

Bayesian Item Response Analysis of Method-of-Payment Habits in Banking Surveys

Saman Muthukumarana¹, Kyle Vincent², Jenna G. Tichon¹

¹Department of Statistics of the University of Manitoba, Winnipeg, Canada ²Independent Consultant Affiliated with San Diego State University Research Foundation, San Diego, USA Email: Saman.Muthukumarana@umanitoba.ca, kyle.shane.vincent@gmail.com, umtichoj@cc.umanitoba.ca

How to cite this paper: Muthukumarana, S., Vincent, K. and Tichon, J.G. (2019) Bayesian Item Response Analysis of Method-of-Payment Habits in Banking Surveys. *Journal of Mathematical Finance*, **9**, 1-10.

https://doi.org/10.4236/jmf.2019.91001

Received: October 22, 2018 Accepted: December 25, 2018 Published: December 28, 2018

Copyright © 2019 by authors and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

http://creativecommons.org/licenses/by/4.0/

CC ① Open Access

Abstract

Customers have a wide variety of choices in selecting a method of payment in modern society due to advancements in technology. In this paper, we investigate the method of payment habits of banking customers using item response models. We consider three binary item response models used in the literature within the Bayesian framework. These models capture the heterogeneity and complexity of customer perception on methods of payment in different capacities, with different features. For this reason, model assessment methods need to be developed for better inferential purposes. We introduce an assessment criterion based on predictive simulations and illustrate the approach using graphical summary measures. The approach is further highlighted using survey data based on consumer payment choices that was conducted for the Federal Reserve Bank of Boston.

Keywords

Bayesian Analysis, Item Response Theory, Markov Chain Monte Carlo, Predictive Simulations

1. Introduction

In modern society, customers can choose various methods of payments in their day-to-day financial activities. Payments by phone or online through financial institutions are easy and convenient. At financial institutions or stores, one can also use a debit card, credit card, cash, cheque, bank draft or money order. These methods come with their own advantages/disadvantages and complexities, so customer response is greatly varied. In this paper, we consider the Bayesian analysis of method of payment habits using Item Response Theory. In psychometrics, item response theory (IRT) is also known as latent trait theory. IRT is widely used in education, psychology and marketing research surveys [1]. The IRT was initiated by three pioneers in three different fields; mathematician Georg Rasch [2], psychometrician Frederic M. Lord [3] and sociologist Paul Lazarsfeld [4]. IRT came to more attention in late 1970 as personal computers gave researchers access to the computing power necessary for applying IRT on complex data.

A wide variety of item response models have been studied in the IRT literature. In this paper, the Rasch model [2] (which is the one-parameter model), the two-parameter model and the three-parameter model [5] are considered within the Bayesian framework. Although a complete review of the literature on these models strikes a difficult task, some of the prominent approaches to the statistical analysis of binary item response data are highlighted in Section 2. One can find a complete overview of these methods and models in [1] [6].

This article is organized as follows. The Rasch model, two-parameter model and three-parameter model are reviewed in Section 2. Prior distributions are then defined on the model parameters. Inference is based on the posterior distribution which is the conditional probability distribution of the parameters given the observed data. Posterior summary statistics are obtained via MCMC methods using R software. In Section 3, we discuss the model selection methods and assess the prediction ability of the models. More specifically, a model assessment criterion is introduced for comparing the observed data against prior predictive output. The uniqueness of the approach is that we advocate a comparison of "features" that are of direct interest by introducing a similarity measure. This is an intuitive and simple approach which is not part of current statistical practice. In Section 4, we analyze a dataset arising from a survey conducted for the Federal Reserve Bank of Boston. We then highlight the suitability of the assessment criterion using graphical summary measures. We conclude with a discussion of the approach in Section 5.

