "Realities and relationships, soundings in social construction," Harvard University Press, Cambridge MA, 1994.
- [30] D. Boje, "Stories of the storytelling organization: A post-modern analysis of Disney as 'Tamara-Land'," *Academy of Management Journal*, Vol. 38, No. 4, pp. 997–1035, 1995.
- [31] M. Berendse, H. Duijnhoven, and M. Veenswijk, "Editing narratives of change, identity and legitimacy in complex innovative infrastructure organizations," *Intervention Research*, pp. 73–89, 2006.
- [32] D. Polkinghorne, "Narrative knowing and the human sciences," State University of New York Press, NY, 1988.
- [33] M. Hatch and D. Yanow, "Organization theory as an interpretive science," *The Oxford Handbook of Organizational Theory*, C. Knudsen and H. Tsoukas, Editors, Oxford University Press, Oxford, UK, 2003.
- [34] K. Verduijn, "Tales of entrepreneurship, contributions to understanding entrepreneurial life," PhD thesis Vrije Universiteit Amsterdam, Amsterdam, 2007.
- [35] P. Ricoeur, "The model of the text: Meaningful action considered as text", *New Literary Society*, Vol. 5, pp. 91–120, 1973.
- [36] C. Oswick, T. Keenoy, and D. Grant, "Discourse, organizations and organizing: Concepts, objects and subjects," *Human Relations*, Vol. 53, No. 9, pp. 1115–1124, 2000.
- [37] D. Grant, *et al.*, eds., "Introduction: Organizational discourse: Exploring the field," *The Sage Handbook of Organizational Discourse*, Sage Publications Inc, Thousand Oaks CA, pp. 1–36, 2004.

- [38] D. Grant, T. Keenoy and C. Oswick, eds, "Discourse + Organization," Sage Publications Ltd, London, UK, 1998.
- [39] N. Frye, "Anatomy of criticism: Four essays," Princeton University Press, Princeton, NJ, 1957.
- [40] K. Burke, "A grammar of motives," University of California Press, Berkeley and Los Angeles, 1969.
- [41] K. Gergen, "An invitation to social construction," Sage, London, 1999.
- [42] Y. Gabriel, "Storytelling in organizations, facts, fictions, and fantasies," Oxford University Press, Oxford, UK, 2000.
- [43] T. Peters and R. Waterman, "In search of excellence, lessons from America's best-run companies," Harper & Row, New York, 1982.
- [44] S. Helmers and R. Buhr, "Corporate story-telling: The buxomly secretary, a pyrric victory of the male mind," *Scandinavian Journal of Management*, Vol. 10, No. 2, pp. 175–191, 1994.
- [45] E. Schein, "Culture: The missing concept in organization studies," *Administrative Science Quarterly*, Vol. 41, No. 2, pp. 229–240, 1996.
- [46] J. Brown, *et al.*, "Storytelling in organizations: Why storytelling is transforming 21st century organizations and management," Butterworth-Heinemann, Burlington MA, 2005.
- [47] A. Brown, P. Stacey, and J. Nandhakumar, "Making sense of sensemaking narratives", *Human Relations*, Vol. 6, No. 8, pp. 1035–1062, 2008.
- [48] K. E. Weick, "Sensemaking in organizations", Sage Publications, London, 1995.
- [49] M. Veenswijk, "Surviving the innovation paradox: The case of Megaproject X," *The Innovation Journal, The Public sector Innovation Journal*, Vol. 11, pp. 2, article 6, pp. 1–14, 2006.
- [50] J. Bruner, "The narrative construction of reality," *Critical Inquiry*, Vol. 18, pp. 1–21, 1991.
- [51] M. Alvesson and D. Kärreman, "Variety of discourse: On the study of organizations through discourse analysis," *Human Relations*, Vol. 53, No. 9, pp. 1125–1149, 2000.
- [52] C. Kohler Riessman, "Narrative analysis," Sage Publications, Newbury Park CA, Vol. 30, 1993.
- [53] B. Czarniawska-Joerges, "A narrative approach to organization studies," Sage Publications Inc, Thousand Oaks CA, Vol. 43, 1998.
- [54] D. Yanow, "Conducting interpretive policy research," Sage Publications Inc, Thousand Oaks CA, 2000.
- [55] D. Boje, "Narrative methods for organizational and communication research", Sage Publications Ltd, London, 2001.
- [56] S. Tesselaar, I. Sabelis, and B. Ligtoet, "Digesting stories: About the use of storytelling in a context of organizational change," in 8th International Conference on Organizational Discourse 2008, London, UK, 2008.
- [57] N. Beech and C. Huxham, "Cycles of identity formation in inter-organizational collaborations," *International Studies of Management & Organization*, Vol. 33, No. 3, pp. 28–52, 2003.
- [58] T. Deal and A. Kennedy, "Corporate cultures: The rites and rituals of corporate life," Penguin, 1982.
- [59] H. Bergson, "The creative mind," Greenwood Press, Westport CT, 1946.
- [60] G. Burrell, "Back to the future: Time and organization, in rethinking organization, new directions in organization theory and analysis," M. Reed and M. Hughes, Editors, Sage Publications Ltd, London, UK, pp. 165–183, 1992.
- [61] G. Burrell, "Time and talk," *Organization*, Vol. 7, No. 3, pp. 371–372, 2000.
- [62] R. Chia, "Essay: Time, duration and simultaneity: Rethinking and change in organizational analysis," *Organization Studies*, Vol. 23, No. 6, pp. 863–868, 2002.
- [63] M. Heidegger, "Sein und zeit [Being and time]," Tübingen D, Max Niemeyer, 1977.
- [64] B. Czarniawska-Joerges and G. Sevón, eds, "Translating organizational change," ed. A. Kieser, de Gruyter Studies in Organization, Walter de Gruyter, Berlin, Germany, Vol. 56, 1996.
- [65] E. Schein, "Organizational culture and leadership," Jossey Bass, 2nd Edition, 1992.
- [66] M. Douglas, "How institutions think", Syracuse University Press, Syracuse, 1986.
- [67] S. Chreim, "The continuity-change duality in narrative texts of organizational identity," *Journal of Management Studies*, Vol. 42, No. 3, pp. 567–593, 2005.
- [68] R. Harré, "Life sentences, aspects of the social role of language," John Wiley & Sons, London, UK, 1976.
- [69] H. Lefebvre, "The production of space," Blackwell Publishers, Oxford, UK, 1991.
- [70] W. R. Scott, "Institutions and organizations," Sage Publications, 1995.
- [71] D. Yanow, "Built space as story: The policy stories that buildings tell," *Policy Studies Journal*, Vol. 23, No. 3, pp. 407–422, 1995.
- [72] D. Yanow, "How built spaces mean, in interpretation and method: Empirical research methods and the interpretive turn," D. Yanow and P. Schwartz, Editors, M. E. Scharpe, Armonk NY, 2006.
- [73] M. Gastelaars, "Talking stuff: What do buildings tell us about an organization's state of affairs?" in 8th International Conference on Organizational Discourse, London, UK, 2008.
- [74] B. Schneider, "The people make the place," *Personnel Psychology*, pp. 437–453, 1987.
- [75] W. Orlikowski, "Using technology and constituting structures: A practice lens for studying technology in organizations," *Organization Science*, Vol. 11, No. 4, pp. 404–428, 2000.
- [76] W. Bijker, "Of bicycles, bakelites, and bulbs, towards a theory of sociotechnical change," The MIT Press, Cambridge, 1995.
- [77] W. Orlikowski, "Sociomaterial practices: Exploring technology at work," *Organization Studies*, Vol. 28, No. 9, pp. 1435–1448, 2007.
- [78] M. Lipsky, "Street-level bureaucracy: Dilemmas of the individual in public services," Russell Sage Foundation, New York, 1980.

- [79] S. Barley, "The alignment of technology and structure through roles and networks," *Administrative Science Quarterly*, Vol. 35, pp. 61–103, 1990.
- [80] J. Bartunek, "The importance of contradictions in social interventions," *Intervention Research*, Vol. 1, pp. 103–113, 2004.
- [81] S. Ybema, "Constructing collective identity: Central, distinctive and enduring characteristics?" in 8th International Conference on Organizational Discourse, London, UK, 2008.
- [82] H. Garfinkel, "Studies in ethnomethodology," Polity Press, Cambridge, 1984.
- [83] E. Wenger, "Communities of practice: Learning, meaning and identity," Cambridge University Press, Cambridge, UK, 1998.
- [84] B. Latour and S. Woolgar, "Laboratory life: The construction of scientific facts," Princeton University Press, 1986.
- [85] K. Golden-Biddle and K. Locke, "Appealing work: An investigation of how ethnographic texts convince," *Organization Science*, Vol. 4, No. 4, pp. 595–616, 1993.
- [86] H. Duijnhoven, "Tales of Security practices within Spanish and Dutch railway operators: Translation, transformation or transgression?" in 8th International Conference on Organizational Discourse, London, UK, 2008.
- [87] T. Watson, "Ethnographic fiction science: Making sense of managerial work and organizational research processes with Caroline and Terry," *Organization*, Vol. 7, No. 3, pp. 489–510, 2000.
- [88] RAVI, "Space for geo-information BSIK knowledge project proposal," RAVI, Amersfoort, 2003.
- [89] A. Schmidt and G. Nieuwenhuis, "Geoportals 'Liberty United'," Project proposal for space for geo information, Wageningen, 2002.
- [90] M. Hoogerwerf and B. Vermeij, "Geoportal framework version 0.2," (Geoloketten Raamwerk versie 0.2), 2005.
- [91] J. Zevenbergen *et al.*, "Connecting the Dutch geo-information network–liberty united," UDMS, Aalborg, Denmark, 2006.
- [92] D. van de Laak, "DVD: Alles draait om geo," *Ruimte voor Geo-Informatie*, Netherlands, 2007.
- [93] A. Bregt and J. Meerkkerk, "Waarheen met de nationale geo-informatie infrastructuur?" *Geo-Info*, Vol. 7/8, pp. 296–301, 2006.
- [94] J. Zevenbergen, *et al.*, "'Geoportal Network': More process catalyst than project," UDMS, Ljubljana, Slovenia, 2009.
- [95] M. Masuch, "Vicious circles in organizations," *Administrative Science Quarterly*, Vol. 30, No. 1, pp. 14–33, 1985.
- [96] P. Edwards, *et al.*, "Understanding infrastructure: Dynamics, tensions, and design," Report of a Workshop on "History & theory of infrastructure: Lessons for new scientific cyberinfrastructures," 2007.
- [97] O. Hanseth, E. Monteiro, and M. Hatling, "Developing information infrastructure: The tension between standardization and flexibility," *Science, Technology & Human Values*, Vol. 21, No. 4, pp. 407–426, 1996.
- [98] A. van Marrewijk, *et al.*, "Managing public-private megaprojects: Paradoxes, complexity, and project design," *International Journal of Project Management*, Vol. 26, pp. 591–600, 2008.

The Transmission of Pricing Information of Dually-Listed between Hong Kong and New York Stock Exchange

Shuangfei LI, Shou CHEN

School of Business Administration, Hunan University, Changsha, China
Email: lsfnhu@163.com

Received July 31, 2009; revised September 5, 2009; accepted October 29, 2009.

ABSTRACT

The study investigates the transmission of pricing information between Hong Kong Stock Exchange and New York Stock Exchange. Using the opening and closing stock prices of these two markets from Jan. 2003 to Apr. 2007 with the method of Seemingly Unrelated Regression, we draw the conclusions that: 1) intraday returns of Chinese dually-listed stocks is influenced more obviously by Hang Seng Index than Dow-Jones Average; 2) transmission of pricing information is only from New York to Hong Kong; 3) intraday returns of stocks from New York Stock Exchange has a remarkable influence on that of the next day in Hongkong market, but the stocks price of Hong Kong Stock Exchange has no relation with which of New York Stock Exchange.

Keywords: Information Transmission, Dually-Listed, Stock Price

1. Introduction

In recent years, there have been quite a number of Chinese companies listed in foreign capital markets. According to the China Securities Regulatory's Statistics, up to early 2006, there are a total of 120 Chinese companies listed in foreign capital markets. The market value of China's A-Shares reaches 9 trillion RMB, while the market value of overseas listed enterprises reaches 11 trillion RMB. At the mean time, Along with China's financial industry is opened to external and to deepen finance reformation, more and more people begin to pay attention to the relationships between Chinese market and overseas securities market. The interdependence of international equity markets has been examined extensively.

The interdependence of international equity markets has been examined extensively. Using stock market indices, Koch and Koch examined relationships between daily closing values of eight national stock indices for the years 1972, 1980 and 1987. Their result suggested that international stock market has grown an more interdependent over time, and increased equity market interdependence has been concentrated primarily between markets in neighboring countries and between markets whose trading hours overlap [1].

Using multivariate GARCH models, Zhao L. and Wang Y. point out there is an asymmetry in predictability of the volatility of A share versus B share. Before the openness to domestic investors of B share in Feb. 2001, the volatility of B share and A share are relatively independent. After that, there is a prominent volatility spill-over effect from A share to B share [2].

However, using the dually-listed stock prices will describe transmission of pricing information better than using the indices [3-4]. Using the daily opening and closing stock prices of seven Japanese corporations that are dually listed on the New York Stock Exchange (NYSE) and the Tokyo Stock Exchange (TSE), Lau and Diltz conclude that market imperfections that may inhibit information transfer between TSE and NYSE stock returns are not readily apparent and that international listings do not give rise to arbitrage opportunities [5].

Bae investigates the transfer of pricing information between the Hong Kong Stock Exchange (HKSE) and the London Stock Exchange (LSE). They find that HKSE overnight returns respond significantly to changes in LSE intraday returns, but the transmission process is not completed at the opening of the SEHK; LSE overnight returns respond significantly to changes in HKSE intraday returns, but the transmission process is not completed at the opening of the LSE, either; the impact is stronger

moving from the LSE to the SEHK [6].

Kalok *et al.* using high frequency data to study how overnight price movements in local markets affect the trading activity of foreign stocks on the NYSE. They found that local price movements affect not only the opening returns of foreign stocks, but also their returns in the first 30-min interval. The magnitude of local price movements is positively related to price volatility of foreign stocks, and this relation is stronger at the NYSE open but weaker afterward. However, local price movements cannot account for intraday variations in trading volume. They interpret the result as evidence that the trading activity of foreign stocks on the NYSE is related more to liquidity trading of US investors and less to local market information [7].

Chinese enterprises have a shorter history in listing in foreign capital markets, Research on the relationships between Chinese Capital Market are in progress. This study is important due to the growing number of firms that choose for their common stock to be traded on foreign stock exchanges. By examining the transmission of pricing information, optimal investment decision for investors from home and abroad can be made.

The remainder of this study is organized as follows. Section 2 describes the data and method used in the study, while Section 3 summarizes the empirical results. Finally, Section 4 presents the conclusions.

2. Data and Method

2.1 Data

Using opening and closing stock prices of Chinese firms which are dually-listed on the Hong Kong and New York stock exchanges over the period of 2003 to April 2007. The firms' New York listings are in the form of Ameri-

can Depository Receipts (ADRs). We eliminate two firms due to unavailability of prices for the entire period. Therefore, the final sample consists of seven firms, the length of time series is 1043. Hong Kong Stock Exchange and New York Stock Exchange opening and closing prices are obtained from www.yahoo.com.cn. (Table 1)

2.2 Method

The specification and methodology used in this study are a refined version of those employed by Lau, S T (1994) and Becker K H. (1990). Intraday and overnight returns are calculated in the following manner. $open_t$ and $close_t$ are the daily opening and closing stock prices, oc_t and co_t are the intraday return and overnight returns:

$$oc_t = 100 * \log(close_t / open_t);$$

$$co_t = 100 * \log(open_t / close_{t-1});$$

All return series are adjusted for dividend payment and stock split. More specifically, we define the overnight return and the intraday return for the HKSE and NYSE as follows:

HK_t^{oc} : The SEHK intraday return on day t ,

HK_t^{co} : The SEHK return for the overnight period ending on day t ,

NY_t^{oc} : The NYSE intraday return on day t ,

NY_t^{co} : The NYSE return for the overnight period ending on day t .

The Hong Kong Stock Exchange opens at 10:00 hours and closes at 16:00 hours, and New York Exchange precedes Hong Kong by half an hour. Thus, on a given trading day, the trading times are overlap, and the

Table 1. The description of sample firms

Firms Name	Exchange Code	Transform Proportion	Industry	Listing Date
SINOPEC SHANGHAI ADS	SHI (0338)	1:100	Energy	07/26/1993
CHINA TELECOM CP LTD	CHA (0728)	1:100	Communication	07/22/2002
BRILLIANCE CHI	CBA (1114)	1:100	Autocar manufacture	10/22/1999
GUANGSHEN RAIL CO LT	GSH (0525)	1:50	Transportation	05/14/1996
ALUMINUM CP CHIN ADS	ACH (2600)	1:100	Metallurgy	12/12/2001
HUANENG POWER INTL	HNP (0902)	1:40	Electric power	01/22/1998
CHINA EASTERN AIRLNE	CEA (0670)	1:100	Aviation	02/05/1997

Note: The ADR in Exchange code is the code on NYSE and corresponding code on HKSE each. Transform proportion is the number which one ADR respond to H share.

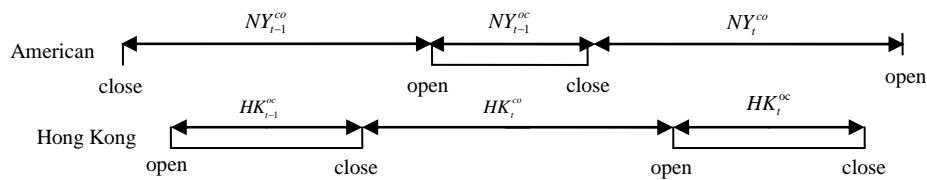


Figure 1. The chronological sequence of events in the New York Stock Exchange (NYSE) and the Hong Kong Stock Exchange (HKSE)

chronological sequence is as shown in Figure 1.

First of all, we give our attention to the transmission of pricing information between NYSE and HKSE. We specify four equations that relate to the transmission of pricing information between market participants in HKSE and NYSE. The first two equations model the transmission of pricing information from the NYSE to the HKSE.

Specifically, we first examine the effect of changes in the intraday NYSE returns, NY_{t-1}^{oc} , on the following day's HKSE overnight return, HK_t^{co} by estimating the following regression specification:

1) Regression Equation (1) illustrates the effect of changes in the intraday NYSE return on day t-1, on the following day's HKSE overnight return,

$$HK_t^{co} = \alpha_1 + \beta_1 NY_{t-1}^{oc} + \varepsilon_{1t} \quad (1)$$

2) Equation (2) illustrates the effect of changes in the intraday NYSE return on day t-1, on the following day's HKSE intraday return,

$$HK_t^{oc} = \alpha_2 + \beta_2 NY_{t-1}^{oc} + \varepsilon_{2t} \quad (2)$$

3) Equation (3) illustrates the effect of changes in the intraday HKSE return on day t-1, on the same day's NYSE overnight return,

$$NY_{t-1}^{co} = \alpha_3 + \beta_3 HK_{t-1}^{oc} + \varepsilon_{3t} \quad (3)$$

4) Equation (4) illustrates the effect of changes in the intraday HKSE return on day t-1, on the same day's NYSE intraday return,

$$NY_{t-1}^{oc} = \alpha_4 + \beta_4 HK_{t-1}^{co} + \varepsilon_{4t} \quad (4)$$

where ε_{it} ($i=1,2,3,4$) is a stochastic disturbance term.

If the one stock exchange responds to changes in returns on the other, then estimation of equation should produce a significant positive slope coefficient.

The empirical specifications (1) and (3) each contain variables that, while not strictly simultaneously determined, nevertheless share a considerable amount of common trading time. Upon consideration of our empirical specification and the types of hypotheses we are testing, we employ a four-step instrumental variables procedure that differs slightly from that of Koch [1] and Lau S.T. [5].

First, we estimate market model regressions for NY_t^{oc} and HK_t^{oc} using ordinary least squares:

$$NY_t^{oc} = \alpha_5 + \beta_5 DJA_t^{oc} + \varepsilon_{5t} \quad (5)$$

$$HK_t^{oc} = \alpha_6 + \beta_6 HSI_t^{oc} + \varepsilon_{6t} \quad (6)$$

Where DJA_t^{oc} is intraday return on the Dow Jones Industrials Index, and HSI_t^{oc} is intraday return on the Hong Kong Hang Seng Index.

Secondly, the resulting parameter estimates are used to create instruments (i.e. predictions), \hat{NY}_t^{oc} and \hat{HK}_t^{oc} for variables NY_t^{oc} and HK_t^{oc} Equations (1) and (3), respectively.

$$\hat{NY}_t^{oc} = \hat{\alpha}_5 + \hat{\beta}_5 DJA_t^{oc} \quad (5')$$

$$\hat{HK}_t^{oc} = \hat{\alpha}_6 + \hat{\beta}_6 HSI_t^{oc} \quad (6')$$

Third, the instrumental variables are substituted into Equations (1) and (3), yielding:

$$HK_t^{co} = \alpha_7 + \beta_7 \hat{NY}_{t-1}^{oc} + \varepsilon_{7t} \quad (1')$$

$$NY_{t-1}^{oc} = \alpha_8 + \beta_8 \hat{HK}_{t-1}^{oc} + \varepsilon_{8t} \quad (3')$$

Finally, the parameters of the system of Equations (1'), (2), (3') and (4) are estimated jointly using the Seemingly Unrelated Regression (SUR) method. By using the estimates of the covariance of residuals across equations, the SUR method improves the efficiency of the parameter estimates.

3. Empirical Results

We believe that the instrumental variables produced in the manner described above are appropriate instruments, as they satisfy the two criteria for valid instruments. Table 2 reports the parameter estimates for the market models specified in Equations (5) and (6), all regression slopes are statistically significant at the one per cent level, indicating that all sample firms from NYSE and HKSE return respond to changes of Share Market Indexes.

For each firm, the results give us an attention that β_6 (0.56~0.87) is more than β_5 (0.97~1.52), which indicates the fluctuation range of Chinese stocks returns of NYSE are narrower than Dow Jones Industrials Index, and which from HKSE are wider than Hong Kong Hang

Table 2. Results of the market model for the NYSE and the HKSE

New York Stock Market				Hong Kong Stock Market			
Stock	α_5	β_5	R^2	Stock	α_6	β_6	R^2
SHI	0.026 (0.73)	0.618*** (13.76)	0.154	0338.HK	0.004 (0.06)	1.475*** (15.72)	0.192
CHA	-0.003 (-0.10)	0.781*** (20.83)	0.294	0728.HK	0.069 (1.41)	1.094*** (16.41)	0.206
CBA	-0.090* (-1.78)	0.759*** (11.99)	0.121	1114.HK	-0.084 (-1.36)	0.998*** (11.88)	0.119
GSH	0.017 (0.40)	0.603*** (11.38)	0.111	0425.HK	-0.137** (-1.98)	1.222*** (13.00)	0.140
ACH	0.011 (0.25)	0.870*** (15.51)	0.188	2600.HK	-0.080 (-1.58)	0.968*** (14.08)	0.160
HNP	0.040 (1.20)	0.590*** (14.18)	0.162	0902.HK	-0.139* (-1.76)	1.214*** (11.26)	0.109
CEA	-0.055 (-1.24)	0.560*** (9.97)	0.087	0670.HK	-0.024 (-0.34)	1.519*** (16.13)	0.200

Note: t-values appear in parentheses.

***Significant at the 1 per cent level. **Significant at the 5 per cent level. *Significant at the 10 per cent level.

Table 3. SUR estimates of Equations (1'), (2), (3') and (4)

Transfer direction	From New York to Hong Kong				From Hong Kong to New York			
Stock	α_7	β_7	α_2	β_2	α_8	β_8	α_4	β_4
SHI	0.039	0.972***	0.040	-0.106*	0.066	0.000	0.066	0.002
(0338.HK)	(0.85)	(10.77)	(0.52)	(-1.74)	(0.95)	(0.05)	(0.96)	(0.55)
CHA	-0.029	0.928***	0.090*	-0.015	0.055	0.002	0.055	0.002
(0728.HK)	(-0.86)	(17.99)	(1.65)	(-0.30)	(1.08)	(0.35)	(1.08)	(0.82)
CBA	0.240***	0.671***	-0.072	-0.098***	0.063	0.002	0.063	0.001
(1114.HK)	(5.83)	(9.95)	(-1.10)	(-2.61)	(0.85)	(0.51)	(0.85)	(0.51)
GSH	0.151***	1.002***	-0.114	0.016	0.088	0.000	0.088	0.000
(0425.HK)	(3.33)	(10.69)	(-1.53)	(0.30)	(1.59)	(-0.06)	(1.60)	(0.07)
ACH	0.049	0.617***	-0.063	-0.015	0.006	0.000	0.006	0.000
(2600.HK)	(0.63)	(5.45)	(-1.14)	(-0.43)	(0.04)	(-0.01)	(0.04)	(0.02)
HNP	0.051	1.138***	-0.104	-0.146**	-0.050	0.000	-0.050	0.000
(0902.HK)	(1.09)	(11.51)	(-1.24)	(-2.05)	(-0.62)	(-0.11)	(-0.61)	(0.15)
CEA	0.241***	1.667***	0.006	0.090*	0.106	0.000	0.106	0.000
(0670.HK)	(4.77)	(14.70)	(0.08)	(1.76)	(1.92)	(0.03)	(1.92)	(0.11)

Note: t-values appear in parentheses.

***Significant at the 1 per cent level. **Significant at the 5 per cent level. *Significant at the 10 per cent level.

Seng Index. Therefore, contrast to Dow Jones Industrials Index, the stocks of mainland firms listed in Hong Kong are sensitivity to Hang Seng Index, which would be an overreact.

The first two columns in Table 3 reports the results of SUR estimates of the effect of changes in the sample firm NYSE intraday return on the following day's sample firm HKSE overnight return. All seven regression slopes are statistically positive significant at the one percent level, indicating that sample firm HKSE returns respond to changes in corresponding firms returns on the NYSE. If the regression slopes of firms, β_7 , higher than one, indicating that stock returns in Hong Kong effected by New York higher than that in New York and vice versa.

The following two columns in Table 3 reports the results of SUR estimates of the effect of changes in the sample firm NYSE intraday return on the following day's sample firm HKSE intraday return. This regression allows us to determine whether the transmission of pricing information has been completed at the opening trade. The insignificance of three of the seven regression slopes suggests that most of the transmission of pricing information is completed at the first trade of the day. The negative significance of three of that suggests that based on the opening price closing price have a reversion. Only one firm have a negative significance slopes at the ten per cent level indicating that pricing information transmission is continue at the all trade of the day.

Associate Equation (1') with Equation (2), as shown in

Table 3 is the transmission of pricing information from New York to Hong Kong. β_7 are higher than β_2 and so are the corresponding t-statistics, which suggests that the transmission from New York to Hong Kong is completed at the first trade of the day.

The regression slopes in Equations (3') and (4) are insignificant, which suggests that the transmission of pricing information from Hong Kong to New York is un conspicuous. This result disagrees with Lau [5] and Bae [6].

4. Conclusions

This study examines the transmission of pricing information between market participants in the HKSE and the NYSE. Using the opening and closing stock prices of seven Chinese firms that are inter-listed on the Hong Kong and the New York stock exchanges over the period of 2003 to April 2007, four regression specifications relating to the transmission of pricing information are examined. The findings of this study suggest that: 1) Contrast to Dow Jones Industrials Index, the stocks of mainland firms listed in Hong Kong are sensitivity to Hang Seng Index. 2) Sample firm HKSE opening returns respond to changes in sample firm NYSE intraday stock performance; the transmission process is apparently completed at the opening of the HKSE; 3) The transmission of pricing information from Hong Kong to New York are un conspicuous.

REFERENCES

- [1] P. D. Koch and W. Koch, "Evolution in dynamic linkages across Daily National Stock Indexes," *Journal of International Money and Finance*, Vol. 10, No. 2, pp. 231–251, 1991.
- [2] L. Y. Zhao and Y. M. Wang, "Information flows and volatility spillover across A share and B share market in China," *Journal of Finance*, Vol. 10, pp. 37–52 (in Chinese), 2003.
- [3] D. Long, J. Payne, and C. Feng, "Information transmission in the Shanghai equity market," *Journal of Financial Research*, Vol. 22, pp. 29–45, 1999.
- [4] C. S. Eun and S. Sabherwal, "Cross-border listings and price discovery: Evidence from US-listed Canadian stocks," *Journal of Finance*, Vol. 58, No. 2, pp. 549–575, 2003.
- [5] S. T. Lau and J. D. Diltz, "Stock returns and the transfer of information between the New York and Tokyo Stock Exchanges," *Journal of International Money and Finance*, Vol. 13, No. 2, pp. 211–222, 1994.
- [6] K. H. Bae, B. Cha, and Y. L. Cheung, "The transmission of pricing information of dually-listed stocks," *Journal of Business Finance and Accounting*, Vol. 26, pp. 709–723, 1999.
- [7] K. Chan, M. Chockalingam, and W. L. Kent, "Overnight information and intraday trading behavior: evidence from NYSE cross-listed stocks and their local market information," *Journal of Multinational Financial Management*, Vol. 10, pp. 495–509, 2000.

Problems Facing Small and Medium Size Contractors in Swaziland

Wellington Didibhuku Thwala¹, Mpendulo Mvubu²

¹Department of Quantity Surveying and Construction Management, University of Johannesburg, Johannesburg, South Africa;

²Department of Construction Management and Quantity Surveying, University of Johannesburg, Johannesburg, South Africa.
Email: didibhukut@uj.ac.za

Received June 20, 2009; revised August 1, 2009; accepted September 19, 2009.

ABSTRACT

The paper explores the problems facing Small and Medium Size Contractors in Swaziland. The construction sector in Swaziland is not only a significant source of direct employment but also a sector which contributes, through its wide range of projects and operations. The paper will also look at the current government initiatives that had been put in place to address the challenges and problems in order to ensure that contractors are successful. There is a high failure rate among small and medium size contractors in Swaziland. These contractors fail for a variety of reasons ranging from lack of adequate capacity to handle the uniqueness, complexity and risks in contracting, lack of effective management, lack of business management, poor record keeping and inadequate technical, financial and contract managerial skills. Drawing on research on small and medium size contractors, the paper used both secondary and primary literature. 100 questionnaires were distributed to different role players in the construction sector in Swaziland. The response rate was eighty seven (87) percent. The paper reveals that the most problems facing small and medium size contractors in Swaziland is lack of access to finance and late payment by government. The paper closes with recommendations and key lessons for the future.

Keywords: Contractor, Construction, Employment, Small and Medium Size, Public Works

1. Introduction

In Swaziland and other countries there seem a general consensus that small enterprises are the mainstay of economic growth and prosperity. Small contractors can be powerful instruments of generating job opportunities; small contractors can perform small projects at different and remote geographical locations that might be unattractive to big firms or too costly using the big firms; low overheads enable small contractors to work at more competitive prices; large number of functional small and medium scale black contractors can help to decentralise the construction industry dominated by established large contractors; the relatively low skills and resources required at this scale can easily lower the entry point for the small and medium size owners to begin to participate in the industry; and a large number of functional Swazi owned contractors can develop a platform for growth and redistribution of wealth in Swaziland. Small and Medium Size contractors in Swaziland are defined as those contractors who do work up to E20 million (Exchange Rate: \$1 = E8). At a time when the public sector and big business are shedding jobs, small businesses are maintaining

real employment growth. The small contractor development programme falls under the Ministry of Public Works and Transport. The main mandate of the ministry is to ensure the provision and maintenance of a sustainable public infrastructure, an efficient and effective seamless transport system and network, regulation for a vibrant construction and transport industry, management of public service accommodation and the provision of meteorological services [1].

2. Research Objectives

The main objective of this paper is to outline the problems faced by small and medium size contractors in the Kingdom of Swaziland.

2.1 Review of South African Literature

2.1.1 Skill Development in the Construction Industry

Historically, the construction industry has largely relied on a core of highly skilled staff (generally white and often expatriate) to supervise a largely semi-skilled and unskilled workforce (generally black). The decline in demand for construction products over the past decades, and associated uncertainty, has seen a reduction in skills

training since the 1980s, and the closing down of industry training institutions in the 1990s. It has been reported that only about 70 percent of the available training capacity is currently being utilized [2]. Skills enhancement in the construction industry faces a very particular challenge since the construction sector employs the fourth highest number of persons having no formal education – after agriculture, households and mining. Industry has expressed a view that skills supplied to the market through the Further Education and Training (FET) system were in many cases not appropriate to their needs, resulting in a skills gap. While industry-based training is seen as better aligned with company-specific requirements, South African Qualification Authority (SAQA) does not accredit some trainers, and some do not issue certificates for training and employment. This tendency limits mobility and career path prospects [2].

The important note was to recognize the improvement of the work skills of all South Africans, which is critical to grow the national economy. The Skills Development Act (SDA) was promulgated to create the structures and framework for the national development strategy. In terms of the SDA employers are obliged to provide formal structured education to their workers, hence in the problem statement one looks at the possibilities of the employed trained people by Group Five being given the opportunity to further their training.

Furthermore, the act encourages partnership in this effect between government, employers, workers, education and training providers, and beneficiary communities. The trained people are the beneficiaries from the community. According to SDA, the needs of employers, the economy and the communities must dictate which skills development should be developed. The part of which skills should be developed lead one to the last objectives in the problem statement to try to identify which skills are often needed in different trades and provinces.

The SDA covers structured, targeted and generic training implying that all training interventions should be planned and managed as projects that is the reason why Group Five has “people at the gate” which is Corporate Social Investment Project. In SDA, employers together with their workers formulate workplace skills plan (WSP) to enable them to realize their employment training targets. All designated employers pay a monthly skills development levy of 1% of their budgeted payrolls to the National Skills Fund (NSF), via South African Revenue services (SARS). Of this amount, the employer can claim back 70% in the form of discretionary grant, provided that they submit WSP and Implementation Report (IR) annually and conduct special training projects.

These levies finance the implementation of the National Skills Development Strategy (NSDS). Construction Education and training Authorities (CETA) receives 10% of the skills levies paid by construction employers

for administration costs, NSF receives 20%, and 70% is available to be claimed back by these contributing employers. However, international trends shows that companies need to spend between 4% and 7% in order to be successful in addressing the current shortages and gaps [3]. Furthermore, there appears to be over-reliance of a number of levels in the micro and provincial economy on the SETA's as being responsible bodies for coordinating the identification of scarce skills in South Africa.

2.2 Small Contractors in South Africa

In South Africa, the contractors enter the market at the lower end and in the general building contracting category, making the sector extremely competitive and unsustainable [4] and the emerging contractor policies intended for black economic empowerment (BEE) are being used as job creation opportunities, which contributes to the over crowding of the emerging market. It is common for black businesses to be based on technical skills which are based on technical skills, which are used to satisfy needs of the community. However, technical competence is no guarantee of business success. Operational (e.g. scheduling and ordering) and business (e.g. planning, financial control and budgeting) skills are vital to the success of any enterprise. Small enterprises contribute positively to the economics of the country and to the survival of large numbers of people.

2.2.1 Skills Shortage in Small Contractors

South Africa is characterized by a systematic under-investment in human capital. This has resulted in a labour force with a skewed distribution of craft skills, career opportunities and work-place experience. While the promulgation of the Skills Development Act of 1997 is commendable, micro enterprises already express concern about the administration costs of recovering levies in the form of grants for training. Furthermore, there are the costs of designing a workplace training programme as an alternative to using external training institutions and the relatively high charges by private training institutions after the closure of the former industrial training boards which had been subsidized through levies from industry [5]. The imbalances of the past with regard to the school curriculum known as “Bantu education” which did not offer much mathematic and science as part of the curriculum hinder the emerging market as these subjects are essential for entry into the engineering and built environment industry. This Bantu system secured the exclusion of black people from participating in the construction industry as they did not have the necessary skills required. According to Matas [6] the Bantu Education Act, Act No. 47 of 1953 established a Black Education Department in the Department of Native Affairs which would compile a curriculum that suited the nature and requirements of black people. The aim was to prevent

Africans receiving an education that would lead them to aspire to positions they wouldn't be allowed to hold in society. Africans were to receive an education designed to provide them with skills to serve their own people in the homelands or work in labouring jobs under whites. On the other hand the job reservation of 1951 applied to Blacks, Coloureds and Indians. The notion behind job reservation was the best, the most highly skilled jobs, should be reserved for whites. The 1951 Native Building Workers Act provided that no Black might be employed as skilled building workers outside of a Bantu area [6]. The above discussed Acts created a skills shortage among the Africans to compete in the labour market.

2.3 Financial Constraints

The high competition among emerging contractors has contributed to increase financial failures of the emerging market, making the market unsustainable. The Construction Industry Development Board [4] states that the large numbers of emerging contractors have moved into higher value public tendering in the 0.5m to 2m market, which is also becoming overly competitive. Statistics South Africa [7] reports that, from 1995 to 2005, about 5907 construction companies were formally liquidated. According to the Construction Industry Development Board [2] states that much more than 90% of the emerging black contractors survive the first five years. According to the SA Construction Industry Status Report [2], 1,400 construction companies were liquidated over the past three years. Emerging contractors feel that the banks are reluctant to deal with them unless exorbitant interest rates and through compulsory business management services. Lack of access to finance both during preconstruction which disqualifies emerging contractors from meeting guarantee and performance bond requirements and during construction which leads to cash-flow problems, incomplete work and even liquidation are financial constraints facing emerging contractors. The inadequacy of external finance at the critical growth/transformation stages of micro enterprises deters the enterprises with growth potential from expanding [8].

2.4 Late Payment by Clients

Small contractors run into problems due to late payments by the clients. Delays with interim and final payments, as well as onerous contract conditions faced by construction firms, can also impose huge constraints on the industry. Many construction firms have suffered financial ruin and bankruptcy because of delays in payment, which are common with government contracts. The unlucky contractor, failing to repay loans in a timely fashion had his business put into liquidation.

2.4.1 Difficulties When Running the Businesses

The following are the lessons that had been learnt in South Africa with regard to the problems facing small contractors in the contractor development programme (Construction Industry Development Board, Department of Public Works, Construction Education and Training Authority [4] are as follows:

- Usually open adverts are placed in the media calling on people to come out and participate; it is very difficult for a selection process to capture those with the proper drive, passion and ability to work as contractors; this brings wrong people in the programs and drives them easily on the way.
- The required academic qualification is usually matric (Grade 12) or less; no prior technical and managerial skills/experience in construction related fields as prerequisites.
- Few matric (Grade 12) holders make rare success; most successful contractors have degree or diploma in construction related field, with 5-10 years technical and managerial work experience.
- Inadequate training done at short period's in-between projects; unsuitable for the contractor's time and project need; inappropriate trainers.
- Clear-cut grading criteria had been elusive; recently CIDB graded/categorized the contractors using some contested criteria; core tech and management staff not stated, this may still lead to contractors getting projects they do not have capacity for.
- The contractors do not seem to understand the nature of complexity and risk in contracting; do not seem to be adequately informed of how to deal with them properly.
- The contractors lack skill, experience and tools to win profitable contract; they either win a grossly under-priced bid, or lose a grossly over-priced one. Cost, price, control program not provided
- They lack own ready finance and access to affordable loan. Due to lack of collateral, any one that gets credit from banks is subjected to high interest and financial risk management charges that make contracts unprofitable.
- In the ambition to grow big and make big profit, most of them take projects they do not have the necessary skills and financial resources to execute.
- The contractors tend not to employ qualified worker; they consider them expensive, but they fail while doing things all by themselves or with cheap, incompetent workers.
- They lack skills to properly program projects resources in monthly segments for healthy cash flow; they are not allowed front load due to lack of trust; they do not know how to prepare documents for timely payment; delayed payment.

- They do not seem to understand terms of contract conditions; do not know how to use applicable contract performance procedure to deal with client; they do not get properly trained in this.

- They are usually considered incapable of doing competent work, which imperils their relationship with the client's agent; they do not seem to know how to use applicable contractual instruments regarding instruction, demand for specific performance, and payment; they are not properly taught; where they know these rules they fail to use them due to fear of being 'red listed'.

- In attempt to make huge profit they cut specified quality, do bad work that falls short of the design standards/specification. Rejection of such works usually leads to none payment, conflict and most times failure of the contractors.

- Those that manage to win profitable contracts get only 2% profit if they are able to successfully complete the project; it seems discouraging.

The studies show that enterprise success is greatly increased by having relatively stable access to markets and access to capital from external sources, and that successful enterprises are characterized by entrepreneurs with a basic level of education, essential technical knowledge and previous industry experience with larger enterprise, and ability to learn new skills, innovate and take risks [9].

2.4.2 Review of Literature – Swaziland

The Ministry of Public Works and Transport [1] is a major contributor in the Swaziland construction industry. The Ministry is Government's implementing agency on behalf of all ministries with regard to all government construction capital projects. The role of the Swaziland Government through the Ministry of Public Works and Transport is to educate the small and medium contractors about government's expectations to the tendering process and information required for qualification. Another important aspect of the dissemination of knowledge to contractors is that of information on the standard of workmanship required in government projects. The Ministry of Public Works and Transport is responsible for initiating; the payment process to contractors by preparing the monthly interim certificates. This is very important to contractors in that the speedy execution of payments will ensure that contractors are paid in time and their cash flow is not stretched to the limit.

2.4.3 The Construction Industry in Swaziland

The Government of the Kingdom of Swaziland, through its 25-year National Development Strategy (NDS) has identified the construction sector as a priority area to impact on improving the social and economic development of the country. However, to maximize the impact of the sector as part of the NDS, it will be necessary to de-

velop a sound national policy framework for the industry to improve its overall effectiveness and efficiency. Fundamental to the policy and in line with the NDS, will be the empowerment of local Swazis within the industry to maximize their participation and consequent impact on the local economy. As far back as 1993, the Government, through the then Ministry of Public Works and Roads took the initiative to organize the Southern Africa Construction Industry Initiative (SACII) on behalf of ten countries of the Southern Africa Development Community (SADC) region. At that time, the overall objectives of the Initiative were to:

- Identify constraints to the development of local construction industries in each participating country within the region;

- Identify specific policy reform to improve the enabling environment for local construction industry growth and development;

- Implement reforms in member countries with Government and donor commitment to local construction industry development.

In general, the main intention of the Initiative was to create jobs, develop local capability and empower local/indigenous companies in the construction industry and is directly in line with the policy framework required by stakeholders in Swaziland. Hence, some of the recommendations and policy options developed as part of the 1993 initiative at national and regional level have been used as a foundation for the framework presented here.

2.4.4 The Swaziland Construction Industry Policy

The vision of the government of the Kingdom of Swaziland, as stated in the Swaziland National Construction Industry Policy 2002 [1] advocates for "a construction industry that maximizes "local participation." The construction industry is not only a considerable source of direct employment, but also an industry which contributes to the overall economy of Swaziland, through its extensive assortment of projects and operations. Since the Swazi economy is unable to provide employment for the ever growing population in Swaziland, focus may be turned to the local construction industry. The contribution of the construction industry to the Swazi economy comes in the form of job creation, development of local capability and the empowerment of local indigenous companies in the construction industry.

The Swaziland Construction Industry Policy intended to create an enabling environment to best meet the needs of the stakeholder in the construction industry in Swaziland and commits Government to various roles, functions and activities. In the past, Government's role has been as a regulator and, in some cases, provider of physical infrastructure with no holistic policy framework and strategic plan to drive the industry. Through the policies de-

veloped, Government intends to refocus its primary role to that of policy and strategy formulation, and regulation of the industry with a reduced direct involvement in the provision of infrastructure and services. The implications are that the capacity and capabilities of the local private sector need to be developed to undertake some of the services previously carried out by Government.

It is also the intention of Government to include other key role players in the co-ordination of the industry and the development of broader national strategies that may otherwise not be achieved successfully. A regulatory role will be retained by Government to ensure safety and quality of service throughout the industry and to monitor progress in achieving the vision and mission of the industry presented here. This will mean, a more focused and skilled national Government that can control, monitor and regulate the relationships between service providers and implementing agencies.

Government also recognizes that the local human resource pool of the construction industry needs to be strengthened to achieve the vision and mission. It is also of concern that there is a significant loss of trained personnel to improved opportunities in other countries. Hence, Government will assume a responsibility for empowerment and capacity enhancement of the local industry through the retention of trained personnel and a general improvement of the resource pool in the industry. However, Government will also seek to involve the private sector in meeting the challenge of growing and retaining the country's human resource base.

2.4.5 Contractor Accreditation Process in Swaziland

A registration of accredited construction enterprises in Swaziland constitutes an essential tool for the industry transformation, for monitoring the performance of enabling environment programmes, and for ensuring compliance with the performance of public-sector projects.

