Scientific Research Publishing

Open Journal of Applied Sciences, 2024, 14, 131-144
https://www.scirp.org/journal/ojapps
ISSN Online: 2165-3925
ISSN Print: 2165-3917

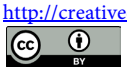
Analysis of the Employment Situation of Non Private Enterprises in Various Regions of China

Junyi Wang

College of Science, Minzu University of China, Beijing, China
Email: wangjunyi0802@163.com

How to cite this paper: Wang, J.Y. (2024) Analysis of the Employment Situation of Non Private Enterprises in Various Regions of China. *Open Journal of Applied Sciences*, **14**, 131-144.
https://doi.org/10.4236/ojapps.2024.141010

Received: December 29, 2023
Accepted: January 27, 2024
Published: January 30, 2024



Abstract

In the past 30 years, Chinese enterprises have been a hot topic of discussion and concern among the general public in terms of economic and social status, ownership structure, business mechanism, and management level. Solving the problem of employment for the people is an important prerequisite for their peaceful living and work, as well as a prerequisite and foundation for building a harmonious society. The employment situation of private enterprises has always been of great concern to the outside world, and these two major jobs have always occupied an important position in the employment field of China that cannot be ignored. With the establishment of the market economy system, individual and private enterprises have become important components of the socialist economy, making significant contributions to economic development and social progress. The rapid development of China's economy, on the one hand, is the embodiment of the superiority of China's socialist market economic system, and on the other hand, it is the role of the tertiary industry and private enterprises in promoting the national economy. Since the 1990s, China's private enterprises have become a new economic growth point for local and even national countries, and are one of the important ways to arrange employment and achieve social stability. This paper studies the employment of private enterprises and individuals from the perspective of statistics, extracts relevant data from China statistical Yearbook, uses the relevant knowledge of statistics to process the data, obtains the conclusion and puts forward relevant constructive suggestions.

Keywords

Correlation Analysis of Employment Numbers, Factor Analysis, Principal Component Analysis, Cluster Analysis

1. Introduction

China is a populous country in the world, and the problem of employment dif-

ficulties is also quite serious. Nowadays, how to solve the employment problem has become an important topic. China's economy is developing rapidly, but it also faces enormous employment pressure. The addition of high-quality labor force every year is constantly increasing, and the population is large, resulting in an unchangeable long-term supply and demand situation. Therefore, it is urgent to study the current employment situation in the field of employment in China. The employment of private and individual enterprises has a profound impact on the employment field in China. Currently, the scale of employment in the individual private economy is constantly expanding, with more than 281 million employees. Of course, as China's economy enters a new normal, employment pressure still exists. The Central Committee of the Communist Party of China and the State Council have established a strategy of prioritizing employment, emphasizing the need to further deepen the reform of the administrative approval system and the commercial system, and vigorously promote "mass entrepreneurship and innovation". Wang Jinshan, Deputy Secretary of the Jilin Provincial Committee of the Communist Party of China, once published an article that included the party building of individual and private economic organizations in the overall party building plan. The article mentioned that for the current situation of individual and private economy in the country, the country should include it in the key construction of the party, and work with organizational departments to study how to improve the current situation. In her article "Analysis of the Employment Situation of Private Enterprises and Individuals", Wu Lizhong takes increasing the number of private enterprises and individuals employed as a breakthrough point to study the current employment situation of private enterprises and individuals in China, providing important basis for macro decision-making. The article points out that with the increase of the total population and the rapid development of the economy and society, the proportion of individual and private economic practitioners in the national employment is gradually increasing. According to statistics, the proportion of self-employed individuals in the private economy among the national workforce increased from 3.5% in 1990 to 32.36% in 2014. In his empirical analysis of the employment situation of individual and private enterprises in China, Lu Yu mentioned that although individual and private enterprises have become the main force in absorbing labor, there are still huge employment disparities in their development. The article empirically analyzes and reveals relevant factors, and proposes corresponding policy recommendations. Meanwhile, with the rapid development of e-commerce, private enterprises have developed rapidly in transportation, warehousing and postal services, finance, information transmission, computer services, and software industries, and their proportion has been increasing year by year. In addition, Sun Chunxiao's research on institutional innovation in the development of private enterprises in China also focuses on the flexible mechanisms and profit seeking market characteristics of private enterprises in China, and has conducted some research to improve the institutional environment of