2. Binary Item Response Models

Let Y_{ij} be the selection of the f^{th} method of payment by the t^{th} customer in a customer satisfaction survey. $Y_{ij} = 1$ indicates that the t^{th} customer prefers the f^{th} method of payment and $Y_{ij} = 0$ indicates that the t^{th} customer is resistant towards the use of the f^{th} payment method. The Rasch model [2] is the simplest model used in the IRT literature for analyzing these types of data. It can be written as a one-parameter logistic response model. In this model, the probability that the t^{th} customer prefers the f^{th} payment method is given by

$$\mathbf{P}\left(Y_{ij}=1 \mid \theta_i, b_j\right) = \frac{\exp\left(\theta_i - b_j\right)}{1 + \exp\left(\theta_i - b_j\right)} = \left(1 + \exp\left(b_j - \theta_i\right)\right)^{-1}$$
(1)

where $i = 1, \dots, n$ denotes the customers and $j = 1, \dots, m$ denotes the payment methods. In (1), θ_i represents the payment habit attitude of the *i*th customer and b_i represents the complexity of the *j*th payment method. We can now use

the Item Characteristic Curve (ICC) to describe the relationship between the payment habit attitude and the probability of selecting a payment method. Note that θ_i is latent and the ICC links the latent θ_i to the probability that a randomly drawn customer with a given attitude will choose the *t*th payment method. The customers with negative attitudes have less of a chance, while the customers with high attitudes are much more likely to select the f^{th} payment method. Note that this probability distribution is a member of the exponential family, so the analysts will enjoy all of the statistical features that accompany the family of exponential models. For example, the model allows for algebraic separation of the customer's attitude parameters and payment method parameters. In frequentist methods, the parameters can be estimated using the conditional maximum likelihood approach. In the ICC of the Rasch model, an increase in attitude levels leads to the same increase in the probability of selecting a payment method since the ICCs are parallel to each other. This means that all payment methods are assumed to discriminate between customers in the same way, so a customer's choice only differs in the payment method's complexity, not in the customer's attitude.

The two-parameter logistic model adds a discrimination parameter to overcome this limitation. In the two-parameter model, the probability that the t^{th} customer prefers the f^{th} payment method is given by

$$\mathbf{P}\left(Y_{ij}=1 \mid \theta_i, a_j, b_j\right) = \frac{\exp\left(a_j\left(\theta_i - b_j\right)\right)}{1 + \exp\left(a_j\left(\theta_i - b_j\right)\right)} = \left[1 + \exp\left(-a_j\left(\theta_i - b_j\right)\right)\right]^{-1}.$$
 (2)

By adding a slope parameter a_j to the model, the items are not equally related to the customer attitudes parameter. The higher the discrimination parameter, the better the item is able to differentiate between low and high attitude levels. The parameter a_j captures how quickly the likelihood of $Y_{ij} = 1$ changes with respect to the customer's attitude. A certain discrimination value is only useful in certain regions of the attitude scale. Since a conditional maximum likelihood approach is not possible in this case, Bock and Aitkin [7] developed an estimation procedure based on marginal maximum likelihood. When $a_j = 1$, this model simplifies to the Rasch model in (1). Note that there is a probit version of the two-parameter model known as the normal ogive model [3]:

$$\mathbf{P}\left(Y_{ij}=1 \mid \theta_i, a_j, b_j\right) = \Phi\left(a_j\left(\theta_i - b_j\right)\right) = \int_{-\infty}^{a_j\left(\theta_i - b_j\right)} \phi(z) \, \mathrm{d}z \tag{3}$$

where $\Phi(.)$ represents the cumulative normal distribution function and $\phi(.)$ is the normal density function.

Note that the models in (1) and (2) do not consider possible random choices of novice customers without any prior payment habits. This might occur when a customer starts to use payment methods for the first time without having any prior payment habits. The three-parameter model introduces an extra parameter c_i to account for this type of behaviour as given by

$$\mathbf{P}\left(Y_{ij}=1 \mid \theta_i, a_j, b_j, c_j\right) = c_j + \left(1 - c_j\right) \frac{\exp\left(a_j\left(\theta_i - b_j\right)\right)}{1 + \exp\left(a_j\left(\theta_i - b_j\right)\right)}.$$
(4)

In (4), the probability of $Y_{ij} = 1$ is given by an unknown guessing probability plus a second term representing the dependency on payment method complexity and the customer's attitude level. When $c_j = 0$, the model simplifies to the two-parameter model in (1). The three-parameter normal ogive model [5] is in the form of

$$\mathbf{P}\left(Y_{ij}=1 \mid \theta_i, a_j, b_j, c_j\right) = c_j + (1-c_j) \Phi\left(a_j\left(\theta_i - b_j\right)\right).$$
(5)

Note that in frequentist parameter estimation, it is assumed that parameters are unknown but fixed. In banking customer surveys, payment method complexities, customer attitudes and discrimination effects may be, rather, random quantities. This is because banks frequently make changes and adjustments to their payment methods and customers adjust their attitudes accordingly. The Bayesian paradigm allows us to treat these parameters as random quantities which is appealing in item response data coming from customer surveys. In the Bayesian settings, these random parameters arise from prior distributions that reflect the uncertainty about the true values of the parameters before conducting the surveys. This prior knowledge can typically arise from prior survey findings.