All construction enterprises engaged in public sector work, or in receipt of State funding training or support functions, will be required to be registered in a manner that will reflect their capacity and performance. The registration process must address the following: the operation of a preference scheme, or approved public tender list, which would reduce industry and public sector cost associated with an all out open tender process at the same time supporting risk management; performance monitoring to enable the promotion of improved contractors and to ensure compliance where standards are violated; and; the targeting of resources to the emerging contractors which are demonstrating progress and the withdrawal of support from those which have graduated or have failed to progress [1].

Contractor grading in Swaziland is one of the tools that is used to regulate the construction sector. Unlike the Republic of South Africa where the contractor accreditation is done by an independent body, the Construction Industry Development Board (CIDB), each department in the Ministry of Public Works and Transport done its own accreditation. The categories start from category 1 to 6 as illustrated on Table 1 for the roads department. Below the table shows the different categories. The small and medium size contractors in Swaziland fall between categories 2 to 6. The tender price category is below R500, 000.00 and less than R20m.

In addition to the contractors summarized in the table above, there are twelve (12) specialist contractors; with four (4) specializing in road marking, six (6) specializing in electrical works (streetlight maintenance) and two (2) specializing in premix road maintenance.

As shown in the Table 2 below there are no foreign contractors under main category M1 and sub-category B, C and D. The table below clearly shows that the Building Contractors are dominated by Swazi people.

Table 1. Civil contractor grading (roads department) in Swaziland

Categories	Local contractors	Foreign contractors	Project Category Eligible to tender for
Category 1	1	2	Locally and internationally funded construction projects above E20 million.
Category 2	1	2	Local and internationally funded construction projects above E10 million but below E20million.
Category 3	4	0	Local and internationally funded construction projects above E5 million but below E10million.
Category 4	6	0	Local and internationally funded construction projects above E1million but below E5million.
Category 5	24	0	Local and internationally funded construction projects above E500, 000.00 but below E3million.
Category 6	158	0	Local and internationally funded construction projects below E500, 000.00.

Source: Ministry of public works and roads- Swaziland: 2007.

Table 2. Contractor grading (buildings department) in Swaziland

Categories	Local contractors	Foreign contractors	Project Category Eligible to tender for
Category M	8	5	Locally and internationally funded construction projects above E20 million.
Category M1	6	0	Local and internationally funded construction projects above E10 million but below E20million.
Category A	11	2	Local and internationally funded construction projects above E5 million but below E10million.
Category B	13	0	Local and internationally funded construction projects above E1million but below E5million.
Category C	20	0	Local and internationally funded construction projects above E500, 000.00 but below E3million.
Category D	91	0	Local and internationally funded construction projects below E500, 000.00.

Source: Ministry of public works and buildings department: 2007.

2.4.6 Problems Facing Small and Medium Size Contractors in Swaziland

The problems facing small contractors are not unique to Swaziland. The vast majority of construction firms are small enterprises that rely on outsourcing personnel as required. This has severely affected skills training and the retention of expertise in the industry as construction workers become highly mobile, walking in and out of the industry, depending on performance in other sectors of the economy. The impact can be seen in the rigid adherence to management techniques and construction practices handed down from colonial times which, as a result of inadequate skills and capacity.

2.5 Research Methodology

The paper used both primary and secondary data. Literature review on small contractors was conducted. A questionnaire was designed with the help of the statistical Consultation Service (STATKON) of the University of Johannesburg. A field survey was carried out on 15 companies; the study target population includes 100 contractors currently registered with the Ministry of Public Works and Transport Roads and Building sections. The Ministry of Public Works and Transport the Ministry of Public Works and Transport serves as a regulatory body for the construction industry in Swaziland. Swaziland is divided into four administrative regions; that are Hhohho, Manzini, Lubombo and Shiselweni. Data was collected between the 1st of June 2007 and 30th July 2007 and 87 respondents were interviewed. A total of one hundred (100) questionnaires were distributed to various contractors registered under categories D, C, B (small) and A and M1 (medium) with the Ministry of Public Works and Transport Buildings Department contractor register and contractors registered under categories 6, 5, 4 (small) and

3 and 2 (medium) with the Ministry of Public Works and Transport Roads Department contractor register. Contractors registered under these categories are regarded as small and medium contractors, respectively. The questionnaire was distributed to senior personnel including directors, project and construction managers. Out of the one hundred (100) questionnaires distributed, eighty-seven (87) were returned. Thirteen (13) questionnaires were not returned by the respondents. (See Figure 1)

The Kingdom of Swaziland is a landlocked country surrounded by the Peoples' Republic of Mozambique in the east and the Republic of South Africa's Kwa Zulu-Natal Province in the south and the Republic of South Africa's Mpumalanga Province in the west and in the north. The Kingdom of Swaziland has a population of around 1,128,814 inhabitants on 17,363 sq km of land.

3. Findings of Research 2009

The problems facing small contractors are not unique to Swaziland. The vast majority of construction firms are small enterprises that rely on outsourcing personnel as required. This has severely affected skills training and the retention of expertise in the industry as construction workers become highly mobile, walking in and out of the industry, depending on performance in other sectors of the economy. The impact can be seen in the rigid adherence to management techniques and construction practices handed down from colonial times which, as a result of inadequate skills and capacity. Delays with interim and final payments, as well as onerous contract conditions faced by construction firms, can also impose huge constraints on the industry. Many construction firms have suffered financial ruin and bankruptcy because of delays in payment, which are common with government contracts. Contemporary research that was conducted in

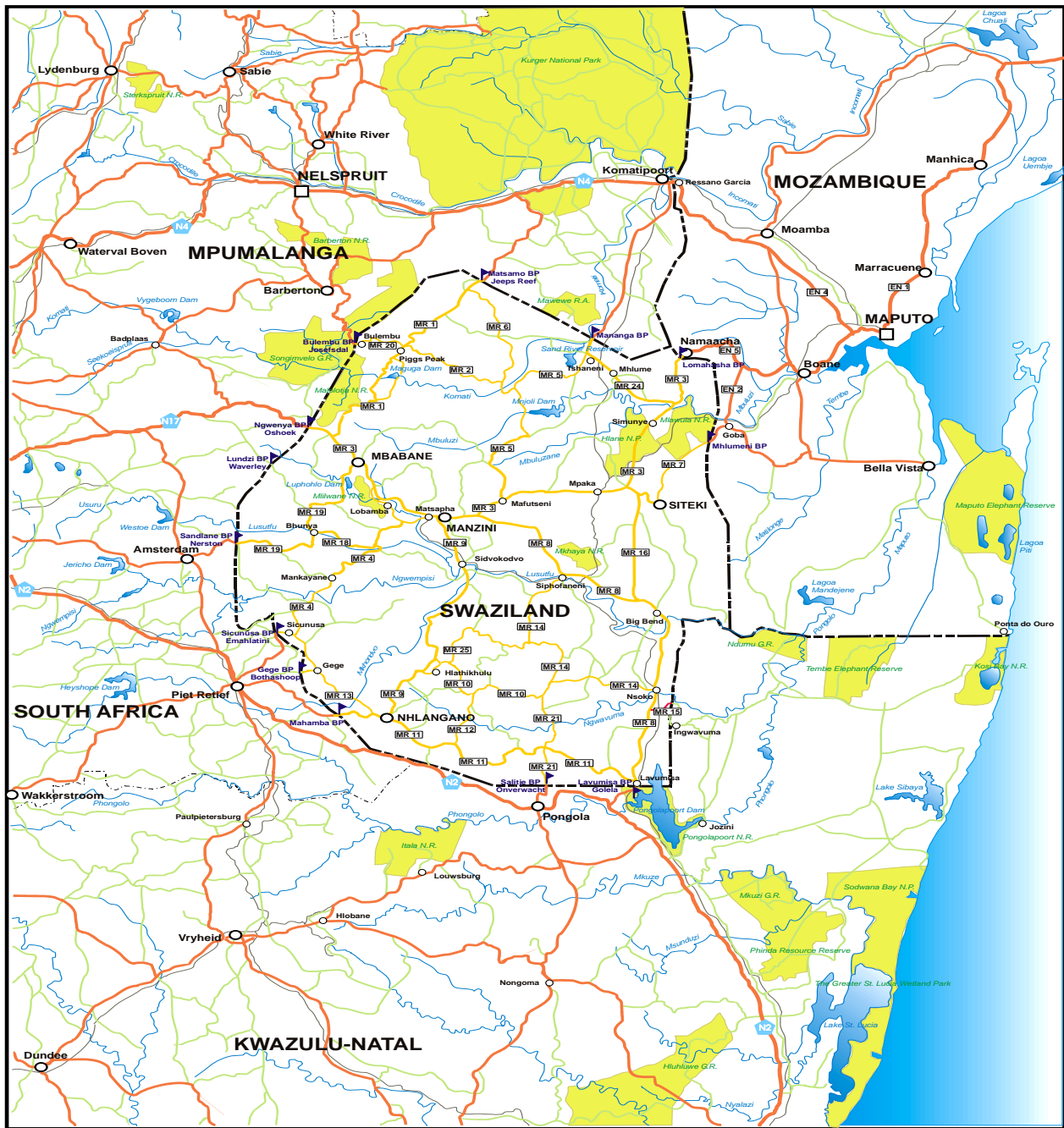


Figure 1. General survey offices, ministry of natural resources, Swaziland, 2009

2007 by the authors revealed the current reasons for the failure of small and medium size contractors in Swaziland. 87 owners of the small and medium size contractors were interviewed. 68% of the contractors are less than four years; 20% are between 5 and 9 years; and 12 % had operated for more than 10 years. There was no contractor that had operated more than 15 years. 63% of the respondents believe that the four major banks in Swaziland have proper systems in place to support small and me-

dium size contractors once they have secured work. On the other hand 37% of the respondents do not believe that the four major banks in Swaziland have proper systems in place to support small and medium size contractors. 33.4% of the respondents think that the current environment within the construction industry in Swaziland is favorable for small and medium size contractors to be successful. On the other hand 66.6% of the respondents believe that the construction industry environment is not

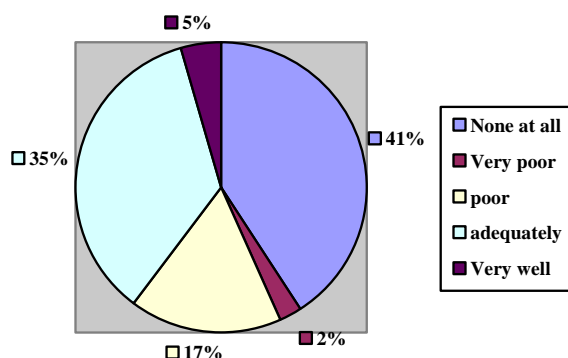


Figure 2. Develop business skills

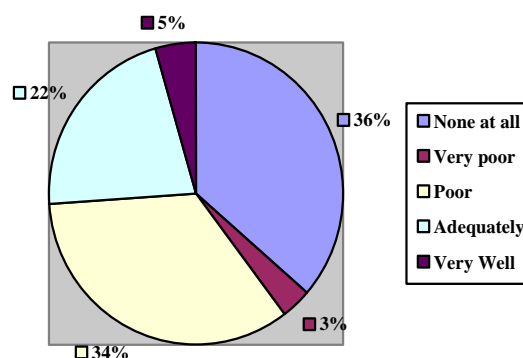


Figure 3. Develop management skills

favorable for the success of the small and medium size contractors.

From the above figure 40% of the respondents are satisfied with business skill development and 60% of the respondents were not satisfied. The figure above shows that the respondents are not satisfied with regard to the development of business skills. (See Figure 2)

From the above figure 27% of the respondents are satisfied with the development of managerial skills. 73% of the respondents are not satisfied with the development of managerial skills. It is clear from the above figure that respondents had not been trained. (See Figure 3)

From the figure below 47% of the respondents are satisfied with the development of technical skills. 53% of the respondents are not satisfied with the development of technical skills. (See Figure 4)

From the research conducted it can be concluded that the relative lack of success among the small and medium size contractors in Swaziland is a results of the following problems which must be addressed in order for the contractors to be successful:

- A lack of resources for either large or complex construction work.
- An inability to provide securities, raise insurance and obtain professional indemnity.

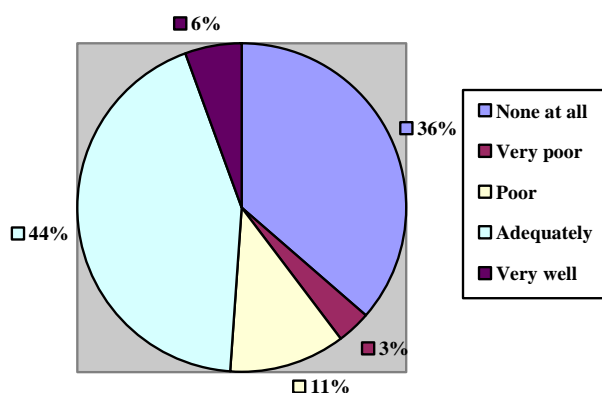


Figure 4. Develop technical skills

- The contracts were inevitably packaged in such a way as to exclude small contractors.
- Inadequacy in technical and managerial skills required in project implementation.
- Lack of continuity in relation to type, scale and location of work
- An inadequate approach and insufficient knowledge, time and experience required for the whole process of finding work, once found, insufficient understanding of the contract documentation and the preparation and submission of tenders.
- Slow and non-payment by government after completing a government project
- Little or inadequate effort has been made to identify the causes of failure among the local contractors in the implementation of construction projects in Swaziland.
- The need for the introduction of a set of rules to govern the construction industry done by an independent body in the Kingdom of Swaziland.

4. Lessons and Recommendations

The literatures that have been reviewed both in Swaziland and internationally demonstrated that the development of Swazi contractors, small and medium sized contractors faced several challenges which must be addressed. In the case of Swaziland the majority of the respondents are dissatisfied with the quality of assistance offered by the government to the small and medium sized contractors in Swaziland. Small enterprises contribute positively to the economics of the country and to the survival of large numbers of people. However, the success of small enterprise is impaired by the common weakness from which many enterprises suffer. A vibrant self-sustaining construction industry is one that is planned and matched with government capital programmes and support schemes aimed at improving the efficiency of the local construction industry. Swaziland is faced with a large challenge of developing infrastructure in the disadvantaged communities, and also upgrading the existing

infrastructure to cope with the high demand. It is recommended that the Swaziland Government must undertake the following measures to address the challenges facing Small and Medium Size contractors:

- Business skills: training in business management skills is key the in managing a construction business effectively and efficiently.
- Management skills: financial management should be emphasized as well as networking with other people with similar businesses
- Good record keeping: financial records should be prioritized and a record of books of accounts on a daily, weekly, monthly and annual basis should be established.
- Well managed cash flow: prepare cash flow forecasts and budgets and prepare a cost-benefit analysis.
- The government should break down big contracts into small contracts so that small contractors with lower grading can qualify to do the job
- The government should assist in training Small and Medium Size contractors on business management
- The government should create a construction advice centre
- The government should establish a Contractor Training Institute to train Small and Medium Size contractors
- Government payment system must be improved
- Government must establish financial institutions that will act as guarantees in order to enable contractors to have access to finance.

5. Conclusions

The literatures that have been reviewed both in Swaziland and internationally demonstrated that the development of Swazi owned contractors, small and medium sized contractors, in particular are face many challenges which need to be addressed in order for the contractors to sustainable in their growth. The majority of the respondents are dissatisfied with the quality of assistance offered by the government to the small and medium sized contractors in Swaziland. The problem lies, however, in the formation of a construction industry in line with the implementation of government programmes that must facilitate the creation of a vibrant self-sustaining construction industry. A vibrant self-sustaining construction industry is one that is planned and matched with government capital programmes and support schemes aimed at improving the efficiency of the local construction industry. The study finds that lack of effective management during their early stages is a major cause of business

failure for small and medium sized contractors. Owners tend to manage their businesses themselves as a measure of reducing operational costs. Poor record keeping is also a cause for start-up business failure. Lack of financial management; lack of entrepreneurial skills; lack of proper training; lack of resources; lack of technical skills, lack of contractual and managerial skills; late payment for work done which are common with government contracts; inability to get credit from suppliers and fronting for established contractors are also contributing factors for the failure of small contractors in Swaziland.

6. Acknowledgements

Research works in this paper are partial supported by the National Research Foundation (NRF) and the University of Johannesburg Research Funding.

REFERENCES

- [1] Ministry of Public Works and Roads, "Swaziland. Swaziland today," 2007, <http://www.gov.sz>.
- [2] Construction Industry Development Board, "South Africa construction industry status report – Synthesis review on the South African construction industry and its development. Discussion document," South Africa, 2004.
- [3] National Advisory Council on Innovation, "The potential impact of skills shortages on the innovative capacity of major capital engineering projects," National Advisory Council on Innovation and Department of Science and Technology, Pretoria, 2003.
- [4] Construction Industry Development Board, Department of Public Works, Construction Education and Training Authority, "Towards a common framework for enterprise growth and sustainability," Sustainable contractor development, National Workshop, Pretoria, 2005.
- [5] A. Kesper, "Failing or not aiming to grow? Manufacturing SMMEs and their contribution to employment growth in South Africa," TIPS Working Paper 15, 2000.
- [6] D. Matas, "Towards the new millennium series: No more: The battle against human rights," Ottawa, Dundurn Press, 1996.
- [7] Statistics South Africa, "Labour force survey for september 2005," Statistics South Africa, Pretoria, 2005.
- [8] M. K. Nissanke, "Financing enterprise development in sub-Saharan Africa," Cambridge Journal of Economics, Vol. 25, pp. 343–367, 2001.
- [9] C. M. Rogerson, "Unpacking tourism SMMEs in South Africa: Structure, support needs and policy response," Development Southern Africa, Vol. 22, pp. 623–642, 2005.

A Projection Clustering Technique Based on Projection

Xiyu LIU¹, Xinjiang XIE², Wenping WANG¹

¹School of Management and Economics, Shandong Normal University, Jinan, China; ²Department of Computer Science and Technology, Shandong Agricultural Administrators College, Jinan, China.
Email: xyliu@sdnu.edu.cn, wenping216@126.com, xjxie@163.com

Received July 24, 2009; revised September 10, 2009; accepted October 20, 2009.

ABSTRACT

Projection clustering is an important cluster problem. Although there are extensive studies with proposed algorithms and applications, one of the basic computing architectures is that they are all at the level of data objects. The purpose of this paper is to propose a new clustering technique based on grid architecture. Our new technique integrates minimum spanning tree and grid clustering together. By this integration of projection clustering with grid technique, the complexity of computing is lowered to $O(N\log N)$.

Keywords: Cluster Analysis, Projection Clustering, Grid Clustering

1. Introduction

Cluster analysis is an important area in data mining which can explore the hidden structures of business databases [1]. Traditionally, cluster analysis can be categorized as three classes. Partitioning method works by constructing various partitions and then evaluating them by some criterion. Hierarchy method creates a hierarchical decomposition of the set of data (or objects) using some criterion. Density-based method is based on connectivity and density functions. Grid-based method is based on a multiple-level granularity structure. Model-based method is to construct a model and to find the best fit model.

Along these lines, many techniques and algorithms have been proposed in the literature. For example, Ester *et al.* [2] present the density-based clustering algorithm which uses an Eps-neighborhood for a point containing at least a minimum number of points. Raphael Bar-Or and Christiaan van Woudenberg [3,4] present a gravity-based clustering method intended to find clusters of data in n -space. The most classical clustering technique is due to Raymond T. Ng and Jiawei Han [5] who developed a CLARANS which aims to use randomized search to facilitate the clustering of a large number of objects. More recent work include agglomerative fuzzy K-Means clustering algorithm by introducing a penalty term to the objective function to make the clustering process insensitive to the initial cluster centers [6].

Among all these clustering techniques, one of the basic

measurements is the Euclidean distance. It requires similar objects to have close values in all dimensions. When similarity between objects in high dimensional space is absent, this kind of technique is often invalid. To solve this problem, dimension reduction and manifold learning is applied [7–9]. Another method for this skewed data is the projection clustering [10]. The main idea of projected clustering is that different clusters may distribute along part of the dimensions. A projected cluster is a subset of data points, together with a subspace of dimensions, so that the points are closely clustered in the subspace.

Different with the above clustering approaches, graph clustering works by transforming the initial working data into a kind of graph. Then graph clustering techniques can be applied to obtain the final clustering. One of these techniques is the Minimum Spanning Tree (MST) based clustering. Although the first MST-based clustering algorithms have been studied for many years, due to its computational efficiency for large databases, it attracts new researches frequently. In a more recent work [11], the authors present a more efficient method based on the divide and conquer approach that can quickly identify the longest edges in an MST so as to save some computations. The experimental results show that their MST inspired clustering algorithm is very effective and stable when applied to various clustering problems. The authors also expect that their algorithms have a $O(N\log N)$ computing time.

In this paper, we propose a new projection clustering technique by Minimum Spanning Tree based on the grid

clustering approach. Basically, our MST-inspired clustering technique works on cells rather than data points directly. This will significantly reduce the size of graph and MST. Due to this reason, our technique has no specific requirements on the dimensionality of the data sets. This is different from some typical projection clustering algorithms [10].

The rest of this paper is organized as follows: Section 2 presents the basic idea of projection clustering. In Section 3, we summarize some basics on MST-based clustering techniques. In Section 4, we propose a flexible grid clustering approach. Clustering is also discussed as an optimization process in this section. Section 5 contains a brief clustering behavior and time complexity analysis.

2. Projection Cell Clustering

Suppose the raw data set is $X = \{x_1, \dots, x_N\}$. Each point has n components by $x = (x^1, \dots, x^n)$. It is contained in a rectangle D_0 in R^n . Generally, we will not cluster the original data set X . Instead, we will consider a derived data set X composed of data cells. However, after we transform the data set into cells, each cell can be considered as a new data point represented by its center. The number of data points in the cell is called the mass of the cell.

More precisely, a cell (grid) is a polyhedron as a complex $x = (D, \text{norm}, c, b, p)$, where D is the polyhedron, norm is a unit vector indicating normal vector of one of its edges, c is the center, b is a boolean value indicating whether the grid is dense or sparse, and p is the number of data points covered by the grid.

In order to simplify symbols, we will use $x \in \mathcal{X}$ as cell data object and $[x] \subset X$ as the data points defined by x in the original data space X . For two objects x, y , the distance is defined as the minimum distance of the two cells

$$\rho(x, y) = \min_{p \in [x], q \in [y]} d(p, q) \quad (1)$$

The diameter of a data object is measurement of its size

$$\sigma(x) = \frac{1}{2} \max_{p, q \in [x]} \rho(p, q) \quad (2)$$

Let $N(x)$ be the set of k -nearest neighbors (KNN) of x including itself, then the number of object in $N(x)$ is $k+1$. The sparseness or thickness of the data object can be measured by the relative location of its k -nearest neighbors

$$\mu(x) = \frac{1}{k} \sum_{z \in N(x)} \rho(z, x) \quad (3)$$

Suppose there is a mass function defined on \mathcal{X} by $m: \mathcal{X} \rightarrow R^+$; $m(x) = \# [x] = p$ (total number of points). Then we define the density of a data point as

¹Notice that the size of this matrix maybe less than N since we are using cells instead of data points.

$$\beta(x) = \frac{1}{\mu(x)} \sum_{x \in N(x)} m(x) \quad (4)$$

Suppose $x \in X$. We use π_i to denote the projection operator in the i -th component, i.e., $\pi_i(x) = \pi_i x = x^i$. Respectively, $d_i(x, y) = d(\pi_i x, \pi_i y)$. For $x, y \in \mathcal{X}$, define $\pi_i x = \{\pi_i z : z \in [x]\}$, $\rho_i(x, y) = \rho(\pi_i x, \pi_i y)$, and $\sigma_i(x) = \sigma(\pi_i x)$. Then we consider projection into the i -th component, the KNN neighbor set $N(x)$ is replaced by $N_i(x) = \{\pi_i x, y_1, \dots, y_k : \pi_i y_i \text{ are the } k\text{-nearest points to } \pi_i x\}$. The corresponding definition of sparseness and density are

$$\mu_i(x) = \frac{1}{k} \sum_{z \in N_i(x)} \rho_i(z, x), \quad \beta_i(x) = \frac{1}{\mu_i(x)} \sum_{x \in N_i(x)} m(x) \quad (5)$$

Now we describe the process of projected clustering. The main idea is that distance between data objects is restricted to subsets of dimensions where object values are dense [10]. This means that we only consider contributions of relevant dimensions when computing the distance between data point and the cluster center.

Different from [10], we use a fixed threshold value to determine the dense and sparse dimensions. Now let $x \in \mathcal{X}$ to be a cell data object. Suppose $\beta_0 > 0$ is a positive threshold determined in the process of gridding process. Define a matrix $\Theta = [\theta_{ij}]_{N \times n}$ by¹

$$\theta_{x,j} = \begin{cases} 1, & \text{if } \beta_j(x) \geq \beta_0 \\ 0, & \text{else} \end{cases}; j = 1, \dots, n \quad (6)$$

By this index matrix we obtain a projected cell distance as follows

$$\rho_\pi(x, y) = \left[\sum_{j=1}^n \theta_{x,j} \theta_{y,j} (\rho_j(x, y))^2 \right]^{\frac{1}{2}} \quad (7)$$

3. Minimum Spanning Trees

Let $G=(V, E)$ be a connected, undirected edge-weighted graph with N nodes as before. $W = [w_{ij}]$ is the weight matrix. For any $P \subseteq V$, we use $Q = G-P$ to denote the subgraph generated by the vertices $V \setminus P$ called a partition of nodes. A spanning subgraph is a subgraph that contains all the vertices of the original graph. A minimal spanning tree of graph G is a spanning graph with no circuits whose weight is minimum among all spanning trees of G .

For a partition P, Q of G , define $\rho(P, Q)$ as the smallest weight among all edges from the cut-set $C(P, Q)$, which is the set of edges connecting P and Q . A link is any edge in $C(P, Q)$ with weight $\rho(P, Q)$. The link set is denoted by $\lambda(P, Q)$ [12].

There are several ways to build Minimum Spanning Tree (MST) from the graph [11]. Two popular ways to implement the algorithms are the agglomerative and the divisive procedures.

The well-known agglomerative procedures are the Kruskal and Prim's algorithms. The first one works by constructing the tree from initial N isolated vertices of the original graph. All the edges are sorted into a non-decreasing order by their weights. For each edge which is not in the tree, if this edge does not form a cycle with the current tree, then we can add this edge to the tree. In the Prim's algorithm, the tree construction starts with a root node. At each step, among all the edges between the nodes in the tree T and those which are not in the tree yet, the node and the edge associated with the smallest weight to the tree T are added.

The second kind of algorithm is the divisive one called the reverse delete algorithm starting with the full graph. Edges are deleted in order of nonincreasing weights based on the cycle property as long as keeping the connectivity of the graph.

Some well-known properties of MST are summarized in the following theorem [12].

Theorem 3.1. The minimum spanning tree $T(G)$ of a graph G has the following properties.

- 1) T contains at least one edge from $\lambda(P, Q)$ for each partition P, Q .
- 2) Each edge of T is a link of some partition of G .
- 3) Let (C_1, C_2) be a partition of G . If $\rho(P, Q) < \rho(C_1, C_2)$ for each partition (P, Q) of C_1 , then $T(C_1)$ forms a connected subtree of $T(G)$.

Once we have the MST, we can obtain the desired clustering by removing inconsistent edges of MST. The simplest way to define inconsistent edges is using weight measure ratio of the edge with average weight of nearby edges in the tree [12]. If the ratio is larger than a threshold, then it is inconsistent. We can determine a stop criteria by the number of clusters, or a minimum size of any cluster by removing edges which can result in two clusters whose sizes are larger than the minimum cluster size.

If we know the number of clusters k , then clustering can start by removing $k-1$ arbitrary edges from the tree, creating a k -partition. Then we can minimize the change of the total weight of the current clusters to obtain the final clustering.

To reduce computation complexity, Xiaochun Wang *et al.* proposed a divide and conquer approach [11]. Given a loose estimate of minimum and maximum numbers of data items in each cluster, they propose an iterative approach for MST clustering algorithm in five steps: 1) Start with a spanning tree built by the sequential initialization (SI). 2) Calculate the mean and the standard deviation of the edge weights in the current distance array and use their sum as the threshold. Partially refine the spanning

tree by running Divisive Hierarchical Clustering Algorithm (DHCA) multiple times until the percentage threshold difference between two consecutively updated distance arrays is below 10^{-6} . 3) Identify and verify the longest edge candidates by running MDHCA until two consecutive longest edge distances converge to the same value at the same places. 4) Remove this longest edge. 5) If the number of clusters in the data set is preset or if the difference between two consecutively removed longest edges has a percentage decrement larger than 50 percent of the previous one, we stop. Otherwise go to Step 3.

However, if the graph size is not large, we can directly get clustering from the graph. When we use the flexible grids technique to obtain the graph, this is often the case. Anyway, the technique of [11] can be applied to further reduce the computing time.

4. Grid Based Spatial Clustering

The grid based clustering uses a multi-resolution grid structure which contains the data objects and acts as operands of clustering performance [1]. For example, the authors [13] propose a gravity based grid which approximates the cell influence by gravity centers. The authors claim that the proposed technique can reduce memory usage and simplify computational complexity with minor losses of the clustering accuracy.

Traditional grids are regular hypercubic grid. This requires the grid construction cover all the data space with the same precision. The second method uses flexible grids, i.e. multi-resolution grids with hypercubic or hyper-rectangular cells having randomly oriented borders [14]. The main clustering technique is a tree based searching with a similarity measure composed of both the density and distance differences [15].

Suppose the data set is $X = [x_1, \dots, x_N] \subset R^n$. It contains in a rectangle D_0 in R^n . A grid is a graph G where each node is a complex $v = (D, norm, c, isCrowded, p)$, where D is the polyhedra, $norm$ is a unit vector indicating normal vector of previous cutting plane, c is a point which lies in the grid acting as its center, $isCrowded$ is 1 or 0 indicating whether the grid is highly populated or not, and p is the number of data points covered by the grid. The initial grid is D_0 with an arbitrary normal vector. In each step, we can define the center of the grid as its geometrical center.

For two nodes $v_i = (D_i, norm_i, c_i, isCrowded_i, p_i)$, $i = 1, 2$ there exists a connecting edge between them. The weight value on this edge is defined as

$$\rho(D_1, D_2) = \begin{cases} 0, & \text{if } \min\{p_1, p_2\} = 0 \\ \rho_\pi(v_1, v_2), & \text{else} \end{cases} \quad (8)$$

The graph is constructed in an iterative way. We start with an initial candidate node $v_0 = (D_0, norm_0, c, isCrowded, p)$ where D_0 is the original rectangle, $norm_0$ is a random selected unit vector, c is the geometrical center of D_0 , $isCrowded = 1$, and p is the total population number. Then at each step, the cell containing more number of points (controlled by a threshold value β_p , or larger enough by diameter) controlled by another threshold value β_d , is split into two subcells by a hyperplane which is orthogonal to the current normal vector. Position of the hyperplane is random. A cell is called crowded if its population is larger than β_p . Otherwise it is called sparse. If we reach a sparse cell, then add this cell to the node set of the graph. If we reach a cell with diameter less than β_d , then add this cell to the node set.

This step continues until each cell has a population less than β_p , or its diameter is smaller than β_d . Table 1 gives the algorithm for the graph construction process.

By this algorithm, we can generate a hierarchical grid together with a resulting graph. When the graph is generated, the clustering will become grouping nodes of the

graph into clusters. A commonly used technique to deal with this problem is the hierarchical clustering [1]. Agglomerative methods iteratively connect vertices that are close to each other with edges. Small clusters are merged with each other building up the hierarchical structure to find the desired larger clustering structure. Divisive methods on the contrary are based on network flows. This is done by iteratively identifying the intercluster edges in a top-down approach.

Once we have completed the graph construction, those nodes in the graph which are not crowded will correspond to vacant area or outliers. Therefore, in order to reduce computing complexity, we first remove all sparse graph nodes with corresponding edges. The resulting graph is $G = (V, E)$ where V is the set of vertices, E the set of weighted edges. An example is shown in Figure 1 with part of its edges.

Now we use $C(X) = \{X_q: q = 1, 2, \dots, k\}$ to denote a clustering of the data set X where O is the set of outliers. Then

$$X = O \cup \bigcup_{C \in C(X)} C \quad (9)$$

Table 1. Flexible grids construction algorithm

Algorithm: Construction of flexible grids

Inputs

$X = \{x_1, \dots, x_N\}$ dataset of N points in R^n .

D : hyper-rectangle containing X

β_p : population threshold value.

β_d : cell diameter threshold value

Outputs

$V = \{v_1, \dots, v_N\}$ set of vertices

Begin

$V(0) = \{\bar{v}_0 = (D_0, n_0, c_0, 1, p_0)\}$. Let $t = 0$.

while $V(t) \neq \emptyset$

for each $v = (D, n, c, isCrowded, p) \in V(t)$

Generate a cutting hyperplane L passing c and with normal vector parallel to n . Cut the current cell v into two subcells

\bar{v}_1, \bar{v}_2 . For each new node, if $p < \beta_p$ or $diam(D) < \beta_d$,

add this new cell to the node set V . Else add it to $V(t+1)$.

Let the new norms be orthogonal to n .

end

$t++$;

end

End

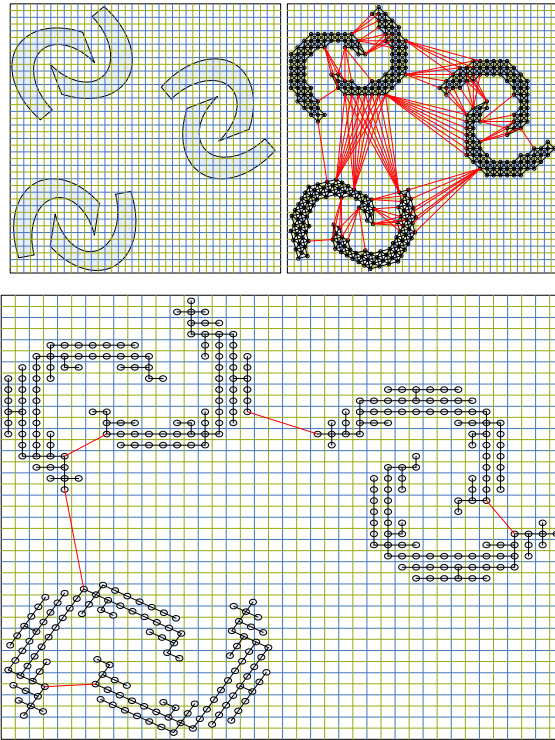


Figure 1. An example clustering area with three clusters

For given population threshold value and diameter threshold value, we can generate the flexible grid and obtain a graph. Then we can generate a minimum spanning tree. Then we can get the final clustering. Figure 1 shows an example of minimum spanning tree corresponding to the data set in Figure 1.

Consequently, for fixed β_d and β_p , the data set X is clustered as $C(X, \beta_d, \beta_p)$. We define $|X_q|$ as the number of cells in X_q . Define the energy of clustering as the sum of intra-cluster variation measure and inter-cluster distance as

$$E(X, \beta_p, \beta_d) = \frac{1}{k} \sum_{q=1}^k \sum_{\substack{v_i \neq v_j \\ v_i, v_j \in X_q}} \frac{1}{|X_q|} \left| \frac{p_i}{\text{diam}(D_i)} - \frac{p_j}{\text{diam}(D_j)} \right|^2 \quad (10)$$

$$+ \frac{1}{\min_{1 \leq i \leq k} \rho(X_i, \bigcup_{j \neq i} X_j)}$$

By this grid based method, the final clustering can be treated without direct computation to the data points and reduce the number of nodes significantly. The only parameters we need to determine beforehand is the cell crowded threshold value and the minimum cell diameter. One way to choose these two parameters is to optimize the energy of clustering.

5. A Performance Study

Suppose that we want to cluster a data set $\Omega \subset R^n$ with N objects. By the graph construction algorithm in the previous section, the data Ω set is split into a cell hierarchy. A minimum spanning tree is constructed associated with the cell graph.

To make things simpler, we will assume that the cutting planes are perpendicular to one of the axis in this section. Moreover, we assume that each cutting plane passes through the geometrical center of the current cell. Therefore, all the cells are rectangles. Then we can easily count the total number of nodes of the graph.

Suppose the original data cube D_0 has a diameter σ_0 , and the initial population threshold is β_p . For some sufficiently large number M , let $\sigma_i = 2^{-i} \sigma_0, i = 1, 2, \dots, M$. Let β_i be another decreasing sequence. By choosing appropriate two sequences, we can optimize the clustering.

For specific population and diameter threshold parameters β, σ , let the induced graph be $G(\beta, \sigma)$ with minimum spanning tree $T(\beta, \sigma)$. A clustering of T is denoted by $C(\beta, \sigma)$ is a set of disjoint subtrees whose nodes set coincide with the original tree. We use $\mathcal{C}(\beta, \sigma)$ to denote the clustering of the data set.

Evidently we have the following properties.

Theorem 5.1. Clusters have two properties.

1) Anti-joint property. If $\sigma_1 > \sigma_2$, then two disjoint clusters in $\mathcal{C}(\beta, \sigma_1)$ are disjoint in $\mathcal{C}(\beta, \sigma_2)$.

2) Monotonicity property. If $\beta_1 > \beta_2$, then a cluster in $\mathcal{C}(\beta_1, \sigma)$ cannot be disintegrated in $\mathcal{C}(\beta_2, \sigma)$.

Now we assume the population threshold β is a constant which determines the granularity of the problem. Therefore we use $\mathcal{C}(\sigma)$ to denote the clustering. Let us split the energy into two parts, the intra-cluster energy $E_{ia}(\sigma)$ and the intercluster energy $E_{ie}(\sigma)$ as follows

$$E_{ia}(X, \sigma) = \frac{1}{k} \sum_{q=1}^k \sum_{\substack{v_i \neq v_j \\ v_i, v_j \in X_q}} \frac{1}{|X_q|} \left| \frac{p_i}{\text{diam}(D_i)} - \frac{p_j}{\text{diam}(D_j)} \right|^2 \quad (11)$$

$$E_{ie}(X, \sigma) = \frac{1}{\min_{1 \leq i \leq k} \rho(X_i, \bigcup_{j \neq i} X_j)} \quad (12)$$

Theorem 5.2. Inter-cluster energy is monotone. That is to say, if $\sigma_1 > \sigma_2$, then $E_{ie}(X, \sigma_1) \geq E_{ie}(X, \sigma_2)$.

Proof: It is clear to see that when $\sigma_1 > \sigma_2$, a cell maybe split into small cells. Therefore, either the set X_i will be smaller which means that the distance $\rho(X_i, \bigcup_{j \neq i} X_j)$ will become larger.

In a recent work [11] the authors propose a divide and

conquer based algorithm consisting of two phases. The first phase includes the sequential initialization and the spanning tree updating, and the second phase uses some technique to locate the longest edges and partitions the obtained approximate minimum spanning tree to form sensible clusters. The authors expect the first phase has $O(fN \log N)$ where f is constant. The average time complexity of the second phase is $O(eN \log N)$ where e is constant. Therefore their expectation of time complexity is $O(N \log N)$. However, our algorithm do provide a time complexity of $O(N \log N)$.

Theorem 5.3. Suppose thresholds are $\beta_p = \beta$ and $\beta_d = \sigma$. Then the time complexity of the flexible grids construction algorithm is $O(N \log N)$.

Proof: At each stage, if a cell $v = (D, n, c, isCrowded, p)$ is sparse, i.e., $isCrowded = 0$, then the cell is a node in the graph. Otherwise, a cutting hyperplane L passing c and with normal vector parallel to n is generated. The current cell v is cut into two subcells \bar{v}_1, \bar{v}_2 . For each new node, if $p < \beta_p$ or $diam(D) < \beta_d$, add this new cell to the node set V .

In this process, the computation of $isCrowded$ and p both have a time of $O(N_i)$ where N_i is the number of data points in the new cell. Ideally, the plane L passes the center of the cell v . Hence $N_i = p/2$ for $i = 1; 2$. If not so, the plane is randomly placed by a uniform distribution which we expect the same property. Let the total time complexity be $T(N)$. Then we have $T(N) = 2T(N/2) + O(N)$. Hence we know that $T(N) = O(N \log N)$.

6. Conclusions

In this paper we present a new projection clustering technique based on grid architecture and minimum spanning tree. The effective using of minimum spanning tree can possibly reduce computing complexity although the construction of the graph and the tree are relatively complicated. The main ingredient here is the application of grid clustering to projection clustering.

Apparently, this research will lead to efficient algorithms. In future work, we will give experimental study on the new technique. This will be lengthy, for the clustering is essentially an optimization process. The best population threshold β_p is to be determined which optimizes the clustering energy presented in Section 3. For this reason, we will present this part of the research in another paper, together with an application.

7. Acknowledgments

This project is carried out under the "Taishan Scholar" project of Shandong, China.

Research is also supported by the Natural Science Foundation of China (No.60873058), the Natural Science Foundation of Shandong Province (No. Z2007G03), and

the Science and Technology Project of Shandong Education Bureau.

REFERENCES

- [1] J. W. Han and M. Kamber, "Data mining concepts and techniques," 2nd Edition, Elsevier, Singapore, 2006.
- [2] M. Ester, H. P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in Proceedings 2nd International Conference on Knowledge Discovery and Data Mining, Portland, OR, pp. 226–231, 1996.
- [3] R. Bar-Or and C. van Woudenberg, "A novel gravity-based clustering method, technical report," Department of Applied Mathematics, University of Colorado, Denver, CO 80202, 2001.
- [4] R. Bar-Or and C. van Woudenberg, "Gene expression analysis with a novel gravity-based clustering method," pp.1–46, December 2001.
- [5] R. T. Ng and J. Han, "Clarans: A method for clustering objects for spatial data mining," IEEE Trans. on Knowledge and Data Engineering, Vol. 14, No. 5, pp. 1003–1016, September/October 2002.
- [6] M. Li, M. K. Ng, Y. M. Cheung, and J. Huang, "Agglomerative fuzzy K-Means clustering algorithm with selection of number of clusters," IEEE Trans on Knowledge and Data Engineering, Vol. 20, No. 11, pp. 1519–1534, 2008.
- [7] Z. He, X. Xu, and S. Deng, "Scalable algorithms for clustering large datasets with mixed type attributes," International Journal of Intelligent Systems, Vol. 20, pp. 1077–1089, 2005.
- [8] Z. F. He and F. L. Xiong, "A constrained partition model and K-Means algorithm," Journal of Software, Vol.16, No.5, pp. 799–809, 2005.
- [9] Z. Y. Zhang and H. Y. Zha, "Principal manifolds and nonlinear dimensionality reduction via tangent space alignment," SIAM Journal of Scientific Computing, Vol. 26, No. 1, pp. 313–338, 2004.
- [10] M. Bouguessa and S. R. Wang, "Mining projected clusters in high-dimensional spaces," IEEE Transaction on Knowledge and Data Engineering, Vol. 21, No. 4, pp. 507–522, 2009.
- [11] X. C. Wang, X. L. Wang, and D. M. Wilkes, "A divide-and-conquer approach for minimum spanning tree-based clustering," IEEE Trans on Knowledge and Data Engineering, Vol. 21, No. 7, pp. 945–958, 2009.
- [12] C. T. Zahn, "Graph-theoretical methods for detecting and describing gestalt clusters," IEEE Trans. Computers, Vol. 20, No. 1, pp. 68–86, January 1971.
- [13] C. H. Li and Z. H. Sun, "A mean approximation approach to a class of grid-based clustering algorithms," Journal of Software, Vol. 14, No. 7, pp. 1267–1274, 2003.
- [14] Marc-Ismael, Akodj'enou-Jeannin, Kav'e Salamatian, and P. Gallinari, "Flexible grid-based clustering," J. N. Kok et al. (Eds.), LNAI 4702, PKDD, pp. 350–357, 2007.
- [15] U. Brandes, M. Gaertler, and D. Wagner, "Experiments on graph clustering algorithms," In ESA, pp. 568–579, 2003.

Revenue and Duration of Oral Auction

Junmin SHI, Ai-Chih CHANG

Department of Supply Chain Management & Marketing Sciences, Rutgers University, Newark, USA.
Email: jshi@andromeda.rutgers.edu, littlei617@gmail.com

Received July 13, 2009; revised August 22, 2009; accepted October 5, 2009.

ABSTRACT

This paper investigates the revenue and duration of a well-known hybrid oral auction (English auction and Dutch auction) that is extensively adopted in practice, for instance the Christie's. Unlike sealed bid auction, oral auction is featured by its complexity of dynamic process. The bidding price varies as a stochastic time series. Therefore, the duration of oral auction as well as its revenue performs randomly. From the seller's perspective, both the revenue and the duration are so important that extra attention and effort should be put on auction design. One of the most important issues is how to choose the starting bid price to maximize its revenue or minimize its duration. In this paper, the bidding process is decomposed into two phases: English auction (descending-bid) phase and the Dutch auction (ascending-bid) phase. For each phase, with the aid of Markov method, we derive the expected revenue and duration as a function of the starting bid. For an oral auction with a large number of bidder and each bidder behaves independently, we provide the limit results of the expected revenue and duration. The results of the auction model can be easily implemented in auction design.