private enterprises in China to adapt to social and economic development. Wang Ruolei selected the data of the individual employment number of private enterprises in the region, used IBM SPSS to cluster the individual employment number of private enterprises in 31 provinces and cities, and then used the discrimination method to verify the clustering results. Simple statistics and research on the number of private enterprises and individual employed in different regions of the country, and find some characteristics of the number of people employed in different industries in different regions from the statistical perspective, so as to understand the distribution of resources in different regions in China [1]. Roche uses the joint equations model to study the influence of trade liberalization on the employment of the tertiary industry and the employment of individual and private enterprises. The results show that although services trade in the proportion of trade and service value is small, little direct impact on employment, but trade liberalization can effectively promote economic growth in the tertiary industry employment, especially for retail, warehousing, construction, residential services, catering, accommodation, electric gas and water supply especially the role of these industries. At the same time, the deepening degree of trade liberalization has also played a significant role in promoting the employment of individual and private enterprises in the manufacturing industry and the tertiary industry [2]. Chen Baoping designed the mathematical model of employment problem evaluation by using the employment number of private enterprises in seven industries as the evaluation index. Using this model, we evaluate the main industries affecting the employment of private enterprises, and analyze the effective ways to expand the employment space in backward areas [3].

Based on the analysis of enterprise employment researchers in our country, this paper will be from the perspective of the provinces and cities, using the correlation analysis, principal component analysis, cluster analysis and multiple statistical analysis method, based on the specific China 31 provinces and cities of private enterprises and individual enterprises employment data, assuming that the objective factors of the provinces and cities of the private enterprise employment personnel no obvious love under the situation, objective analysis of some provinces and cities in private enterprises and individual enterprises in employment differences.

2. Data Source

2.1. Source

The relevant data on the employment numbers of individual and private enterprises selected in this article comes from the 2017 China Statistical Yearbook. Several representative indicators were selected from various indicators, including the manufacturing industry (x_1), construction industry (x_2), transportation, warehousing, postal industry (x_3), wholesale and retail industry (x_4), accommodation and catering industry (x_5), leasing and business services industry (x_6) in

31 provinces and cities across the country in 2016, Employment numbers in residential services, repair, and other service industries (x_7).

2.2. Variable Description

Manufacturing: refers to the industry in the era of mechanical industry, which transforms manufacturing resources (materials, energy, equipment, tools, funds, technology, information, and manpower, etc.) into large-scale tools, industrial products, and consumer products that can be used and utilized by people through the manufacturing process according to market requirements. It directly reflects the productivity level of a country.

Construction industry: refers to the material production sector in the national economy that is engaged in the survey, design, construction of construction and installation projects and the maintenance of existing buildings. Its development is closely related to the scale of fixed assets investment.

Transportation, Warehousing, and Postal Industry: Transportation is the engineering field that studies the layout and construction of railway, highway, waterway, and air transportation infrastructure, the use of transportation vehicles, transportation information engineering and control, and transportation operation and management. Warehousing refers to the general term for the storage, safekeeping, and related storage activities of materials through a warehouse. The postal industry refers to the business activities of China Post Group Corporation and its affiliated postal industry, which provide basic postal services such as mail delivery, postal summary confidential communication, and postal agency.

Wholesale and retail industry: It is an important link in the process of socialized large-scale production, and a guiding force that determines the speed, quality, and efficiency of economic operation. It is one of the most market-oriented and fiercely competitive industries in China.

Accommodation and Catering Industry: Refers to the industry that provides accommodation, catering, and various comprehensive services. The catering industry is a food production and operation industry that integrates timely processing, commercial sales, and service-oriented labor to provide consumers with various alcoholic beverages, food, consumption venues, and facilities.

Leasing and Business Services: It is an important component of the productive service industry, mainly providing services for production and business activities, including machinery and equipment leasing, cultural and daily necessities leasing, and business services. The leasing and business service industry is the result of deepening social division of labor, which reduces transaction costs and improves production efficiency through continuous improvement of specialization.

Resident services, repair and other service industries: mainly including home services, daycare services, washing and dyeing services, hair and beauty services, bathing services, health services and other service industries. **Table 1** provides explanations for each variable.

Table 1. Meaning of variables.

Variable	Subhead
x_1	manufacturing industry
x_2	construction
x_3	transportation, storage and postal services
x_4	wholesale and retail
x_5	accommodation and catering
x_6	leasing and business services
x_7	residential services, repair and other services

3. Analysis Methods

3.1. Method Introduction

Correlation analysis is the use of an umerical indicator to measure the correlation between a random variable y and a set of random variables $X=\left(x_1,x_2,\cdots,x_p\right)'$, and it is interesting to know to what extent this correlation can be achieved. Correlation analysis is mainly used to analyze the degree of correlation between indicators. Set up

$$E\begin{pmatrix} y \\ x \end{pmatrix}=\begin{bmatrix} \mu_y \\ \mu_x \end{bmatrix},\ V\begin{pmatrix} y \\ x \end{pmatrix}=\begin{bmatrix} \sigma_{yy} & \sigma'_{yy} \\ \sigma_{xy} & \Sigma_{xx} \end{bmatrix},\ \begin{pmatrix} y \\ x \end{pmatrix}\ \text{relation matrix of}\ \begin{bmatrix} 1 & \rho'_{xy} \\ \rho_{xy} & R_{xx} \end{bmatrix}$$