We now describe the Bayesian formulation of the models in (1), (2), and (4). The likelihood of the observed data y is

$$L\left(\underline{y} \mid \underline{P}\right) = \prod_{i=1}^{n} \prod_{j=1}^{m} \left(P_{ij}\right)^{y_{ij}} \left(1 - P_{ij}\right)^{1 - y_{ij}}$$
(6)

where $P_{ij} = P(Y_{ij} = 1)$ is given by (1) in the Rasch model, (2) in the two-parameter model and (4) in the three-parameter model. Note that in our data set \underline{y} is an $n \times m$ matrix of 0's and 1's. We remark that customers must be informative and different payment methods may contain information in different capacities. We can quantify this payment method specific information using the Fisher Information of the data. Conditioning on the item parameters a_j, b_j and c_j , it is easy to show that the Fisher information for estimating θ_i is

$$I\left(\theta_{i}\right) = \sum_{j=1}^{m} \frac{\left(P_{ij}'\left(\theta_{i}\right)\right)^{2}}{P_{ij}\left(\theta_{i}\right)\left(1 - P_{ij}\left(\theta_{i}\right)\right)}.$$
(7)

The Fisher information for the three-parameter model is

$$I(\theta_{i}) = \sum_{j=1}^{m} a_{j}^{2} \frac{(1 - P_{ij}(\theta_{i})) (P_{ij}(\theta_{i}) - c_{j})^{2}}{P_{ij}(\theta_{i}) (1 - c_{j})^{2}}.$$
(8)

The Fisher information for the two-parameter model and one-parameter model can be obtained from (8) by substituting $c_j = 0$ and $a_j = 1$, respectively.

We assign the following prior distributions to the primary parameters of interest:

$$\begin{split} &\log(a_k) \sim \mathrm{N}\big(\mu_a, \sigma_a^2\big) \\ &b_k \sim \mathrm{N}\big(\mu_b, \sigma_b^2\big) \\ &c_k \sim \mathrm{Unif}\left(0,1\right) \\ &\theta_i \sim \mathrm{N}\big(\mu_\theta, \sigma_\theta^2\big) \end{split} \tag{9}$$

The customers are assumed to be sampled independently from a population, and a normal prior density is specified for the attitude parameters with mean μ_a and variance σ_a^2 a priori. A common normal prior is assumed for the discrimination and complexity parameters. The discrimination parameter is restricted to be positive with mean one which indicates a moderate level of discrimination. We do not want complexity parameters to be characterized extremely simple or complex, so we set the mean as zero indicating an average level of complexity. Both variance parameters σ_a^2 and σ_b^2 are fixed to be one. Note that one can assign suitable hyper-priors, such as an inverse Gamma for σ_a^2 and σ_b^2 . The guessing parameter arises from a U(0,1). These priors reflect the state of our knowledge about the parameters before we look at the data. The likelihood tells us how likely it is to observe the current data if the parameters of interest have their current values. The posterior reflects the state of our knowledge about the parameters after we have observed the data.

Note that our likelihood contains multidimensional parameters which lead to a high-dimensional posterior distribution. In the case of complex posteriors, simulation procedures are often used to sample variates from the posterior. The most widely used sampling method is Markov chain Monte Carlo (MCMC). In MCMC, a Markov chain is constructed which has the posterior as its stationary distribution. We implement MCMC for the models described in this paper using R software.