Keywords: Duration, Dutch Auction, English Auction, Oral Auction, Revenue

1. Introduction

As a powerful and well-known tool in business markets, auction plays an important role in selling objects especially for antiques and art. With a long history around the world, auctions are very common for the commodities such as tobacco, fish, cattle, racehorses, and anything that has a market of multiple people interested in purchasing. The main reason why auction is so common is that a group of people are interested in buying the same object, and thereby offering their individual bids on the object. Serving as a tool that takes all the interested buyers into one game, auction decides the winner (usually the highest bidder) of the game.

Recently, the auction theory has been well developed systematically in practice and academy as well. Generally speaking, there are four types of auction that are used for the allocation of a single item: These four standard auctions are the English auction, the Dutch auction, the First-Price Sealed-Bid auction, and the Second-Price Sealed-Bid auction. The context for each type of auction is explained briefly as follows. We refer the interested reader to [1].

• **Open ascending-bid auctions (English auctions)** is commonly referred to as oral outcry auctions, in which the price is steadily raised by the auctioneer with bidders dropping out once the price becomes too high. This con-

tinues until there remains only one bidder (the highest bidder) who wins the auction at the current price.

• **Open descending-bid auctions (Dutch auctions)** in which the price starts at a level sufficiently high to deter all bidders and is progressively lowered until a bidder indicates to buy at the current price. The bidder wins the auction and pays the price at which he or she bid.

• **First-price sealed-bid auctions** in which bidders place their bid in a sealed envelope and simultaneously hand them in to the auctioneer. The envelopes are opened and the bidder with the highest bid wins, paying a price at which he or she bid.

• **Second-price sealed-bid auctions (Vickrey auctions)** in which bidders place their bid in a sealed envelope and simultaneously hand them to the auctioneer. The envelopes are opened and the bidder with the highest bid wins, but paying at the *second highest* bid.

Revenue management is the most crucial topic in auction design and its application. One of the most remarkable results in auction theory is the revenue equivalence theorem, which was first introduced by [2]. Two auctions are said to be "*revenue equivalent*" if they produce the same expected sales price. This is an important issue to a seller who wants to hold an auction to sell the item for the highest possible price. If one type of auction is found to generate higher average sales revenue, then that type

auction will obviously be preferred by the sellers. In other words, the revenue equivalence theorem states that, if all bidders are risk-neutral bidder and have independent private value for the auctioned items, then all four of the standard single unit auctions have the same expected sales price (or seller's revenue).

To analyze and formulate a dynamic auction, most literature adopts the stochastic process approach. For example, an online auctions problem is studied by [3] and they design a $(1 - 1/e)$ competitive (optimal) algorithm for the online auction problem. Vulcano, G. *et al.* analyze a dynamic auction, in which a seller with C units to sell faces a sequence of buyers separated into T time periods [4]. They assume each group of buyers has independent, private values for a single unit. Buyers compete directly against each other within a period, as in a traditional auction. For this setting, they prove that dynamic variants of the first-price and second-price auction mechanisms maximize the seller's expected revenue. E. J. Pinker, *et al.* analyze the current state of management science research on online auctions [5]. They develop a broad research agenda for issues such as the behavior of online auction participants, the optimal design of online auctions, and so on. Most recently, Z. J. Shen and X. Su provide a detailed and up-to-date review of customer behavior in the revenue management and auction literatures and suggest several future research directions [6].

In the auction study, an increasing number of empirical studies apply a structural econometric approach within the theoretical framework of the *independent private values model* and the *common value model* [7,8]. In common value auctions, a bidder's value of an item depends entirely on other bidders' values of it. By contrast, in private value auctions, the value of the good depends only on the bidder's own preferences. In auction design, efficiency and optimality are the two primary goals: the former focuses on the social welfare of the whole seller-bidder system, whereas the latter emphasizes the revenue-maximizing from the seller's perspective [9].

Markov theory was developed by the Russian mathematician A. A. Markov. The theory provides a foundation for modeling a stochastic process whose future state depends solely on its current state and is completely independent of its past states. This property is well known as memorylessness [10]. Markov process has been widely applied to model the auction process. For example, S. Park, *et al.* devise a new strategy that an agent can use to determine its bid price based on a more tractable Markov chain model of the auction process [11]. They show that this strategy is particularly effective in a "seller's market". A. Segev, *et al.* model an online auction in terms of a Markov process on a state space defined by the current price of the auctioned item and the number of bidders that were previously "bumped" [12]. They first convert an online auction into a small-to-medium sized auction.

Then the transition probability matrix of states is derived and the price trajectory of the small-scale Markov process is obtained. Finally, the final price prediction can be determined based on the obtained transition probability matrix.

Duration of auction is another factor under consideration in auction design. D. Reiley, *et al.* show that the length of the auction positively influences the auction price [13]. To the best of our knowledge, except the aforementioned reference, there are very few literatures considering the duration of auction. At this point, one effort of this paper is to bridge the gap via deriving the duration as a function of the starting bid.

Unlike sealed bid auction, oral auction is featured by its dynamic complexity of the bidding process. In practice, oral auction is more widely preferred than sealed bid auction. For instance, Christie's has auctioned off artwork and personal possessions mostly via oral auctions [14]. Christie's was founded in London, England, on 5th December 1766 by James Christie. Christie's soon established a reputation as a leading auction house, and took advantage of London's new found status as the major centre of the international art trade after the French Revolution. Christie's has held the greater market share against its longtime rival, Sotheby's, for several years and is currently the world's largest auction house by revenues. In addition to Christie's, a variety of world famous auction organizations adopt oral auction in their business.

The bidding process of the oral auction under study is explained as follows. The auctioneer begins the auction with an announced starting bid. This bid is referred to as the starting bid. Then the auctioneer will ask the bidders for their response by open cries. If nobody responds for the bid, then the auctioneer announce "Going once" for a short while. If there is still no response from bidders, then "going twice" is announced for another short while. If no response again, then the auctioneer deduces the bid, and ask the bidder for their response. The similar process continues until there is bidder responding to the revised bid. Such bid-decreasing phase will be stopped since all bidders shall response to a revised bid while it gets low enough. Once a bidder response to a revised bid, the auctioneer increases the bid and asks for the response with "going once, going twice" as aforementioned crying out process. The process proceeds and stops until there is no response within the sequential announcement of "going once, going twice and gone". In other words, once no bidder is willing to raise the revised bid, the object is "hammered down", and the last bidder (with the highest bid) wins the auction.

From the bidding process described above, we can see the oral auction is composed of two phase: *descending-bid* (Dutch auction) and *ascending-bid* (English auction), and thereby the oral auction is referred to as a *hybrid auction*. For the hybrid oral auction and from the

practice point of view, the final bid and the total time spent on bidding are two important factors investigated in auction design. The final bid of the auction is referred to as the revenue and the total time of bidding is referred to as the duration. Each of them can be employed to evaluate the efficiency of the auction. Maximum revenue or minimum duration can be obtained via choosing an optimal starting bid. In this paper, it is of our interest to derive the expected revenue and duration as a function of the starting bid. To this end, we shall first formulate an oral auction model based on Markovian property. Then we decompose the hybrid oral auction into two phases: descending-bid phase and ascending-bid phase. For each phase, we derive the Markov transition matrixes which are referred to as the downward for the descending-bid phase and upward matrix for the ascending-bid phase. With the aid of Markov approach, we finally obtain the revenue and duration of the auction as a function of the starting bid.

This paper makes the following contributions. First, it presents an exploratory analysis of the hybrid auction and obtains a closed form expressions for the auction revenue and duration. Secondly, for a large group of bidders, the limit performance has been analyzed. These results can be applied directly in practice as an aid in auction design.

The remaining of this paper is organized as follows. The model of oral auction is formulated in § 2. The expected revenue is derived in § 3, and the duration is derived in § 4. Finally, § 5 concludes the paper.

2. Model Formulation

Let the discrete bid levels be denoted by $\{P_i : i = 0, 1, 2, 3, \dots\}$,

where $P_0 = 0$ and $P_i < P_{i+1}$. Let S and F denote the starting bid and the final bid, respectively. Let B denote the lowest bid level over the bidding process, where $B \leq S$ and $B \leq F$. For each bid, let T denote the whole period length of announcement by the auctioneer (briefly referred to as the *announcement period* for each bid). Let $N_i(T)$ denote the number of responds to the bid P_i from bidders within announcement period T . In the descending phase, if $N_i(T) = 0$ for bid P_i , then the auctioneer will decrease the bid, otherwise, the descending-bid phase will stop and proceed to the ascending-bid phase. In the ascending phase, if $N_i(T) > 0$ for bid P_i , the auctioneer will increase the bid, otherwise, the auction will end with the final bid $F = P_i$. Figure 1 depicts a sample path of the oral auction bidding process.

From the auction process described above, we have some following conclusions.

1) All the items could be auctioned off since the bidders are willing to take the auction at a low enough price, say P_1 . Therefore, the final bid of the auction is at least P_1 .

2) Within the bidding process, there are two bidding phase: descending and ascending. In the descending phase, bidding price decreases from start price S to the lowest bid B . In the ascending phase, bidding price increases from the lowest bid B to the final bid F .

The random variable $N_i(T)$ corresponding to bid P_i governs the bidding process. If $N_i(T) > 0$, namely, there are some bidders willing to take the auction with bid P_i , then the auctioneer will revise the bid and increase

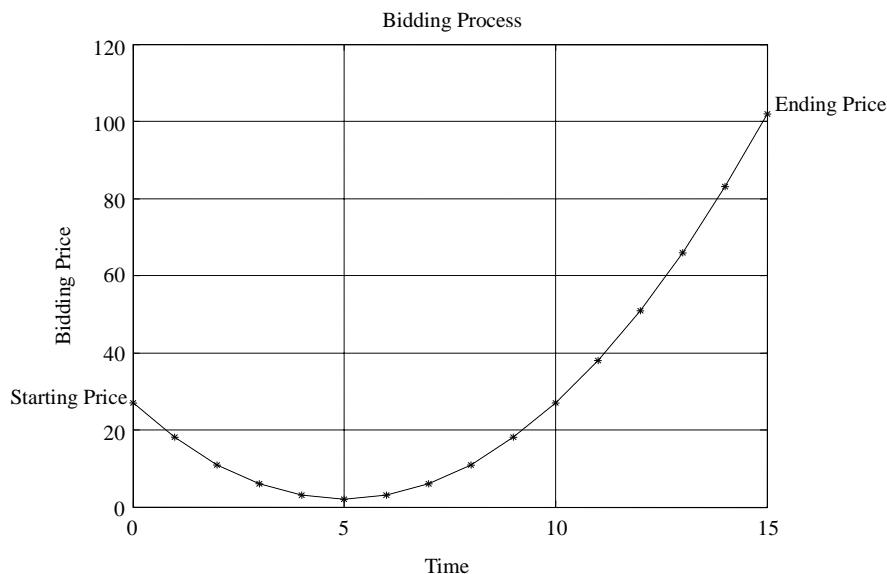


Figure 1. A sample path of the bidding process

from P_i to P_{i+1} . Let

$$\Pr\{N_i(T) = 0\} = q_i \quad (1)$$

then,

$$\Pr\{N_i(T) > 0\} = 1 - q_i \quad (2)$$

Since each q_i is given and constant, the transition probability from P_i to P_{i+1} or P_{i-1} is solely determined by q_i . It implies that the bidding process is a Markovian.

From previous discussion, we can see there are three typical processes which are possibly incurred in practice.

Case 1. Descending-bid (Dutch auction)

In the descending-bid process, the bidding is monotonously decreasing from starting price S to final bid F . Figure 2 depicts a sample path of such process.

Case 2. Ascending-bid (English auction)

In the ascending-bid process, the bidding is monotonously increasing from starting price S to final bid F . Figure 3 depicts a sample path of such process.

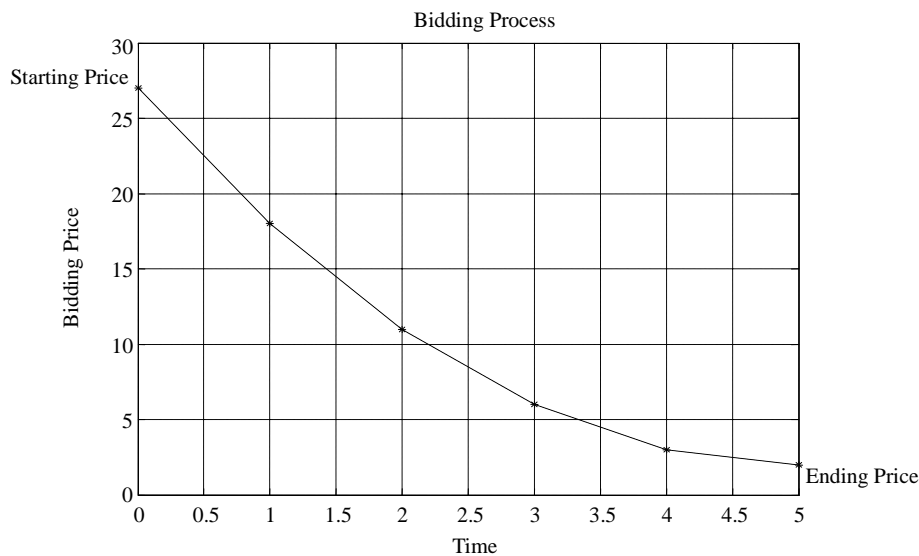


Figure 2. A sample path of the descending-bid process

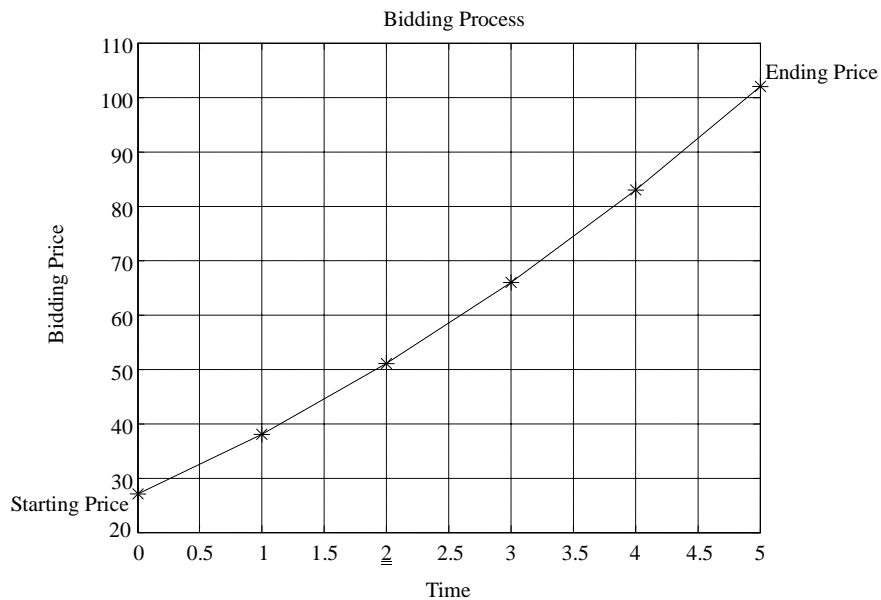


Figure 3. Ascending-bid process

Case 3. Hybrid bidding

As shown in Figure 1, the bidding is first decreasing from starting bid S to the lowest bid B , and then increasing from the lowest bid B to the final bid F .

As we can see, the descending-bid (described in case 1) and ascending-bid (described in case 2) processes are trivial cases of the hybrid bidding process. The hybrid auction could be decomposed into descending-bid phase and ascending-bid phase, which gives an idea to analyze oral auction.

2.1 Descending Phase

The descending phase is a Markov process and its one-step transition matrix is given by

$$\mathbf{M}_- = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots \\ q_2 & 1-q_2 & 0 & 0 & \cdots \\ 0 & q_3 & 1-q_3 & 0 & \cdots \\ 0 & 0 & q_4 & 1-q_4 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \quad (3)$$

To see this, we consider the states and their one-step transition over the descending-bid process.

For any $i > 1$,

$$[\mathbf{M}_-]_{(i,i-1)} = \Pr\{N_i(T) = 0\} = q_i \quad (4)$$

and

$$[\mathbf{M}_-]_{(i,i)} = \Pr\{N_i(T) > 0\} = 1 - q_i \quad (5)$$

Since any bidder is willing to take the auction at price P_1 , we must have $q_1 = 0$. Therefore

$$[\mathbf{M}_-]_{(1,1)} = \Pr\{N_1(T) > 0\} = 1 \quad (6)$$

For any other states where $j \neq i$ and $j \neq i-1$, we have

$$[\mathbf{M}_-]_{(i,j)} = 0 \quad (7)$$

In summary, we have \mathbf{M}_- given by Equation (3).

2.2 Ascending Phase

The ascending phase is a Markov process and its one-step transition matrix is given by

$$\mathbf{M}_+ = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \cdots \\ 0 & q_2 & 1-q_2 & 0 & 0 & \cdots \\ 0 & 0 & q_3 & 1-q_3 & 0 & \cdots \\ 0 & 0 & 0 & q_4 & 1-q_4 & \cdots \\ 0 & 0 & 0 & 0 & q_5 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \quad (8)$$

To see this, we consider the states and their one-step transition over the descending-bid process.

For any $i > 1$,

$$[\mathbf{M}_+]_{(i,i)} = \Pr\{N_i(T) = 0\} = q_i \quad (9)$$

and

$$[\mathbf{M}_+]_{(i,i+1)} = \Pr\{N_i(T) > 0\} = 1 - q_i \quad (10)$$

Since any bidder is willing to take the auction at price P_1 , we must have

$$[\mathbf{M}_+]_{(1,1)} = \Pr(N_1(T) > 0) = 1 \quad (11)$$

and

$$[\mathbf{M}_+]_{(1,2)} = \Pr(N_1(T) = 0) = 0 \quad (12)$$

for any other states where $j \neq i$ or $j \neq i+1$,

$$[\mathbf{M}_+]_{(i,j)} = 0 \quad (13)$$

In summary, we have \mathbf{M}_+ given by Equation (8).

3. Revenue of Oral Auction

In this section, we shall derive a functional expression for the revenue of oral auction as a function of the starting bid.

Given the starting bid P_k , let the expected revenue of the auction be denoted by

$$R(P_k) = \mathbb{E}[F | S = P_k] \quad (14)$$

In the following, we consider the revenue in descending-bid, ascending-bid and hybrid auctions.

3.1 Descending-Bid Phase

During the descending-bid process with the starting bid $S = P_k$, the probabilities for $F = P_k$, P_{k-1} , P_{k-2} , ..., P_1 are provided in Table 1.

Note that

$$[(\mathbf{M}_-)^{k-l}]_{(l,k)} = q_k q_{k-1} \cdots q_l \quad (15)$$

Therefore, in ascending-bid process

$$\begin{aligned} P(F = P_l | S = P_k) &= q_k q_{k-1} \cdots q_{l+1} (1 - q_l) q_{l+1} \\ &= (1 - q_l) q_{l+1} [(\mathbf{M}_-)^{k-l-1}]_{(l+1,k)} \end{aligned} \quad (16)$$

Accordingly, the expected revenue for the ascending-bid auction with the starting bid $S = P_l$ is given by

Table 1. Final bid and its probability over the descending-bid phase

Final bid F	$P(F S = P_k)$
P_k	$(1 - q_k)q_{k+1}$
P_{k-1}	$q_k(1 - q_{k-1})q_k$
\vdots	\vdots
P_{l+1}	$q_k q_{k-1} q_{k-2} \cdots q_{l+2} (1 - q_{l+1})q_{l+2}$
P_l	$q_k q_{k-1} \cdots q_{l+1} (1 - q_l)q_{l+1}$

Table 2. Final bid and its probability over the ascending-bid phase

Final bid F	$P(F S = P_l)$
P_l	$(1 - q_l)q_{l+1}$
P_{l+1}	$(1 - q_l)(1 - q_{l+1})q_{l+2}$
\vdots	\vdots
P_{m-1}	$(1 - q_l) \cdots (1 - q_{m-1})q_m$
P_m	$(1 - q_l) \cdots (1 - q_m)q_{m+1}$

$$R_-(P_k) = \mathbb{E}[F|S = P_k] = \sum_{l=1}^k P_l (1 - q_l) q_{l+1} [(\mathbf{M}_-)^{k-l-1}]_{(l+1,k)} \quad (17)$$

3.2 Ascending-Bid Phase

In the ascending process with the starting bid $S = P_l$, the probabilities for $F = P_l, P_{l+1}, \dots, P_m$ are provided in Table 2.

Note that

$$[(\mathbf{M}_+)^{m+1-l}]_{(l,m+1)} = (1 - q_l)(1 - q_{l+1}) \cdots (1 - q_m) \quad (18)$$

therefore,

$$\Pr(F = P_m | S = P_l) = (1 - q_l) \cdots (1 - q_m) q_{m+1} = q_{m+1} [(\mathbf{M}_+)^{m+1-l}]_{(l,m+1)} \quad (19)$$

Accordingly, the expected revenue for the ascending-bid auction with the starting bid $S = P_l$ is given by

$$R_+(P_l) = \mathbb{E}[F|S = P_l] = \sum_{m=l}^{\infty} P_m \cdot \Pr(F = P_m | S = P_l) = \sum_{m=l}^{\infty} P_m q_{m+1} [(\mathbf{M}_+)^{m+1-l}]_{(l,m+1)} \quad (20)$$

3.3 Hybrid Auction

In the hybrid auction process with the starting bid $S = P_k$, the probability for the process with the lowest bid level $B = P_l$ and the final bid $F = P_m$ is given by, where $l \leq k$ and $l \leq m$,

$$\Pr(B = P_l, F = P_m | S = P_k) = q_k q_{k-1} \cdots q_{l+1} (1 - q_l) \cdots (1 - q_m) q_{m+1} = [(\mathbf{M}_-)^{k-l-1}]_{(l+1,k)} [(\mathbf{M}_+)^{m+1-l}]_{(l,m+1)} q_{m+1} \quad (21)$$

Given the starting bid $S = P_k$, the expected revenue is given by

$$\begin{aligned}
R(P_k) &= \mathbb{E}[F|S = P_k] = \mathbb{E}\left[\mathbb{E}[F|S = P_k, B = P_l]\right] \\
&= \sum_{l=1}^k \left[(\mathbf{M}_-)^{k-l}\right]_{(l,k)} \sum_{m=l}^{\infty} P_m q_{m+1} \left[(\mathbf{M}_+)^{m+1-l}\right]_{(l,m+1)} \\
&= \sum_{l=1}^k \sum_{m=l}^{\infty} P_m q_{m+1} \left[(\mathbf{M}_-)^{k-l}\right]_{(l,k)} \left[(\mathbf{M}_+)^{m+1-l}\right]_{(l,m+1)}
\end{aligned} \tag{22}$$

3.4 An Example of Auction with a Large Number of Bidders

In this subsection, we consider an example of oral auction where there are a large number of potential bidders. We assume that each bidder responds to the bid independently and the probability of responding to bid P_i over the announcement period is small. To begin with,

we give a limit theory as follows.

Lemma 1: Let X be a binomial random variable with parameters (n, p) , then X approached to Poisson random variable with parameter $\lambda = np$ as n gets large and p gets small.

Proof: since X is binomial, we have

$$\begin{aligned}
\Pr(X = i) &= \frac{n!}{i!(n-i)!} p^i (1-p)^{n-i} \\
&= \frac{n(n-1)\cdots(n-i+1)}{i!} \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i} \\
&= \frac{n(n-1)\cdots(n-i+1)}{n^i} \frac{\lambda^i}{i!} \left(1 - \frac{\lambda}{n}\right)^n / \left(1 - \frac{\lambda}{n}\right)^i
\end{aligned}$$

Then, for a large enough n and a small enough p , we have the following limits,

$$\lim_{n \rightarrow \infty} \frac{n(n-1)\cdots(n-i+1)}{n^i} = 1, \quad \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda},$$

$$\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^i = 1.$$

$$\text{Finally, } \Pr(X = i) = \frac{\lambda^i}{i!} e^{-\lambda}. \text{ (Q.E.D.)}$$

In view of Lemma 1, for a large n and small p_i , the number of responding bidders within a unit time interval is a Poisson random variable with arrival rate $\lambda_i = n \cdot p_i$. It follows that the number of responding

bidders within the announcement period T , $N_i(T)$ follows a Poisson distribution, that is $Poisson(\lambda_i T)$. Since $\lambda_i = n \cdot p_i$, therefore $N_i(T) \sim Poisson(n p_i T)$. Let further $\alpha = nT$, then

$$N_i(T) \sim Poisson(\alpha p_i) \tag{23}$$

Accordingly,

$$q_i = \Pr(N_i(T) = 0) = \exp(-\alpha p_i) \tag{24}$$

and $\Pr(N_i(T) > 0) = 1 - \exp(-\alpha p_i)$. Substituting Equation (24) into Equation (21), we have

$$\begin{aligned}
\Pr(B = P_l, F = P_m | S = P_k) \\
= \exp(-\alpha p_{m+1} - \alpha \sum_{i=l+1}^k p_i + \sum_{j=l}^m \ln(1 - \exp(-\alpha p_j)))
\end{aligned} \tag{25}$$

Substituting Equation (25) into Equation (22), we finally have

$$R(P_k) = \sum_{l=1}^k \sum_{m=l}^{\infty} P_m \exp(-\alpha p_{m+1} - \alpha \sum_{i=l+1}^k p_i + \sum_{j=l}^m \ln(1 - \exp(-\alpha p_j))) \tag{26}$$

Figure 4 depicts the sketch of the functional relationship between the expected revenue and the starting bid. Graphically, the peak point in the curve represents the optimal starting bid as well as its corresponding expected revenue.

4. Duration of Oral Auction

In this section, we consider the expected length of the oral auction. In practice, each bid is announced for at most 3 times. For each bid, let T_1 , T_2 and T_3 denote

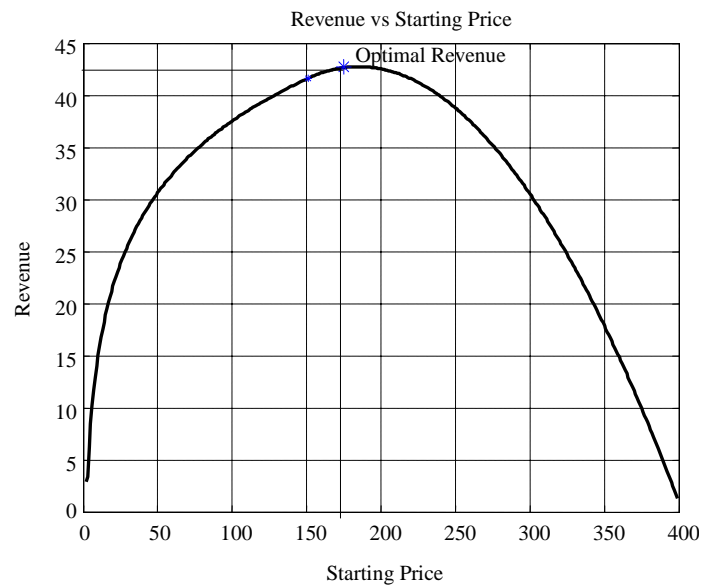


Figure 4. Revenue vs starting bid

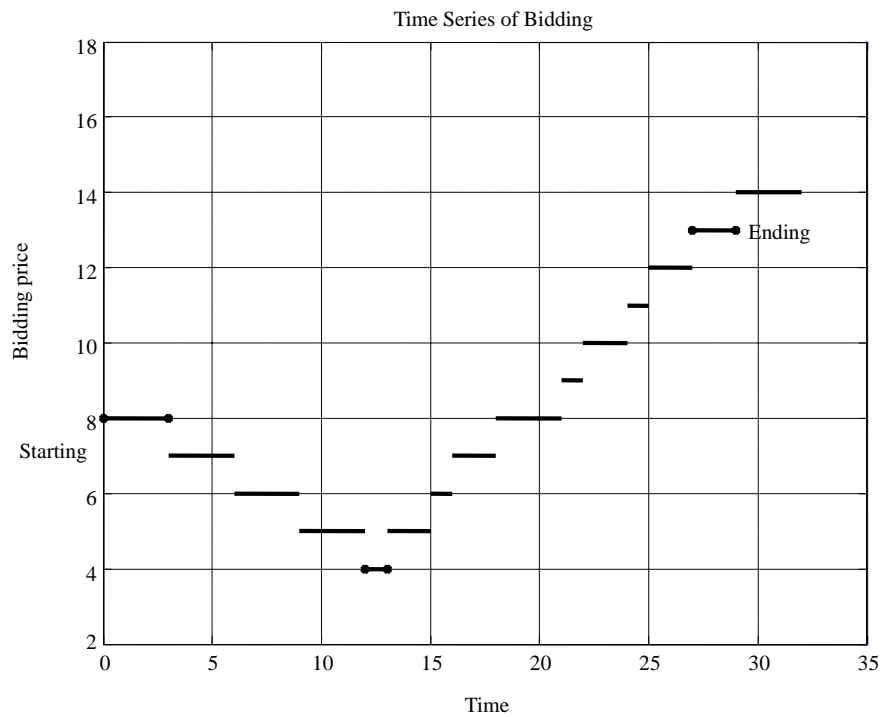


Figure 5. Time evolution of a bidding process

the time length of the periods between the announcements of the bid beginning, “going once” “going twice” and “gone”, respectively. Let $T = T_1 + T_2 + T_3$. For example, when the auctioneer announces the bid P_i , if there is no response up to T_1 , that is $N_i(T_1) = 0$, then the auctioneer announces “going once”; if there is still no

response up to $T_1 + T_2$, that is $N_i(T_1 + T_2) = 0$ then “going twice” will be announce; if there is no response up to $T_1 + T_2 + T_3$, that is $N_i(T) = 0$, then “gone” is announced. Figure 5 depicts a sample path of the time evolution of the bidding process in term of a step function.

Let X_i denote the time length for bid P_i , where X_i can take value of T_1 , $T_1 + T_2$, or $T_1 + T_2 + T_3$. For any bid P_i , let $q_{i,j}$ denote the probability of there is no response during period T_j , where $j = 1, 2, 3$, that is

$$q_{i,j} = \mathbf{P}(N_i(T_j) = 0) \quad (27)$$

Lemma 2. For any bid P_i , the following holds

$$\begin{aligned} \mathbb{E}[X_i \mid N_i(T) > 0] \\ = \frac{T_1(1 - q_{i,1}) + (T_1 + T_2)q_{i,1}(1 - q_{i,2}) + Tq_{i,1}q_{i,2}(1 - q_{i,3})}{1 - q_i} \end{aligned} \quad (28)$$

Proof. The conditional probabilities are given as

$$\begin{aligned} \mathbf{P}(X_i = T_1 \mid N_i(T) > 0) &= \frac{\mathbf{Pr}(N_i(T_1) > 0)}{\mathbf{Pr}(N_i(T) > 0)} = \frac{1 - q_{i,1}}{1 - q_i} \\ \mathbf{P}(X_i = T_1 + T_2 \mid N_i(T) > 0) &= \frac{\mathbf{Pr}(N_i(T_1) = 0) \mathbf{Pr}(N_i(T_2) > 0)}{\mathbf{Pr}(N_i(T) > 0)} \\ &= \frac{q_{i,1}(1 - q_{i,2})}{1 - q_i} \end{aligned}$$

and

$$\begin{aligned} \mathbf{P}(X_i = T_1 + T_2 + T_3 \mid N_i(T) > 0) \\ = \frac{\mathbf{Pr}(N_i(T_1) = 0) \mathbf{Pr}(N_i(T_2) = 0) \mathbf{Pr}(N_i(T_3) > 0)}{\mathbf{Pr}(N_i(T) > 0)} \\ = \frac{q_{i,1}q_{i,2}(1 - q_{i,3})}{1 - q_i} \end{aligned}$$

Then, the proof is completed by the definition of conditional expectation. (Q.E.D)

Let $T_{k \rightarrow l \rightarrow m}$ denote the time length of the bidding process with the starting bid $S = P_k$, the lowest bid $B = P_l$ and the final bid $F = P_m$. Thereby

$$T_{k \rightarrow l \rightarrow m} = \sum_{i=k}^{l-1} X_i + \sum_{j=l}^m X_j + X_{m+1} \quad (29)$$

Note that there is no response for any bid along the descending-bid process. Then, Equation (29) can be simplified as

$$T_{k \rightarrow l \rightarrow m} = (k - l + 1)T + \sum_{j=l}^m X_j \quad (30)$$

Therefore, the duration for the auction with starting bid P_k is,

$$D(P_k) = \mathbb{E} \left[\mathbb{E} \left[T_{k \rightarrow l \rightarrow m} \mid S = P_k, B = P_l, F = P_m \right] \right]$$

$$\begin{aligned} &= \mathbb{E}_{B=P_l} \left[\mathbb{E}_{F=P_m} \left[(k - l + 1)T + \sum_{j=l}^m X_j \mid S = P_k, B, F \right] \right] \\ &= \mathbb{E}_{B=P_l} \left[(k - l + 1)T + \mathbb{E}_{F=P_m} \left[\sum_{j=l}^m X_j \mid N_j(T) > 0, B, F \right] \right] \\ &= \sum_{j=l}^m \sum_{l=1}^k \mathbf{Pr}(B = P_l, F = P_m \mid S = P_k) \\ &\quad \left[(k - l + 1)T + \sum_{j=l}^m \mathbb{E}[X_j \mid N_j(T) > 0] \right] \\ &= \sum_{j=l}^m \sum_{l=1}^k \left[(\mathbf{M}_-)^{k-l-1} \right]_{l(l+1,k)} \left[(\mathbf{M}_+)^{m+1-l} \right]_{l(l,m+1)} q_{m+1} \\ &\quad \left[(k - l + 1)T + \sum_{j=l}^m \mathbb{E}[X_j \mid N_j(T) > 0] \right] \end{aligned} \quad (31)$$

where $\mathbb{E}[X_j \mid N_j(T) > 0]$ is given by Equation (28).

There, the first equation above holds by the definition and the conditional expectation. The second equation holds by Equation (30). The third equation holds follow-

ing the definition of conditional expectation. The last equation follows Equation (21).

5. Conclusions and Discussions

This paper studies the revenue and duration of an oral auction, which has a hybrid structure of English auction and Dutch auction. Our effort is to derive the revenue and duration of the auction as a function of the starting bid. To this end, we decomposed the bidding process into two phases: English auction (descending-bid) phase and the Dutch auction (ascending-bid) phase. For each phase, we first gave the one-step transition matrix and the formula for revenue and duration are obtained consequently. For an oral auction with a large number of bidder and each bidder behaves independently, we also derived the limit results of the expected revenue and duration.

The results obtained can be implemented in practice directly. In particular, the probability of bidder responding to a bid can be statistically estimated from the observed data. Therefore, the one-step Markov transition matrix can be computed directly. The one-step transition matrix for each phase can be used to compute the expected revenue and duration. From the seller's perspective, the optimal starting bid is of great interest and it can be obtained numerically by some basic searching algorithm. With the formula for revenue and duration, we may take their ratio to evaluate the efficiency of the oral auction. This ratio accounts for the revenue as well as the time, and thereby provides a comprehensive evaluation.

Our model is formulated based on Markov assumption, that is, the bidder behaves only according to a function of the bid level. It does not depend on the bidder's previous behavior as well as the other bidders' behavior. Although Markov process models provide a mathematical approach to predict online auction prices, estimating parameters of a Markov process model in practice is a challenging task. For example, S. Chou, *et al.* propose a simulation-based model as an alternative approach to predict the final price in online auctions [15]. To study the oral auction with bidder inter-dependent behavior, we can extend our model to a multi-space Markov model, in which each state space represents the bidding price for each bidder. This leads to a new topic of further research.

It is commonly assumed that the customer behavior is exogenous. For example, market size is often represented using a demand distribution (e.g., the newsvendor model). However, in our real world of oral auction, all bidders do, at some point, actively evaluate alternatives and make choices. This suggests that bidders' decision is jointly effected together. Thereby, "customer behavior" should be introduced to auction design. In our view, it is impor-

tant to adopt a micro-perspective on such bidding interactions. This requires a high-resolution lens to zoom in on the incentives and decision processes of bidders at their individual level.

REFERENCES

- [1] C. W. Smith, "Auctions: The social construction of value," University of California Press, Berkeley, Los Angeles, 1990.
- [2] W. Vickrey, "Counter speculation, auctions, and competitive sealed tenders," *Journal of Finance*, Vol. 16, pp. 8–37, 1961.
- [3] N. Buchbinder, K. Jain, and J. Naor, "Online primal-dual algorithms for maximizing ad-auctions revenue," *Lecture Notes in Computer Science*, 4698, pp. 253–264, 2007.
- [4] G. Vulcano, G. V. Ryzin, and C. Maglaras, "Optimal dynamic auctions for revenue management," *Management Science*, Vol. 48, No. 11, pp. 1388–1407, 2002.
- [5] E. J. Pinker, A. Seidmann, and Y. Vakrat, "Managing online auctions: Current business and research issues," *Management Science*, Vol. 49, No. 11, pp. 1457–1484, 2003.
- [6] Z. J. Shen and X. Su, "Customer behavior modeling in revenue management and auctions: A review and new research opportunities," *Production and Operations Management*, Vol. 16, No. 6, pp. 713–728, 2009.
- [7] K. Hendricks and H. J. Paarsch, "A survey of recent empirical work concerning auctions," *Canadian Journal of Economics*, Vol. 28, No. 2, pp. 403–426, 1995.
- [8] J. Laffont, "Game theory and empirical economics: The case of auction data," *European Economic Review*, Vol. 41, No. 1, pp. 1–35, 1997.
- [9] R. L. Zhan and Z. J. Shen, "Optimality and efficiency in auctions design: A survey," *Working Paper*, 2006.
- [10] S. M. Ross, *Stochastic Processes*, 2nd Edition, Wiley, 1995.
- [11] S. Park, E. H. Durfee, and W. P. Birmingham, "Use of Markov chains to design an agent bidding strategy for continuous double auctions," *Journal of Artificial Intelligence Research*, Vol. 22, pp. 175–214, 2004.
- [12] A. Segev, C. Beam, and J. Shanthikumar, "Optimal design of internet-based auctions," *Information Technology and Management*, Vol. 2, pp. 121–163, 2001.
- [13] D. Reiley, D. Bryanz, N. Prasad, and D. Reevesz, "Pennies from eBay: The determinants of price in online auctions," *The Journal of Industrial Economics*, Vol. 55, pp. 223–233, 2007.
- [14] C. Mason, "The Art of the steal: Inside the Sotheby's-Christie's auction house scandal," *Berkley Trade*, 2005.
- [15] S. Chou, C. S. Lin, C. Chen, T. R. Ho, and Y. C. Hsieh, "A simulation-based model for final price prediction in online auctions," *Journal of Economics and Management*, Vol. 3 No. 1, pp. 1–16, 2007.

Informal Financing of Small – Medium Enterprise Sector: The Case of Greece

Panagiotis Petrakis^{1,*}, Konstantinos Eleftheriou²

¹Department of Economics, University of Athens, Stadiou Street, Athens, Greece; ²Department of Economics, University of Piraeus, Karaoli & Dimitriou Street, Piraeus, Greece.
Email: ppetrak@econ.uoa.gr, kostasel@otenet.gr

Received July 17, 2009; revised August 23, 2009; accepted September 29, 2009.

ABSTRACT

In this paper, we attempt to find a “channel” through which Greek economy can exhibit a relative “resistance” in a credit crunch. For this purpose, we specify an error correction model so as to test the relationship between corporate bank loans and commercial papers comprised of post-dated cheques and bills of exchange. The results show that corporate bank loans and cheques - bills of exchange are substitutes. This finding combined with the fact that in Greece, the issuance of these papers is positively connected with the informal economic activity which in turn rises during economic downturns, has a strong economic implication regarding the ability of Greek economy to partly “amortize” the shocks connected with the current financial crisis.

Keywords: Corporate Finance, Credit Crunch, Shadow Financing

1. Introduction

Is there an interrelation between bank loans and commercial papers (cheques, bills of exchange) as a source of external debt financing for firms in Greek economy, and if yes, are they substitutes or complements? Which is the economic intuition between such an interrelation and can it offer a safety net to the current credit crunch? These are the main crucial questions we try to answer in this paper.

One of the main factors which determine the level of “resistency” of an economy in a bank credit crunch is the ability of the economic system to create multiple “channels” of financing and exploit them properly. In modern economies, firms have a variety of debt financing tools at their disposal. However, each of these tools has a different rank in firm’s preferences. According to the traditional “pecking order” hypothesis of corporate finance [1], borrowing firms prefer to finance their debts through external resources (securities, bank loans) rather than equity issuance. Equity issuance is less preferred since the funds it provides are generally limited by the scale of expenditures (dividends) and it is considered by investors, as a “bad”

signal for the economic performance and viability of the firm. Hence, firms mainly choose between bank borrowing and debt securities issuance, when it comes to finance their corporate expenditures. Greenspan [2,3], emphasized the importance of such a choice under a credit crisis regime. More specifically, he suggested that there is a rate of substitutionality between the market of bank loans and that of bonds which smoothes the negative impact that a financial crisis has on real economy. On the other hand, Holmstrom and Tirole [4], stressed that “multiple avenues of intermediation” (availability of the aforementioned sources of external debt financing) for corporations are characterized by complementarity. Their analysis is based on a principal-agent problem with monitoring costs. When the supply of intermediary capital falls due to a credit crunch, the quantity of informed (banks) finance which is available to firms decreases. This also means that less uninformed (securities) finance can be attracted since the level of monitoring undertaken is lower¹. The findings of Holmstrom and Tirole [4], were empirically verified for U.S. economy by Davis and Ioannidis [5].

Gertler and Gilchrist [6], argued that the salutary effects stemming from the substitutionality between the main alternative “channels” of corporate debt financing are limited when the market is dominated by small firms. This occurs, since large firms have access to short-term sources of credit (e.g. commercial papers market) unavailable to

^{*}This paper is based on an ongoing research project titled: “Economic Growth and Development in the Greek Economy”. We would like to thank an anonymous referee for useful comments and suggestions. Any remaining errors are ours.

¹Uninformed investors are less willing to offer their funds when the level of monitoring connected with the informed finance is low.

Table 1. Greek commercial papers in circulation and bank finance (in million euros)

Year	Bounced cheques	Unpaid bills of exchange	Total	Credit delinquency rates (%)	Nominal estimated amount of cheques and bills of exchange in circulation	Yearly adjustment of (D)	Nominal domestic MFI loans to domestic enterprises
	(A)	(B)	(A)+(B)	(C)	(D)	(E)=(D)/3	(F)
2004	1024.8	169.2	1194	2.64	45227.3	15075.8	4587
2005	1464.4	180.7	1645.1	3.04	54115.1	18038.4	5716.6
2006	1202.1	188.1	1390.2	4.02	34582.1	11527.4	5376.9
2007	921.9	177.5	1099.4	3.57	30795.5	10265.2	13095.3
2008	1291.3	170.2	1461.5	3.73	39182.3	13060.8	15488.7

small firms and therefore they respond more effectively to a cash flow squeeze. In Greece, the market of commercial papers as a source of short-term financing has not been adequately developed. Instead, there is a market of post-dated cheques² and bills of exchange. The maturity period of a post-dated cheque is not the date of issue but the due date specified by the drawer.

Transactions through post-dated cheques involve high risk for the payee. Therefore, the operation of this “quasi-commercial” papers market is based on long-term relationships (mutual trust) between engaged parties. This characteristic can be proven quite beneficial for an economy during business cycle downturns. According to a survey conducted by International Monetary Fund in 2006 [7], countries with a higher degree of relationship-based lending (low degree of arm’s length transactions) may experience a less sharp decrease in the level of nonresidential business fixed investments during a downward phase of the business cycle. The rationale behind this conclusion is that the lender gives a greater weight to the long-run gains from maintaining an existing relationship with a borrower and thus he provides a short-term assurance that financing will be available in case of a credit crisis. Another advantage of the market of post-dated cheques and bills of exchange compared with the traditional market of commercial papers is that small firms have access to it. One more interesting feature of the Greek market of commercial papers is its positive relation with the size of shadow economy. The fact that post-dated

cheques can be endorsed and transferred by the payee means that the “traces” of a transaction cannot be tracked very easily by tax authorities. Hence, firms have an incentive to evade taxes by issuing iconic invoices. A recent work by Schneider [8], shows that Greece had and still has the largest informal economy between 21 OECD countries over the last twenty years³.

This result is an indication for the expected large size of the Greek “quasi-commercial” papers market. The above analysis implies that if there is substitutability between bank loans and post-dated cheques and bills of exchange, then Greek economy may have an arrow left in its quiver against the current financial crisis.

The rest of the paper is organized as follows. In the next section we set out our empirical methodology and give our main empirical results. Section 3 concludes.