3. Prediction Abilities of the Models

In complex models, it is important that the proper model is selected for the data. In principle, the Bayesian approach to model selection is straightforward and the diagnostics such as AIC [8], BIC [9] and DIC [10] have been proposed and are often used for this purpose. However, the practical implementation of this approach often requires careful investigation as there is potential inconsistency in model selection depending on which of the diagnostics was used. In IRT literature, inconsistencies and inaccuracies have been found among model selection methods under various simulated conditions [11]. In our application, it is highly important that the model selected from competent models is capable of predicting realistic results as it helps to identify future directions in terms of customer behaviour on use of payment methods. For this purpose, one can generate datasets from the posterior predictive distribution [12] of the model and compare it against the observed data using appropriate features. The posterior predictive

density is defined as

$$f(y|x) = \int f(y|\theta) \pi(\theta|x) d\theta$$
(10)

where x is the observed data, $f(y|\theta)$ is the sampling density and $\pi(\theta|x)$ is the posterior density. Model assessment then involves a comparison of the future values y versus the observed data x. A major difficulty with the posterior predictive method in (10) concerns double use of the data [12]. Specifically, the observed data x is used both to fit the model giving rise to the posterior density $\pi(\theta|x)$ and then is used in the comparison of y versus x. For this reason, one can sample "model variates" y from the prior predictive density

$$f(y) = \int f(y|\theta)\pi(\theta)d\theta$$
(11)

where $\pi(\theta)$ is a proper prior density. It is then a matter of deciding how to compare the *y*'s against the observed data matrix *x*. For this purpose, we define a measure called the "Similarity Measure" (SM) as

$$SM(Observed, Predicted) = SM(x, y) = \sum_{i=1}^{n} \left(x_i^T y_i + x_i^{c^T} y_i^c \right)$$
(12)

where x and y are $n \times m$ matrices of observed and predicted data. The values x_i^c and y_i^c are the complements of the t^{th} row of x and y, respectively. A simple comparison of y's against the observed data matrix x can be easily carried out through the calculation of a Similarity Measure. Note that SM(x, y) in (12) is a combination of *correct matches* and *incorrect matches* which provides a meaningful summary measure. We illustrate the SM for the simulated and observed data in the next section.

4. Data Analysis

We consider a dataset arising from the 2010 Federal Reserve Bank of Boston Survey of Consumer Payment Choices (SCPC) to illustrate the models described in this paper. The SCPC is a household survey that aims to measure the banking and spending habits of Americans. Our analysis focuses on responses to the following five methods-of-payments that were made at least once in the past year; cash, mobile phone, money order, traveler's check, and a non-bank online service (e.g. PayPal). Our analysis also focuses on responses to whether the individuals have access to the following four methods-of-payment; general purpose pre-paid cards, merchant specific pre-paid cards, contactless credit card, and contactless debit card.

The dataset consists of n = 1275 customers responding to m = 9 questions on the payment methods. Observations made on payment method frequencies reveals that payment by cash is the most popular among customers as it has the highest proportion of selection, 0.767. The payment method via a traveler's check is the least used method by customers with a proportion of 0.039. The posterior estimates of the parameters under the three models are given in **Table** 1. **Figure 1** shows the ICC curves of the different payment methods under the three models. The plots indicate the heterogeneity among customers in selecting payment methods. **Figure 2** shows the behaviour of the total Fisher information for estimating θ conditioning on the estimates of item parameters in the three models.

We now look at the prediction ability of the models using the similarity measure defined in (12). We generate 100 prior predictive datasets and combine them with the observed data matrix. We then calculate the similarity measures

among these 101 datasets. Note that there are $\binom{101}{2} = 5050$ similarity

measures in which 100 of these are between the observed dataset and the generated datasets. **Figure 3** gives the histogram of these similarity measures highlighting the similarity measures between the observed dataset and the generated datasets in the red colour. Note that this plot is capable of displaying the variation among predictive outputs and the variation between the observed dataset and the predictive outputs as well. It is clear that the plot under the threeparameter model reports the highest similarity measures with a good mixing behaviour indicating that the three-parameter model has the highest prediction ability.









Figure 3. Histogram of similarity measures on 100 simulated datasets.

	1-Parameter Model	2-Parameter Model		3-Parameter Model		
Payment method	b_j	a _j	b_{j}	a_j	b_j	c_{j}
1) Cash	-1.39611	0.29256	-4.15003	0.30529	-2.88671	0.21441
2) Prepaid, general	1.31810	1.55940	1.02058	2.88494	0.96973	0.05409
3) Prepaid, merchant	1.23686	1.04710	1.21966	1.56446	1.26589	0.07882
4) Contactless credit	2.45606	1.09327	2.34532	1.47048	2.64573	0.06291
5) Contactless debit	2.80338	1.10585	2.65162	1.40421	2.97977	0.04783
6) Money order	2.07308	0.39323	4.69576	0.89827	4.31405	0.11820
7) Travelers check	3.57827	0.49224	6.72113	1.01373	5.40572	0.03515
8) Online	1.03717	0.20037	4.42880	0.77891	12.41687	0.29183
9) Mobile	3.53573	0.74481	4.57102	1.12737	4.44243	0.03160

Table 1. The posterior estimates of payment method parameters in IRT models.

5. Discussion

We have presented a Bayesian analysis of item response data with an application to the method-of-payment habits in banking surveys. We also presented a mechanism for assessing the prediction ability of the models and advocate that model selection should be based on the prediction ability of the models. If the model is able to express the uncertainty in the data via model parameters, there is no need to assess the prediction abilities. However, it is difficult to determine the sampling and parameter spaces as the true model behind the observed data is unknown. Our approach is based on the prior predictive simulations which allow us to compare the prediction ability of the models relative to the observed data.

In our data analysis of the SCPC, we found that the three-parameter model possessed the best prediction ability based on our assessment criterion. It indicates that the payment method complexities, customer attitudes and discrimination effects are important in explaining the method-of-payment habits of banking customers. We have not considered possible covariate effects and longitudinal aspect of the data in this paper. This is a topic of future research as this is an ongoing survey.

Acknowledgements

The authors thank anonymous reviewers whose comments helped to improve the manuscript.

Funding

Muthukumarana has been partially supported by a Discovery grant from the Natural Sciences and Engineering Research Council of Canada.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- Fox, J.P. (2010) Bayesian Item Response Modeling: Theory and Applications. Springer, New York. <u>https://doi.org/10.1007/978-1-4419-0742-4</u>
- [2] Rasch, G. (1960) Probabilistic Models for Some Intelligence Tests and Attainment Tests. Danish Institute for Educational Research, Copenhagen.
- [3] Lord, F.M. and Novick, M.R. (1968) Statistical Theories of Mental Test Scores. Addison-Wesley, Reading.
- [4] Lazarsfeld, P.F. and Henry, N.W. (1968) Latent Structure Analysis. Houghton Mifflin, Boston.
- [5] Birnbaum, A. (1968) Some Latent Trait Models and Their Use in Inferring an Examinee's Ability. In: Lord, F.M. and Novick, M.R., Eds., *Statistical Theories of Mental Test Scores*, Addison-Wesley, Reading.
- [6] Hambleton, R.K., Swaminathan, H. and Rogers, H.J. (1991) Fundamentals of Item Response Theory. Sage Press, Newbury Park.
- Bock, R.D. and Aitkin, M. (1981) Marginal Maximum Likelihood Estimation of Item Parameters: Application of an EM Algorithm. *Psychometrika*, 46, 443-459. <u>https://doi.org/10.1007/BF02293801</u>
- [8] Akaike, H. (1973) Information Theory and an Extension of the Maximum Likelihood Principle. In: Petrov, B. and Csáki, F., Eds., *Second International Symposium on Information Theory*, Akadémiai Kiadó, Budapest, 267-281.
- [9] Schwarz, G. (1978) Estimating the Dimension of a Model. *The Annals of Statistics*, 6, 461-464. <u>https://doi.org/10.1214/aos/1176344136</u>
- [10] Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and van der Linde, A. (2002) Bayesian Measures of Model Complexity and Fit (with Discussion). *Journal of the Royal Statistical Society Series B*, 64, 583-639. <u>https://doi.org/10.1111/1467-9868.00353</u>
- Kang, T. and Cohen, A.S. (2007) IRT Model Selection Methods for Dichotomous Items. *Applied Psychological Measurement*, **31**, 331-358. https://doi.org/10.1177/0146621606292213
- [12] Gelman, A., Meng, X.L. and Stern, H.S. (1996) Posterior Predictive Assessment of Model Fitness via Realized Discrepancies. *Statistica Sinica*, 6, 733-807.