2. Quantitative Analysis

In order to conduct our analysis, we obtained monthly data over 2004–2008 (more precisely from 2004/07 to 2008/12) for: 1) bounced cheques and unpaid bills of exchange (in million euros) from Hellenic Credit Profile Databank (Tiresias Bank Information Systems S.A.), 2) Consumer Price Index (CPI) from General Secretariat of National Statistical Service of Greece, 3) outstanding balances (in million euros) of domestic Monetary Financial Institutions (MFI) loans to domestic enterprises and 4) interest rates on euro-denominated loans without a defined maturity by domestic MFIs to euro area non-financial corporations. Data for 3) and 4) were obtained from Bank of Greece (Bulletin of Conjunctural Indicators).

Moreover, we used firms’ credit delinquency rates from ICAP Group, in order to calculate the value of commercial papers (cheques and bills of exchange) in circulation⁴.

⁴We assume that firms’ credit delinquency rate is a good approximation of the percentage of bounced cheques and unpaid bills of exchange to commercial papers in circulation.

²Post-dated cheques facilitate the interindustry financial relations without being recognized as a formal financial instrument, since cheques are officially defined as a bill of exchange payable on demand. This kind of financial instruments usually “covers” under-the-table real sector financial transactions (shadow economy transactions).

³Schneider used the Multiple industries and multiple courses procedure (MIMIC) (for an overview see Aigner *et. al* [9]) and currency demand approach (see Schneider [10]) in order to obtain his estimates about the size of the shadow economy.

Table 2. Summary statistics: Monthly *RL*, *RCP* and *RR* data from July 2004 to December 2008

Summary statistics	<i>RL</i>	<i>RCP</i>	<i>RR</i>
No. of observations	54	54	54
Mean	0.746884	1.135214	0.038768
Median	0.620918	1.020567	0.038218
Maximum	4.184875	3.069930	0.052121
Minimum	-2.276165	0.608280	0.026796
Standard deviation	1.102762	0.450345	0.006904
Skewness	0.409230	1.772342	0.158600
Kurtosis	4.146723	7.768323	2.069039
Jarque-Bera	4.465909	79.42880	2.176434
J-B P-value ^ν	0.107211	0.000000	0.336817

^ν J-B P-value is the probability that a Jarque-Bera statistic exceeds (in absolute value) the observed value under the null hypothesis of a normal distribution. The negative minimum value of *RL* implies that the amount of new loans given is less than the part of the past loans which are paid off.

Table 3. Stationarity tests

	ADF test (lags)	
Variables in levels		
<i>RL</i>	-2.93	(2)
<i>RCP</i>	-0.52	(5)
<i>RR</i>	-2.59	(3)
Variables in first difference		
<i>RL</i>	-12.14***	(1)
<i>RCP</i>	-5.92***	(2)
<i>RR</i>	-6.65***	(0)

Notes: Boldface values denote sampling evidence in favour of unit roots. ***Signifies rejection of the unit root hypothesis at the 1% level of significance. The numbers in parentheses for the ADF test are the optimal lag lengths, which are determined using AIC (Akaike Information Criterion). Trend and constant were included in the test equation.

Three variables are constructed from the above data: *RL*, *RCP* and *RR*, where *RL* and *RCP* are the real new domestic MFI loans to domestic enterprises (this variable is constructed by taking the first differences of 3) and deflating by the CPI) and the real estimated amount of new cheques and bills of exchange issued each month⁵ respectively. *RR* denotes the real interest rate on *RL* (*RR* is derived by subtracting inflation rate from 4). Inflation rate is derived by the use of CPI). The variables *RL* and *RCP* are expressed in billion euros. In Table 1, we present a comparison of the two alternative sources of firm's financing examined in this paper; bank loans and the "parallel financial system" of posted-dated cheques and bills of exchange. As we note from Table 1, the Greek "quasi-commercial" papers market plays an important role (almost the same as bank loans) in corporate financing. Moreover,

⁵We get *RCP* by dividing 1) with the product of firms' credit delinquency rates and CPI and adjusting the result in yearly basis (the amount of cheques and bills in circulation needs to be adjusted in yearly basis, since it is a common practice for the bearer of a cheque to accept a new cheque (bill) of the same amount instead of cashing it in, when the maturity period ends. Hence, if we assume that the mean maturity period of a post-dated cheque (bill) is four months, we have to divide the total amount of Greek commercial papers by three).

it can be easily ascertained from Table 1, that the most important component of the Greek commercial papers market is that of cheques.

Table 2 shows the descriptive statistics of the variables under consideration. For *RL* and *RR* series displayed in Table 2, we do not reject the hypothesis of normal distribution at the 10% significance level. The first step of our analysis is to test whether *RL*, *RCP* and *RR* are stationary. Table 3 reports unit root test statistics of the augmented Dickey and Fuller test [11]. The results in Table 3, indicate that all series are non-stationary and contain a unit root. In order to examine whether they are integrated of order one, *I*(1), we perform the augmented Dickey-Fuller (ADF) test on first differences. The results suggest that all variables are stationary in first differences.

Engle and Granger [12] argued that even if a set of economic series is not stationary, there may exist some linear combinations of the variables that are stationary. If the separate series are *I*(1) (i.e. non-stationary in their levels but stationary in their first differences) but a linear combination of them is *I*(0), then these series are cointegrated. If series are cointegrated, an error correction model (ECM) is appropriate for modeling their relation, as suggested by Engle and Granger. More specifically, if a

group of variables is non-stationary (random walks), then by regressing one variable against the others can lead to spurious results in the sense that conventional significance tests will tend to indicate a relationship between the variables when in fact none exists. This problem can be solved if we use in our modeling process the first differences of the above variables after verifying that these differences are stationary (integrated of order one, I(1), variables). However, even though this approach is correct in the context of univariate modeling [e.g. Autoregressive – Moving Average (ARMA) processes], it is inadvisable when we try to examine the relationship between variables. The main drawback of this, in other respects, statistically valid approach is that it has no long-run solution (common problem in pure first difference models). More specifically, one definition of the long run that is employed in econometrics implies that variables have converged upon some long term values and are no longer changing. Hence, all the first difference terms will be zero and by simply regressing the one against the others gives results which say nothing about whether the variables under consideration have an equilibrium relationship. However, this problem can be overcome by using a combination of the first differenced and lagged levels of cointegrated variables. This formulation is known as an error correction model. Through this model, we can examine the short run dynamic relationship between the variables under consideration by taking into account their deviations from their equilibrium/long run relationship (residuals of the cointegrating regression).

In order to test for cointegration, we use the maximum likelihood methodology proposed by Johansen [13]. According to Johansen a Vector Autoregression (VAR)

model of order p , can be written as follows:

$$\Delta Y_t = \mu + \Gamma(L)\Delta Y_{t-p} + \Pi Y_{t-1} + u_t \quad (1)$$

where Y_t is the vector of RL , RCP and RR , $\mu = (\mu_1, \mu_2, \mu_3)'$, $\Gamma(L)$ is a polynomial of order $p-1$, u_t is a vector of independent Gaussian errors with zero mean and covariance matrix Ω , Δ is the first difference operator and Π is a matrix of the form $\Pi = \alpha\beta'$ where α and β are $3 \times r$ matrices each with rank r , with β being the matrix of the r cointegrating vectors (i.e. the columns of β represent the r cointegrating relations) and α being the matrix of adjustment coefficients. As stated above, the existence of cointegration has implications about the way we should model the relationships between RL and RCP , RR .

The results of Johansen cointegration test are presented in Table 4. Since Johansen's procedure is sensitive to the lag length of the Vector Autoregression (VAR) (Banerjee *et al.* [14]), we determine the lag length by using the appropriate criteria.

The max eigenvalue statistic supports the existence of one cointegrating vector. More specifically, the cointegrating equation is:

$$RL = -1.23RCP - 35.52RR \quad (2)$$

This finding implies that there is a long run equilibrium relationship between RL , RCP and RR . Equation (2) indicates that the real value of newly issued Greek commercial papers in circulation and the real interest rates of corporate bank loans are negatively correlated to the real amount of corporate bank loans (in the long-run), with the estimated coefficients of -1.23 and -35.52, respectively.

Table 4. Johansen cointegration test

Hypothesized # of cointegrated equations (r)	Max eigenvalue statistic	Critical values at 5%
None	31.99**	21.13
At most 1	4.45	14.26
At most 2	2.88	3.84

Note: ** Indicates the rejection of the hypothesis about the number of cointegrated equations at the 5% level. The sequential modified LR test statistic, the final prediction error (FPE) and the Schwarz information criterion indicate that the optimal lag length of the VAR is equal to one. Moreover, the VAR residual Portmanteau test for autocorrelations does not reject the null hypothesis of no residual autocorrelations.

Table 5. TSLS estimates of the ECM

Parameter	Coefficient	t-value
μ	-0.075	-0.364
γ_0	-2.66**	-2.049
δ_0	-56.92	-0.39
θ_1	-1.2***	-7.026
R^2	0.56	

Note: *** and ** denote statistical significance at 1% and 5%, respectively. The following series were used as instruments: ECT_{t-1} , ECT_{t-6} , ECT_{t-12} , ΔRCP_{t-1} , ΔRCP_{t-4} , ΔRL_{t-12} , ΔRR_{t-1} , ΔRR_{t-12} , constant.

Table 6. Diagnostic tests

	Value of test statistic	P-value
<i>JB</i>	5.295	[0.071]
<i>Reset test</i>	0.455	[0.504]
Hansen's <i>J</i> statistic test	5.449	[0.363]
<i>LM</i> ₁ / <i>LM</i> ₂ test	0.071/1.212	[0.789/0.545]
<i>BDS test</i>		
Dimension		
2	-0.0053	[0.7908]
3	-0.01389	[0.2847]
4	-0.00601	[0.5855]
5	-0.00222	[0.8633]
6	-0.00329	[0.3745]

Note: Figures in brackets represent asymptotic *P* – values associated with the tests. *JB* denotes the Jarque-Bera normality test of errors. The Reset test tests the null hypothesis of functional form misspecification. *LM*₁ / *LM*₂ is the Lagrange multiplier test for first and second order serial correlation (under the null there is no serial correlation in the residuals up to the specified order). Hansen's *J* statistic test is a general version of the Sargan test, a test of overidentifying restrictions (under the null hypothesis the overidentifying restrictions are satisfied). For the relationship between *J* statistic and Sargan test see Murray [15]. Finally the BDS [16] test tests the null hypothesis that the errors are independently and identically distributed (In our test we set the value of distance between the pair of the elements of a time series, equal to 0.7. Since our sample is relatively small, we use bootstrap *P* - values).

As *RL*, *RCP* and *RR* are cointegrated, it is necessary to specify an ECM in order to examine the short-run relationship of these variables. We specify an error correction model of the following type:

$$\Delta RL_t = \mu + \theta_1 ECT_{t-1} + \gamma_0 \Delta RCP_t + \delta_0 \Delta RR_t + \varepsilon_t$$

$$t = 1, 2, \dots, T \quad (3)$$

where *ECT* is the error correction term and ε_t is a disturbance term. Since ΔRCP_t and ΔRR_t current values appear in the above equation, Ordinary Least Squares (OLS) estimation produces inconsistent estimators. In order to overcome this problem, we apply a two stage least squares (TSLS) estimation procedure. Table 5 presents the TSLS estimates of Equation (3). Moreover, we check the specification of our estimated model by performing various diagnostic tests. These tests are reported in Table 6. Our results indicate that the ECM seems to be quite well specified and free from specification error.

As we note from Table 5, the coefficient of the *ECT* has the correct sign, is statistically significant and is rather large indicating rapid adjustment of *RL*, *RCP* and *RR* to the proceeding imbalance (ECT_{t-1}) in a short run period. In order to investigate the relationship between the provision of corporate bank loans and issuance of cheques and bills, we turn our attention to the short-run elasticity.

Table 5, indicates that γ_0 is negative and statistically significant. More specifically, if the issuance of new Greek commercial papers increases by 100 million euros then the provision of corporate bank loans decreases, ceteris paribus, by 266 million euros within a month. This finding implies that there is substitutability between bank

loans and post-dated cheques and bills of exchange. If we take into consideration that during a credit crunch, there are bank credit shortages, there is a tendency for informal economy to increase [17] and the fact that (as mentioned above) there is a positive relation between the size of Greek commercial papers market and that of shadow economy, then we conclude that Greek economy may alleviate the negative impact of economic downturns caused by financial crisis.

3. Conclusions and Policy Implications

In our analysis, we showed that the Greek market of cheques and bills of exchange can serve as a substitute for bank loans. By combining this result with the distinctiveness in the structure of the commercial papers market in Greece, which allows small firms to have access to short-term funding and the close connection of the size of this market with that of informal economy, we can argue that Greek economy can partly “amortize” the shocks connected with the current financial crisis. Thus, monetary policy exhibits only indirect effects on real enterprise sector. In case of an interest rate fall, someone would expect that the positive effects will find the way out to the real sector. On the other hand, when interest rates rise during a credit crunch due to the segmentation of the financial sector (low interbank and interindustry trustiness), the enterprise sector will substitute the absence of financial credit with interindustry financing. Therefore, the IMF argument, that economies with low degree of arm's length transactions can smooth the negative credit crunch effects and regenerate economic activity during the easing of a credit crisis, is confirmed.

The main deficiency of our analysis is that, although we

have developed our arguments about the positive relation between Greek market of cheques and bills of exchange and informal economic activity by citing the appropriate references, we have not explicitly included the underground economy in our analysis. The estimation and the inclusion of a variable indicating the size of the informal economy in our model can further enhance the robustness of our analytical results. Moreover, the use of dummy variables which will capture the relevant effects during periods of economic turbulence and the expansion of our dataset so as to include the latest available data, will also reinforce our conclusions. All these issues can be considered as topics for future research.

REFERENCES

- [1] S. Myers, "The capital structure puzzle," *Journal of Finance*, Vol. 39, pp. 575–592, 1984.
- [2] A. Greenspan, "Do efficient financial markets mitigate financial crises?" Speech to the Financial Markets Conference of the Federal Reserve Bank of Atlanta, 1999.
- [3] A. Greenspan, "Global challenges," Speech to the Financial Crisis Conference, Council on Foreign Relations, New York, 2000.
- [4] B. Holmstrom and J. Tirole, "Financial intermediation, loanable funds, and the real sector," *The Quarterly Journal of Economics*, Vol. 112, pp. 663–691, 1997.
- [5] P. Davis and C. Ioannidis, "External financing of US corporations: Are loans and securities complements or substitutes?" *Economics and Finance Discussion Papers*, Brunel University, 2004.
- [6] M. Gertler and S. Gilchrist, "Monetary policy, business cycles, and the behavior of small manufacturing firms," *The Quarterly Journal of Economics*, Vol. 109, pp. 309–340, 1994.
- [7] World Economic Outlook, "How do financial systems affect economic cycles?" *International Monetary Fund*, 2006.
- [8] F. Schneider, "The size of the shadow economy in 21 OECD countries (in % of 'official' GDP) using the MIMIC and currency demand approach: From 1989/90 to 2009," Working paper, Johannes Kepler University, 2009, (http://www.economics.unilinz.ac.at/members/Schneider/files/publications/ShadowEconomy21OECD_2009.pdf).
- [9] D. Aigner, F. Schneider, and D. Ghosh, "Me and my shadow: Estimating the size of the US hidden economy from time series data," In W. A. Barnett, E. R. Berndt, and H. White, (edition), *Dynamic Econometric Modeling*, Cambridge University Press, Cambridge (Mass.), pp. 224–243, 1988.
- [10] F. Schneider, "Size and measurement of the informal economy in 110 countries around the world," *The World Bank: Rapid Response Unit*, Washington D.C., 2002.
- [11] D. A. Dickey and A. M. Fuller, "Likelihood ratio statistics for autoregressive time series with a unit root," *Econometrica*, Vol. 49, pp. 1057–1072, 1981.
- [12] R. F. Engle and C. W. G. Granger, "Co-integration and error correction representation, estimation and testing," *Econometrica*, Vol. 55, pp. 251–276, 1987.
- [13] S. Johansen, "Statistical analysis of cointegration vectors," *Journal of Economic Dynamics and Control*, Vol. 12, pp. 231–254, 1988.
- [14] A. Banerjee, J. J. Dolado, F. D. Hendry, and G. W. Smith, "Exploring equilibrium relations in econometrics through state models: Some monte carlo evidence," *Oxford Bulletin of Economics and Statistics*, Vol. 48, pp. 253–278, 1986.
- [15] M. Murray, "Avoiding invalid instruments and coping with weak instruments," *Journal of Economic Perspectives*, Vol. 20, pp. 111–132, 2006.
- [16] W. A. Brock, W. Dechert, and J. Scheinkman, "A Test for independence based on the correlation dimension," *Econometrics Reviews*, Vol. 15, pp. 197–235, 1996.
- [17] *The Economist*, "The black market: Notes from the underground", 2 April 2009.

Knowledge Innovative Organization: The Effect of Constant Organization Renewal

Abu Saleh Md. Sohel-Uz-Zaman¹, Umana Anjalin²

¹United International University, Dhaka, Bangladesh; ²The University of Asia Pacific, Dhaka, Bangladesh.
Email: sohelzaman@gmail.com, umana@uap-bd.edu

Received May 28, 2009; revised July 7, 2009; accepted August 14, 2009.

ABSTRACT

The term 'knowledge' is getting increased importance as the economy is becoming progressively knowledge intensive. Contemporary view of innovation management is acknowledging knowledge as the most effective source of innovation. Success of the business depends on how successfully knowledge is used for innovation. As a result, organizations are increasingly embracing knowledge strategy for their innovation scheme. Incorporating knowledge in the innovation process; building innovative capabilities and identifying essential features of a knowledge innovative organization are the key concerns of this study. Knowledge is a context specific notion. In this case study, the use of knowledge for innovation has been analyzed from an individual organization point of view so that contextual elements of a knowledge innovative organization can be identified more precisely along with the institutional implications. Thus, this article would be able to provide a good number of meaningful indications in using knowledge for innovation; it would also shed light on constant capability building for innovation; nurturing of creativity and developing of knowledge infrastructure of a knowledge innovative organization. In addition, this study will be revealing the true nature of knowledge strategy and its application in the context of a developing country.

Keywords: Innovation, Knowledge, Knowledge-Innovative Organization (KIO), Knowledge Strategy, Competitive Advantage, Capacity Development

1. Introduction

Nowadays innovation is one of the mostly used business strategies. In the knowledge era, organizations are acknowledging the necessity of knowledge in their business strategy and attempting to incorporate knowledge in their innovation process. In the competitive global markets, firms are trying to satisfy their customers with their creativity and innovation. Organizations are in an endless struggle of differentiating themselves by continuously introducing new products. The irony is that the market is saturated with many similar (me, too) products. Consumers do not find any significant difference. They (consumers) are indifferent to the products. Relentless effort for innovation, in one hand, is flooding the market with similar products or closely resembled ones and on the other hand, a huge amount of financial and other resources (e.g. physical, technological, intellectual, etc.) are being drained in promoting these products. Firms are being deprived from gaining any benefit from their innovation. In these circumstances, a newer perspective has emerged to gain maximum benefit from both the concepts – in

novation and knowledge. This new perspective can be termed as 'knowledge innovation'. Knowledge is the key driver for innovation and organization renewal in the form of capacity development.

2. Objective, Scope and Importance of the Study

Innovation is not merely a strategy; it is the result of total organizational effort. Similarly knowledge is a holistic phenomenon. The concept of knowledge innovation conceives both the ideas: knowledge and innovation in a common platform so that either knowledge can be used for innovation or innovation can be more intelligent with the help of knowledge. Indeed, there is a question about how these two concepts can be fitted together. In this connection, this case study will make an attempt to resolve this question in the light of an experience of a medium sized organization in a developing country's (Bangladesh) context.

The objectives of this study are to know –

- The process of incorporating knowledge into the in-

novation process;

- The method for developing innovative capability with the help of knowledge; and
- The essential factors that make organizations knowledge innovative.

In order to attain the above-mentioned objectives, a number of issues like business strategy, leadership and management style, culture, people management, work system, etc. are required to be examined critically so that a general understanding regarding a knowledge innovative organization can be made.

Managing knowledge for innovation is relatively a new idea. So far, most of the researches have been done in developed countries' scenario. In the era of globalization, competition is widespread; competitors are appearing from cross borders. Businesses of developing countries are facing severe competition. Local businesses are becoming more concerned for their survival. In order to fine-tune with the global business competitiveness, a number of organizations in many developing countries¹ (Brazil, India, China, etc.) are attempting to pursue knowledge strategy. To follow this trend of globalization, a number of organizations in Bangladesh are also trying to foster knowledge strategy in their business practices. In this study, an organization from Bangladesh has been drawn as sample to understand the application of knowledge in the typical indigenous condition.

It is commonly believed that business strategy is a deliberate effort. This is not always true in the case of knowledge strategy; it could be unconsciously followed by the companies. All the organizations, irrespective of industry, region and/or economy are repositories of a certain set of knowledge. This knowledge-base works as a strategy springboard in the form of culture, style, pattern, leadership and so on. Examining an organization's culture, work pattern, management and leadership style, etc. is also an important part of this case study.

3. Methodology

It has already been mentioned that this research has adopted the *case study* approach. Data and information of this study are basically primary in nature and have been analyzed qualitatively. This *exploratory study* has made an attempt to explore all the relevant issues with this study and address each of those accordingly. Although, this is a *snap-shot* study yet this short-term perspective did not hamper the reliability and dependability of the data and information, as well as the findings.

4. Managing Knowledge for Innovation

There is a new kind of influence in the global economy –

¹The Policy Forum on Using Knowledge for Development: a unique policy conference jointly organized by the World Bank Institute, the British Council and the OECD, 17 April 2001, Wilton Park, UK. Brazil, China and India participated in this conference in order to share knowledge strategies. www.oecd.org.

knowledge. Many scholars have indicated the importance of knowledge for organizational competitiveness and sustainability. Drucker comments “We are moving to a society in which the basic resource of economy is knowledge, instead of capital, labor and natural resources [1]”. Stewart believes that “Knowledge has become the preeminent economic resources – more important than raw material; more important, often than money. Considered as an economic output, information and knowledge are more important than automobiles, oil, steel, or any of the products of the industrial age [2]”. It (knowledge) is the most strategically important resources of the business [3]. Knowledge is being used more in product design and development, production and distribution. The competitive advantage of the firms can be undercut overnight due to the new knowledge of the competitors. Market leadership, size, name, recognition structure no longer guarantee the survival [4] even the skill and core competency cannot ensure that the firms will be able to capitalize on the new opportunity and tackle new challenges. It is a complete new world with new realities before the marketers.

In the knowledge era, managing knowledge can be an opportunity in new product development, i.e. innovation. There is a broad agreement in favor of using knowledge for innovation. In many instances knowledge and information have been used as major impetus of new product development [5]. According to Verona, the design of new product development work is anchored in knowledge management [6]. Innovation is frequently considered as the primary purpose of knowledge management [7].

Knowledge is also considered as the key impetus for new capability development. Knowledge strategy allows business firm to improve its learning capability to combine knowledge-based capability and make better use of them [8]. Another strategic aspect of knowledge strategy is the continuous interaction with the customers and other related parties. Garcia-Murillo and Annabi opine that customer knowledge is a part of organization knowledge [9]. Through close interaction with the customers, an in-depth understanding is developed which ultimately converts into organization's business knowledge. Constant interaction increases new customer knowledge-base.

5. Concerns in Using Knowledge for Innovation

There are few challenges in incorporating knowledge into the innovation process. Neither the literature of innovation management nor knowledge management has focused much on the process of incorporating knowledge in the innovation mechanism. Consequently, there is no clear direction for imparting knowledge into the innovation process; no precise knowledge strategy for innova-

tion; no suggestion regarding knowledge infrastructure and system.

There is a common understanding that innovation is the result of research and development activities. This is partially true. According to knowledge-based view, creativity and innovation can result from regular business activities and distributed organizational knowledge-base (DOKB). There is another misconception that knowledge is fundamentally related with the new offer development. But the fact is that developing new technology; building new capability; formulating new process and system are within the jurisdiction of innovation.

Inherently, innovation is uncertain and thus the knowledge innovation is also uncertain [7]. Innovation is perceived as the single most destructive force in an industry [10]. Any form of innovation, such as knowledge innovation may devastate organizations existing capabilities, expertise and knowledge-base.

The roots of all these challenges have institutional implications and the mitigation of these challenges is also dependant on the institutional initiatives. Success depends on how efficiently an organization is managing its knowledge resources. Perez and Pablos consider that it is necessary to manage knowledge effectively in the new economy, because the achievement of sustainable competitive advantage depends on firm's capacity to develop and deploy its knowledge-based resources [11]. Institutional approach is the innate aspect in both the concepts – knowledge and innovation.

6. Case Study: Knowledge Program of Amber Food Limited (AFL)

Amber Food Limited is one of the promising regional companies in Bangladesh. Currently, the business is offering 55 varieties under 10 product categories. Though it is a relatively new entrant in comparison to other players in the market, yet it could manage a spectacular growth in recent years, soon after its inception.

6.1 The Early Days – Business Idea and Groundwork

The idea of establishing a food processing business was conceived after observing the scenario of a favorable environment for agro-based business. Traditionally, Bangladesh is a harvest economy due to the factors like fertile soil, good irrigation system, favorable climate, ample water resources and so forth. In addition to these, entrepreneurs observed a number of institutional supports (e.g. soft loan from Bangladesh Krishi Bank –BKB, tax holiday and tax rebate, and free counseling) from the Government side for the agro-based business as well. The critical success factor in this type of industry is uninterrupted supply of ingredients. In ensuring a smooth procurement of ingredients, the company decided to de-

velop a partnership with the local producers (suppliers of the ingredients). It made a special agreement with the farmers in their locality to buy their produce. This had two-fold implications – the business was in a relatively safer condition in procuring ingredients and on the other hand, the farmers had a confirmed buyer.

The entrepreneurs spent two years for planning their business and grooming themselves. During these two years, they tried their best to know market and customer preferences; acquire business exposure and identify key success factors. In the early days, the main challenges were acquiring processing technology, machineries, equipment and skilled workers. They participated in a 6-month course of food processing organized by a private agro-business development NGO (non-government organization) in Bogra. During this course, they realized that it is simply essential to incorporate customer preference in new product development. Otherwise, product would not be desirable to the customers. Side-by-side, they were looking for skilled workers. Fund was not much of a problem for them. They had abundant fertile lands in their villages. They used those lands as collaterals to the bank. All machineries and equipment were procured from China. By the year 1997, the venture was all set to start their business.

6.2 Commencement of the Business, Contemporary Market Scenario and Initial Business Strategy

In the year 1998, Amber Food Limited was formally established. In the late 90's, the market was full of imported products from neighboring countries like India, China, Pakistan, Thailand and Bhutan, etc. There were several local players (e.g. Pran, BD Food, Sezan, etc.), too. It was noticed that none of the existing market players were serving in the markets of North Bengal. It appeared to be a potential opportunity for AFL.

Being the consumers in a developing country, Bangladeshis are price-sensitive. Most of the business practices are based on cost-minimization strategy. AFL planned to pursue that along with the emphasis on differentiation, i.e. a hybrid strategy. Economic endowment of upper middle class in Bangladesh though was steady but demonstrated a slow pace of progress. There was a steady growth in demand of differentiated products.

6.3 Business Development in AFL

The motto of the business was to provide the best value to its customers. People were at the heart of business development strategy. Strategic agenda of AFL were – collecting information about customers; incorporating customers' requirement into the product developing process; reaching new customers with the new range of products and configuring internal state of the business in

order to support the business strategy.

AFL, on a part-time basis appointed a few students of Rajshahi University to collect information on customers. In collecting data, they targeted two types of respondents – groceries and direct consumers. The part-time workers interviewed groceries and customers in Rajshahi, Bogra, Rangpur and Dhaka city at random. They interviewed only those customers who were buying food products in the market. Within three months of time, ample information was collected. Initially, data was processed manually and later on, electronically. All collected information was categorized in terms of preference, item, price, timing, quantity and frequency. These data and information had a great tactical significance. AFL not only could assess the existing market demand but also assess the future demand as well. In addition, the company could enjoy some other benefits from its knowledge strategy – a direct communication with customers; first-hand market information; better understanding about the distribution channels; new product ideas and so on. In June 1998, AFL started its production and products were marketed in August 1998. Within few months, AFL started getting positive responses from its customers. By 2000, AFL

products were available in all major towns in North Bengal and Dhaka city. Since it was a regional company, that is why, it was not possible to assess the exact market share but it definitely became to be the number one company in North Bengal. Also, its presence in Dhaka was strongly felt in terms of market share. Altogether, the overall business performance was satisfactory.

Enthusiastic workers were at the heart of its internal innovation mechanism. Management believed that a person with the right attitude and aptitude can bring about a great change in his/her organization. AFL gave the full freedom to its workers in coming up with new/improved product and process idea. It allowed its workers to spare some time in doing experiments with their innovative ideas. From the motivation point of view, it was a good initiative. By doing and learning, the workforce of AFL became a truly flexible workforce. Employees' performance, creative efforts and team-based productivity were directly related with its compensation package. There was another success in AFL; it could develop an integrated support system with the help of logistics and supply chain, technology and management. (See Figure 1, Figure 2 and Table 1)

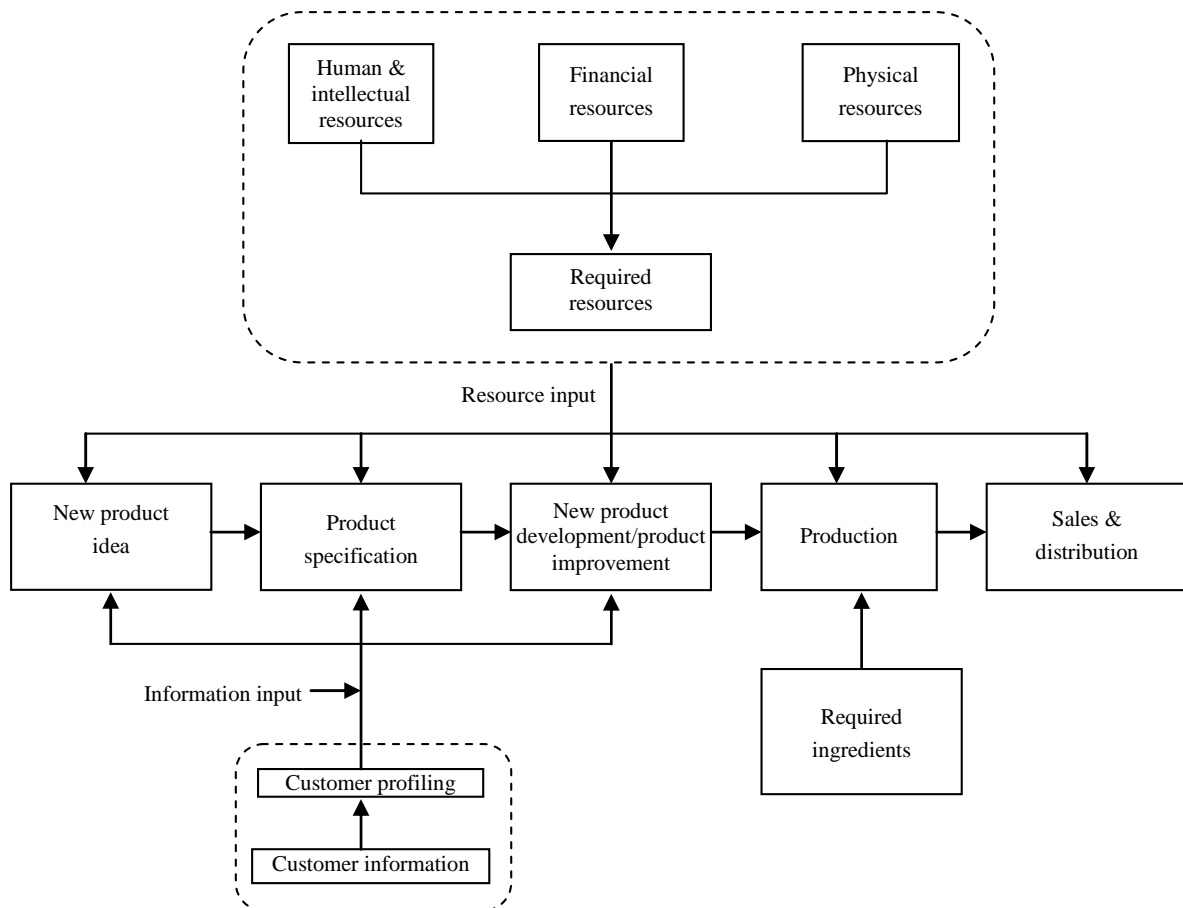


Figure 1. Product development process in AFL

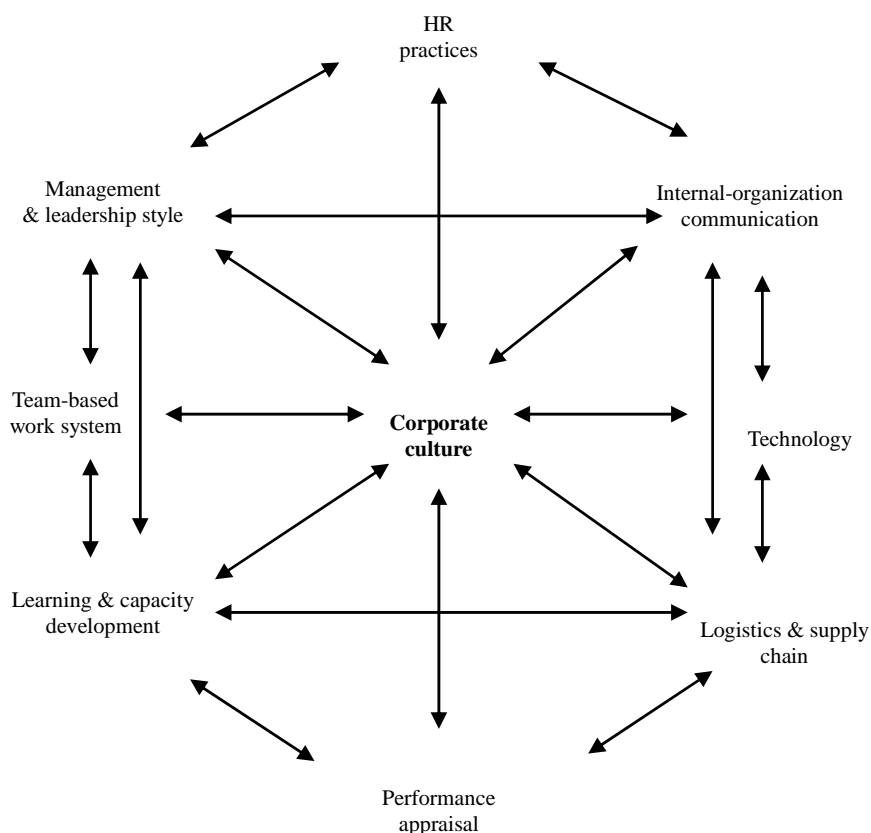


Figure 2. Internal enablers of AFL's business development strategy

AFL kept a keen eye over the changing market phenomena. Therefore, it not only knew what the customers wanted but also to some extent, what the customers might need in the near future. For example, there was an increasing awareness about health and hygiene and increasing demand for dairy products. Its new dairy products became very popular just after the launching. These types of processed and packaged dairy products were not available in Bangladesh. Similarly, there was a growing demand for instant cooking spices as the number of office going female population was increasing with the increase of dual earning families. The increased employment of female workers caused a decreasing trend in home-making service, especially, in the urban areas, since traditional way of preparing spice is cumbersome and time consuming. AFL's instant spices turned into one of the mostly demanded items, especially in the North Bengal region. By considering the changing lifestyle, it (AFL) introduced two new products – mango bar and pickle lozenge under the 'Snacks' category. In 2005, two more new products—tea and energy drink were added in its range under the 'Tea and Beverage' category. Presently, AFL is actively considering expanding its market in other major cities (e.g. Khulna, Chittagong and Barisal) in Bangladesh within the soonest possible time.

6.4 Looking Beyond the Horizon

AFL considered that in Bangladesh, food processing companies have a swell future since there is a strong traditional agricultural infrastructure and availability of attractive incentives for the entrepreneurs. SAPTA, among the SAARC countries, had opened a new horizon for export business. 'Pran' (one of the key market players in Bangladesh), had started exporting in the neighboring countries. This phenomenon inspired AFL to do the same. At the end of 2005, AFL decided to expand its market to other neighboring countries like Nepal, India, Afghanistan, etc.

7. Facts and Findings

The case content can be perceived from two pints of view:

- To see how all internal components, particularly human resources, technology, logistics, systems and management style, etc. have been integrated for the purpose of innovation; and
- To assess critically the extent the knowledge initiatives along with the knowledge infrastructure were effective in making a knowledge-innovative organization.

7.1 Arrangements of Internal Components in AFL

Internal management system of AFL was basically people oriented. People from inside and outside were integrated into its product development process. Customers contributed in the form of data and information (mainly product idea and preference) and internal stakeholders contributed directly by producing products as per customers' requirements. Each piece of information was used either in developing new products or product improvement, which also instigated new cycles of learning. Eventually, new competencies were developed, which was a clear indication of a learning culture in AFL. Management relied more on its employees; direct supervision was less; rate of error reduced significantly and more experiments were done on new ideas. These were nothing but the outcomes of imparting knowledge into its (AFL) processes and systems. These had two major benefits - product diversity and cost minimization.

In AFL, there was a collective leadership approach. For developing new products, work teams took the basic initiatives and other supporting teams provided support to the new product development initiatives. In case of product improvement, both the product development teams and supporting teams worked together. In coordinating their (teams) activities, there was no formal approach. It depended on the nature of the initiatives and the situation. AFL followed functional structure. By reducing hierarchy, introducing team-based work system and encouraging informal communication system, this functional structure was made dynamic. Trust, involvement and co-operation were the basic elements of AFL culture.

Technology was the inseparable part in AFL. Its new product development ability was primarily based either on adoption of new technology or customizing technology. Products were produced, bottled and packaged by the automatic machines. Technology enabled AFL in attaining economies of scale and in maintaining enhanced product differentiation capability. In analyzing data and information, gradually it started using modern information technology.

Although there was no deliberate HR policy, yet healthy HR practices were there in AFL (See Figure 3). Training and development was replaced by continuous learning; compensation system and performance evaluation were linked with new product development and existing product improvement. There was no set rule or policy for recruitment and selection. However, ever since its establishment, it was recruiting and selecting enthusiastic workers. Therefore, experience of the new workers was irrelevant. Management believed that if a person had a positive attitude towards work, then he/she would become a good performer. Since it was a somewhat flatter organization; the scope of career development was comparatively narrow. Annual increment system could nevertheless, resolve this problem to some extent. Reward system, performance evaluation and compensation management were interlinked. Performance parameters in AFL were: participation, initiative, motivation, multi-skilling, teamwork, idea generation, quality of work and on time delivery. Being a medium scale organization, there was not much distinction between HR and administrative functions and therefore, there was less complicity in one sense.

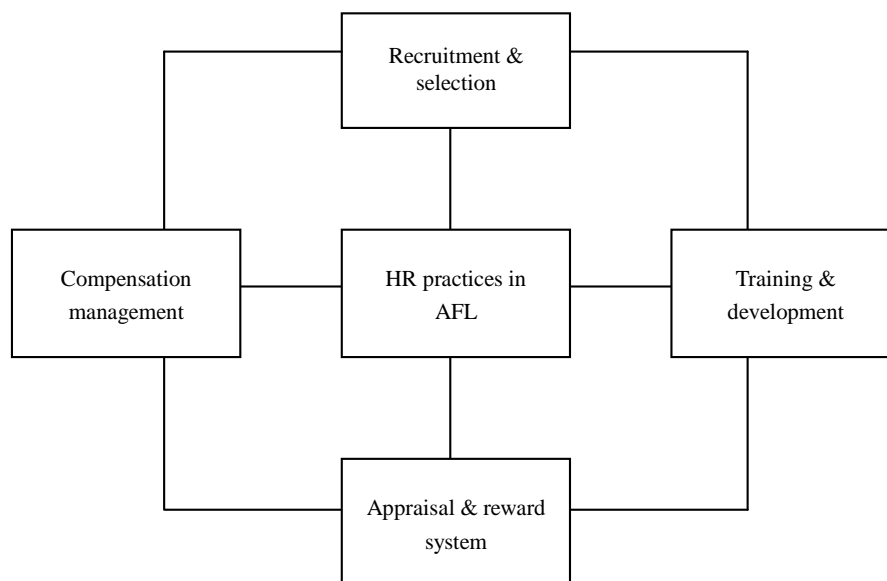


Figure 3. Interlinked HR functions in AFL

Perhaps the most active enabler was its supply chain. AFL designed its supply chain in such way so that all required supplies and logistics were available on time. The business was very much careful from the very beginning in ensuring the availability of ingredients. Equal emphasis was also given on the distribution of products. Other logistical support like repairing and maintenance, inventory management, etc. were befitting with the operation system.

7.2 Critical Analysis of the Knowledge Strategy

According to Van de Van, innovation is a critical mechanism by which firms secure a place in the competitive world of the future [12]. From this point of view, innovation strategy of the business was pretty successful. AFL designed its innovation strategy with the help of knowledge in order to attain competitive advantage. Its (AFL's) key business strategy was to incorporate knowledge, especially, customer knowledge in new product development and to improve existing products by which it could secure its position in the market, and create a competitive advantage. The success of business was on the extent a business could understand its customers and incorporate customer information into its product development process.

AFL had a general understanding of its customers' taste and preference along with their buying behavior. It used to monitor environment relentlessly. Each new piece of knowledge was working as an impetus or a drive for new product development or product improvement. As a result, the firm's product line was extending. This not only could add value to the business but also to the customers. Knowledge only can add value if it is linked with actions, tasks and purposes [7]. Here in AFL, knowledge was linked with collecting customer information, new product development, improving existing products, developing new competencies and satisfying customers. While improving new product and process, knowledge was also adding value to the business. It (knowledge) adds value when it (knowledge) is applied in order to improve, change, or develop specific tasks and activities [13]. AFL improved its existing process and modified its infrastructure whenever necessary as per the merit of the information and knowledge. Overall, it can be said that knowledge was adding value quite noticeably in AFL.

Barton opines that there is a clear linkage in between firm's core capability and new product development [14]. According to Dougherty, product innovation is a primary means of corporate renewal [15]. When AFL used to receive any information from the market, it tried to fit that information into the organization, to see if it was relevant. For instance, when it noticed that the number of working women was increasing in all major cities in Bangladesh, it decided to launch packaged cooking spice.

Before that, the business assessed itself whether it would be feasible. It (AFL) analyzed its existing skill, competency and technology. It was found that some new machineries, few additional skilled laborers, raw materials and packaging materials were needed. For packaging and printing it used its old machines; for distribution it used the same distribution channel and therefore, the overall cost could be minimized significantly. Besides, while launching a new product, AFL acquired new capabilities. Every new product enabled AFL to enhance its existing capacity and/or acquire new capability. Every new product idea was the function of new information on customer and the situation. Eventually, the firm could renew itself along with maintaining its growth in the market. Danneel argues that organizations need to renew themselves continuously for their survival and prosperity in dynamic environments [16].

The most substantial advantage of KM strategy is that it allows improvements of firm's learning capability and its ability to combine knowledge-based capabilities as well as make better use of them [8]. In AFL, individual and organization learning took place simultaneously. Organization learns when a new body of knowledge is imparted with its existing knowledge-base. By processing customer information, noticing changes in the environment and by acquiring new technology, the business was learning. At the same time, employees were learning by experimenting with new ideas, doing tasks, working in a team, directly participating in the product development process and so forth.

Innovation was not a sudden result or outcome rather it was the result of systematic organizational efforts and initiatives. In the knowledge era, 'knowledge' resources and knowledge-based innovation mechanism is at the core of creativity and innovation. Knowledge management strategy refers to the overall change of process, a form of organizational renewal and a focus on innovation through the creation, transmission, and application of new knowledge, which is within the resource-based view of the firm [17]. This holistic approach of knowledge-based innovation or knowledge-driven innovation clearly indicates about the alignment of organizational resources. AFL's management and leadership style was people oriented; employees were encouraged to nurture their creativity and innovativeness; the key HR functions were inter-related and integrated with its new product development and product improvement; high emphasis was given on team-based work; formal and informal communication; imparting knowledge through learning and innovation, etc. These clearly indicate an institutional approach within AFL, which, at the same time implies the alignment of resources, particularly: the intellectual resources in the line of business strategy. A dynamic knowledge infrastructure was thereby created spontaneously. (See Figure 2 and Table 1)

Table 1: Internal enablers of knowledge strategy in AFL

Enablers	Results
HR practices	Overall employee satisfaction
Performance appraisal	Achievement oriented workforce
Management & Leadership style	Empowerment, participation and involvement
Team-based work system	Decentralization and leadership development
Inter-organization communication	Shared vision
Learning & capacity development	Capacity development and organizational renewal
Technology	Technology alignment and utilization
Supply chain and logistics	Smooth procurement and effective support system

**Figure 4. Essential elements of a knowledge innovative organization (KIO)**

The most significant finding is that - knowledge strategy in AFL was not deliberate. It was the result of incorporating customer preference into its product development process. Initial thinking of focusing on customers' requirements exerted a breakthrough impact over AFL in terms of organization development, innovation, capability development, and so forth. Knowledge strategy in AFL emerged automatically. In order to sustain the benefits, the business continued to pursue this (KM) strategy which made it a total knowledge innovative organization. All these practices created a total knowledge oriented culture in AFL. This knowledge-based culture was actually playing the key role in making the organization innovative.

8. Knowledge Innovative Organization

There is a traditional view that service organizations are primarily considered as knowledge-intensive organization [7]. According to Alvesson [18], knowledge-intensive firms are the companies where most of the works are intellectual in nature and well-educated, and qualified employees are the major part of the workforce. But in making organization knowledge innovative, it is essential to go one step ahead. Knowledge has to be put at the heart of the organization so that it can play a role of a catalyst in guiding business activities, create a knowledge infrastructure, develop knowledge capability through human and institutional development. In AFL, it

has been observed that in due course of time, it (AFL) could develop a knowledge infrastructure for innovation. This infrastructure consisted of various essential elements like HR best practices, integrated performance appraisal system, management and leadership style, team-based work system, inter-organization communication, learning and capacity development, technology, supply chain and logistics. Core business processes were: autonomy (decentralization of authority), skill development, research and development and socialization. From the experience of AFL case, in general, it is possible to identify several basic elements of a knowledge innovative organization. These are as follows:

- 1) People (internal and external) integration through socialization.
- 2) Organizational renewal in the form of capacity development.
- 3) Alignment and utilization of knowledge resources.
- 4) Continuous research and development.
- 5) Knowledge-based infrastructure and logistic system.

The central essence of knowledge innovative organization can be as follows (See Figure 4):

- Employee participation is necessary in business development process with a common vision of innovation.
- Multidisciplinary, self-managed and motivated teams can create diversified knowledge and skill, and ensure the best use of those.
- Successful accomplishment of knowledge strategy mainly depends on the autonomy of the people, socialization with the consumers, flexible organization structure in addition to research and development.
- Knowledge management strategy is involved with constant change. Here, the change is not imposing, rather its natural in the form of capacity development and overall organizational renewal, which is now-a-days termed as human and institutional development.
- Knowledge infrastructure is simply necessary to ensure all logistical support towards knowledge strategy. In addition, there is no specific condition(s) for the successful implementation of knowledge strategy yet organization culture and leadership have a great influence over it.

This knowledge innovative organization has other crucial connotations, too. There are several concerns in using knowledge for innovation. Earlier, it has been discussed that a totalistic institutional approach can resolve these concerns (challenges in using knowledge for innovation). It (AFL) developed its own functional definition of knowledge management. To AFL, knowledge management is imparting customer knowledge into the innovation process through internal capacity development.

9. Conclusions

AFL embraced a holistic approach of knowledge strategy. New product development and product improvement were not the sole jurisdictions of marketing department rather all of its units were equally responsible. Sales and marketing was involved in understanding customers' requirements; production department was responsible to produce as per requirements of the customers; human resource was engaged in continuous people development; resources along with logistics were aligned in supporting innovation process; leadership and management was inspiring in fostering teamwork and creativity. AFL's formal and informal systems were playing a key supportive role in creating a social network inside and outside the business.

It is widely believed, since knowledge is a very contemporary thought, therefore, the application of this strategy is limited within the developed countries. But from this case, it is revealed that knowledge has no boundary; it is applicable irrespective of the economy, country and region. There is another conviction that a huge investment is needed in technology, modern equipment and process. But the fact was that, in KM, technology is more involved in the business development process, product development and improvement of mechanism and systems. The central issue is the extent to which organizations are tailoring it (knowledge) as per their requirements. If an organization can customize knowledge strategy according to its need, certainly it would result into long-term competitive advantage. According to Liebeskind, the concept 'knowledge' incorporates and integrates capabilities, structured information and application of technologies in the improvement of products and process and thus, has become a source of competitive advantage [19].

REFERENCES

- [1] P. F. Drucker, "The post capitalist society," Butterworth-Heinemann, Oxford, 1993.
- [2] T. A. Stewart, "Intellectual capital: The new wealth of organizations," Doubleday, New York, 1997.
- [3] M. H. Zack, "Developing a knowledge strategy," California Management Review, Vol. 41, No. 3, pp. 125-145, 1999.
- [4] T. Housel and A. H. Bell, "Measuring and managing knowledge," McGraw-Hill Irwin, New York, 2001.
- [5] M. Grant, "Toward a knowledge-based theory of the firm," Strategic Management Journal, Winter Special Issue, Vol. 17, pp. 109-122, 1996.
- [6] G. A. Verona, "Resource-based view of product development," Academy of Management Review, Vol. 24, No. 1, pp. 132-142, 1999.

- [7] S. Newell, M. Robertson, H. Scarbrough, and J. Swan, "Managing knowledge work," Palgrave, New York, 2002.
- [8] B. Kogut and U. Zander, "Knowledge of the firm, cognitive capabilities and replication of technology," *Organisation Science*, Vol. 3, No. 3, pp. 383–397, 1992.
- [9] M. Garcia-Murillo and H. Annabi, "Customer knowledge management," *The Journal of the Operational Research Society*, Vol. 53, No. 8, pp. 875–884, 2002.
- [10] J. Kay, "Foundations of corporate success," Oxford University Press, Oxford, 1993.
- [11] J. R. Perez and P. O. Pablos, "KM and organizational competitiveness: A framework for human capital analysis," *Journal of Knowledge Management*, Vol. 7, No. 3, pp. 82–91, 2003.
- [12] A. M. Van de Van, "Central problems in the management of innovation", *Management Science*, Vol. 32, No. 5, pp. 590–607, 1986.
- [13] R. McDermott, "Why information technology inspired but cannot deliver knowledge management," *California Management Review*, Vol. 41, No. 4, pp.103–117, 1999.
- [14] D. Leonard-Barton, "Core capabilities and core rigidities: A paradox in managing new product development," *Strategic Management Journal*, Summer Special Issue, Vol. 13, pp. 111–125, 1992.
- [15] D. Dougherty, "A practice-centered model of organizational renewal through product innovation," *Strategic Management Journal*, Summer Special Issue, Vol. 13, pp. 77–92, 1992.
- [16] E. Danneels, "The dynamics of product innovation and firm competences," *Strategic Management Journal*, Vol. 23, No. 12, pp. 1095–1121, 2002.
- [17] W. M. Cohen and D.A. Levinthal, "Absorptive capacity: A new perspective on learning and innovation," *Administrative Science Quarterly*, Vol. 35, No.1, pp. 128–135, 1990.
- [18] M. Alvesson, "Knowledge work: ambiguity, image and identity," *Human Relations*, Vol. 54, No. 7, pp. 863–886, 2001.
- [19] J. P. Liebeskind, "Knowledge, strategy and theory of firm," *Strategic Management Journal*, Winter Special Issue, Vol. 17, pp. 93–107, 1996.

Study on Option Price Model of the Transaction of Information Commodities

Changping HU, Xianjun QI

Center for Studies of Information Resources, Wuhan University, Wuhan, China.
Email: xianjun.qi@gmail.com

Received August 25, 2009; revised September 29, 2009; accepted November 1, 2009.

ABSTRACT

The option is viewed as an important tool of setting price in accordance with objective benefits and actual effectiveness of information commodities, as ensures that the holder of the option right acquires more benefits in the favorable market and reduces losses in the adverse market. And the information market, as a kind of typical monopolistic and competitive market, is especially adequate for introducing the option theory to set more reasonable price and further to improve the operational efficiency of information market. Hence, it is quite necessary to apply the option theory to the transaction of information commodities. In this paper, after analyzing the applicability of the option theory to the information market and studying the option price of information commodities, the authors put forward the option price model of the transaction, followed by a case study to demonstrate the validity of the model.

Keywords: Information Commodities, Market Mechanism, Option Pricing, Transaction Model

1. Introduction

With the development of commodity economy, incomplete market information, which creates difference value between the cost and the profit of information searching, and which has its special value, gradually evolves into a form of independent commodity [1]. And information commodities inevitably promote the emergence and the development of information market [2]. As the product of information commercialization, information market is not only an independent and physical commodity market, but also an invisible factor market included in other markets [3]. Therefore, in the narrow sense, information market refers to as some places, where people trade information commodities at a certain period of time; while in the broad sense, it refers to all the relationships between the supply and the demand of information commodities, involving the whole course and the all fields of the circulation of information commodities, from the production to the consumption.

Information market, like the material market, could only operate formally with the complete market systems and flexible market mechanisms, namely, the supply and demand mechanism, the price mechanism, the competitive mechanism, risk mechanism, management mechanism and other mechanisms [4]. And it is the price mechanism that serves as the most effective way of resources allocation in the market economy. The main

reason is that the price, as the most sensitive signal in the market, could accurately and timely reflect the imbalance between the supply and the demand of information commodities distributed to many different directions of utilization to maximize the effectiveness. Therefore, pricing is the primary task in the transaction of information commodities.

Regarding the price of information commodities, there have been several pricing methods such as multiple pricing [5], bundle pricing, discriminatory pricing, Ramsey pricing and integrated pricing strategies etc. [6]. All of them are classified as Net Present Value (NPV) methods, based on the traditional pricing theories, neglecting the impact of market uncertainty on the transaction of information commodities. Since information commodities have the characteristics of monopoly, externalities, public product properties and a shorter life cycle, information asymmetry are more common and information market are more uncertain. Thus the above pricing methods do not always make right prices identical with the real value of information commodities [7].

In the middle of the last century, many researchers and managers found that discounted cash flow methods, represented by NPV, led to ineffective resources allocation [8]. Shortly afterwards many scholars, like Louis Bachelier, Paul Samuelson, Fischer Black, Myron Scholes, Robert Merton and so on, began to seek other solutions and put forward several transaction models under the

uncertainty conditions, which made a significant contribution to option research. After analyzing those models, it is easy to draw the conclusion that the option theory had two obvious advantages in resolving the transaction of information commodities: the one is that the option theory dealt with the uncertainty under the condition of observing the principles of the classic economics and mathematics; the another is that the option method, replacing the inaccuracy with the complexity, is almost applicable to every market.

Moreover, in the information market, the option of information commodities can objectively reflect the impact of the uncertainty of information market and the flexibility of decision-making on the price and the value of information commodities, by assessing the value of options, by analyzing the costs and profits of the both parties and by setting a reasonable price, so as to achieve the optimal allocation of information resources.

2. The Applicability of the Option Theory to Information Market

2.1 The Option Theory

The option theory had been gradually developed in the long-term practice of solving the uncertainty of the price in the future market. Essentially, an option is also called a option contract, which stipulates the right of the seller or the buyer to choose to sell or buy a certain number of underlying assets such as portfolio, financial assets and physical assets at a fixed price within a prescribed time. Thus, the option is no mere a financial derivatives, but rather an objective right to choose. Obviously, the option is a tool of transaction and pricing in the future market [9].

Previous studies suggested that the real option is derived from the financial option. Its core idea is that investors should maximize the value of the project by means of some market-based methods instead of subjective probability methods and utility functions when they determine the value of investment opportunities and choose the optimal investment strategy. Simply, the real option is a kind of realistic option defined by the option concept. According to different holders of the right, options are divided into a call option and a put option: the former is the right that the option buyer purchases underlying assets in terms of stated price and quantity within validity of the option contract, and the latter refers to the right to sell the underlying assets under the same conditions; While according to the difference of exercise time, options are divided into American option and European option: the former specifies buyers can exercise options on or prior to any given trading day, and the latter is just the opposite. Actually, the buyers of American call option can flexibly choose the best time that is most advantageous to them, to exercise options according to the market shift and the actual demand.

The most distinctive feature of the option is that an option grants the exercise right, rather than the obligation, to the option buyer in the form of the contract. That is, option buyers may make profits from the transaction when the price of the underlying asset is higher than the exercise price, on the contrary, they just avoid more losses (aside from the value of the option) by giving up their option rights. That is why the method of the option is specifically suitable for the high-risk and strong flexible market. Especially, in the market with asymmetric information, the option can help to reduce the risk caused by the uncertainties of the supply and the demand, to add the effectiveness of information commodities, and further to reduce transaction costs by improving the market transaction.

2.2 The Option Features of Information Commodities

Some experts on the theory of the real option, like Dixit, maintained that the irreversibility, the uncertainty and the selectivity of the investment were just theoretic premise of the research on real options [10]. In addition, as the main form of information resources in the market economy, information commodities also fulfill the given precondition: firstly, most information commodities are technology-based or experience-based commodities, requiring consume a majority of intangible capital that is a part of submersion capital, difficult to recover invested costs; secondly, as invisible values of information commodities, which often attach to material commodities, do not directly create values, and useful values usually depend on their own values and on the specific use conditions, information market fill with many uncertainties that are associated with the price, the quantity of supply and demand, the quality as well as effectiveness and that reduce the accuracy of assessing the real value of information commodities; thirdly, the production and the transaction of information commodities at different times may result in quite different profits and risks due to the long production time, the short life cycle, and the unpredictable factors in information market. In a word, the option of information commodities in the information market, an invisible market, where both of the time and the form of the transaction are extremely flexible, could be regarded as the American option. The option holders of information commodities maybe obtain more market opportunities or reduce less losses by changing the operating scale, investing periodically, delaying, giving up the option, and so on.

2.3 The Function of the Option in Information Commodities Transaction

In the information market, various market mechanisms are regulating the whole transaction course and the all economical relationships to realize market equilibrium.

According to western economics, market equilibrium refers to that interrelated economic subjects, who start from their maximal welfare, interact on each other continuously to adjust market parameters until the whole economic system come up to a stable state. In theory, the above market equilibrium is a sort of “point equalization” that could realize Pareto Optimality. However, in the real information market, the price of information commodities cannot automatically regulate to a balance between supply and demand, but to a certain range deviating from each other. Therefore, information market equilibrium is a kind of “domain equalization”, which is not able to fully realize ideal Pareto Optimality. The main reason is that information commodities have the following three characteristics of indivisibility, externalities and uncertainty [11], and the first two issues could be solved by pricing with marginal cost and by separating the social costs (profits) from individual costs (profits), requiring use options to solve the third one to achieve the transaction in the uncertain information market.

It is well known that the option promotes the transaction mainly by diverting, avoiding risks and regulating price. Researches indicated that the option model could simulate optimal information system [12], and that the option method could help to solve the problem of venture capital investment of information technologies [13]. Yet, in the uncertain information market, the price cannot objectively reflect the real value of information commodities. Furthermore, it is difficult for two parties to reach an agreement owing to the higher risk and the wider fluctuation margin, which leads low market efficiency. However, with an option granting exercise right instead of the obligation, option holders could transfer risks to their counterparts. And this way of transaction provides a special insurance, namely, reducing losses under the adverse condition and obtaining more profits under the favorable condition in virtue of the option of information commodities, and is bound to ease the worries of traders, reduce transaction costs, simplify the transaction process and further to promote the circulation of information commodities.

3. The Option Price of Information Commodities

The theory of option price is one most important achievements of modern finance theory. In 1997, having put forward the theory of option price, Myron Scholes, a Professor of Stanford university and Robert C. Merton received Nobel Prize for Economics of the year. And then the formula of option price—Black-Scholes, which is view as one of successful models of applied mathematics in social science domain, has been used widely for its simple operation and relatively accurate results [14]. Although it was proposed aiming at the European

option at first, it could be quickly extended into American option that is able to pay the dividend in the option period [15], and then:

The call option price is:

$$V_t = SN(d_1) - Ke^{-\delta t}N(d_2) \quad (1)$$

The put option price is:

$$P_t = Ke^{-\delta t}N(-d_2) - SN(-d_1) \quad (2)$$

$$\text{Where, } d_1 = \frac{\ln(S/K) + (f + \delta^2/2)t}{\delta\sqrt{t}},$$

$$d_2 = d_1 - \delta\sqrt{t}, \quad N(d) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^d e^{-\frac{t^2}{2}} dt, \quad N(d_1) \text{ and } N(d_2)$$

respectively stands for the cumulative distribution function of standard normal distribution of d_1 and d_2 , whose value may be acquired by inquiring the normal distribution table; S is the current price of underlying assets; K is the exercise price of underlying assets; f is the risk-free profit margin, δ is the profitability of underlying assets.

In the information market, however, information commodities are mainly divided into two categories: the first one, independent of physical products, could be exchanged in a transaction without transferring the original physical carriers, such as knowledge and decision; the second one, integrated with physical carriers, carries out a transaction which often accompanies by the exchange of physical commodities, for example, a book or a CD-ROM. Clearly, for the latter, the real value of information commodities is difficult to separate from in the whole value. Then this paper is aimed at the first information commodity, which is independent of physical carriers. Moreover, the meaning of each variable of information market is quite different from those of general physical market. For example, the effective period of call options of physical commodities does not ran over the time of market transaction, meanwhile, information commodities are on the other hand. Although these produced information commodities would be marketed immediately, their effectiveness cannot accomplish right now. As thus people do not regard variable prices within a short time as the foundation for the decision whether to purchase information commodities. Instead, they mainly consider the following two issues in the long run: how much effectiveness and risks would information commodities bring about? Could they achieve expected profits? Now, if the full time from purchasing to using is assumed to be an effective option period, represented by t , accordingly, δ represents the value fluctuation ratio of information commodities (expressed by standard deviation of expected rate of return), S represents the expected cash value of information commodities, K represents the cash value of information commodities,

V_t represents the expected benefits of information commodities. Evidently, if V_t was less than zero or was lower than the expected, buyers would not purchase the commodities.

4. The Transaction Model of Option Price

Absolutely, the price is the most sensitive signal of market change and correspondingly, the price mechanism is the most effective transaction mechanism in the market economy. Under the function of the option, the price could objectively reflect the supply and the demand of information commodities, the competitive situation, the management performance, as well as the risk index of development and utilization. And it helps to improve the operational efficiency of information market and further to promote the optimal allocation of information resources. In other words, the option devotes to achieve the rational allocation of information commodities by pricing.

However, most information markets serve as seller's market, in which information asymmetry is very common, just for buyers' lack of important information including commodities, markets and seller's credibility. Hence, sellers often have a comparative advantage during the transaction, which is carried out mostly based on the price set by them. In this paper, taking a seller's market as an example, the authors constructed the option price model of the transaction of information commodities and then analyzed how the option optimized the model in order to improve the transaction in the information market.

4.1 The Seller Price

In the information market, P_s is used to stand for the price set by the seller (a reasonable price, instead of exorbitant price). And the seller consider the two main determinants of production costs (C_p) and the dividend of expected additional profits (C_H), where C_p includes three parts: the cost of material carriers (C_1), the consumption of active labors (C_2) and the opportunity cost (C_3), namely, $C_p = a_1C_1 + a_2C_2 + a_3C_3$, where, a_1 , a_2 and a_3 , whose value is between 0 and 1, respectively stand for the corresponding weight of the cost. They are affected by some factors like the way of specific transaction, negotiation skills, the effectiveness of commodities and the expected values of both parties.

In addition, C_H is determined by the two parts: the expected additional benefits (R) and the proportional coefficient of the dividend (L). Thus, R is equal to the sum of additional economic profits caught by information commodities of each year. However, information commodities usually represent themselves in the form of techniques, knowledge, recipes and other intangible assets, which cannot create economic values directly, but enhance productivity and benefits by improving production processes. As a result, it is very hard to assess how much economic values and effectiveness on earth the information commodities brought about in the short time. Therefore, R is just estimated value or reference value of the industry during transaction. Besides, the inflation is not neglected, and it is indispensable that cash flows should be discounted to the present by a given discount rate r . Thus, additional profits received by using the method of net present value (NPV), are as follow:

$$R = \sum_{t=1}^N \frac{b_1 \cdot b_2 \cdot b_3 (G_t - G_0)}{(1+r)^t} \times WT \quad (3)$$

Where, b_1 is cost factor, b_2 is the factor of quality & performance, b_3 is the sales factor, G_t is the production capacity of the buyer after t years, G_0 is the production capacity of the buyer before purchasing the information commodity, r is the discount factor of the currency, W is the profit of per unit of production after purchasing and T is the monopoly coefficient of the information commodity. And the proportional coefficient of the dividend (L) have to balance the profits of both sides and depend on the funding of the corresponding supporting facilities, noted as C_M , namely, the proportional coefficient of the dividend (L) is equal to the ratio of C_p :

$$L = \frac{C_p}{C_M} \times 100\% = \frac{C_p}{Q' \cdot c} \times 100\% \quad (4)$$

In the above formula, C_M is the application cost of the buyer, assessed by experts, namely, $C_M = Q' \cdot c$, Q' is the production of the buyer after using the information commodities, c is the application cost of the buyer after using the information commodities. And the formula of the proportional coefficient of the dividend (L) and the seller price P_s are as follows:

$$C_H = L \cdot R = \frac{C_p}{Q' \cdot c} \cdot \sum_{t=1}^N \frac{b_1 \cdot b_2 \cdot b_3 (G_t - G_0)}{(1+r)^t} \times WT \times 100\% \quad (5)$$

$$P_s = C_p + C_H = (a_1C_1 + a_2C_2 + a_3C_3) + \frac{C_p}{Q' \cdot c} \cdot \sum_{t=1}^N \frac{b_1 \cdot b_2 \cdot b_3 (G_t - G_0)}{(1+r)^t} \times WT \times 100\% \quad (6)$$

Obviously, the upper limit of the seller price:

$$P_{S \max} = (a_1 C_1 + a_2 C_2 + a_3 C_3) + \sum_{t=1}^N \frac{b_1 \cdot b_2 \cdot b_3 (G_t - G_0)}{(1+r)^t} \times WT \quad (7)$$

The bottom limit of P_S is: $P_{S \min} = a_1 C_1 + a_2 C_2 + a_3 C_3$, where P_S ought to meet the equation: $P_{S \min} \leq P_S \leq P_{S \max}$.

4.2 The Transaction Price

For buyers, the lower the transaction price is, the more profits they obtain. However, their expected price can not be lower than the lowest expected price of the seller, or it cannot reach a deal. Meanwhile, buyers set a fuzzy value of expected price which is generally higher than the bottom limit of the seller price provided that the transaction is concluded, that is, the two parties would soon reach a deal at a transaction price P when the seller price was absolutely propitious to achieve the expected return of the buyer or much lower than the expected price. Nevertheless, it is rare. Instead, there is a more general situation that the seller price is so high that the consumer surplus of the buyer is less than the expected return. In this case, whether to reach a deal or not depends on whether the consumer surplus of the buyer come up to the expectation at the lowest price set by the seller. Then represented by letters, it is showed as follows. Substitute the lowest price of the seller ($P_{S \min}$) into the formula of the option price to compute the value of V_t , which is not less than the expected return to reach a deal after the bargain, otherwise, the deal is failed.

$$\varphi = \frac{V}{S} = N \left[\frac{(f + \delta^2 / 2) \sqrt{t}}{\delta} \right] - e^{-f t} N \left[\frac{(f - \delta^2 / 2) \sqrt{t}}{\delta} \right] \quad (8)$$

In fact, it is very complicated to introduce the model of Black-Scholes to compute comprehensive evaluation of the option, an absolute value of quantities. To simplify the calculation, it is advisable to transform the above formula into the expression of relative options. Furthermore, the difference between the present value of future profits (S) and further out-of-pocket (K) determines whether the buyer purchases the certain information commodities or not. And then we can compute the option price (V) of the critical point (i.e. $S=K$), then the ratio of the price (V) and the value of profits (S) is just the impact factor of the option price, noted as φ . Now, substituting the equation of $S=K$ to formula 1, we can see:

The traditional methods of evaluation regarded cost-income ratio as the relative value without regard to the option. When options that exist objectively are considered, the impact of the option price on the value of information commodities is expressed as $\frac{S}{K} \cdot \varphi$, and then the relative evaluation value of the real options (g)

similarly is estimated as follows:

$$g = \frac{S}{K} \left(1 + \frac{S}{K} \varphi \right) \quad (9)$$

It's crucial, therefore, to compare the value of g with 1, that is, if $g > 1$, the buyer would purchase the commodity; otherwise, the buyer would give up the transaction when $g < 1$.

5. Case Study

Here, we assume that some company (A) has developed the new generation of ERP, whose costs of material carrier are \$50 million, the costs of active labor are \$250 million, the opportunity costs are \$300 million, and which could add a profit of \$100million per year within the service life of 10 years, when the monopoly coefficient equal to 0.2. In the meantime, another company (B) that would purchase the ERP of A, needs to cost \$500 million to allocate the corresponding equipments and staff. At this time, the fluctuation ratio of the value of the ERP system δ is 30% ($\delta=30\%$), and risk free rate (f) is 8% ($f=8\%$).

For company A: the cost of production $C_p=50+250+300=600$ (million dollars); the rate of profit commission $= 600 / (600 + 500) = 0.545$; the expected profit is 1000 million dollars; the expected profit commission $C_H = (1000+600) \times 0.545 \times 0.2 = 174.4$ (million dollars); Then reasonable price of A is $P_S=600+174.4=774.4$ (million dollars), and the floor price accepted by A is \$600 million while its highest price should be $1000+600=1600$ million dollars, i.e., the range of transaction price P meeting the requirements of $600 < P \leq 1600$ (million dollars).

What if we use the method of DCF (Discounted Cash Flow) to compute these parameters?

Well, the minimum amount company B need to pay is $600+500=1100$ million dollars, and correspondingly, its profit is $100 \times 10 = 1000$ million dollars. Clearly, in that case, B would not purchase the ERP system since income is less than the expense.

6. Conclusions

The real option is not so much a tool as a way of thinking, which is greatly able to provide people with a scientific, dynamic and strategic analysis framework, and which is appropriate not only for evaluating the value and the price but for arranging optimal actions to avoid risks. Absolutely, it is the option with above excellent features that has outstanding advantages in the information mar-

ket, where the option can effectively promote reasonable distribution of information commodities in the aspects of the time, the space and the type in order to maximize their effectiveness. Therefore, from the standpoint of economic profits, the option mainly regulates the benefits of market players to optimize the price model in the virtue of the price mechanism.

In a word, this paper proposed that information commodities had the characteristics of the option, that options played an important role in commodities transaction, and that the classic model of the option price was applied to set the option price of information commodities. Additionally, taking for an example of information commodities, we analyzed how the call option exerted its influence to buyers' decision-making, as is of importance in guiding the activities of information economy. However, there are several problems to be further elaborated in the future, for instance, combining the option theory with the game theory to investigate some new laws and features of information commodities in the information market and further to improve the operational efficiency.

7. Acknowledgments

This paper is supported by the major project of National Social Science Foundation of China under Grants 06J2D0032.

REFERENCES

- [1] H. Q. Ma and C. Zong, "Analysis on the implementation efficiency problems and countermeasures about the policy and regulations for the network information resources construction and location," *Library and Information*, Vol. 10, pp. 31–35, 2006.
- [2] X. J. Zha, "The allocation of resources and sharing of information," Wuhan University Press, Wuhan, 2008.
- [3] F. C. Ma, G. Li, and X. J. Zha, "Information resources management," Wuhan University Press, Wuhan, 2001.
- [4] W. Y. Xu and L. N. Li, "Study on information resources effective allocation under the conditions of market economy," *Journal of Information*, Vol. 11, pp. 128–131, 2005.
- [5] L. Huan and Y. Jiang, "Research on the Multi-prices Pricing Model of Information Product," *Finance and Economics*, Vol. 4, pp. 63–66, 2002.
- [6] F. C. Ma and X. G. Wang, "Pricing strategies and methods of information commodities," *Information Studies: Theory & Application*, Vol. 3, pp. 285–287, 2003.
- [7] F. Ancarani and V. Shankar, "Price levels and price dispersion on the internet: A comparison of pure play internet, bricks-and-mortar, and bricks-and-clicks retailers," Business Research Center Working Paper, 2002.
- [8] L. Trigeorgis, "Real options: Managerial flexibility and strategy in resource allocation," MIT Press, 1996.
- [9] H. L. Yu, "Financial options and real options: Comparison and application," Shanghai University Finance & Economics Press, Shanghai, 2003.
- [10] A. K. Dixit and R. S. Pindyck, "Investment under uncertainty," Princeton University Press, Princeton, 1996.
- [11] P. Mirowski and E. M. Sent, "Science bought and sold: Essays in the economics of science," University of Chicago Press, Chicago, 2002.
- [12] J. A. Campbell, "Real options analysis of the timing of IS investment decisions," *Information & Management*, Vol. 39, pp. 357–364, 2002.
- [13] M. Benaroch, "A real options perspective of forthcoming," *Journal of Management Information Systems*, Vol. 1, pp. 23–31, 2002.
- [14] F. Black and M. Scholes, "The pricing of options and corporate liabilities," *Journal of Political Economy*, Vol. 4, pp. 81–93, 1973.
- [15] J. C. Hull, "Options, futures and other derivative securities (2th)," Prentice-Hall, Englewood Cliffs, NJ, 1993.

A Personalized Recommendation Algorithm Based on Associative Sets

Guorui JIANG, Hai QING, Tiyun HUANG

School of Economics and Management, Beijing University of Technology, Beijing, China.
Email: jianggr@bjut.edu.cn

Received August 11, 2009; revised September 13, 2009; accepted October 24, 2009.

ABSTRACT

During the process of personalized recommendation, some items evaluated by users are performed by accident, in other words, they have little correlation with users' real preferences. These irrelevant items are equal to noise data, and often interfere with the effectiveness of collaborative filtering. A personalized recommendation algorithm based on Associative Sets is proposed in this paper to solve this problem. It uses frequent item sets to filter out noise data, and makes recommendations according to users' real preferences, so as to enhance the accuracy of recommending results. Test results have proved the superiority of this algorithm.

Keywords: Frequent Itemsets, Associative Sets, Collaborative Filtering, Recommendation Technology

1. Introduction

How to help users quickly and effectively access to the information they really need when facing abundant resources becomes a challenging task and also a hot topic of current academic study. Personalized recommendation system is one of the effective tools to solve this problem. A helpful method is to develop intelligent recommendation system to provide personalized service [1], that is to recommend products to users according to their preferences or demands, so as to help them finish the purchasing process.

Nearest neighbor collaborative filtering approach is a recommendation technique that is the most widely used right now [2]. Its basic idea is to generate recommendations for target users according to the rating data of nearest neighbors that have given similar ratings. As items' (movies, music, etc.) ratings given by the nearest neighbors are quite similar to those given by target users, items' ratings given by target users can be estimated by the weighted average of the ratings given by the nearest neighbors. The advantage of collaborative filtering approach is that it can adapt to the rapid updating of users' information. It calculates the tightness among users according to the latest data every time, so as to make recommendations. However, the consequent disadvantage is that it is quite slow to get K nearest neighbors within large amounts of data. Meanwhile, results would not be satisfactory when sparse data is dealt with, especially for new products and new users. At the same time, its scalability is not very good [3].

On the basis of traditional collaborative filtering algorithm, our paper proposes a personalized recommendation algorithm based on Associative Sets. This algorithm first supposes user rating matrix as transaction sets, while every transaction is a user's rating set. Then it generates frequent itemsets through frequent itemsets generation algorithm, puts frequent itemsets into a series of Associative Sets according to one user's rating record, and performs collaborative filtering among Associative Sets so as to improve the accuracy and scalability of the algorithm.

2. Traditional Collaborative Filtering Algorithm and Its Analysis

2.1 Traditional Collaborative Filtering Algorithm

Collaborative filtering algorithm is the most widely used approach in personalized recommendations, which can forecast target users' interests and preferences according to neighbor users' interests and preferences. It first finds neighbors that have the same preferences with target users under the help of statistical techniques, and then makes recommendations to target users according to their neighbors' preferences. It includes three stages [4]:

1) Representation

Inputting data can usually be expressed as an $m \times n$ user rating matrix, where m represents the number of users, n represents the number of items, and R_{ij} represents the rating given by user i to item j . Such ratings can have several scales just as Table 1.

Table 1. User/item rating matrix

User	Item					
	I_1	I_2	\dots	I_j	\dots	I_n
U_1	R_{11}	R_{12}	\dots	R_{1j}	\dots	R_{1n}
U_2	R_{21}	R_{22}	\dots	R_{2j}	\dots	R_{2n}
\dots	\dots	\dots	\dots	\dots	\dots	\dots
U_i	R_{i1}	R_{i2}	\dots	R_{ij}	\dots	R_{in}
\dots	\dots	\dots	\dots	\dots	\dots	\dots
U_m	R_{m1}	R_{m2}	\dots	R_{mj}	\dots	R_{mn}

2) Neighbor Generation

For user u , generate a “nearest neighbor” set according to the level of similarity between neighbors. The calculation of similarity values between neighbors can be performed through vector space similarity calculation methods that are widely used currently, such as cosine method, pearson similarity method and so on. There are two ways to determine neighbors, one is to determine the similarity threshold through cosine method first, and then select users whose similarity values are greater than the similarity threshold as neighbor users; the other is to determine the number of neighbor users N first, and then select the first N users whose similarity values are greater as neighbor users.

3) Recommendation

As “nearest neighbor” set is generated, we can forecast one certain user’s rating for each item, and then make recommendations to that user according to the level of forecasting ratings.

2.2 Problem Analysis

Traditional collaborative filtering algorithm considers users’ entire historical information as its preference information and uses such information to find its nearest neighbors. However, users’ preferences are often formed exploringly and progressively in reality. It is a historical progress, and during this progress, users often try many times and become stable gradually, so as to form their real interests. Even though users’ interests have already been formed, they would sometimes try other items in daily search process for various reasons. Such items cannot be seen as their interests and the supporting evidences for recommendations. Therefore, if we want to gain real preference information of users, we must filter out the occasional search information to reduce interference. Find nearest neighbors according to users’ real preference information, and then make recommendations, while the results of recommendations can become better.

3. A Personalized Recommendation Algorithm Based on Associative Sets

As we have analyzed above, we can gain associative items through frequent itemsets. These associative items constitute the foundations of different interests. As for current users, we use their entire information to filter associative items, and then merge the associative items after filtering to form their interest sets.

3.1 Algorithm Descriptions

1) Set all items as $I = \{I_1, I_2, I_3, \dots, I_n\}$, n as the number of items. See every user’s rating record as one item of transaction, $T_i \subseteq I$, which represents the rating set of user i , wherein $i \in \{1, 2, 3, \dots, m\}$, m represents the number of users. So, user rating matrix can be seen as transaction set $T = \{T_1, T_2, T_3, \dots, T_m\}$.

2) Use Apriori algorithm to generate the frequent itemsets F of transaction set T , whose support level is support.

- In the first iteration of the algorithm, each item of I is a member of the set of candidate 1-itemsets. The algorithm simply scans all of the transactions T in order to count the number of occurrences of each item.

- Select the candidate 1-itemsets, which satisfies minimum support *support*, to consist the set of frequent 1-itemsets L_1 .

- Use $L_1 \times L_1$ to generate a candidate set of 2-itemsets, and prune using apriori property---All nonempty subsets of a frequent itemset must be frequent. Then, scan all of the transactions T in order to count the number of occurrences of each item in candidate set of 2-itemsets.

- Select the candidate 2-itemsets, which satisfies minimum support *support*, to consist the set of frequent 2-itemsets L_2 .

- Constantly use $L_{k-1} \times L_{k-1}$ to generate a candidate set of k -itemsets, and prune it. Then, scan all of the transactions T in order to count the occurrence of each

item in candidate set of k-itemsets. Select the candidate k-itemsets, which satisfies minimum support *support*, to consist the set of frequent k-itemsets L_k .

- If the candidate set of k-itemsets is null, all frequent itemsets are gained.

3) Gain rating items I^* of current user a , and merger the frequent itemsets, which contains some items of I^* and also the number it contains is more than parameter num in F , as associative sets C .

4) Use Pearson correlation coefficient algorithm to calculate the similarity between user a and any other user b in associative sets C .

$$w^C(a, b) = \frac{\sum_{j \in C} (R_{aj} - \bar{R}_a^C)(R_{bj} - \bar{R}_b^C)}{\sqrt{[\sum_{j \in C} (R_{aj} - \bar{R}_a^C)^2][\sum_{j \in C} (R_{bj} - \bar{R}_b^C)^2]}} \quad (1)$$

where j is the item in associative sets C , R_{aj} is the rating given by user a to item j , R_{bj} is the rating given by user b to item j , \bar{R}_a^C and \bar{R}_b^C are average ratings of user a and user b separately.

5) For a , arrange all the users according to the value of $w^C(a, b)$, and select the first M users that have greater values as neighbor users $Neighbor_a$ of user a .

6) Forecast the rating of user a to item j . The forecasting formula is:

$$P_{a,j} = \bar{R}_a + \frac{\sum_{b \in Neighbor_a} w^C(a, b)(R_{b,j} - \bar{R}_b)}{\sum_{b \in Neighbor_a} w^C(a, b)} \quad (2)$$

where $P_{a,j}$ is the forecasting rating of user a to item j , $R_{b,j}$ is the rating of user b to item j , \bar{R}_a and \bar{R}_b are average ratings of user a and user b to all items.

7) Arrange items according to the value of $P_{a,j}$, and select the first N items that have greater $P_{a,j}$ as recommendatory items.

3.2 Algorithm Explanations

1) In addition to the original rating records, there are four other parameters in this algorithm: support for the calculation of frequent itemsets *support*, threshold for the selection of associative sets from frequent itemsets *num*, number of nearest neighbors M and number of items that are recommended to users N . Wherein, *support* and *num* are used to determine Associative Sets, but what is the right combination needs to be tested. Usually, different data sets have different proper combination of *support* and *num*. Therefore, it will take more time to learn this algorithm.

2) Step 1 and step 2 in algorithm description are mainly used to generate frequent itemsets, which will take much more time. However, as it is performed offline, instant recommendations cannot be influenced.

3) As frequent itemsets have to be merged (Step 3) before collaborative filtering, it will take more time online than traditional algorithm will take, but its accuracy can be improved greatly. As the duration of merging frequent itemsets relies on the number of frequent items, to reduce frequent itemsets through offline activities can shorten online duration. In addition, because associative filtering items for every user are somewhat less than all the items, the duration of collaborative filtering process itself will be reduced. Through optimization, online duration of algorithm can be reduced accordingly.

4) Frequent itemsets include the complete set of frequent itemsets, the closed frequent itemsets, maximal frequent itemsets and so on [5]. The frequent itemsets used in this paper are maximal frequent itemsets, which can reduce the number of frequent itemsets greatly. If other frequent itemsets are used, we can calculate the importance of different items when calculating nearest neighbors with the help of support when merging frequent itemsets, which can improve the accuracy further more.

4. Test Process

4.1 Data Set and Evaluation Standard

Data set MovieLens is used to test this algorithm, which is provided by the GroupLens research lab at the University of Minnesota. The data was collected through the MovieLens web site (movielens.umn.edu) during a seven-month period. MovieLens includes 100000 records of ratings given by 943 users to 1682 movies. A rating is a number from 1 to 5, optionally supplemented by the number of seconds which the user spent reading the movie. Users are encouraged to assign ratings based on how much they liked the movie, with 5 highest and 1 lowest. Each user has given ratings to 20 moves at least. You can get the data set at www.grouplens.org.

Average Absolute Error (MAE) is used to evaluate the forecasting accuracy of this algorithm. MAE is the deviation average of the actual value and the predictive value of the ratings given by all users to the items. The lower the value of MAE is, the better the recommendations are. Supposing user rating set is $\{p_1, p_2, \dots, p_N\}$, and the actual user rating set is $\{q_1, q_2, \dots, q_N\}$, MAE is defined as follows [6]:

$$MAE = \frac{\sum_{i=1}^N |p_i - q_i|}{N} \quad (3)$$

4.2 Test Results and Remarks

We will compare Associative Sets Based Collaborative Filtering (ASBCF) and traditional User Based Collaborative Filtering (UBCF) during the test process. In order to verify the results, we will test in two dimensions.

One is to test in different sparse degrees. Here we select the first 200 users and first 500 items in MovieLens rating records, and then deduct some records every time randomly. In the end, we gain rating records under

200*500, 13270 pieces, 7976 pieces, 4525 pieces and 2162 pieces separately, and their sparse degrees are 0.13, 0.08, 0.043, 0.021662 separately, see Figure 1.

From the results, we can see that in different sparse degrees, MAE of ASBCF is lower than that of UBCF, 5.2206% in average. That is to say, ASBCF performs better in every sparse degree than UBCF. However, as the data is sparse extremely, ASBCF's accuracy will be reduced accordingly.

The other is to compare values of MAE with different numbers of nearest neighbors. Here we select 10, 20, 30, 40 and 50 nearest neighbors, and the test results can be seen in Figure 2.

From Figure 2, we can see that ASBCF also performs better than UBCF with all kinds of nearest neighbor numbers. However, along with the increasing of neighbor number, their gap becomes smaller and smaller. Maybe it is because ASBCF can find nearest neighbors more ef-

fectively than UBCF, and when the number of neighbors increases, users that are a little further from current user are also selected, which can increase error. That is to say, ASBCF is more effective than UBCF.

5. Conclusions

This paper proposes a personalized recommendation (collaborative filtering) algorithm based on Associative Sets. It generates a series of frequent itemsets through frequent itemsets generation algorithm, and then filters out some noise items that have little relevance with users by merging, so as to make collaborative filtering algorithm more effective. It is proved that this algorithm is better than traditional algorithm in recommendation accuracy. Although it takes more time to generate frequent items, it will not influence instant recommendations, as the generation can be performed offline. Support of frequent itemsets owns one kind of new information, which represents different items' importance. If such information is used in collaborative filtering, forecasting accuracy will be improved, and this is the breakthrough point for further research.

REFERENCES

- [1] J. B. Schafer, J. A. Konstan, and J. Riedl, "E-Commerce recommendation applications," *Journal of Data Mining and Knowledge Discovery*, pp. 115–153, 2001.
- [2] J. Breese, D. Hecherman, and C. Kadie, "Empirical analysis of predictive algorithms for collaborative filtering," In *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence (UAI-98)*, pp. 43–52, 1998.
- [3] L. Zhao, N. J. Hu, and S. Z. Zhang, "Algorithm design for personalization recommendation systems," *Journal of Computer Research and Development*, pp. 986–991, August 2002.
- [4] Y. Li, L. Liu, and X. F. Li, "Research on personalized recommendation algorithm for user's multiple interests," *Journal of Computer Integrated Manufacturing Systems*, pp. 1610–1615, December 2004.
- [5] J. W. Han and M. Kamber, "Data mining concepts and techniques (Second Edition)," Ming Fang, Xiaofeng Meng translated, China Machine Press, pp. 149–161, March 2007.
- [6] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item based collaborative filtering recommendation algorithms," In *Proceedings of the Tenth International World Wide Web Conference*, pp. 285–295, 2001.

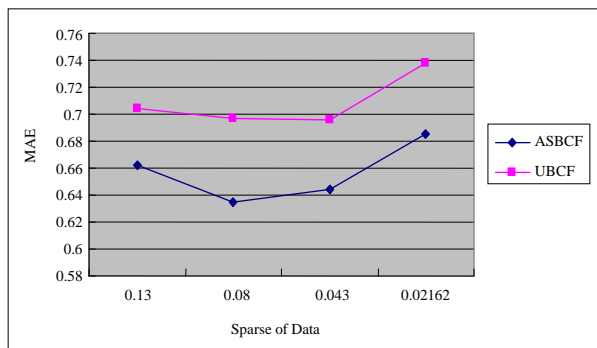


Figure 1. MAE of ASBCF and UBCF under different sparse degrees

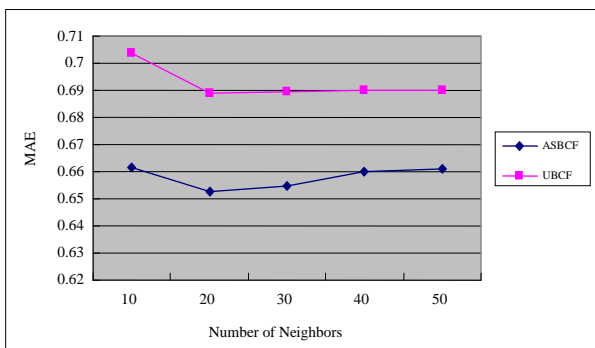


Figure 2. MAE of ASBCF and UBCF with different numbers of nearest neighbors

Gender Differences in Satisfaction with the Type of Work University Employees Do: Evidence from the University of Botswana

Thabo T. Fako¹, Stoffel R. T. Moeng¹, Ntonghanwah Forchheh²

¹Department of Sociology, University of Botswana, Botswana; ²Department of Statistics, University of Botswana, Botswana.
Email: fakott@mopipi.ub.bw, Moengsrst@mopipi.ub.bw, forchehn@mopipi.ub.bw

Received July 15, 2009; revised August 27, 2009; accepted October 3, 2009.

ABSTRACT

The study investigated the extent to which male and female employees of a University differ in various attributes and attitudes and in the level of satisfaction with the type of work they do, and further established factors that might help explain these differences. A stratified random sample of 360 academic and administrative staff of the University of Botswana was collected. Findings indicate that differences between males and females in the level of satisfaction were due to certain negative work experiences such as gender discrimination, tribalism and racism, nepotism and favoritism, and due to certain sources of stress from the immediate supervisor, demands of work on private life and from domestic responsibilities. These negative experiences contributed to lower levels of satisfaction among women than among men. Considering all factors that might explain different job satisfaction among employees at a university, five attributes stood out as significant, namely: belonging to an older age group; understanding the competing demands of teaching and research responsibilities; taking advantage of technological advancement at work; perceiving equity in the distribution of the workload; feeling that there was always enough to do at work.

Keywords: Job Satisfaction, Gender, University Employees, Multivariate Analysis, Botswana

1. Introduction

The main objective of this study was to determine whether or not there were significant gender differences in the level of satisfaction with the type of work that employees of a university do. The second objective was to determine whether or not the difference (if it exists) in satisfaction with the type of work that male and female employees of a university do could be explained by other factors within and outside the work environment. Finally we sought to identify key attributes that could be used to predict the likelihood of job satisfaction among university employees.

In recent years, there has been a consistent call at international, regional, national and institutional levels for gender to be integrated into all decision-making processes [1,2]. In recognition of the importance of gender in the workplace, the University of Botswana established the Gender Policy Programme Committee (GPPC) to create gender awareness and facilitate the incorporation of gender into policies, academic programs and administrative procedures [3]. However, systematic empirical research that seeks

to understand gender differences and inform policy decisions in a university environment is still at its infancy.

Several studies have reported higher levels of job satisfaction among male employees compared to female employees. In a study of academics across eight nations, [4] found that male academics tended to be more satisfied with most aspects of their jobs than female academics. Cooper and Kelly [5] found that significantly more female head teachers in secondary and higher education schools suffered job dissatisfaction than their male counterparts. In Botswana, Maphorisa [6] found that male teachers in secondary schools reported higher levels of job satisfaction than females with the same length of teaching experience.

It has been argued that male workers are more likely to be satisfied with their jobs than female workers [7] especially in organizations where males have more opportunities for advancement or where females are the last to be employed and the first to be fired [8]. In many societies females still generally face greater challenges of access to education, access to mentors and sponsors, and demands related to home and family responsibilities as well

as lack of female role models, all of which could lead to lower job satisfaction.

Amaro *et al.* [9] found that having children was negatively associated with personal and professional satisfaction among Hispanic women professionals, managers and business owners. Sekaran [10] found that men had higher levels of job satisfaction because they spent less time in childcare and household activities and therefore, had more discretionary time to spend on job related activities. Women on the other hand, experienced significantly lower levels of satisfaction when they spend greater amounts of discretionary time on job related activities as opposed to childcare and household activities. Hanson and associates [11] found that rural women with a smaller number of children who devoted resources not needed for child rearing to attaining occupational goals had higher levels of job satisfaction. Demands from both the home and the work environment are likely to affect the performance and satisfaction of male and female employees differently. Given the patriarchal system in which married female employees may have less time to spend on professional self development and research and subsequent publication in favor of domestic activities, it would seem that the scales are tipped in favor of male output, career prospects and job satisfaction.

While most of the literature reviewed suggests that men are more likely to be satisfied with their jobs than women in most occupations, a few studies have found females to be more satisfied with their work than males [12–15]. For example, Rondeau [16] found that female physicians were more satisfied with their jobs than their male counterparts. Gardner *et al.* [17] found that female anaesthetists in South Africa were more satisfied with their jobs than their male counterparts. Other studies have found no differences in the level of job satisfaction between men and women [18,19].

2. Conceptual Framework

Job satisfaction is one of the most widely discussed and enthusiastically studied constructs among scholars in the behavioral and organizational sciences [20]. By the early 1970's, more than 4,000 articles had been published on the topic [21]. Yet, job satisfaction is among the least understood phenomena in organizations today [22].

The concept of job satisfaction consists of at least three dimensions: 1) a subjective perception of the job by the individual; 2) the degree to which the job is perceived as pleasurable; 3) the difference between what employees want out of a job and what they actually get from the job. As a subjective perception, job satisfaction represents a worker's own evaluation of his or her job [23]. It is how the worker feels about his or her work [22] or about various facets of the work environment [24]. Job satisfaction is largely a present-oriented response to the current situa-

tion or conditions at work [25]. It reflects favorable or unfavorable views employees have toward the work itself [26] and aspects of work such as pay, supervision, benefits and so on [15]. It is simply the degree to which a person feels satisfied by their job [27].

Job satisfaction has also been defined as a pleasurable or positive emotional reaction and state of mind resulting from the appraisal of one's job or job experiences [28,29], the tasks one has to perform at work, as well as the physical and social conditions of the workplace [30,31]. It is a positive emotional and cognitive attitude that results from the pleasure a worker derives from component facets of the job [32]. It consists of positive feelings at the end of the job [33], and reflects the degree to which individuals like their jobs [34,35] or simply enjoy their occupations [36], experience a sense of self-growth and accomplishment in their work and would choose the same job again if given an opportunity [37].

Job satisfaction has also been defined in terms of discrepancies between desires and achievements at work. It is as a combination of cognitive and affective reactions to the differential perceptions of what an employee wants to receive compared with what he or she actually receives [38]. It is a function of the perceived relationship between what one wants from one's job and what one perceives the job offers [39], or "what is expected (or desired) and what is received". Job satisfaction is determined by the difference between the amount of some valued outcome that a person receives and the amount of the outcome he feels he should receive [20]. If one expects little and gets little, they would be satisfied. Similarly, if one expects a lot and gets a lot, they would be satisfied. If one expected a lot and got little, they would be dissatisfied.

It has been reported that many employees are not satisfied with the type of job they do [26,40]. Kristiina [41] found that six out of ten workers were planning to leave their current employer for other pursuits within the next two years. In Denmark, Pors and Johannsen [42] found that 25 percent of library employees left their employment before serving three years due mainly to dissatisfaction with their jobs. Dissatisfying job conditions motivate employees to engage in behaviors aimed at reducing frustration and anxiety, and at improving working conditions, the standard of living and equality [43]. Dissatisfied workers may be disruptive, go on strike and may even resort to violence and cause physical harm to other employees [44].

3. The Research Problems

It is known that low levels of job satisfaction can have a negative effect on morale [45], employee relations [26], employee performance, organizational functioning [46], organizational efficiency and productivity. Low job satisfaction or outright dissatisfaction with a job is likely to

evoke an array of negative and potentially damaging personal and professional consequences [28] such as frustration, deterioration of mental and physical health [47], withdrawal, absence, lateness, sickness, accidents [26], intra-organizational conflict [47], thinking about quitting and retirement [48], examining the costs and benefits associated with leaving a job, and labour turnover [49]. Intentions to quit can be very costly to organizations [50]. A recent "Ernst & Young" survey calculated that the cost of replacing a high-level employee may be as much as 150 percent of the departing employee's salary [41].

In educational institutions pressures from academic and organizational reform, working with challenging students, increased staff-student ratios, increased administrative duties, time constraints, economic difficulties, dwindling financial resources from both internal and external sources, funding shortages, etc, increase the level of stress and reduce job satisfaction [51]. In Africa, repressive regimes, poor management, mediocrity and complacency have led to the deterioration of universities and tertiary education in general. For more than two decades, universities and other educational institutions in Africa have, to varying degrees, experienced overstretched financial and physical resources, deterioration of existing facilities and infrastructure, lack of a maintenance culture, insipient decay in the fabric of institutions and a decline in quality and excellence [52]. These conditions are bound to strain human resources and result in low levels of job satisfaction among employees.

In addition, advances in information technologies have resulted in information overload due to the increased use of e-mail and the Internet. While technology is desired and can make work more interesting and satisfying, it has also meant that workers are never really far from the office, and can be reached anytime on cell phones. Those who fail to catch up with technological developments run the risk of frustration and poor performance at work, which may lead to reduced job satisfaction [47].

Given the importance of job satisfaction on employee productivity, it is necessary to periodically take a snapshot of the extent of job satisfaction among employees of major organizations, especially after major changes are introduced in the organization. At the University of Botswana, two such major changes were the restructuring of the administrative system and the restructuring of the academic system through the introduction of *semesterisation* in 2002. It is therefore, opportune to investigate the extent to which employees of a major national institution are satisfied with their job, and to determine the role of gender as a key factor.

4. Methods

The target population for the study comprised of all employees with officially allotted office space. This included all academic staff, all managerial (executive, senior and

middle level administrative) staff including the secretaries and support staff of sections and departments whose names appeared in the university telephone directory. Thus the lower end of the hierarchy (who did not appear in the telephone directory) such as messengers, gardeners, and other junior level staff, were not included in the study. All 1460 staff members were identified within this study population of which 928 worked in academic faculties while the remaining 532 (36.4%) worked in non-academic units.

For sampling purposes, the study population was stratified into enumeration units. The average size of the enumeration units was 11 employees. Broadly, each enumeration unit was a well defined administrative or academic department with a staff establishment of approximately 11 employees (such as the Department of Sociology), a combination of small Departments that had substantially less than 11 employees (such as Educational Technology and Home Economics) or a sub-division of a larger department that had substantially more than 11 employees (such as the Department of English).

The study targeted 25 percent of employees, which translated into approximately 3 staff members from each Enumeration Unit. The 10 to 12 staff members within each unit were numbered sequentially and simple random sample approach used to select three staff members from each unit. In order to minimize non-response, all staff members were formerly notified of the study, and its benefit to them, and two weeks were allowed for data collection, with follow-ups made during the two weeks.

5. Measurement of Variables

A questionnaire was used to record a variety of information about employee attributes and work-related attitudes and experiences. The main variable of our research interest, "satisfaction with the type of work that employees did", was measured by asking each respondent to rate the level of satisfaction with the type of work that he or she did on a scale from zero to 10. Zero represented the lowest level of satisfaction and 10 represented the highest level of satisfaction. For analysis, the level of satisfaction with the type of work employees did was divided into three ordinal categories: 1=High, 2=Moderate and 3=Low, which corresponded to scores of 7-10, 4-6 and 0-3 respectively.

Background variables were measured by appropriately phrased questions on respondent's gender, age, number of dependent children, level of education, type of contract, years of service, etc. Negative experiences at work were measured by asking respondents to indicate on a scale from zero to 10 how often they experienced tribalism and racism, xenophobia, gender discrimination, nepotism and favoritism. Stress from different sources was measured by asking respondents to indicate on a scale from zero to 10 how often they experienced stress from im-

mediate supervisors, domestic responsibilities, demands of work on private life, relations with subordinates, and from the appropriateness of their qualifications. Various aspects of the work environment were measured by questions asking respondents to indicate, on a four-point scale ranging from strongly agree to strongly disagree, the extent to which: technology at work was up to date; they felt misplaced in their job; they experienced competing and conflicting work demands; there was tribalism in recruitment and promotion; they made a valued contribution at work; there was autonomy at work; the level of supervision was adequate; and there was clarity in their job description or responsibilities. For purposes of analyses each of these variables was recoded into an ordinal variable, as shown in attached tables.

6. Data Analysis Methods

Preliminary analysis was directed at determining basic characteristics of the selected employees as well as determining the extent to which they were satisfied with the type of work that they did. Chi-squared tests of association were used to investigate whether there were statistically significant¹ differences between male and female employees with respect to their background characteristics, work-context experiences, the chemistry of social relations at work, sources of stress and satisfaction with the type of work that they did.

Further analysis were aimed at investigating whether the association between gender and satisfaction with the type of work employees did depended on, and could be explained by other factors. For these analyses, variables that were found to have a statistically significant relationship with gender were introduced one at a time as controls for the bivariate relationship between gender and satisfaction with the type of work employees did. The use of controls was important because some of the factors that on their own could explain differences in employee's level of satisfaction were also associated with gender. Suppose for example that older employees tend to be more satisfied than younger employees, but more females tend to be younger employees than males. Then in order to determine the extent to which satisfaction depends on gender, one would need to exclude the mediating effect of age on satisfaction.

One way of doing so (i.e. adjusting or controlling for age) is to do different comparisons of males and females within younger respondents and within older respondents (within the age variable) and for all respondents combined (overall). The Mantel-Haenszel chi-square test of association [53] was employed to assess whether the association between gender and satisfaction with the type of work employees did still remained statistically significant when

controlling for significant background variables, work-context variables and variables measuring the chemistry of relations at work. The decision as to whether an association was statistically significant was measured using the likelihood (the p-value of the test) that the null hypothesis of no association could be true in the general population. The smaller the p-value is, the lower the likelihood that the null hypothesis is true and vice versa. The traditional cut-off of 5 percent ($p=0.05$) was used.

A model for predicting the probability that an employee would be satisfied given their gender, while controlling for the moderating factors such as age, marital status, etc. was developed. Competing models were the ordinal regression and a proportional odds model of the logistic family [54,55], since the dependent variable, "satisfaction with type of work employees do" had three ordinal levels (high, moderate and low).

The proportional odds model involved recoding the dependent variable into two dichotomous variables, Y_1 and Y_2 . Where, for a given employee, Y_1 equal to 1 if the employee's level of satisfaction was high and 0 otherwise; and Y_2 equal to 1 if the employee's level of satisfaction was high or moderate and 0 if it was low. Hence the models provided a framework for comparing employees who were highly satisfied with those who were not, and also comparing employees who were not satisfied (low level of satisfaction) with those that were satisfied (high or medium level of satisfaction). The fitted model, as implemented in SAS Version 8, produces separate intercept parameters and the same regression parameters for the two logits². The model results include estimates of the coefficients of each factor level in the model and their standard error as well as corresponding Wald statistic for determining the level of significance of the factor level, the level of significance of the Wald statistics, the odds ratio and its 95% confidence interval. The odds ratio is particularly useful in comparing the relative likelihood that an employee with a given attribute would be satisfied relative to an employee with a reference attribute. For example, the odd-ratio corresponding to "middle age", gives the likelihood that a middle aged respondent would be satisfied compared to a young employee (the reference category for age), when all other factors in the model are considered. The stepwise variable selection procedure with backward elimination was used to derive the most parsimonious model. The adequacy and goodness of fit of the fitted models were evaluated using the score test and the global Wald statistics respectively.

6.1 Preliminary Findings

The sample consisted of 360 university employees. There were 233 (65.8%) males and 121 (34.2%) females who indicated their gender. Out of 355 who indicated their age, 20.6% were categorized as old (50 years or older),

¹Subsequently, a "significant difference" shall imply "a statistically significant difference" in this paper.

48.5% were categorized as middle aged (35-49 years) and 31.0% were categorized as young (35 years or younger). Many respondents (46.1%) had one or two dependent children, 34.4% had three or more dependent children while 19.4% had no dependent children.

Most respondents (57.9%) were employed on a permanent basis while the remainder (42.1%) were employed on renewable contract (mostly for two years). Respondent's educational background included 37.4% with a doctorate degree, 36.0% with a Masters degree and 26.6% with qualifications below a Masters degree (i.e., post graduate diploma and Bachelor's degree or lesser qualifications). The majority of respondents (84.4%) worked in the six academic Faculties (Business, Education, Engineering, Humanities, Science and Social Science) and 74.4% described their current responsibilities at work as a combination of teaching, research, service and some administrative duties.

The mean rating for satisfaction with type of work respondents did was 7.29 (median=8). When level of satisfaction with the type of work employees did was broken into three ordinal categories, 73.0% of respondents were classified as being highly satisfied with the type of work they did, 21.0% as moderately satisfied and only 6.0% were classified as having a low level of satisfaction with the type of work they did.

6.2 Association between Gender and Other Socio-Demographic Factors

Bivariate analysis began with the determination of whether male and female employees in the sample differed in terms of their background characteristics (socio-economic and demographic). The results (see Table 1) indicate that males and females differed significantly in many socio-economic and demographic characteristics.

It was found that males were significantly older than females ($p < 0.001$). A significantly higher proportion (73.5%) of males was married, compared to 55.9 percent of females. By contrast, 81.0 percent of the females were citizens compared with about 51.9 percent of the males. A significant proportion of males had acquired higher educational qualifications than females ($p < 0.001$). A high percentage of the males (83.8%) had acquired a Masters or doctorate degree compared to 53.7 percent of the females. Other significant differences between men and women were observed regarding number of years working for the university, number of years working in the current position, type of contract, and salary level.

The extent of satisfaction among males was significantly higher than that among females difference was further confirmed ($p = 0.014$) when the level of satisfac-

tion was broken into the three ordinal categories used throughout the study.

While 78.1 percent of the males were highly satisfied with the type of job they did, 63.6 percent of the females were highly satisfied with the type of work they did. This indicated that males were more highly satisfied with the type of work they did than females.

6.3 Bivariate Association between Gender and Satisfaction

After identifying background factors that had a statistically significant relationship with gender, we adjusted for their possible effects on the relationship between gender and job satisfaction by including each one of them as a control or mitigating variable. Table 2 shows a summary of results.

When controlling for age, a statistically significant association between gender and satisfaction with the type of work was observed only among middle-aged employees ($p = 0.023$). While 81.0 percent of middle-aged males were highly satisfied with the type of work they did, a significantly lower percentage (63.0%) of middle-aged females were highly satisfied.

The lack of any gender differences in satisfaction among younger or older respondents meant that when adjusting for age, the association between gender and satisfaction became non-significant statistically ($p = 0.051$).

When controlling for the effects of marital status, gender continued to be associated with satisfaction with the type of work employees did ($p = 0.021$). However, the significant difference was found only between married respondents ($p = 0.010$) and not among single respondents ($p = 0.599$). The results show that while 82.9 percent of married males were highly satisfied with the type of work they did, a significantly lower proportion (64.5%) of married females were highly satisfied. Similarly, males were more likely to be highly satisfied than females when controlling for number of dependent children. The difference was between men and women who had three or more children ($p = 0.008$).

When controlling for years of service to the university, the relationship between gender and satisfaction with the type of work employees did remained significant ($p = 0.024$).

The significant difference between male and female respondents ($p = 0.006$) was found among those with 10 or more years of service. There were no significant differences between male and female respondents who had worked for nine or fewer years at the university. Similarly, a significant difference between males and females ($p = 0.006$) was found among those with 10 or more years in their current position in the university.

²Multiple logistic regression is similar to the usual multiple linear regression model, but is used when the dependent variable is dichotomous instead of continuous.

Table 1. Association between gender and selected background variables

Socio-Demographic Factor	percent Among			Chi-sq	p-value
	Male	Female	Overall		
Age				20.5	<0.001
Young (34 years or less)	23.8	44.6	31.0		
Middle (35 – 49 years)	50.2	44.6	48.5		
Old (50 years and above)	26.0	10.7	20.6		
Marital				10.6	0.001
Single	26.5	44.1	32.1		
Married	73.5	55.9	67.9		
No. of Dep. Children				1.4	0.496
None	20.2	17.4	19.2		
1-2	41.6	38.0	40.4		
3 or more	38.2	44.6	40.4		
Citizenship				28.5	<0.001
Citizens	51.9	81.0	61.9		
Non Citizens	48.1	19.0	38.1		
Education				52.7	<0.001
Doctorate	47.6	18.5	37.4		
Masters	36.2	35.3	36.0		
Post Graduate	2.2	3.4	2.6		
Bachelors	7.9	11.8	9.4		
Other	6.1	31.1	14.6		
Years of Service in UB				8.7	0.013
0 - 3 years	35.5	24.2	31.4		
4 - 9	37.2	53.3	43.1		
10 or more	27.3	22.5	25.5		
Year at Current Position				14.7	0.001
0 - 3 years	42.5	35.5	40.8		
4 - 9	30.0	49.6	36.4		
10 or more	27.5	14.9	22.8		
Overall Working Experience				15.9	0.000
0 - 3 years	26.2	43.8	33.1		
4 - 9	38.2	38.0	37.8		
10 or more	35.6	18.2	29.2		
Type of Contract				28.6	<0.001
Permanent & Pensionable	44.6	73.3	54.7		
Fixed 2 yrs or more	49.8	20.8	39.7		
Temporary Full Time	3.9	4.2	3.9		
Other	1.7	1.7	1.7		
Salary per month				23.4	<0.001
Low (P7000 or less)	29.1	54.6	37.9		
Medium (P7001 – P11,000)	43.6	32.8	39.9		
High (11,001 and above)	27.3	12.6	22.3		

Table 2. Association between gender and satisfaction with the type of work respondents did, controlling for background variables

Factor Controlled for	Level of Satisfaction				Association between gender and Satisfaction within each category		Overall Association Controlling for factor	
	High		Moderate					
	percent Among		percent Among					
	Male	Female	Male	Female	Chi-sq	p-value	Chi-sq	p-value
Age							3.8	0.051
Young	63.6	59.3	29.1	29.6	0.4	0.517		
Middle	81.0	63.0	14.7	29.6	5.2	0.023		
Old	86.7	84.6	10.0	15.4	0.0	0.924		
Marital							5.3	0.021
Single	64.4	61.2	30.5	30.6	0.3	0.599		
Married	82.9	64.5	12.2	27.4	6.6	0.010		
Children							7.7	0.006
None	70.2	66.7	25.5	33.3	0.0	0.959		
1-2	78.4	63.0	15.5	28.3	2.7	0.100		
3 or more	82.0	63.0	14.6	25.9	7.0	0.008		
Citizenship							3.4	0.067
Citizens	72.7	59.2	19.8	32.7	2.8	0.097		
Non Citizens	83.9	82.6	14.3	8.7	0.6	0.440		
Education								0.266
Doctorate	85.3	72.7	10.1	18.2	1.9	0.168		
Masters	75.9	83.3	18.1	16.7	1.9	0.174		
Below MA	62.2	47.3	35.1	38.2	3.5	0.060		
Years at UB							5.1	0.024
0 – 3 years	81.7	69.0	13.4	27.6	1.0	0.327		
4 – 9	67.4	64.1	27.9	29.7	0.3	0.609		
10 or more	87.3	59.3	7.9	25.9	7.5	0.006		
Year at Post							6.3	0.012
0 – 3 years	77.8	67.4	17.2	27.9	0.9	0.330		
4 – 9	72.9	63.3	21.4	28.3	1.3	0.262		
10 or more	84.4	55.6	12.5	27.8	7.7	0.006		
Type of Contract							3.3	0.068
Pensionable	71.8	58.0	20.4	34.1	2.3	0.127		
Contract	83.5	84.0	15.7	4.0	1.1	0.299		
Salary Level							4.1	0.042
Low	68.2	53.9	25.8	35.4	2.8	0.093		
Medium	77.8	82.1	18.2	15.4	0.4	0.551		
High	90.3	53.3	4.8	33.3	8.0	0.005		

Table 3. Association between gender and satisfaction with the type of work respondents did, controlling for characteristics of the work environment

Factor Controlled for	Level of Satisfaction				Association between gender and Satisfaction within each category		Overall Association Controlling for factor	
	High		Moderate					
	percent Among		percent Among					
	Male	Female	Male	Female	Chi-sq	p-value	Chi-sq	p-value
Adequacy of technology							4.9	0.027
Agree	81.9	71.0	14.2	22.6	2.9	0.091		
Disagree	70.1	58.2	23.4	30.9	2.1	0.150		
Feeling Misplaced							3.9	0.049
Agree	58.3	28.0	29.2	48.0	3.8	0.051		
Disagree	80.4	73.7	15.8	22.1	1.21	0.271		
Teaching takes time for research							0.6	0.427
Agree	82.0	75.0	13.2	23.2	0.3	0.611		
Disagree	66.7	62.9	27.1	25.7	0.4	0.532		
Tribalism in recruitment & promotion							4.5	0.034
Agree	63.2	56.5	26.3	34.8	0.1	0.717		
Disagree*	81.0	66.7	15.8	25.0	6.29	0.012		
Making a valued contribution							6.3	0.012
Agree*	78.9	62.9	18.1	29.9	8.9	0.003		
Disagree	66.7	61.9	12.5	23.8	0.0	0.939		
Autonomy at work							6.3	0.012
Agree	79.9	67.0	17.5	28.6	5.1	0.024		
Disagree	70.3	51.7	13.5	27.6	1.4	0.238		
Level of supervision							7.8	0.005
Agree*	79.1	64.5	15.9	27.1	6.4	0.011		
Disagree	75.0	57.1	20.8	35.7	1.4	0.230		
Clarity of responsibilities							7.7	0.005
Agree*	85.7	69.7	13.1	27.0	9.5	0.002		
Disagree	56.1	48.4	28.1	29.0	0.71	0.400		

When controlling for the effects of salary level, gender continued to be associated with satisfaction with the type of work employees did ($p=0.042$). The significant difference ($p=0.005$) between male and female satisfaction was among those with a high salary. While 90.3 percent of males with a high salary were highly satisfied with the type of work they did, only 53.3 percent of the females with a high salary were highly satisfied with their work. There were no significant differences in satisfaction between men and women with a low or medium level salary.

Table 3 shows that the association between gender and job satisfaction remained significant after controlling for

the differences in perceptions of aspects of the work environment. Males who experienced autonomy at work were significantly more satisfied with the type of work they did than their female counterparts ($p=0.024$). Other similarly significant differences in the level of job satisfaction between men and women were among those who felt that: they were making a valued contribution at work ($p=0.003$); the level of supervision at work was adequate ($p=0.011$); there was no tribalism in recruitment and promotion ($p=0.024$); their job descriptions and responsibilities were clear ($p=0.002$). Men and women who did not perceive these positive aspects of the work environ-

ment did not differ significantly in the level of satisfaction with the type of work that they did.

Men tended to be more satisfied than women with the type of work they did regardless of whether or not they felt misplaced at work ($p=0.051$ for those feeling misplaced; $p=0.271$ for those who did not feel misplaced). Similarly, men tended to be more satisfied than women regardless of whether or not they felt that technology at work was adequate ($p=0.150$ for those who felt that technology was not adequate; $p=0.091$ for those who felt that technology was adequate).

There were no significant differences in satisfaction between men and women who felt that teaching did not allow enough time for research ($p=0.611$) and between men and women who felt that teaching allowed enough time for research ($p=0.532$).

Table 4 shows that the association between gender and job satisfaction remained significant after controlling for negative work experiences and sources of stress. Males who often experienced stress from demands of work on private life were more highly satisfied than females who often experienced stress from demands of work on private life ($p=0.006$).

Men were more satisfied with the type of work they did regardless of whether they experienced racism or tribalism ($p=0.019$ for those who often experienced racism or tribalism; $p=0.022$ for those who did not often experience racism or tribalism). Men also tended to be more satisfied than women with the type of work they did, whether or not they experienced xenophobia (overall $p=0.002$).

Men and women who often had negative work experiences had a similar level of satisfaction with the type of work they did. Significant differences in satisfaction were found among employees who did not often have negative work experiences. For example, males who did not often experience gender discrimination were significantly more satisfied with the type of work they did than females who did not experience gender discrimination ($p=0.003$). Other similarly significant gender differences in the level of job satisfaction were among those who did not experience: nepotism and favoritism ($p=0.038$); stress from the immediate supervisors ($p=0.001$); stress from domestic responsibilities ($p=0.039$); stress from relations with subordinates ($p=0.005$) and stress from the inappropriateness of their qualifications ($p=0.005$ for moderately inappropriate qualifications; $p=0.032$ for very inappropriate qualifications).

6.4 Estimating the Probability that an Employee would be Satisfied Given Gender

The stepwise regression procedure was used to analyze the observed differences in satisfaction with the type of

work that respondents did, when controlling for the potential confounders simultaneously. The variables that were removed by the stepwise regression procedure did not significantly add to our ability to predict satisfaction with the type of work respondents did (residual chi-square statistic 26.7; $p=0.370$). The goodness of fit of the proportional odds model was statistically significant (Wald Chi-square=41.5; $df=11$; $p<0.001$ and the Score Test statistic=10.5; $d.f.=11$; $p=0.483$). The estimates of the final prediction model are given in Table 5.

The coefficient of gender in the proportional odds model was not significant, which suggests that, when modeling the effects of several variables simultaneously, there were no significant differences in satisfaction with the type of work that males and females did. The results show that employees in the old age category were almost 3 times more likely to be satisfied with the type of work they did than employees in the young age-group category (Odds Ratio=2.8). Employees who felt that teaching did not allow enough time for research were two-and-a-half times (Odds Ratio=2.6) more likely to be satisfied with the type of work they did than those who felt that teaching allows enough time for research).

Employees who felt that the level of technology at work was up to the standard it should be, were more than twice as likely to be satisfied with the type of work they did than those who felt that the level of technology at work was below standard (Odds Ratio=2.4). Employees who felt that there was equity in the distribution of the work load in their departments were more than twice as likely to be satisfied with the type of work they did than those who felt that there was no equity in the distribution of the workload in their departments (Odds Ratio=2.3, 95percent confidence interval=(1.26, 4.38). Employees who felt that they were underutilized were about 40% as likely to be satisfied with the type of work they did as those who felt that they were given enough to do at work (Odds Ratio=0.4, 95percent confidence interval=(0.20, 0.73). Given the above factors, other variables, including number of years in the organization, relations with subordinates, feeling misplaced at work, etc., were found not to be important predictors of satisfaction with the type of work that employees did.

7. Discussions

This study sought to contribute to an understanding of gender differences in job satisfaction. It builds on studies that have found higher levels of job satisfaction among male employees [4,5,7,56], studies that have found higher levels of job satisfaction among female employees [13,16,17], and studies that have found no difference in the levels of job satisfaction between male and female employees [18,19].

Table 4. Association between gender and satisfaction with the type of work respondents did, controlling for negative work experiences & for sources of stress

Factor Controlled for	Level of Satisfaction				Association between gender and Satisfaction within each category		Overall Association Controlling for factor	
	High		Moderate					
	percent Among		percent Among					
	Male	Female	Male	Female	Chi-sq	p-value	Chi-sq	p-value
Racism and Tribalism							9.2	0.002
Often	74.4	33.3	18.6	50.0	5.6	0.019		
Not Often	79.3	66.7	16.5	25.9	5.3	0.022		
Xenophobia							5.3	0.022
Often	83.8	50.0	10.8	40.0	3.4	0.066		
Not Often	77.2	67.3	18.1	25.0	3.2	0.074		
Gender Discrimination							8.8	0.003
Often	83.3	81.3	16.7	6.3	0.2	0.629		
Not Often	78.1	60.6	17.0	31.7	8.7	0.003		
Nepotism and Favoritism							6.1	0.014
Often	63.2	48.5	29.0	36.4	1.8	0.183		
Not Often	81.0	68.6	15.3	25.6	4.3	0.038		
Immediate Supervisor							4.3	0.038
Often	50.0	65.2	22.7	21.7	1.5	0.221		
Not Often	81.0	63.8	16.7	29.8	10.5	0.001		
Domestic Responsibilities							6.4	0.011
Often	73.0	50.0	17.2	36.7	2.2	0.141		
Not Often	79.2	68.6	17.7	24.4	4.2	0.039		
Demands of Work on Private Life							6.7	0.010
Often	97.1	44.1	14.0	41.2	7.5	0.006		
Not Often	78.0	71.4	18.3	22.6	1.5	0.221		
Relations with Subordinates							9.9	0.002
Often	70.0	42.9	20.0	42.9	2.1	0.144		
Not Often	80.4	65.4	15.9	26.9	7.7	0.005		
Appropriateness of Qualifications							4.5	0.034
High	79.9	82.5	15.9	14.3	0.2	0.619		
Medium	72.7	44.7	25.5	47.4	7.8	0.005		
Low	88.9	33.3	0.0	38.9	4.6	0.032		
Medium	77.8	82.1	18.2	15.4	0.4	0.551		
High	90.3	53.3	4.8	33.3	8.0	0.005		

Table 5. Maximum likelihood estimates of the logistic model for satisfaction with the type of work employees do

Parameter	Estimate	Standard Error	Wald Chi-Square*	P-value	Odds Ratio	95percent CI for odds Ratio	
Intercept 1	-0.102	0.523	0.04	0.846			
Intercept 2	1.892	0.562	11329.00	0.001			
Gender	-0.041	0.340	0.02	0.903	1.0	0.5	1.9
Age							
Middle age	0.580	0.351	2.74	0.098	1.8	0.9	3.5
Old age	1.025	0.489	4.39	0.036	2.8	1.1	7.3
Length of Service							
Worked for 4 to 9 years at UB	-0.603	0.363	2.75	0.097	0.5	0.3	1.1
Worked at least 10 years at UB	-0.562	0.455	1.53	0.217	0.6	0.2	1.4
Stress from Relations with subordinates	0.524	0.551	0.90	0.342	1.7	0.6	5.0
Feeling technology is up to standard	0.886	0.314	7.95	0.005	2.4	1.3	4.5
Feeling under utilized	-0.969	0.337	8.28	0.004	0.4	0.2	0.7
Feeling misplaced in department	-0.544	0.425	1.63	0.201	0.6	0.3	1.3
Feeling equity in distribution of workload	0.854	0.317	7.24	0.007	2.3	1.3	4.4
Teaching does not allow time for research	0.938	0.339	7.63	0.006	2.6	1.3	5.0

The study investigated the level of satisfaction with the type of work that male and female employees of a university did; and sought to explain any differences found. Analyses began by exploring the bivariate association between gender and satisfaction with the type of work respondents did, and was improved by adopting conditional and multivariate analyses that shed more light on the initial association. The study has shown that although the majority of university employees in the sample were highly satisfied with the type of work that they did, males were significantly more satisfied than females.

The study has also shown that male and female employees differ in terms of many factors that affect employee satisfaction such as age, marital status, citizenship, years of service at the institution, and so on. In order to gain more understanding of the relationship between gender and job satisfaction, it was important to determine and then control for these factors in the analyses. This approach provided greater insight into the differences between males and females with certain background characteristics. For example, middle-aged males were more satisfied with the type of work they did than middle-aged females. Married males were more satisfied with their jobs than married females. But there were no differences in job satis-

faction between single males and single females. Males who had 3 or more dependent children were more likely to be satisfied than women who had 3 or more dependent children. Males with a high salary were more satisfied with the type of work they did than females with a high salary.

Males who had worked for 10 or more years at the university were more satisfied with their jobs than women who had worked for the same number of years. This is consistent with findings that workers with less than ten years of work experience have less job satisfaction than those with ten or more years [38]. It is also consistent with studies that have found that job satisfaction increases progressively with length of service [16,57].

Controlling for work environment variables, negative experiences at work and for sources of stress enabled for a deeper understanding of how a positive or negative work environment and work experiences could affect male and female employees differently. It was observed that female employees tended to be less satisfied than males regardless of whether or not they: experienced racism, tribalism or xenophobia; felt misplaced at work; or felt that technology at work was inadequate. It was also observed that females who experienced certain positive aspects of the work environment were significantly less

satisfied with their jobs than their male counterparts. These positive experiences included: autonomy at work; making a valued contribution; adequate level of supervision; lack of tribalism in recruitment and promotion; and having clearly spelt out job descriptions and responsibilities. Men and women who perceived or experienced these positive aspects of the work environment had a similar level of satisfaction with the type of work they did.

The study identified certain negative work-related experiences that affected job satisfaction. These negative experiences included: gender discrimination; nepotism and favoritism; stress from the immediate supervisor; stress from domestic responsibilities; stress from relations with subordinates; and stress from the inappropriateness of one's qualifications. Men and women who often had any of these negative experiences had a similar level of satisfaction with the type of work they did. Thus positive aspects of the work environment as well as negative (stressful) experiences had similar effects on male and female satisfaction with the type of work they did. Only one negative work experience was associated with females being less satisfied with their jobs than males. Females who often experienced stress from demands of work on private life were less satisfied than their male counterparts.

Multivariate analyses, using the likelihood ratio criteria, showed that age-group was the most important predictor of satisfaction with the type of work that respondents did. This is consistent with several studies that found that older workers were more satisfied with their jobs than younger workers [7,57,58]. Other important factors in the multivariate model involved aspects of the work itself. Respondents who felt that teaching did not allow enough time for research were significantly more likely to be satisfied with the type of work they did than those who did not perceive conflict between teaching and research responsibilities. Perhaps those who were actively involved in both teaching and research experienced the competition for time, energy and effort between the two activities of a practicing academic. Such employees would have had an understanding of the importance of research and teaching in the university value and reward systems and may have been rewarded for successfully juggling the two very important aspects of academic life, which is a prerequisite for liking an academic job.

The model further shows that employees who took advantage of technological advancement derived higher satisfaction with the type of work they did. Similarly, employees who experienced equity in the distribution of the workload in their departments were generally more satisfied with their jobs. Those who did not have much to do at work (i.e. those who felt under-utilized) derived less satisfaction with the type of work they did than those who had adequate work to do, regardless of whether they were male or female.

8. Conclusions

Job satisfaction is an important factor for any organization to achieve among its employees. It should help to maintain stability and productivity in the operations of academic institutions and of other organizations. Dissatisfaction with work has been linked to absenteeism, poor performance and staff turnover at United Kingdom universities [59]. It may lead to stress and ultimately to burnout if allowed to continue unabated [60]. It is, therefore, important for university managers and policy makers to identify factors that they can address in order to improve levels of job satisfaction among their workers.

This study addressed the question whether gender can help predict or determine employee satisfaction with the type of work they do. Bivariate analysis suggest that gender is an important factor that affects levels of job satisfaction, and should therefore be taken seriously by managers. Conditional analysis based on adding one control variable at a time shed further light on the relationship between gender and job satisfaction by specifying the conditions under which men and women have different levels of satisfaction. Multivariate analysis showed that gender ranked lower than many other factors in explaining the level of satisfaction with the type of work that university employees did, and the final model did not include gender as a predictive variable.

The most parsimonious model of satisfaction with the type of work that university employees did included the following five predictive attributes: 1) belonging to an older age group; 2) understanding the competing demands of teaching and research responsibilities; 3) taking advantage of technological advancement at work; 4) perceiving equity in the distribution of the workload; 5) feeling that one has enough to do at work. Once the above five factors are given, gender and other variables, including individual background characteristics, measures of personal health and well-being, attributes of the home environment, attributes of the work environment, characteristics of the job itself, and level of job commitment that have been found to be associated with job satisfaction in this and other studies, were no longer important predictors of satisfaction with the type of work that employees do in a university.

The study has made a contribution to research methodology by illustrating that conclusions based on bivariate analyses alone have some limitations. Thus while managers should continue to be sensitive to the importance of gender differences in the workplace, they should be aware that other factors, which they can control or have influence over, determine whether or not employees are satisfied with the type of work that they do, and need to be addressed. Policies aimed at improving job satisfaction should, among other things, ensure that the distribution of the workload is transparent and equitable. Supervisors should regularly evaluate the workload of those who

work under their leadership and ensure that they are not under-utilized or overworked.

This study has several limitations. First, since the study used a cross-section and correlation design that identified significant associations, it cannot be used to make causal inferences among variables. Future studies should adopt longitudinal designs that that would enable cause and effect inferences to be drawn. Second, the study relied on self-reporting and not observations of behavior or in-depth interviews. This reduces the ability to interpret observed associations between variables and improve understanding of the findings. While the findings are interesting, generalizations to university employees and to other professional groups must be investigated further or made with caveats.

REFERENCES

- [1] J. C. Kwesiga and E. N. Ssendiwalu, "Gender mainstreaming in the university context: Prospects and challenges at Makerere University, Uganda," *Women's Studies International Forum*, Vol. 29, No. 6, pp. 592–605, 2006.
- [2] Mookodi and D. Fuh, "Finding the missing male in gender discourses in Botswana," *Pula: Botswana Journal of African Studies*, Vol. 18, No. 1, pp. 31–42, 2004.
- [3] E. M. Alexander, "Beyond inequalities 2005: Women in Botswana," Southern African Research and Documentation Centre, Women in Development Southern Africa Awareness Program; Women's NGO Coalition (Botswana). Botswana WNGOC/SARDC, 2005.
- [4] J. Lacy and B. A. Sheehan, "Job satisfaction among academic staff: An international perspective," *Higher Education*, Vol. 34, No. 3, pp. 305–322, 1997.
- [5] L. Cooper and M. Kelly, "Occupational stress in head-teachers: A national UK study," *British Journal of Educational Psychology*, Vol. 63, No. 1, pp. 130–143, 1993.
- [6] J. K. Maphorisa, "Job satisfaction of Batswana secondary school teachers," Unpublished M Ed Thesis, University of Botswana, 1997.
- [7] O. Siu, "Predictors of job satisfaction and absenteeism in two samples of Hong Kong nurses," *Journal of Advanced Nursing*, Vol. 40, No. 2, pp. 218–229, 2002.
- [8] D. Bendall-Lyon and T. L. Powers, "The impact of gender differences on change in satisfaction over time," *Journal of Consumer Marketing*, Vol. 19, No. 1, pp. 12–23, 2002.
- [9] H. Amaro, N. F. Russo, and J. Johnson, "Family and work predictors of psychological well-being among Hispanic women professionals," *Psychology of Women Quarterly*, Vol. 11, No. 4, pp. 505–521, 1987.
- [10] U. Sekaran, "Understanding the dynamics of self-concept of members in dual-career families," *Human Relations*, Vol. 42, No. 2, pp. 97–116, 1989.
- [11] S. L. Hanson, G. T. Cornwell, G. F. DeJong, and C. S. Stokes, "Consequences of involuntary low parity for women's perception of homemaker and work roles: Findings from a 24-year longitudinal study," *Sociology and Social Research*, Vol. 68, No. 3, pp. 326–349, 1984.
- [12] V. Baxter, "The differential effects of technology, individual attributes, and perceived job rewards on job satisfaction," *Sociological Spectrum*, Vol. 10, pp. 249–269, 1990.
- [13] T. Oshagbemi and C. Hickson, "Some aspects of overall job satisfaction: A binomial logit model," *Journal of Managerial Psychology*, Vol. 18, No. 4, pp. 357–367, 2003.
- [14] E. Terpstra and A. L. Honoree, "Job satisfaction and pay satisfaction levels of university faculty by discipline type and geographical region," *Education*, Vol. 124, No. 3, pp. 528–539, 2004.
- [15] Thilmany, "Formula for job satisfaction," *American Society of Mechanical Engineering-CIME*, Michigan, 2004.
- [16] K. V. Rondeau and L. H. Francescutti, "Emergency department overcrowding: The impact of resource scarcity on physician job satisfaction," *Journal of Healthcare Management*, Vol. 50, No. 5, pp. 327–340, 2005.
- [17] S. V. Gardner, M. F. M. James, and N. R. Evans, "Gender issues among South African anesthetists," *South African Medical Journal*, Vol. 92, No. 9, pp. 732–736, 2002.
- [18] J. K. Eskildsen, K. Kristensen, and A. H. Westlund, "Work motivation and job satisfaction in the Nordic countries," *Employee Relations*, Vol. 26, No. 2, pp. 122–136, 2004.
- [19] A. Schnee and F. Reitman, "The impact of gender as managerial careers unfold," *Journal of Vocational Behavior*, Vol. 47, No. 3, pp. 290–315, 1995.
- [20] J. Cranny, P. C. Smith, and E. F. Stone, "Job satisfaction: How people feel about their jobs and how it affects their performance," New York. Lexington books, 1992.
- [21] T. T. Fako and N. Forchheh, "Job satisfaction among nurses in Botswana," *Journal of the Sociological Association of South Africa*, Vol. 31, No. 1, pp. 10–21, 2000.
- [22] R. W. Rowden and C. T. Conine, "The impact of workplace learning on job satisfaction in small US commercial banks," *Journal of Workplace Learning*, Vol. 17, No. 4, pp. 215–230, 2005.
- [23] S. M. Donohue and J. S. Heywood, "Job Satisfaction and gender: An expanded specification from the NLSY," *International Journal of Manpower*, Vol. 25, No. 2, pp. 211–238, 2004.
- [24] A. Freund, "Commitment and job satisfaction as predictors of turnover intentions among welfare workers," *Administration in Social Work*, Vol. 29, No. 2, pp. 5–21, 2005.
- [25] L. Evans, "Understanding teacher morale and job satisfaction," *Teaching and Teacher Education*, Vol. 13, No. 8, pp. 831–845, 1997.
- [26] W. Moyle, W. Skinner, W. Rowe, and C. Gork, "Views of Job satisfaction and dissatisfaction in Australian long-term care," *Journal of Clinical Nursing*, Vol. 12, No. 2, pp. 168–176, 2003.
- [27] J. Garcia-Bernal, A. Gargallo-Castel, M. Marzo-Navarro, and P. Rivera-Torres, "Job satisfaction: Empirical evidence of gender differences," *Women in Management Review*, Vol. 20, No. 4, pp. 279–288, 2005.
- [28] A. D. Koustelios, "Personal characteristics and job satisfaction of Greek teachers," *International Journal of Edu-*

- cational Management, Vol. 15, No. 7, pp. 354–358, 2001.
- [29] Petty, E. W. Brewer, and B. Brown, "Job satisfaction among employees of a youth development organization," *Journal of Child and Youth Care Forum*, Vol. 34, No. 1, pp. 57–73, 2005.
- [30] M. M. Gruneberg, "Understanding job satisfaction," Macmillan Press, London, 1979.
- [31] R. (Jr.) Schermerhorn, J. G. Hunt, and R. N. Osborn, "Basic organizational behavior," John Wiley & Sons, New York, 1995.
- [32] V. M. Dendinger, G. A. Adams, and J. D. Jacobson, "Reasons for working and their relationship to retirement attitudes, job satisfaction and occupational self-efficacy of bridge employees," *International Journal of Aging and Human Development*, Vol. 61, No. 1, pp. 21–35, 2005.
- [33] A. Aksu and A. Aktas, "Job satisfaction of managers in tourism: Cases in the Antalya region of Turkey," *Managerial Auditing Journal*, Vol. 20, No. 5, pp. 479–488, 2005.
- [34] L. A. Flowers, "Job satisfaction differentials among African American faculty at 2-year and 4-year institutions," *Community College Journal of Research and Practice*, Vol. 29, No. 4, 317–328, 2005.
- [35] S. Tyson and T. Jackson, "The essence of organizational behavior," Prentice Hall, New York, 1992.
- [36] Pollard, "Job satisfaction among newswriters: The influence of professionalism, perceptions of organizational structure, and social attributes," *Journalism and Mass Communications Quarterly*, Vol. 72, No. 3, pp. 682–697, 1995.
- [37] T. Oshagbemi, "The influence of rank on the job satisfaction of organizational members," *Journal of Managerial Psychology*, Vol. 12, No. 8, pp. 511–519, 1997.
- [38] C. M. Hellman, "Job satisfaction and intent to leave," *The Journal of Social Psychology*, Vol. 137, No. 6, pp. 677–689, 1997.
- [39] D. Lund, "Organizational culture and job satisfaction," *Journal of Business & Industrial Marketing*, Vol. 18, No. 3, pp. 219–236, 2003.
- [40] L. Levy-Garboua and C. Montmarquette, "Reported job satisfaction, what does it mean," *Journal of Socio-Economics*, Vol. 33, No. 2, pp. 135–151, 2004.
- [41] K. Hyrkas, "Clinical supervision, burnout, and job satisfaction among mental health and psychiatric nurses in Finland," *Issues in Mental Health Nursing*, Vol. 26, No. 5, pp. 531–556, 2005.
- [42] N. O. Pors and C. G. Johannsen, "Job satisfaction and motivational strategies among library directors," *New Library World*, Vol. 103, No. 1177, pp. 199–208, 2002.
- [43] R. J. Burke, "Organizational values, work experiences and satisfactions among managerial and professional women," *Journal of Management Development*, Vol. 20, No. 4, pp. 346–354, 2001.
- [44] P. E. Spector, "Job satisfaction: application, assessment, causes and consequences," Sage Publications, CA, 1997.
- [45] H. C. Koh and E. H. Y. Boo, "Organizational ethics and employee satisfaction and commitment," *Management Decision*, Vol. 42, No. 5, pp. 677–693, 2004.
- [46] R. W. Gimbel, S. Lehrman, M. A. Strosberg, V. Ziac, J. Freedman, K. Savicki, and L. Tackley, "Organizational and environmental predictors of job satisfaction in community based HIV-AIDS services organizations," *Social Work Research*, Vol. 26, No. 1, pp. 43–55, 2002.
- [47] S. A. Sparks, K. J. Corcoran, L. A. Nabors, and C. A. Hovanitz, "Job satisfaction and subjective well being in a sample of nurses," *Journal of Applied Social Psychology*, Vol. 35, No. 5, pp. 922–938, 2005.
- [48] L. B. Rosenfed, J. M. Richman, and S. K. May, "Information adequacy, job satisfaction and organizational culture in a dispersed-network organization," *Journal of Applied Communication Research*, Vol. 32, No. 1, pp. 28–54, 2004.
- [49] M. E. Capella and J. D. Andrew, "The relationship between counselor job satisfaction and consumer satisfaction in vocational rehabilitation," *Rehabilitation Counseling Bulletin*, Vol. 47, No. 4, pp. 205–214, 2004.
- [50] Harrington, N. Bean, D. Pintello, and D. Mathews, "Job satisfaction and burnout: Predictors of intentions to leave a job in a military setting," *Administration in Social Work*, Vol. 25, No. 3, pp. 1–16, 2001.
- [51] P. Krueger, K. Brazil, L. Lohfeld, H. G. Edward, D. Lewis, and D. E. Tjam, "Organisation specific predictors of job satisfaction: Findings from a Canadian multi-site quality of work life cross-sectional survey," *BMC Health Services Research*, Vol. 2, No. 6, pp. 1–6, 2002.
- [52] Teferra and P. G. Altbach, (eds.), "African higher education: An international reference handbook," Indiana University Press, Bloomington, 2003.
- [53] J. L. Fleiss, "Statistical methods for rates and proportions," John Wiley and Sons, New York, 1981.
- [54] P. McCullagh and J. A. Nelder, "Generalized linear models," Second Edition, Chapman & Hall, London, 1989.
- [55] W. N. Venables and B. D. Ripley, "Modern applied statistics with S-Plus," Springer, New York, 1999.
- [56] A. Crossman and B. Abou-Zaki, "Job satisfaction and employee performance of Lebanese banking staff," *Journal of Managerial Psychology*, Vol. 18, No. 4, pp. 368–376, 2003.
- [57] C-K. Cheung and S. A. Scherling, "Job satisfaction, work values and sex differences in Taiwan's organization," *Journal of Psychology*, Vol. 133, No. 5, pp. 563–575, 1999.
- [58] K. B. Ang, C. T. Goh, and H. C. Koh, "The impact of age on the job satisfaction of accountants," *Personnel Review*, Vol. 22, No. 1, pp. 31–39, 1993.
- [59] T. Oshagbemi, "Personal correlates of job satisfaction: empirical evidence from UK universities," *International Journal of Social Economics*, Vol. 30, No. 12, pp. 1210–1232, 2003.
- [60] L. C. Pearson and W. Moomaw, "The relationship between teacher autonomy and stress, work satisfaction, empowerment and professionalism," *Educational Research Quarterly*, Vol. 29, No. 1, pp. 38–54, 2005.

A Study of the Joint Advertising Channels

Ming LEI, Shuguang SUN, Dan YANG

Department of Management Science and Information System, Guanghua School of Management, Peking University, Beijing, China.
Email: leiming@gsm.pku.edu.cn, sunshuguang@gsm.pku.edu.cn, yangdan@gsm.pku.edu.cn

Received July 22, 2009; revised September 1, 2009; accepted October 11, 2009.

ABSTRACT

The study of joint advertising channels decision-making is a very difficult issue. It mainly focuses on ads investment game between a manufacturer and a retailer in a vertical supply chain. With the rise of the game theory, scholars have been starting to investigate this issue in the frame of game theory in recent years. The paper improves the existing models and introduces single-period and multi-period modified models. The paper obtains the closed-form solutions to the single-period model and the simulation results of the multi-period model. We also examine the implications of these results and obtain some insights into the real practice.

Keywords: Joint Advertising Channel, Game Theory, Simulation Method

1. Introduction

The cooperative advertising has been an important and interesting issue in the channel management literature. The paper adopts the frame of the two-level supply chain: the upstream firm (a manufacturer) and the downstream firm (a retailer). When pushing a new product into market, both the manufacturer and the retailer need to advertise. Meanwhile, the manufacturer will provide the retailer some incentives. In this paper, we investigate how much the manufacturer and the retailer invest on the ads respectively and how much the manufacturer should pay for the retailer as a subsidy.

Actually, the study of the cooperative advertising has a long history. However, prior to the application of game theory, the papers had been viewing this question as a pricing problem, not a game. The examples of such kind are Jeuland and Shugan, Moorthy, Ingene and Parry [1-3].

Currently, more and more scholars adopt the frame of game theory to study the cooperative advertising. All these studies differ from each other by the hypothesis on forms of the ads expense and subsidy rate and by the game type.

As for the form of the ads expense and subsidy rate, Dant and Berger assumed that the manufacturer provided the retailer a constant subsidy for per product sold [4]. Doraiswamy *et al.* also adopted such an assumption [5]. Under this hypothesis, the manufacturer charges a lower wholesale price. Bergen and John proposed that the manufacturer and retailer share the ads cost together [6]. Under this assumption, the manufacture compensates the

retailer for the ads cost. However, they thought that the retailer chooses the total amount of ads. It is somewhat unreasonable because the manufacturer usually hold the leading position nowadays. In the study of Steffen, Simon and Georges, ads are divided into two different types: long and national, short and local. And it is also assumed that the manufacturer and retailer have the right to make ads of these two kinds. The manufacturer compensates the retailer for the ads costs via two constant subsidy rates, respectively [7]. But this assumption is a little complicated because the retailer is declined to make the ads of the first kind. Bergen and John distinguished between national ads and local ads [6]. Zhimin Huang and Susan distinguished between ads made by the manufacturer and ads made by the retailer. And they also assumed that the retailer is compensated for her ads investment [8].

Most papers adopt the complete information static or dynamic game models. For example, Zhimin Huang and Susan used Cournot and Stackelberg model, while other papers only adopted Stackelberg model where the manufacturer acts as the leader. Although the static game model is simple to deal with, the cooperative advertising has a long-term effect therefore a multi-period game is necessary [8].

The paper employs the dynamic programming method to study the multi-period dynamics model. Unfortunately, there are many limitations on the application of such method so few papers have adopted the multi-period dynamics model. Steffen distinguished the “long-term and national” ads from the “short-term and local” ads. This

paper assumed that the ads of the first kind have a long-term effect: the ads affect the manufacturer's reputation and the market demand later. To solve the model, the paper adopted some simplified function forms therefore made the demand equation unconvincing [7]. To obtain a better understanding of cooperative advertising in the frame of game theory, the paper provides a single-period and multi-period models, on the basis of Huang and Susan, and Steffen. We get a closed-form solution to the single-period model and simulate the multi-period model. In the end, the paper explains the results from economics perspective.

2. Modeling

2.1 Single-Period Model

On the basis of Huang and Susan, and Steffen, the paper introduces a game theory model of the joint ads [7,8]. This model is a Stackelberg model with the manufacturer as the leader.

2.1.1 Players

There are two players: the manufacturer and the retailer. The players are rational persons with complete information. The decision variables of the manufacturer are the ads investment (b) and subsidy rate (t). The decision variable of the retailer is her ads investment (a).

The subsidy the manufacturer compensates for the retailer is . Therefore, the profit functions are:

$$\pi_m = \rho_m Q - ta - b \quad (1)$$

$$\pi_r = \rho_r Q - (1-t)a \quad (2)$$

where m and r stand for the manufacturer and the retailer respectively, ρ stands for the profit ratio subtracting the ads cost, and Q is the sales volume.

2.1.2 Relationship between Market Demand and Ads Investment

On the basis of Huang and Susan, and Steffen, we assume that the demand function satisfies:

$$Q(a, b) = (\alpha - \beta a^{-\gamma} b^{-\delta}) \sqrt{G} \quad (3)$$

where G is the reputation of the manufacturer, q and a are the ads investments of the manufacturer and retailer respectively, t is the subsidy rate, and $\alpha, \beta, \delta, \gamma$ are all positive constants.

In addition, we assume that the reputation of the manufacturer satisfies the following dynamics equation:

$$\Delta G = \lambda b - \mu G \quad (4)$$

These two equations are explained as follows:

1) The demand function implies that the market demand is affected by two factors: the reputation and cur-

rent ads investment. The maximum value of the current market demand is and can be affected by the reputation [7,8]. It is widely assumed that the marginal revenue of the ads decreases, so the paper does not adopt the linear function.

2) In the reputation dynamics equation, the paper adopts the form of exponential decay. The model of such kind is generally used in depicting decay trend [9–11].

2.1.3 Payoffs

It is reasonable to propose that the objective of the retailer is to maximize her current profits. Although the reputation of the manufacturer has a long-term effect, the retailer cannot gain profits brought by the long-term effect. Admittedly, to enhance the manufacturer's reputation will increase the long-term sale volume therefore benefit the retailer indirectly. However, the profit gained by the manufacturer far outweighs that gained by the retailer. Meanwhile, the manufacturer can take several measures to grab away these indirect profits gained by the retailer [12]. For example, the manufacturer can change retailers or increase the commission. So the payoff of the retailer is π_r .

The objective of the manufacturer is to maximize the payoff:

$$\text{Max: } w = (1 - \phi) \tau \Delta G + \phi \pi_m \quad (5)$$

where τ is a positive constant, adjusting the units of the two terms in (5). The manufacturer is concerned about both the current profits and the reputation. The manufacturer can benefit from the current product sales and the enhancement of reputation: the enhancement of reputation can increase the product sales in future. So the manufacturer wants to maximize G and π simultaneously [13]. However, these two goals conflict because to increase G means to advertise more therefore to reduce current profits. The paper introduces a weight factor ϕ to reconcile the conflict. And ϕ is exogenous in the paper. It is somewhat unreasonable. In the second model later in the paper, we discuss how the manufacturer selects the most proper ϕ .

2.2 Multi-Period Model

In this multi-period model, there are still two parties: the manufacturer and the retailer and they are rational person with complete information. And similar to the single-period model, this model is a Stackelberg model with the manufacturer as the leader.

The hypothesis of the market demand, profit functions, reputation dynamic equation is the same with that of the single-period model. The only difference is that we need a footnote n to denote the n th period. For example, the reputation dynamics equation is

$$\Delta G = \lambda b_n - \mu G_{n-1}$$

or

$$G_n = \lambda b_n + (1 - \mu) G_{n-1}$$

The goal of the retailer is still to maximize its profit at each period, i.e. $\pi_r(n)$. However, the goal of the manufacturer changes a little, for the long-term profit is in the consideration of the manufacturer.

2.2.1 Goal of the Manufacturer

We suppose that the goal of the manufacturer is

$$\text{Max: } W = (1 - \phi) \tau G_N + \phi \sum_{n=0}^{N-1} \left((1 + \kappa)^{N-n-1} \pi_{mn} \right) \quad (6)$$

where n stands for the n th period, k means the discounting rate, N is the total periods, G_N stands for the long-term interests, $\phi \sum_{n=0}^{N-1} \left((1 + \kappa)^{N-n-1} \pi_{mn} \right)$ is the current profits at the end of N th period [13,14]. ϕ is the weight factor adjusting the units of the two terms in (6). Different from the single-period model, we do not introduce ϕ_n for n th period. Instead, the paper adopts an overall weight factor ϕ for the model. The manufacturer needs to choose ϕ_n for each period.

2.2.2 Hypothesis of the Long-Term Strategies

To make the model solvable, the paper introduces two hypotheses.

Hypothesis 1: Let ω be the weight factor for each period. And the N -period decision-making question can be converted into a question selecting the optimal ω -path. What we need to do is just to solve the single-period optimization problem with the weight factor ω in each period. All these single-period solution under ω -path consist the set Ω . So another statement of this hypothesis is that if there is a solution sequence to question (6), the solution sequence must lie in the set Ω .

This hypothesis is reasonable because it is naturally accepted that there exists a weight factor ω to adjust the long-term and short-term interests. Actually, this is our hypothesis in the first model. And this hypothesis is tenable in the management practice: if the long-term interests conflicted with the short-term interests, the managers would allocate resources between them based on a weight factor.

The role of the hypothesis 1 is to convert a complicated multi-period optimization question into a route choice and single-period optimization questions. It is difficult to employ the dynamics programming method to solve the multi-period optimization question mathematically. That is why there are few papers that adopted the multi-period game theory model to deal with the cooperative ads question. It is true that some papers did adopt

dynamics programming, but these papers made many unreasonable assumptions. In contrast, this paper does not sacrifice the reasonability for the simplification. We just confine the solution to the question into some forms. Fortunately, the hypothesis is reasonable from management practice perspective.

Hypothesis 2: the optimal ω -path is monotonous and convergent to ϕ .

This hypothesis is to restrict the types of the optimal ω -path making the question solvable. It is reasonable to assume that the optimal ω -path is convergent to ϕ : ω and ϕ are two measures weighing the long-term and short-term interests. At the end of the multi-period game, the short-term measures should approximate the long-term measures. And when $N=1$, ω must be equal to ϕ .

Why do we assume that the optimal ω -path is monotonous? In the real practice, how to choose ω depends on the strategy of the firm. ω means that the firm how to balance the market against profits. Generally speaking, there are three types of the strategies:

1) Market-leading strategy: in prophase, the goal of the firm is to sacrifice profits for expending market. So $\omega < \phi$ in prophase. In anaphase, ω Increases and approximates to ϕ .

2) Profit-leading strategy: contrary to the market-leading strategy, the profit-leading strategy seeks profits in prophase. So $\omega < \phi$ at this stage, and ω decreases and approximates to ϕ in anaphase.

3) Equalitarian strategy: the firm has no preference towards profits or market. During the whole process, $\omega = \phi$.

Obviously, the hypothesis is justified under these three strategies.

3. Analysis of the Models

3.1 Analysis of the Single-Period Model

For this model, we can get the closed-form solution. We substitute (3) into (2) and then get the first order condition of the retailer:

$$\frac{d\pi_r}{da} = \gamma \rho_r \beta \sqrt{G} b^{-\delta} a^{-(\gamma+1)} - (1-t) = 0$$

$$\text{Therefore,} \quad a = \left(\frac{\gamma \rho_r \beta \sqrt{G}}{(1-t)b^\delta} \right)^{\frac{1}{\gamma+1}} \quad (7)$$

Next, we substitute (7) into (1) and get:

$$\pi_m = \rho_m \alpha \sqrt{G} - \eta (1-t)^{\gamma/(\gamma+1)} b^{-\delta/(\gamma+1)} - \epsilon t (1-t)^{-1/(\gamma+1)} b^{-\delta/(\gamma+1)} - b \quad (8)$$

$$\text{where} \quad dw = (1 - \phi) \tau \lambda db + \phi d\pi_m = 0,$$

$$dw = (1 - \phi)\tau\lambda db + \phi d\pi_m = 0.$$

The first order condition for the manufacturer is:

$$dw = (1 - \phi)\tau\lambda db + \phi d\pi_m = 0 \quad (9)$$

$$\text{So, } b^* = \left[\left(\frac{\delta\phi}{\phi - \tau\lambda + \phi\tau\lambda} \right)^{\gamma+1} \gamma^{-\gamma} \beta (\rho_m - \gamma\rho_r) \sqrt{G} \right]^{1/(\delta+\gamma+1)},$$

$$\text{when } \tau\lambda < \frac{\phi}{1-\phi} \quad (10)$$

$$b^* = +\infty, \text{ when } \tau\lambda \geq \frac{\phi}{1-\phi} \quad (11)$$

$$t^* = \frac{\rho_m - (\gamma+1)\rho_r}{\rho_m - \gamma\rho_r}, \text{ when } \frac{\rho_m}{\rho_r} > \gamma+1 \quad (12)$$

$$t^* = 0, \text{ when } \frac{\rho_m}{\rho_r} \leq \gamma+1 \quad (13)$$

We notice that when $(1-\phi)\tau\lambda$ is large enough ($\tau\lambda(1-\phi) \geq \phi$), the manufacturer will increase the ads investment as much as possible. $1-\phi$ is the marginal value of the reputation and λ is marginal reputation per ads. $\tau\lambda(1-\phi) \geq \phi$ means it is more profitable for the manufacturer to increase the ads investment. Therefore, the manufacturer will spare no effort to produce. (12,13) demonstrate that the manufacturer will not compensate for the retailer as an incentive, until the marginal profit of the manufacturer is large and the marginal profit of the retailer is small ($\rho_m/\rho_r > \gamma+1$), for the manufacturer benefits more from ads than the retailer does.

3.2 Analysis of the Multi-Period Model

As discussed above, the retailer holds the same behaviors as in the single-period model. As for the manufacturer, under the hypothesis 1, the N-period optimization problem is converted into a problem selecting the optimal ω -path and a series of single-period problems. And these single-period questions are almost the same with that in the first model. The only difference is that the parameter is ω , rather than ϕ .

Hence, the key part of the analysis of the multi-period model is how to select the optimal ω -path. We will use the simulation method to solve this problem [13].

3.2.1. Selecting the Optimal ω -Path

Selecting the optimal ω -path is a part of the decision of the manufacturer. In the real practice, the manufacturer will calculate the profit W of the final period and then compares the profits of the all possible paths to select one to maximize W . Now, we use computer programs to select the optimal ω -path.

According to the hypothesis 2, the process of selection is divided into two steps:

Step 1: given the origin of the path, we try various approaching methods.

Step 2: select another origin, and repeat step 1.

As Figure 1 demonstrates, we experiment five different paths given two origins. They correspond to the three types of strategy respectively: the two paths above the horizontal line correspond to the profit-leading strategy; the two paths below the horizontal line correspond to the market-leading strategy; the horizontal line corresponds to the equalitarian strategy. The arrows mean changing the origins. Also as shown in Figure 1, we adopt two approximating methods: linear approximation and exponential approximation [13]. To be specific, the exponential approximation employs the function:

$$\omega = \phi - \Delta\phi * \frac{e^{-x} - e^{-1}}{1 - e^{-1}}, \quad x = \frac{n}{N-1}$$

Of course, these two approximation methods fail to cover all possibilities. However, we just want to find an approximate solution, not the precise solution. And the approximate solution is accurate enough for the real practice. If greater accuracy is needed, we can try more approximation functions and the origins.

3.2.2 Parameters Valuating

There are thirteen parameters: $\alpha, \beta, \gamma, \delta, \rho_m, \rho_r, \lambda, \mu, \phi, \tau, \kappa, N, G_0$. Their meanings and values in the simulation are demonstrated in the following Table 1.

These parameters can be divided into three groups:

Demand and profit: $\alpha, \beta, \gamma, \delta, \rho_m, \rho_r$

Market indicator: λ, μ

Long-term strategy: $\phi, \tau, \kappa, N, G_0$

It is true that to get the best evaluation, we need to adopt the data in the real economy. However, there are too many limitations for us to do so. Another difficulty is that several parameters always change in the real practice, for example, ρ , but in the paper, they are viewed as constants.

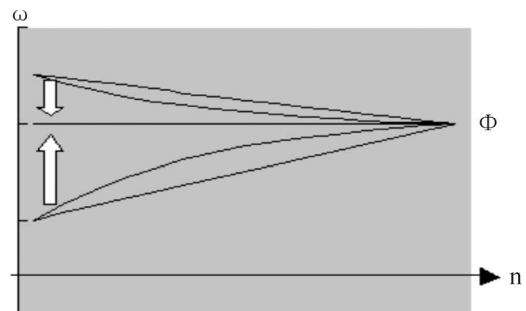


Figure 1. Various approaching methods

Table 1. Parameters in the multi-period model

parameter	value	implication
α	10	market capacity
β	5	ads effect coefficient
γ	0.5	ads effect coefficient of the retailer
δ	0.3	ads effect coefficient of the manufacturer
ρ_m	0.18	manufacturer's profit
ρ_r	0.1	retailer's profit
λ	2	reputation effect coefficient of the manufacturer
μ	0.2	reputation decay coefficient
ϕ	0.8	profit weight of the manufacturer
κ	0.06	subsidy rate of the manufacturer
τ	1	conversion coefficient of reputation into profit
N	10	total periods of the game
G_0	10	initial reputation of the manufacturer

Hence, we only can estimate them on the basis of the common sense and some constraints in the sense of mathematics and economics.

1) Ads investment is of medium term decision. Firms always make a decision every half or a year. The manufacturer's ads are national and large-scaled lasting one year; the retailer's ads are local and short-term lasting a quarter. Therefore, we assume that the each period lasts half a year.

2) The product life cycle is always 5 years so we assume that $N=10$. N is an important factor affecting the long-term strategy. We will relax the restrictions on N later in the paper.

3) The subsidy rate should be equal to the capital cost. We assume that the capital cost is 12%. Because each period lasts half a year, the subsidy rate is 6% each period.

4) The ads effect coefficient of the retailer γ is larger than that of the manufacturer δ . It means that the retailer's ads have more influence in local market than the manufacturer's. The ads of manufacturer take effect in terms of reputation increase.

5) According to (12), $\rho_m / \rho_r > \gamma + 1$ should hold.

3.2.3 Optimal Long-Term Strategy:

Now we study what long-term strategy the manufacturer should choose in the multi-period game and how the parameters affect the optimal long-term strategy. According to the classification of parameters, $\phi, \tau, \kappa, N, G_0$ affect the long-term strategy. In addition, λ, μ may influence the strategy because they affect the market demand.

By simulation, we find out that the game equilibrium paths of the market-leading and profit-leading strategies are as follows:

- 1) The equilibrium path of market-leading strategy;
- 2) The equilibrium path of profit-leading strategy.

With the evaluation of the parameters, the optimal strategy is the market-leading strategy (see Figure 2 and Figure 3). And we also adjust the value of the seven parameters that affect the long-term strategy slightly, the following results are gained:

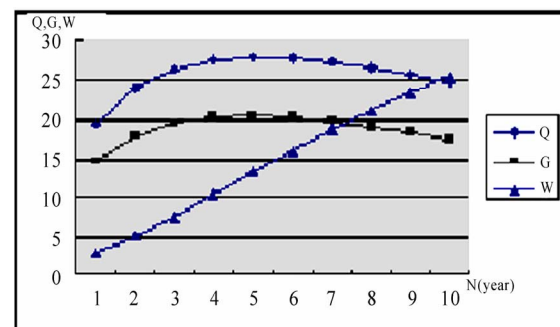


Figure 2. Demand, reputation, and manufacturer's profit (market-leading)

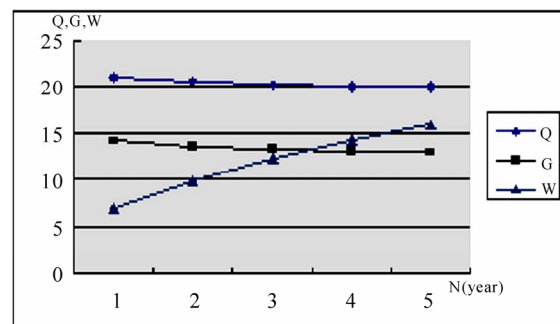


Figure 3. Ads and profit (market-leading)

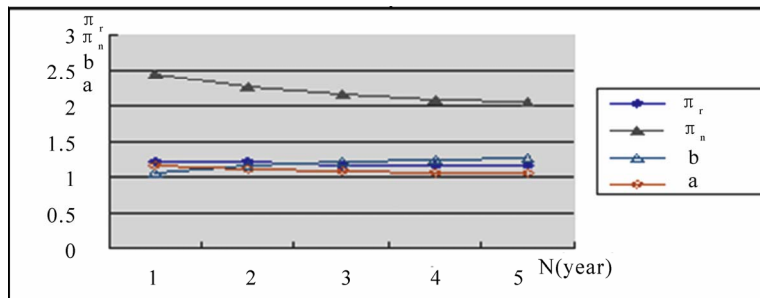


Figure 4. Demand, reputation, and manufacturer's profit (profit-leading)

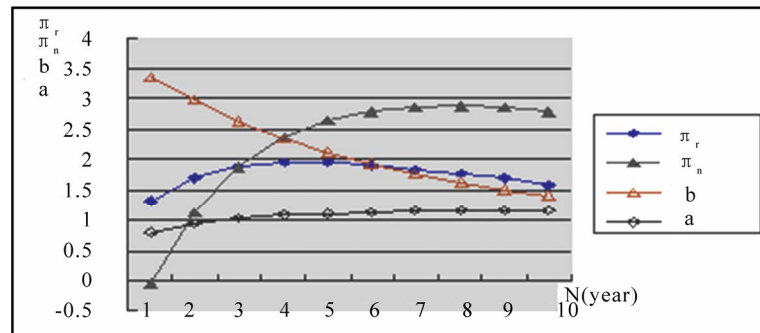


Figure 5. Ads and profit (profit-leading)

1) In most cases, the optimal strategy is the market-leading strategy.

2) In many cases where we had thought the profit-leading strategy is dominant, it proves not to be the optimal, although the final profit W is almost the same.

3) In few cases where the profit-leading strategy is dominant (see Figure 4, Figure 5), the value of N is small.

These conclusions seem a little astonishing. However, there are indeed profound implications behind them. In the next section, we will discuss these results in detail.

3.2.4 Parameters' Effect on Game Equilibrium

In this section, we study the parameters $\gamma, \delta, \rho_m, \rho_r, \lambda, \mu$ how to affect the game equilibrium when the manufacturer operates according to the optimal strategy.

The following are the results of the simulation, where the values of a, b, t, π_m, π_r are mean values of all periods. The value of W is calculated at the final period. The benchmarks are from Table 2.

In the process of simulation, we find out that if a parameter affects the game equilibrium, the effect is monotonous. So we only experiment those cases with increasing values of the parameters.

4. Discussion

Now we explore the implications of the results above and get some suggestions on marketing management. We ne-

ed to point out that the assumptions on profit functions, demand functions, and ads enhancement equation are the same in the single-period model and the multi-period model. These two models take the long-term effect of ads into consideration. So we can gain some results about the sales profit ratio and ads long-term effect.

4.1 Sales Profit Ratio and Ads Long-Term Effect

In the models of this paper, the retailer's ads have only short-term effect and the effect coefficient is γ ; the manufacturer's short-term ads coefficient is δ . The sales profit ratios of the retailer and manufacturer are ρ_r and ρ_m respectively.

According to the closed-form solution (7) and (10,11) to the single-period model, the short-term ads effect coefficients (γ and δ) do not have a monotonous influence on the ads investments of the manufacturer and retailer. When the ads have more influence on the sales profits, the more ads will be made. Simultaneously, the more the ads affect the sales profits, the less ads investment is needed to achieve the given goal. Table 2 shows that the ads investment of the retailer and manufacturer (a and b) increase in γ and δ .

The manufacturer's sales profit ratio ρ_m is positively correlated with her ads investment. This relationship is supported by the closed-form solution (10,11)

Table 2. Parameters' effect on game equilibrium

	a	b	t	π_m	π_r	W
Benchmark	1.07	2.17	0.23	2.22	1.75	25.50
$\gamma=0.6$	1.13	1.92	0.17	2.35	1.54	26.64
$\delta=0.4$	1.01	2.70	0.23	2.48	2.23	28.62
$\rho_m = 0.2$	1.18	2.38	0.33	2.81	2.01	31.76
$\rho_r = 0.11$	1.04	2.11	0.12	2.29	1.85	26.20
$\lambda=2.2$	1.16	3.43	0.23	2.74	2.68	30.76
$\phi = 0.85$	1.04	1.30	0.23	1.87	1.10	22.34
$\mu=0.3$	0.99	2.03	0.23	1.61	1.39	18.17

and the Table 2. When the sales profit ratio increases, the manufacturer will sell more products and therefore make more ads. From the simulation results, we find out that the retailer's ads investment increases in the sales profit ratio. It is because the increase of the sales profit ratio leads to the increase of the compensation for the retailer's ads and hence to increase the retailer's ads investment.

Now, we turn our attention onto the subsidy rate. According to the closed-form solution (12,13) to the single-period model, $t = 1 - 1 / (\rho_m / \rho_r - \gamma)$. So t increases in ρ_m / ρ_r and decreases in γ . It is to say: the smaller the manufacturer's sales profit ratio is compared with the retailer's sales profit ratio and the larger the retailer's ads effect coefficient is, the lower the manufacturer will compensate for the retailer. This conclusion is also supported by the simulation results of the multi-period model. The conclusion can be explained as follows:

The product sales benefit two players therefore they have incentives to advertise to increase the product sales. However, their profit-cost structures are different. The player (the manufacturer) who benefits more from ads wants another player (the retailer) to advertise more by taking measures to spur her. When the difference between their sale profit ratios becomes larger, the manufacturer will pay more for the retailer. And if the retailer has a large ads effect coefficient, she does not need many ads to achieve her sales goal or the retailer has strong incentives to advertise. So the manufacturer will need not to pay much to the retailer.

The closed-form solution (13) also shows an exception: when $\rho_m / \rho_r \leq \gamma + 1$, the manufacturer will not pay for the retailer. It is because in this situation, the retailer gains a high profit or is expert in advertising. So the retailer may pay for the manufacturer conversely.

Consequently, the ads investment of the manufacturer

is positively correlated with her sales profit ratio. The ads investment of the retailer is positively correlated with her sales profit ratio and the subsidy rate from the manufacturer. The subsidy rate is positively correlated with the quotient of the two sales profit ratios and negatively correlated with the retailer's ads effect coefficient.

4.2 Reputation

In the models, the reputation functions via enhancing the market capacity indirectly and the reputation is affected by the ads. The relevant variables are $\lambda, \mu, 1 - \phi$. According to the solution (10,11) to the single-period model, ϕ affects the manufacturer's ads monotonously: the increase of ϕ will lead to the decrease of the ads investment of the manufacturer b . This relationship is also shown in the simulation solution to the multi-period model: when ϕ increases from 0.8 to 0.85, b decreases from 2.17 to 1.30. It is because ϕ is the weight factor of profit and profit conflicts with the market which needs ads support. If the manufacturer wants to increase the current profit, the investment for the long-term ads must be reduced.

As shown above, when λ becomes larger, the manufacturer will advertise more. This conclusion is supported by the solution (10,11) to the single-period model and the simulation results of the multi-period model: when λ increases from 2.0 to 2.2, b increases from 2.17 to 3.43. This result is explained as follows:

The manufacturer's ads have two effects: enhancing current product sales and enhancing the reputation and therefore promoting future sales. In the models, these two functions do not conflict. When λ becomes larger, the manufacturer can gain more reputation. On the other hand, the current ads effect stays the same. Hence the manufacturer will advertise more.

The parameter μ is the decay coefficient of the reputation. From the simulation results of the multi-pe-

riod model, it is found out that when μ increases, the average profits of the manufacturer and retailer decrease. It is because the market increases slower or decays faster in this situation.

According to (11), when $(1-\phi)\tau\lambda \geq \phi$, the manufacturer will advertise as much as possible. Why does it happen? $(1-\phi)\tau$ is the marginal profit of the reputation and λ is the marginal reputation of the ads, so $(1-\phi)\tau\lambda$ is the marginal value of the ads. Meanwhile, ϕ is the marginal value of profit. The inequality $(1-\phi)\tau\lambda \geq \phi$ means that ads can bring the manufacturer more value than current profits do.

4.3 Long-Term Game Strategy

The long-term game strategy is a concept that the paper introduces to explain the simulation results of the multi-period model. This concept is employed to determine the weight factor of profit each period.

Generally speaking, there are several variables to affect the long-term strategy: $\mu, \phi, \kappa, N, G_0$. However, it seems that only the variable N affects it according to the simulation results. Actually, the market-leading strategy is called the market penetration strategy and profit-leading strategy is called market skimming strategy. According to the marketing theory, if a firm plans to develop in long term, the market penetration is optimal. The firm should sacrifice profit in prophase to increase the market share, or else, the firm adopts the policy of "high price and high profit" in prophase and then withdraws from the market. Of course, there are no metaphase or anaphase plans in this situation.

That is why the optimal long-term strategy is always the market-leading strategy in the multi-period model when N is large. There is no market withdrawal mechanism in the models. So when the game duration is long, the manufacturer has to adopt the market-leading strategy. This analysis justifies assumptions and the simulation method in models.

5. Conclusions

In this paper, we explore the interactions of ads investments between the manufacturer and retailer. Two types of game theory are employed to achieve our goal. In the first model: single-period model, we get a closed-form solution while in the multi-period model, we have to resort to the simulation methods to get some numerical solutions. The paper provides some interesting conclusions. The manufacturer's reputation has influence on the market demand and it is affected by the long-term ads investment only by the ads investment of the manufacturer, while the increase of the reputation is beneficial to the retailer too. At the same time, the manufacture must

pay for the retailer to spur her ads investment. In such a context, the interaction between these two firms is complex.

The main results of the paper are:

- 1) The manufacturer's ads investment is positively correlated with her sales profit ratio and the retailer's ads investment is positively correlated with her sales profit ratio and the subsidy rate from the manufacturer.
- 2) The smaller the manufacturer's sales profit ratio is compared to the retailer's sales profit ratio and the larger the retailer's ads effect coefficient is, the lower the manufacturer will compensate for the retailer.
- 3) The increase of the manufacturer's profit weight and will lead to the decrease of the ads investment of the manufacturer, while as reputation effect coefficient becomes larger, the manufacturer will advertise more.
- 4) When the game duration is long, the manufacturer will adopt the market-leading strategy.

6. Future Study

6.1 Improving the Models

In this paper, we adopt the simulation method that imposes little limitation on modeling. We can introduce more complicated models: 1) there are more than two players; 2) the assumptions on the long-term ads effect are a little simple: attributing all long-term effect factors to the manufacturer's reputation. The future work can take other factors into consideration; 3) the future researches can introduce the information asymmetry into models. For example, the manufacturer does not know the sales profit ratio and the ads effect coefficient of the retailer, or the retailer knows little about the product quality and after service of the manufacturer.

6.2 Applying the Models and Simulation into Specific Situations

The simulation relies heavily on the valuation of the parameters. If we can get data from the firms in real practice and then simulate on the basis of the data, the results will be more convincing. The implications of these new results will bring us several new findings.

REFERENCES

- [1] A. P. Jeuland, and S. M. Shugan, "Managing channel profits," *Marketing Science*, Vol. 2, No. 3, pp. 239–272, 1983.
- [2] K. S. Moorthy, "Managing channel profits: Comment," *Marketing Science*, Vol. 6, No. 4, pp. 375–379, 1987.
- [3] C. A. Ingene and M. E. Parry, "Channel coordination when retailers compete," *Marketing Science*, Vol. 14, No. 4, pp. 360–377, 1995.
- [4] R. P. Dant and P. D. Berger, "Modeling cooperative ad-

- vertising decisions in franchising,” *Journal of the Operational Research Society*, Vol. 49, No. 9, pp. 1120–1136, 1996.
- [5] K. Doraiswamy, T. W. McGuire, and R. Staelin, “An analysis of alternative advertising strategies in a competitive franchise framework,” *Proceedings of Educator's Conference of American Marketing Association*, pp. 463–467, 1979.
 - [6] M. Bergen and G. John, “Understanding cooperative advertising participation rates in conventional channels,” *Journal of Marketing Research*, Vol. 34, No. 3, pp. 357–369, 1997.
 - [7] S. Jørgensen, S. P. Sigue, and G. Zaccour, “Dynamic cooperative advertising in a channel,” *Journal of Retailing*, Vol. 76, No. 1, pp. 71–91, 2000.
 - [8] Z. Huang and S. X. Li, “Co-op advertising models in manufacturer-retailer supply chains: A game theory approach,” *European Journal of Operational Research*, Vol. 135, No. 3, pp. 527–544, 2001.
 - [9] Deal and R. Kenneth, “Optimizing advertising expenditures in a dynamic duopoly,” *Operations Research*, Vol. 27, No. 4, pp. 682–692, 1979.
 - [10] Chintagunta, K. Pradeep and D. Jain, “A dynamic model of channel member strategies for marketing expenditures,” *Marketing Science*, Vol. 11, No. 2, pp. 168–188, 1992.
 - [11] Chintagunta and K. Pradeep, “Investigating the sensitivity of equilibrium profits to advertising dynamics and competitive effects,” *Management Science*, Vol. 39, No. 9, pp. 1146–1162, 1993.
 - [12] K. C. Sen, “Advertising fees in the franchised channel,” *Journal of Marketing Channels*, Vol. 4, No. 1, pp. 83–101, 1995.
 - [13] S. P. Wang, K. L. Judd and C. Gung, “Solving a savings allocation problem by numerical dynamic programming with shape-preserving interpolation,” *Computers & Operations Research*, Vol. 27, pp. 399–408, 2000.
 - [14] L. Tesfatsion, “Introduction to the special issue on agent-based computational economics,” *Journal of Economic Dynamics and Control*, Vol. 25, No. 3–4, pp. 281–293, 2001.

Grey Incidence Relation Analysis and Granger Causality Tests of the Income Level and Economic Growth – Case Study on Gansu Province, China*

Bing XUE^{1,2}, Xingpeng CHEN^{2*}, Weiwei ZHANG², Jing WANG², Xiaojia GUO², Yong GENG¹

¹Circular Economy and Industrial Ecology Group, Institute of Applied Ecology, Chinese Academy of Sciences, Shenyang, China;

²College of Earth and Environmental Sciences, Lanzhou University, Lanzhou, China.

Email: xuebing.china@yahoo.com.cn

Received July 14, 2009; revised August 30, 2009; accepted October 7, 2009.

ABSTRACT

In order to keep the economic growing, the Chinese government released series of public policies with regard to stimulate consumption and expand domestic demand. This paper, based on the series data of GDP, Per Capita Annual Disposable Income of Urban Households (PCAD), and Per Capita Annual Net Income of Rural Households (PCAN) of Gansu province from 1978 to 2007, analyzed the relationship and causality of the PCAD and PCAN to GDP by using the methodologies called Grey Incidence Relation and Granger Causality Tests. The outcomes show that: the incidences relation of PCAD and PCAN to GDP are prominent, and the trend of the prominent concerning PCAD to GDP is climbing; the PCAD and PCAN are the Granger causality to GDP, which means the GDP could increase 0.7337% unit due to the 1% unit increase of PCAN. Instead, the GDP only could increase 0.4817 % unit due to the 1% unit increase of PCAD. The conclusion indicates that to improve the net income of rural households is a priority selection to stimulate the economic growth, and the governments should rethink the role of the farmers and the agriculture issue.

Keywords: Grey Incidence Relation, Granger Causality Test, Income Level, Economic Growth, Gansu Province

1. Introduction

Over the past 30 years, from the beginning of economic reforms in 1978-2007, China has experienced a steady and high economic growth [1], of, measured in gross domestic product (GDP), on average 9.8% per year. In 2008, due to the change of worldwide economic situation resulted from the global financial crisis, the economic growth rate reduced to 9.0% [2], and would continue to reduce to 6.5% in 2009 [3]. China's future economic growth has received much attention. What could be the driving forces of China economic growth in future? In 2009, Chinese government, in order to keep the economic growing, has released series of public policies with regard to stimulate consumption and expand domestic demand, In the theory concerning the regional development, one of the research issues is the relationship between the per capita income and regional economic growth [4-6]. Some scientists thought that the consump-

tion payout of the households is the driving force of the regional economic growth, and instead, some others thought that the relation between per capital income and regional development is interaction [5-9].

With the goal oriented to figure out the relationship between per capita income and economic growth, this paper takes Gansu province as a sample, based on the methodologies called Grey Incidence Relation Analysis (ab. GRA) and Granger Causality Test (ab. GCT). We studied the incidence relation of the per capita annual disposable income of urban households (ab. PCAD), and per capita annual net income of rural households (ab. PCAN) to regional gross domestic product (ab. GDP), followed by clarifying the causality of the PCAD's growth to GDP growth, as well as the PCAN's to GDP's.

2. Research Area and Indices Chosen

2.1 Research Area

Gansu province is in the northwestern China, which located at 92.13-108.46 E and 32.31-42.57 N. The land-forms in Gansu are complicated and varied. There are

*Sustentation Fund: National Natural Science Foundation of China [No.: 40871061]. Corresponding Author: Prof.Dr.CHEN Xingpeng.

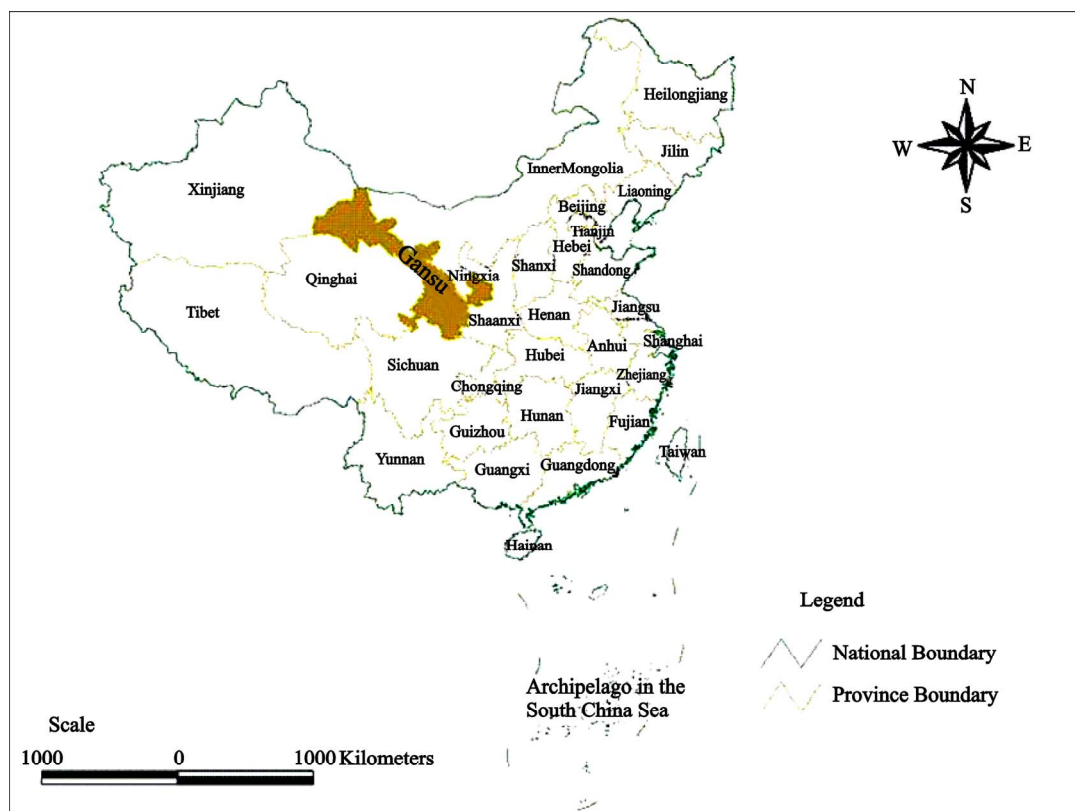


Figure 1. Sketch map of Gansu province

450 rivers in Gansu, among which 78 rivers have a yearly runoff of over 100 million cubic meters each. Gansu has a semiarid climate with plenty of sunshine, strong radiation and the temperature varies greatly from day to night. The annual average temperature is between 0°C-14°C dropping from the southeast to the northwest.

Gansu has been a multi-ethnic province since ancient times. Among its total population of 26 million, the minority population takes up 2.199 million. With non-ferrous metals, energy, petrochemicals, machinery and electronics, building materials, food and textile as its mainstay, Gansu has a relatively reasonable and complete industrial system. Although modern industry appeared laggardly and had a weak fundament, nowadays it develops fast.

2.2 Indices Chosen

This paper chose the following three indices as analyzed objectives: per capita annual disposable income of urban households, and per capita annual net income of rural households, and regional gross domestic product. The temporal series of the indices is from 1978 to 2007 and all the data sourced from <Statistic Yearbook for New China of 55 Years (Branch of Gansu Province)>, and <Gansu Statistic Yearbook: 2006, 2007, and 2008>.

3. Temporal Differences of the Incidence Relation

3.1 Methodology: Grey Incidence Relation Analysis

The grey incidence relational analysis (ab. GRA) applies to explore the qualitative and quantitative relationships among abstract and complex sequences and to capture their dynamic characteristics during the development process [10]. The GRA could make use of the relatively small data sets and do not demand strict compliance to certain statistical laws, simple or linear relationships among the observable variables [10, 11]. Thus, the GRA can analyze a grey system that is of poor, incomplete and with uncertain information. The modeling processes are as following [11]:

Step 1: To establish the reference series and comparative series.

Reference Series: $X_0^{(k)} = [x_0^{(1)}, x_0^{(2)}, x_0^{(3)}, \dots, x_0^{(k)}]$;

Comparative Series: $X_i^{(k)} = [x_i^{(1)}, x_i^{(2)}, x_i^{(3)}, \dots, x_i^{(k)}]$;
 $(i=1, 2, 3, \dots, n)$;

Step 2: To remove anomalies associated with different measurement units and scales by dimensionless processing, such as the initial-value processing, which is appro-

prate for data that varies with time. In the initial-value processing, the elements in each sequence is divided by the first component.

$$X_j^{(k')} = X_j^{(k)} / X_j^{(0)}$$

Step 3: To calculate the relational coefficient $L_{0i}^{(k)}$.

$$L_{0i}^{(k)} = (\Delta_{\min} + \lambda \Delta_{\max}) / (\Delta_i^{(k)} + \lambda \Delta_{\max})$$

Wherein : $L_{0i}^{(k)}$ is to used as expressing the relative distance between two factors,

$$\Delta_i^{(k)} = |X_0^{(k)} - X_i^{(k)}|, (i=1,2,3,\dots,n);$$

Δ_{\min} is the minimum number amongst the series $\Delta(k)$, and Δ_{\max} is the maximum number amongst the series $\Delta(k)$. λ is the distinguishing coefficient used to adjust the difference of the relational coefficient, usually, λ equals to 0.5 in a grey system.

Step 4: To calculate the grey relational grade R_{0i} .

Here, assuming each point has a sequence of equal weight, thus, the grey relational grade R_{0i} equals to the average of $L_{0i}^{(k)}$. The relational grades are numerical measures of the influence of factors on the reference values, and have numeric values between 0 and 1. Generally, $R_{0i} > 0.6$ indicates a notable incidence under the assumption that $\lambda = 0.5$.

The grey relational grade (R) is simultaneously computed corresponding to each performance characteristics. It reveals the relative variations between two factors indicating magnitude and gradient in a given system.

3.2 Outcomes of the Calculation Based on GRA

This paper takes the historical data of GDP from 1978 to 2007 as reference series, historical data PCAD & PCAN both from 1978 to 2007 as comparative series. According to the calculation processes shown in 3.1, the results are

as following: (See Table 1)

The results show that, during the period from 1978 to 2007, both of the two grades are bigger than 0.6, which means PCAN and PCAD have notable incidence to GDP. The GRA grade of PCAN (R_{PCAN}) to GDP is 0.864, higher than the GRA grade of PCAD to GDP (R_{PCAD}) which is 0.814.

Compared the R_{PCAN} with R_{PCAD} in different periods (Figure 2), we can found that the trend of R_{PCAN} is durative climbing, and the R_{PCAD} appears wave shape. Particularly, during the period of 1978-1989 and 1990-1999, PCAN has incidence relation to GDP. PCAD is not related to GDP during the period of 1990-1999.

Due to the grey relational grade could only indicate the incidence relation between two series but can not to clarify the causality, thus, we did further research to study the causality between PCAN and GDP, and PCAD to GDP as well.

4. Granger Causality Test and the Outcomes

4.1 Methodology: Granger Causality Test

Granger causality is a technique for determining whether one time series is useful in forecasting another. A time series X is said to Granger-cause Y if it can be shown, usually through a series of F-tests on lagged values of X (and with lagged values of Y also known), that those X values provide statistically significant information about future values of Y [12,13]. The function is as following:

$$F(X_{n+1} | \Omega_n) \neq F(X_{n+1} | (\Omega_n - Y_n))$$

Here, Ω_n means all the information concerning X and Y .

Table 1. Results of the GRA grade of PCAD, PCAN to GDP

	1978-2007	1978-1989	1990-1999	2000-2007
R_{PCAN}	0.864	0.487	0.591	0.684
R_{PCAD}	0.814	0.648	0.552	0.707

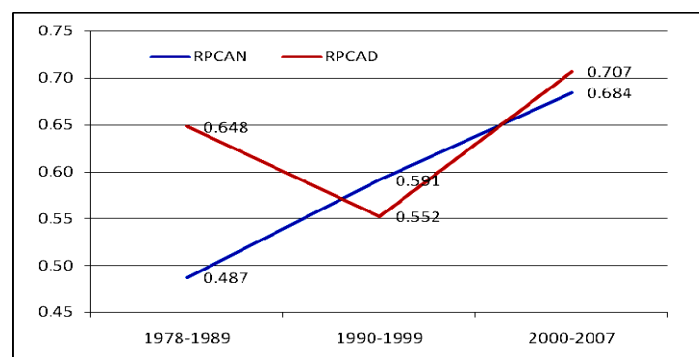


Figure 2. Temporal analysis of the GRA grade

Table 2. Causality tests of GDP to PCAN and PCAD

Null Hypothesis	L	F-Sta.	Prob.
GDP does not Granger Cause PCAN	3	0.711	0.557
PCAN does not Granger Cause GDP		5.006	0.010
GDP does not Granger Cause PCAD	4	0.488	0.744
PCAD does not Granger Cause GDP		6.420	0.002
GDP does not Granger Cause PCAD	3	0.633	0.602
PCAD does not Granger Cause GDP		3.236	0.044
GDP does not Granger Cause PCAD	4	0.697	0.604
PCAD does not Granger Cause GDP		4.233	0.015

Table 3. Results of the generalized difference analysis

Variables	Coeff.	Std. Err	T-Sta	Prob
LNRURAL	0.7337	0.1043	7.0332	0.0000
LNURBAN	0.4817	0.0876	5.4978	0.0000
C	-2.4008	0.3138	-7.6493	0.0000
AR(1)	1.1820	0.1960	6.0306	0.0000
AR(2)	-0.4604	0.1976	-2.3300	0.0289
R-squared	0.9987	Mean dependent var		6.0219
Adjusted R ²	0.9985	S.D. dependent var		1.1488
S.E. of regression	0.0452	Akaike info criterion		-3.1942
Sum squared resid	0.0470	Schwarz criterion		-2.9563
Log likelihood	49.7188	F-statistic		4350.749
DW stat	1.7002	Prob(F-statistic)		0.0000

4.2 Outcomes from the Tests Results

Based on the software *Eviews 3.2*, we tested the causality. The results are in Table 2 below.

The results show that, considering the notable lever equals to 10%, assuming the lags equals to 3 years or 4 years, under these scenarios, both the PCAN and PCAD are the Granger Cause GDP, which means there is single-causality between the two variables and GDP.

In order to quantitative analyze the influence degree of the two variables and GDP; we use the Generalized Difference Analysis to find out the influence degree by using the software *Eviews 3.2* under the program of Granger Causality Tests. See chapter 4.3

4.3 Generalized Difference Analysis and the Outcomes

Before analyzing the differences, in order to remove the errors caused by the self-related series, we processed the

series by taking the following change as following function:

$$X_t^* = X_t - \rho X_{t-1}$$

The complete results are in following Table 3.

The function of PCAD, PCAN, and GDP is:

$$LNGDP = (-2.4008) + (0.7337 * LNRURAL) + (0.4817 * LNURBAN)$$

Based on this function, we can find that, from 1978 to 2007, 1% unit increase of PCAN result in 0.7337% increase of GDP, but compared to PCAD, 1% unit increase of that only result in 0.4817% increase of GDP.

5. Conclusions

This paper initiated to indicate the relationship between the economic growth and income level, and to find the scientific foundation to public policy making. The main conclusions are as following:

GRA and GCT are appropriate methodologies to clarify the relation and grade. This research shows that the

GRA could be used to find out the incidence relation between the variables and the GCT is appropriate to test the causality and find out the influence degree. Hereby, based on the working progress when we studied the methodologies, we would like to point out seriously that the incidence relation does not equals to causality, which means, we cannot define the causality relation only according to the calculation results based on GRA.

In the past years, the urban inhabitants shared the major outcomes due to the “Reform and Open” policy from 1978. On national level, averagely, per capita annual net income of rural households improved from 343.4 Yuan to 13,786 Yuan, instead, per capita annual net income of rural households improved from 133.6 Yuan to 4,140 Yuan, the gap between PCAN and PCAD in 2007 is 3.33 times. Nowadays, Chinese government released series of public policies with regard to improve the citizen income level, which aims to move forward the economic growth. This research indicates that to improve the rural household’s income level is an urgent issue for China. The outcomes also proved that the national policy undertaking in the coming years of “improve the rural income and expand the domestic consumption” has scientific base and should be taken exactly in the next decade.

REFERENCES

- [1] Z. G. Wen and J. N. Chen, “A cost-benefit analysis for the economic growth in China,” *Ecological Economics*, Vol. 65, pp. 356–366, 2008.
- [2] A. H. Carsten, “China’s economic growth 1978–2025: What we know today about china’s economic growth tomorrow,” *World Development*, Vol. 36, pp. 1665–1691, 2008.
- [3] World Bank, “China 2009 growth forecast to 6.5%,” *China Quarterly Update*, pp. 9, March 18, 2009.
- [4] R. Clarke, “More evidence on incomer distribution and growth [J],” *Journal of Development Economics*, Vol. 47, No. 2, pp. 403–427, 1995.
- [5] W. X. Zhou, “Case study of the relation between household’s income and economic growth [J],” *Economic Science*, Vol. 1, pp. 40–47, 2002.
- [6] N. Zhao and Y. B. Wei, “Research on the income level and economic development,” *Rural Economy, Science and Technology*, Vol. 19, pp. 26–27, 2008.
- [7] E. Podrecca and G. Carmeci, “Fixed investment and economic growth: New results on causality,” *Applied Economics*, Vol. 33, pp. 177–182, 2001.
- [8] Y. Li and A. Oberheitmann, “Challenge of rapid economic growth in china: Reconciling sustainable energy use, environmental stewardship and social development,” *Energy Policy*, Vol. 37, pp. 1412–1422, 2009.
- [9] T. Persson and G. Tabellini, “Is inequality harmful for growth?” *Theory and Evidence, American Economic Review*, Vol. 84, pp. 600–621, 1994.
- [10] T. Persson and G. Tabellini, “Economic growth,” *Quarterly Journal of Economics*, Vol. 77, pp. 465–490, 1994.
- [11] S. He, Y. Li, and R. Z. Wang, “A new approach to performance analysis of ejector refrigeration system using grey system theory,” *Applied Thermal Engineering*, Vol. 29, pp. 1592–1597, 2009.
- [12] G. X. Xu, “Statistic, forecast and decision making,” Shanghai: Shanghai Finance and Economy University Press, pp. 207, 2005.
- [13] Y. F. Cao, “Reviews on granger causality test,” *World Economic Research*, Vol. 2, pp. 16–21, 2005.

The Development of Services in Customer Relationship Management (CRM) Environment from ‘Technology’ Perspective

Gaurav K. Agrawal¹, Daniel Berg²

¹Decision Sciences and Engineering Systems, Rensselaer Polytechnic Institute, Troy, USA; ²Decision Sciences and Engineering Systems, Rensselaer Polytechnic Institute, Troy, USA.

Email: gkagrawal@engineer.com, bergd@rpi.edu

Received May 15, 2009; revised July 2, 2009; accepted August 21, 2009.

ABSTRACT

The service sector is receiving much deserved attention resulting from its inevitable role in a country's economic development. Despite all the efforts gaps such as the relationship between technological advances and service development are yet to be revealed from the perspective of new applications that organizations want to develop and implement. This paper explores opportunities using a comprehensive model (and CRM, as an example) that can be used to extend the research relating service development to the technology development aspects of the market.

Keywords: Service Systems, Service Hierarchy, Customer Relationship Management, Service Development, Technology Complexity

1. Introduction

Presently, extremely competitive and globalized markets demand economic globalization as a ‘must do’ activity for the competitors to maintain their niche in the (service) market. Technology development is seen as a solution, allowing organizations to enhance their service portfolios using latest technological advancements. Also, new technology development triggers the service development process to be more customer intensive with customized service offerings. With the shorter maturity times of (different services and hence) the service organizations [1], it became evident for businesses to pursue the technology development and its implementation aspects at a faster pace. It is established that ‘technology’ plays an inevitable role in the service development process [2–4].

With the increasingly changing perception (of customers) and fierce competition (among business organizations) at the international level, organizations lean towards using customer relationship management (CRM) like approaches. CRM is known to be a strategic approach where organizations implement new processes enabling them to create long term profitable relationship with their customers. The use of CRM approach in different industrial sectors (manufacturing, services, construction, extraction, and mining) requires associated

organizations to be technology compatible to implement them. This paper discusses the need and impact of implementing different technologies (with varying levels of complexities) in the course of developing (new) services from the customer relationship management perspective. CRM is proven to be helpful in organizing and analyzing activities (e.g., marketing, sales, customer services, etc.) in an organization to keep the overall organizational efforts useful to both the customers’ and the organizations’ growth and development.

A systems level approach is used, in this paper, to identify the level of technical complexity that needs to be employed in developing and operating different services. A Service System is defined as a collection of different (service-sub) systems, and their placement with well defined roles and responsibilities. The customer relationship management concept (at large), its applicability and relationship to service development processes and levels of technology being used to develop them is discussed as well.

2. Service Systems and CRM

In simple terms a system can be a group of several interacting elements consisting input parameters, a process to manipulate them and output(s). A service system on the other hand is said to have - customers, a physical or vir-

tual facility to house the required hardware and/or infrastructure, and service providers. Services can be distinguished from other activities based on their specific characteristics, a.k.a. *CHIPS* which represent **C**oproduction, **H**eterogeneity, **I**ntangibility, **P**erishability and **S**imultaneity [5]. Similarly, the concept of service package consisting **D**elivery, **I**nfrastructure and **P**roduct a.k.a. *DIP* can be used in defining the service systems [5].

In case of a simple service system, e.g. a restaurant (as shown in Figure 1), it is observed that to offer end services to their immediate customers, it would need to use several other services. This would require the main service system (restaurant in this example) to interact with other service (sub) systems, e.g., transportation services, banking services, insurance services, credit card services, recreational services, etc. The visibility of systems (including its sub-systems) varies based on their access to customers and their placement in the system.

A closer look on such a system clarifies that all the sub-systems (e.g. credit card services, transportation services, insurance services, etc.) providing their services to a master service provider (restaurant in above example), can actually be the master service providers to other service (sub) systems. While analyzing an end service provider, it is clearly observable that all the sub-systems can be given a specific role and placement in its service system hierarchy. While on a global level all the service

systems are expected to have some (moderate to extensive) level of interaction with other systems in the hierarchy subject to their role and placement in it (as shown in Figure 2).

Every member in a service system hierarchy (i.e. a service sub-system) consists of their own target customers, business strategy, development & expansion plans, software and hardware requirements, infrastructural needs, and marketing policies. Such a diverse set of differentiating factors among service sub-systems makes it crucial to study the primary and secondary effects of them on their customers [6] and observe their relationship with each other. The interactive relationship amongst the systems in whole can be said to have a significant impact on setting the market trends. The change in (sub) systems' operating and developing strategies creates a dynamic impact on all other (sub) systems, due to their close and regular interaction needs.

The role of customers in a business is as important as the business itself. All businesses (including manufacturing, services, agriculture, mining and construction) operate to fulfill their customers' direct or indirect requirements in a best way suitable for their mutual development. Services, although, have the unique characteristics of having the mandatory presence of customers as the co-producers [7]. In the midst of fast paced, globally competitive environment and continually growing need

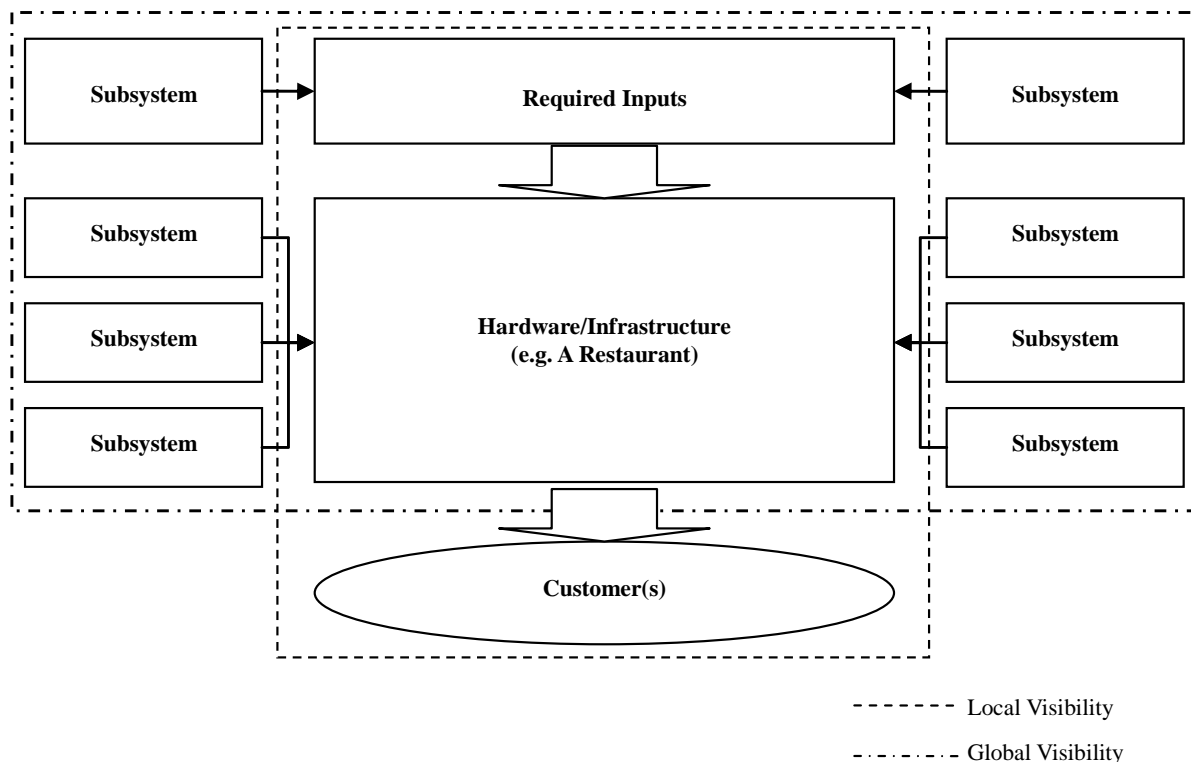


Figure 1. Example of a complete service system, using subsystem approach

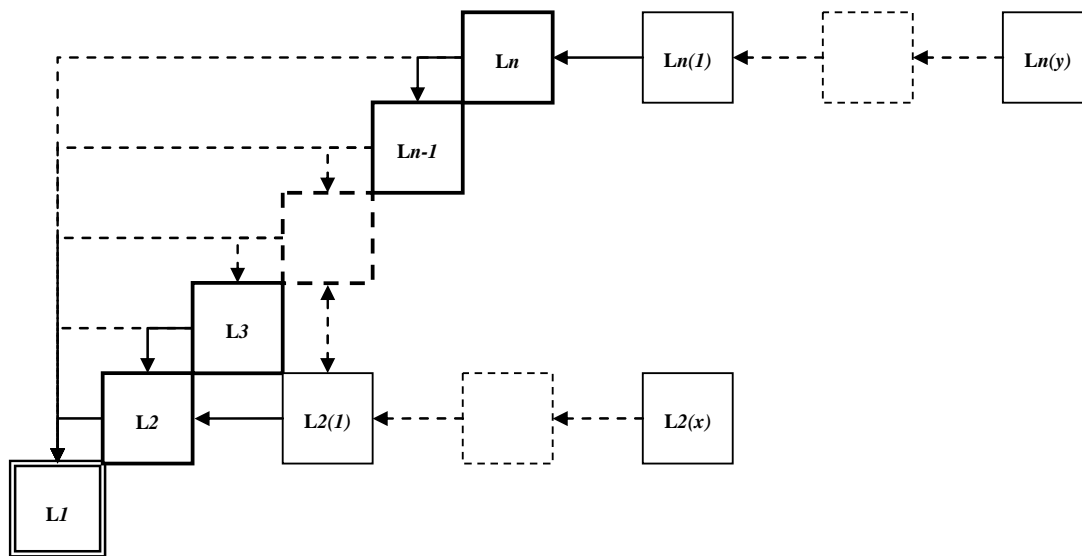


Figure 2. Master service system (L1) with its first and second level subsystems

of economic globalization, it has become evident for the businesses to not only look for new market segments to expand but also work harder to retain their existing customer base. Different organizations/businesses choose a variety of strategies to earn the loyalty of their customers. This includes examples such as sales follow ups, free upgrades, quick complaint responses, free home deliveries, holiday packages, personalized call center responses, etc. Based on the size, need and type of businesses, organizations opt either for a single or a combination of several strategies to expand their customer base while serving their existing customers within their best abilities. This whole process is named as the "Customer Relationship Management" or CRM. Payne and Frow [8] defined CRM as –

"CRM is a strategic approach that is concerned with creating improved shareholder value through the development of appropriate relationships with key customers and customer segments. CRM unites the potential of relationship marketing strategies and IT to create profitable, long-term relationships with customers and other key stakeholders. CRM provides enhanced opportunities to use data and information to both understand customers and cocreate value with them. This required a cross-functional integration of processes, people, operations, and marketing capabilities that is enabled through information, technology, and applications [8]."

Several organizations are dedicating their resources to deliver an optimized (software) system to serve the CRM needs of different organizations. Microsoft®, ORACLE®, SAP, Salesforce.com® are some of the major players in the CRM arena. Aside pure service providing organizations, other industrial sectors receive a large portion of their revenues from the service activities they offer.

Some good examples in the manufacturing sector are – GM, IBM, etc., where a large portion of their annual revenue comes from their service counterparts. In a similar fashion other sectors (i.e. mining, agriculture, and construction) though may not qualify as pure service organizations but do have a large sum of their revenues coming from the services they offer. Embedded relationships, like these, between non-service organizations and their embodied service providing networks make the requirements of CRM like interfaces more crucial and hard to ignore to keep them operational.

3. Technology, Business Process Outsourcing and CRM in Services

The concept of CRM is not as new as its accepted need and demand for different application packages in the market (in last decade or so). The concept of building customer relationships is as old as any business in the industrial world. With the increasing competition among different business (organizations) types, technology used, originating subcontinents, etc. their survival becomes highly dependent on their capabilities to sustain their competitive positioning and customer base. Introduction of new technologies (e.g. internet and IT based tools) not only reduced the distance among customers and business owners but also brought the opportunities of sub contracting the business functions overseas with much reduced costs and enhanced effectiveness. The concept of Business Process Outsourcing (BPO) had emerged in the past decade to provide an edge to different (mainly US based) business owners in terms of gaining the strategic advantage over their non US counterparts. The concept of BPO was added to the multidimensional scenario of the industrial world as the low cost solution of mainly its

information technology (IT) and finance operational needs. BPO did not only introduce the low cost IT and Finance business solutions to the corporate world but also paved the way for the global strategic management of technological innovation. With the intercontinental business ties, customer relationship management strategies took shape in the form of new software packages and alliances formed between different corporate giants to retain their target customer base to survive the competition. Some examples can be given as the recent mergers of AT&T and Cingular, Sears and K-Mart, and talks of acquiring Yahoo! by Microsoft are the steps forward to maintain their niche in the market and expand their customer base.

Services being the largest contributor to the United States' (and many other countries') Gross Domestic Product (GDP), plays a decisive role in directing the development of various tools and technologies in the market [9]. It is an inevitable fact that 'technology' plays a significant role in the course of service development process and needs to be seen as the driving dimension in the service development process [2–4]. The contribution of services had continued to rise (in the vicinity of ~80%) in US GDP in last decade or so, as shown in Figure 3.

It is evident from the contribution of services that services cover a wide spectrum of activities responsible for the economic development. Because of their embedded nature, services can sometimes be hard to distinguish from other activities (i.e., mining, agriculture, construction and manufacturing). A more detailed literature on service characteristics can be found in Agrawal, Fitzsimmons and Fitzsimmons [2,7]. Several attempts

have been made to classify the services, some of the most accepted service classifications in the literature are Schmenner, Bitner, Agnihotri *et al.* [10–12].

As stated earlier in Section 2 (with Figure 1, and 2) a (service) system exhibits a dynamic hierarchy constituting several sub systems. The dynamic behavior of sub-systems directs (large) organizations in strategizing and implementing their CRM activities and vice versa. Based on the specifics of the business, e.g., their strategic motivation and vision; different businesses respond to fluctuations in the market differently. The changes in the market are largely dependent on either the changing perception and needs of customers or a result of the erratic behavior of sub-systems in the system hierarchy. The changing behavior of sub-systems can be explained using varying needs of customers as well, which is a result of introduction of new technologies, products and services in the market. New products and services offered by different organizations drives customers' ever-changing expectations [13] and hence a demand in the market motivating organizations to develop and deliver new services, and hence completing the circle.

Factors such as, market orientation, global strategic growth, technology innovation, management structure, and service blueprints are the driving factors, used in explaining the variations in the service systems development [2,4]. The increasingly embedded sub-systems and complexity therein, calls for the implementation of CRM solutions for organizations to assure their (long term) existence in the market. CRM solutions are equally applicable to all segments of the industry, including the service sector and goods sector.

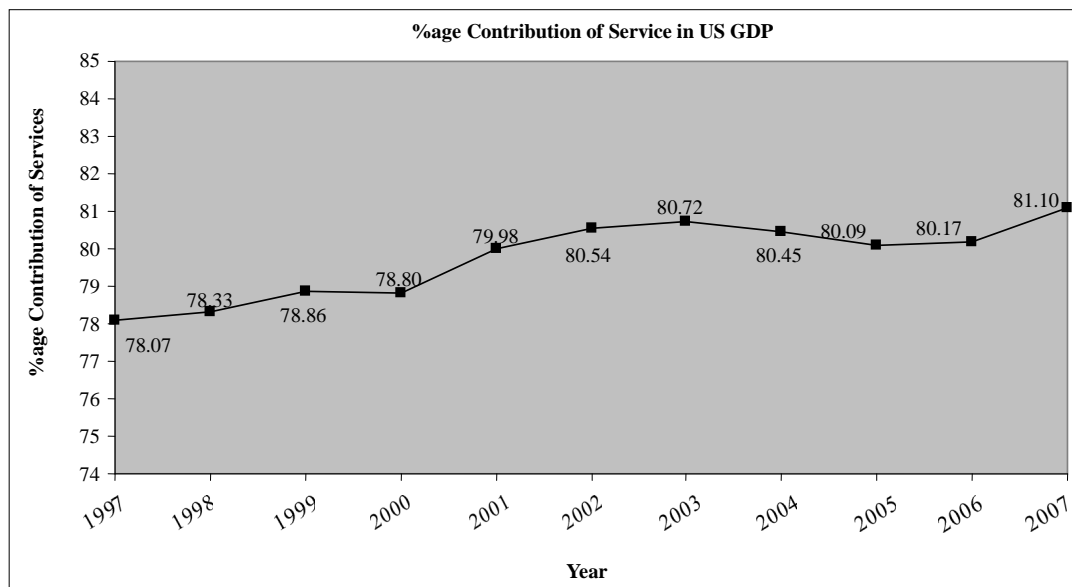


Figure 3. Percentage contribution of service in the US Gross Domestic Product (GDP) source: Bureau of economic analysis

		Degree of Interaction and Customization	
		Low	High
Degree of Labor Intensity	Low	Service Factory <ul style="list-style-type: none"> • Airlines • Trucking • Hotels • Resorts & Recreation 	Service Shop <ul style="list-style-type: none"> • Hospitals • Auto Repair • Other Repair Services
	High	Mass Service <ul style="list-style-type: none"> • Retailing • Wholesaling • Schools • Retail aspects of commercial • banking 	Professional Service <ul style="list-style-type: none"> • Doctors • Lawyers • Accountants • Architects

Figure 4. Service-process matrix [10]

4. The Service Development Process, Technology and CRM

Service organizations adopt different approaches to develop their services based on the level of technology they choose to use in the course of developing and delivering their services. As discussed in the earlier sections, it is becoming customary for the organizations to implement CRM strategies to stay competitive in the market. Businesses face tough competition from their long existing or low cost international contenders in the market. Although using latest technology and tools, some are able to expand their business networks with uniquely customized services, e.g. OnStar[®] services by General Motors Corporation (GMC). These services added another dimension to already existing global positioning systems (GPS) in the automobile market by providing live customer support to their customers. Even though GM faces great competition in the automobile market, OnStar[®] services gave them an edge in the market to sell not just their products but their customized services as well. To offer such technically intensive services, organizations need to choose the appropriate personnel, infrastructure, hardware and software to be able to implement that specific level of technology to not only developing but also delivering such services.

Among different service businesses (even for the similar kind of services) the level of technical complexity chosen by different organizations to develop and deliver their services causes them to pursue the service development process differently [2]. In a similar fashion, implementing CRM applications in a technology intensive environment would demand a higher level of developmental efforts than a comparatively lower level of technology intensive environment. Reinartz *et al.* [14] linked CRM process economic performance with the Informa-

tion Technology as a critical moderator. Based on the Schmenner's Service-Process matrix (as shown in Figure 4) to classify the services [10], a new model is suggested as the Service-Process-Technology (SPT) matrix (shown in Figure 5) that shall be used in relating the service development process to the level of technology used. The service development process and the level of technology that organizations select to use in developing (and delivering) their services can be visualized and explained using the SPT matrix.

Schmenner used labor intensity (*labor to capital ratio*), degree of interaction and customization to develop the Service-Process matrix. Using these factors any service organizations can be nominated to one of the four quadrants in the matrix based on the specific characteristics it reflects. As an extension to Service-Process matrix the SPT matrix is proposed with the technology to be used as the 3rd dimension in it. Research studies have established that the level of technology, organizations use to develop their services, affects their service development process immensely [2–4].

The SPT matrix can be used in explaining the relationship between different service systems based on the level of technology they have used to develop their systems. Although only 3 layers are shown in the SPT matrix, it is possible to have a larger number of layers based on the technical complexity and details involved in the analysis. Reinartz *et al.* [14] stated the need to allocate resources to different tiers of customers based on the economic value (or position in the Service Hierarchy) they have. In an embedded environment, the relationship among different systems can become highly cross linked and ambiguous. Such relationship among (sub) systems makes it essential for the sub (service) systems to keep up with the technology advancements in either master or related sub systems to retain their customer base, which

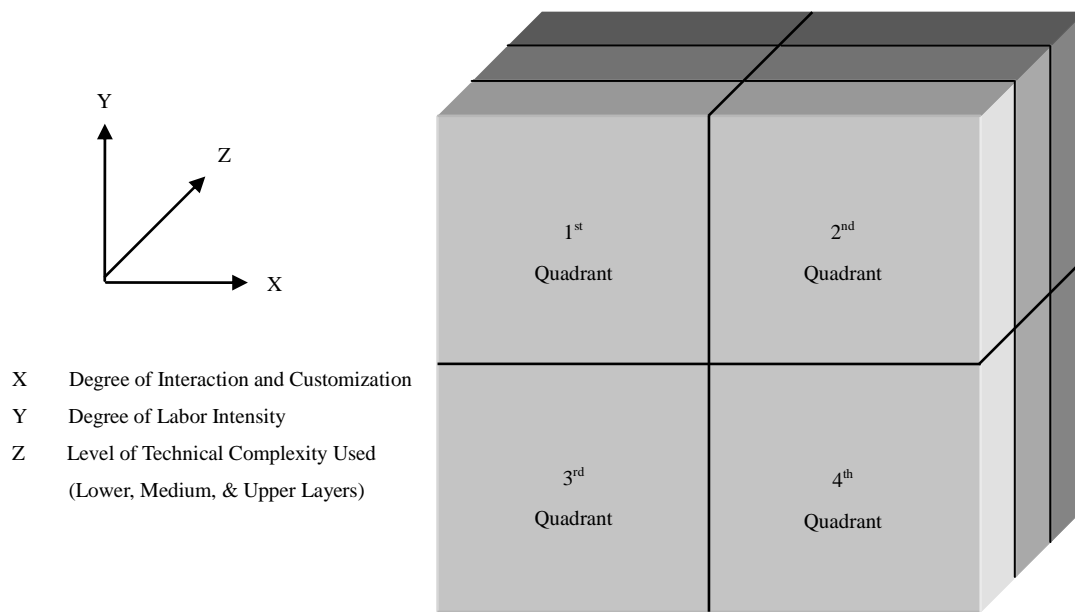


Figure 5. Service-Process-Technology (SPT) matrix

aggressively pursues the technology upgrades available in the market [15].

The implementation of applications such as CRM requires target systems to have some minimum basic technological infrastructure available to them to be able to use CRM applications successfully. As discussed earlier in Section 2, each member in the system hierarchy is capable of affecting other member systems due to their embedded structure and cross relationships. To sustain such minimum requirements, organizations may need to upgrade their existing facilities and hence causing a chaos in the system hierarchy requiring other (attached) subsystems to keep up with them. The relationship between placement of organizations and their subsystems in different 'technology' layers (or level in hierarchy) and the changes/upgrades they need to pursue in order to implement new applications (e.g., CRM) can be explained using SPT matrix like tools. Although, the definition and number of layers in the model may differ for different business types, the matrix features and applicability remains unchanged. As discussed in Section 3, the concept of BPO is playing an important role in the current economic scenario of the United States by giving corporate firms an opportunity to outsource their technical needs to low cost overseas organizations. Because a big part of technology requirements are being outsourced, it is evident that local US companies are being affected greatly and are facing harsh competition in the market to retain their customer base. CRM practices are proven to be helpful in such situations [16]. The organizations outsourcing their operations overseas are implementing and taking advantage of CRM practices as well. With the

infrastructural internationalization, organizations with their establishments in the United States are unable to gain any advantage over the organizations that are using overseas resources, in expanding their customer base.

Two organizations situated in different continents, but in the same business, may qualify in different quadrants of Service Process matrix based on the characteristics they reflect in their business models. Research has established that hardware and software (or infrastructural) requirements for different organizations (or quadrants in the Service Process matrix) differ significantly from each other in the course of their service development process. Using the z-axis in the SPT matrix (i.e., 'Technology' layer) it is possible to distinguish among different organizations who operate similar businesses but in different quadrants based on the similarities in the technological advancements they pursued over time to implement the CRM like applications. Once the most suitable layer for a system (using the service provider, customer base, technology being implemented, and the quadrants it resides in) is identified, the specific requirements for those organizations can be predicted in terms of the infrastructural updates they would need in the process of implementing an application system such as CRM.

Bitner *et al.* [17] established (using their Technology Infusion matrix) that technology plays a crucial role in customers' satisfaction in the service encounter process. Bitner used the Technology Infusion matrix as the framework explaining the improvements in the service encounter experience using the available technology effectively. Using the SPT matrix, the change in customers' expectations can be explained subject to an organiza-

tion's shift from one 'technology' layer to another in terms of the drivers, i.e., customization/flexibility, effective service recovery and spontaneous delight as stated by Bitner *et al.* [17] in their technology infusion matrix. Also, with the customers' inevitable role in the service delivery process, the organizations' efforts to shift within different technology layers (or quadrants in SPT matrix) can be explained based on the shift in customers' expectations and requirements over time.

5. Conclusions

Despite all the attention in services arena, the role of technology in the service development process has not received much attention from researchers, even though the technology development is established to have strong link with the services arena [9]. The need and advent of new applications such as CRM led us to explore the opportunities to study the service organizations using a system based approach and establish the important relationship among different (sub) systems therein.

This paper presented the SPT matrix as one of the solutions that can be used to explain the relationship between different service systems using technology as a common connecting factor among them. Technology being the driving force for (service) organizations to compete in the global markets amid all the intercontinental low cost solutions, it is necessary to have such models available to be able to differentiate among them based on the level of technological advances they wish to pursue over time to stay competitive in the business.

REFERENCES

- [1] L. J. Menor, M. V. Tatikonda, and S. E. Sampson, "New service development: Areas for exploitation and exploration," *Journal of Operations Management*, Vol. 20, pp. 135–157, 2002.
- [2] G. K. Agrawal, "The development of services: Role and impact of technology in the service development process," *Decision Sciences and Engineering Systems*, Rensselaer Polytechnic Institute, Troy, New York, 2006.
- [3] G. K. Agrawal and D. Berg, "Technology in the service development process: A missing dimension," *International Journal of Services Technology and Management*, Vol. 8, No. 2–3, pp. 107–122, 2007.
- [4] G. K. Agrawal and D. Berg, "Role and impact of 'technology' in the service development process: A research study," *International Journal of Services Technology and Management*, Vol. 9, No. 2, pp. 103–121, 2008.
- [5] I. S. Thukral, "Quality in services through design: An integrated approach," *Decision Sciences and Engineering Systems*, Rensselaer Polytechnic Institute, Troy, NY, 1995.
- [6] B. Edvardsson and J. Olsson, "Key concepts for new service development," *The Service Industries Journal*, Vol. 16, No. 2, pp. 140–164, 1996.
- [7] J. A. Fitzsimmons and M. J. Fitzsimmons, "Service management: Operations, strategy, information technology," McGraw-Hill Irwin, New York, NY, 2004.
- [8] A. Payne and P. Frow, "A strategic framework for customer relationship management," *Journal of Marketing*, Vol. 69, pp. 167–176, October 2005.
- [9] J. B. Quinn, "Technology in services: Past myths and future challenges," *Technology in Services: Policies for Growth, Trade and Employment*. B. R. Guile and J. B. Quinn. Washington, D.C., National Academy Press, pp. 16–46, 1988.
- [10] R. W. Schmenner, "How can service businesses survive and prosper?" *Sloan Management Review*, Vol. 27, No. 3, pp. 21–32, 1986.
- [11] M. J. Bitner, "Service and technology: Opportunities and paradoxes," *Managing Service Quality*, Vol. 11, No. 6, pp. 375–379, 2001.
- [12] S. Agnihotri, N. Sivasubramaniam, and D. Simmons, "Leveraging technology to improve field service," *International Journal of Service Industry Management*, Vol. 13, No. 1, pp. 47–68, 2002.
- [13] R. Adner and D. Levinthal, "Demand heterogeneity and technology evolution: Implications for product and process innovation," *Management Science*, Vol. 47, No. 5, pp. 611–628, 2001.
- [14] W. Reinartz, M. Krafft, and W. D. Hoyer, "The customer relationship management process: Its measurement and impact on performance," *Journal of Marketing Research*, Vol. 41, pp. 293–305, August 2004.
- [15] D. A. Norman, "The invisible computer," The MIT Press, Cambridge, MA, 1998.
- [16] R. S. Winer, "A Framework for customer relationship management," *California Management Review*, Vol. 43, No. 4, 2001.
- [17] M. J. Bitner, S. W. Brown, and M. L. Meuter, "Technology infusion in service encounters," *Journal of the Academy of Marketing Science*, Vol. 28, No. 1, pp. 138–149, 2000.



Journal of Service Science & Management (JSSM)

ISSN 1940-9893 (Print), 1940-9907 (Online)

www.scirp.org/journal/jssm/

JSSM is an international multidisciplinary journal with the emphasis laid on the service innovation in the global economy and entrepreneurship, the latest management technologies. It also explores the contributions of knowledge discovery and information management research and applications. The goal of this journal is to keep a record of the state-of-the-art research and promote the fast moving service science and management technologies.

Editor-in-Chief

Prof. Samuel Mendlinger **Boston University, USA**

Editorial Board (According to Alphabet)

Dr. Sónia Bentes	Institute of Accounting and Management of Lisbon, Portugal
Dr. Hengjin Cai	Wuhan University, China
Dr. Chia-Fen Chi	National Taiwan University of Science & Technology, Taiwan, China
Dr. Tsan-Ming Choi	Hong Kong Polytechnic University, Hong Kong, China
Dr. Yong Deng	Shanghai Jiao Tong University, China
Dr. Jerry Fjermestad	New Jersey Institute of Technology, USA
Dr. Leung Kit-nam Francis	City University of Hong Kong, Hong Kong, China
Dr. Javier Sánchez García	Universitat Jaume I, Spain
Dr. Dai-Ping Hu	Shanghai Jiao Tong University, China
Dr. Guang-Wei Hu	Nanjing University, China
Dr. Jongsu Lee	Seoul National University, Korea
Dr. Patrick L. Leoni	University of Southern Denmark, Denmark
Dr. Patrick C. H. Ma	Hong Kong Polytechnic University, Hong Kong, China
Dr. Mehron Sepehri	Sharif University of Technology, Iran
Dr. Chun-Shan Shen	Chinese Academy of Sciences, China
Dr. Kan-Liang Wang	Xi'an Jiao Tong University, China
Dr. Shu-Liang Wang	Wuhan University, China
Dr. Xu-Song Xu	Wuhan University, China
Dr. Fengqi You	Carnegie Mellon University, USA
Dr. Tingsheng Zhao	Huazhong University of Science and Technology, China
Dr. Hong-Ping Zhu	Huazhong University of Science and Technology, China

Subject Coverage

All manuscripts must be prepared in English, and are subject to a rigorous and fair peer-review process. Accepted papers will immediately appear online followed by printed in hard copy. The journal publishes the highest quality, original papers included but not limited to the fields:

- Service Science
- Business Intelligence
- Operational Research
- Computational Economics
- Financial Engineering
- Decision Support System
- Business Process Re-engineering
- Data Mining and Knowledge Discovery
- Innovation and Entrepreneurship
- Risk Management
- Quality Management
- Project Management
- Supply Chain Management
- Software Engineering Management
- Environment and Energy Management
- Knowledge Management and Semantic Web
- Information System Management
- Customer Capital Management
- Human Resources Management

We are also interested in:

Short reports—Discussion corner of the journal:

2-5 page papers where an author can either present an idea with theoretical background but has not yet completed the research needed for a complete paper or preliminary data.

Case studies—Rather than present a classical paper.

Book reviews—Comments and critiques.

Notes for Intending Authors

Submitted papers should not have been previously published nor be currently under consideration for publication elsewhere. Paper submission will be handled electronically through the website. All papers are refereed through a peer review process. For more details about the submissions, please access the website.

Contact Us:

E-Mail: jssm@scirp.org