The maximum correlation coefficient between the linear function $l'x$ of y and x is called the complex correlation coefficient y between and x, coef $\rho_{y\cdot x}$ or $\rho_{y\cdot 1,2,\cdots,p}$, It measures the degree of correlation between variable y and a set of variables $x_1,x_2,\cdots,x_p$, The square of the correlation coefficient between y and $l'x$

$$\rho^2\left(y,l'x\right)=\frac{Cov^2\left(y,l'x\right)}{V\left(y\right)\cdot V\left(l'x\right)}=\frac{\left(\sigma'_{xy}l\right)^2}{\sigma_{yy}\cdot l'\Sigma_{xx}l}\le\frac{\left(\sigma'_{xy}\Sigma_{xx}^{-1}\sigma_{xy}\right)\left(l'\Sigma_{xx}l\right)}{\sigma_{yy}\cdot l'\Sigma_{xx}l}=\frac{\sigma'_{xy}\Sigma_{xx}^{-1}\sigma_{xy}}{\sigma_{yy}}$$

Uality $l=\Sigma_{xx}^{-1}\sigma_{xy}$ is obtained from Cauchy's inequality. If it is taken, then the above equality sign holds. So the complex correlation coefficient between y and x is

$$\rho_{y\cdot x}=\max_{l\ne 0}\rho\left(y,l'x\right)=\rho\left(y,\sigma'_{xy}\Sigma_{xx}^{-1}x\right)=\sqrt{\frac{\sigma'_{xy}\Sigma_{xx}^{-1}\sigma_{xy}}{\sigma_{xy}}}=\sqrt{\rho'_{xy}R_{xx}^{-1}\rho_{xy}}$$

Partial correlation coefficient is $\Sigma_{11\cdot 2}=\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}=\left(\sigma_{ij\cdot k+1,\cdots,p}\right)$ defined as the partial x_1 covariance matrix x_2 at a given time, where the non diagonal elements are called the partial covariance and the diagonal elements are called the deviation.

$$\rho_{ij\cdot k+1,\cdots,p}=\frac{\sigma_{ij\cdot k+1,\cdots,p}}{\sqrt{\sigma_{ii\cdot k+1,\cdots,p}\sigma_{jj\cdot k+1,\cdots,p}}},1\le i,j\le k \tag{1}$$

is called the (p-k)-order partial correlation coefficient between x_i and x_j

given x_2 time. This article uses Pearson correlation coefficient for analysis, and the correlation coefficient table is as follows:

3.2. Data Analysis

According to Table 2 and Table 3, the correlation coefficient between x_3 and x_5 reached 91.6%, the correlation coefficient between x_3 and x_7 reached 94.88%, and the correlation coefficient between x_5 and x_7 reached nearly 96.68%. Therefore, it can be concluded that the correlation coefficient between x_3 wholesale and retail industry and x_5 accommodation and catering industry, x_3 wholesale and retail industry and x_7 resident service, repair and other service industry, x_5 accommodation and catering industry and x_7 resident service, repair and other service industry is relatively high. Obviously, there is inevitably a connection between wholesale and the catering industry, as the catering industry must wholesale food, daily necessities, and other commodities. And residential services and any other service industry must rely on wholesale and retail for their livelihoods. Accommodation and catering are essential in every industry, so the high correlation coefficient between them and residential services, repairs, and other service industries is easy to understand.

Table 2. Pearson correlation coefficient.

Category	x_1	x_2	x_3	x_4	x_5	x_6	x_7
x_1	1.00	0.87 <0.001	0.74 <0.001	0.74 <0.001	0.68 <0.001	0.79 <0.001	0.77 <0.001
x_2	0.87 <0.001	1.00	0.64 <0.001	0.73 <0.001	0.60 <0.001	0.84 <0.001	0.64 <0.001
x_3	0.74 <0.001	0.64 <0.001	1.00	0.78 <0.001	0.92 <0.001	0.76 <0.001	0.95 <0.001
x_4	0.74 <0.001	0.73 <0.001	0.78 <0.001	1.00	0.69 <0.001	0.74 <0.001	0.78 <0.001
x_5	0.68 <0.001	0.60 <0.001	0.92 <0.001	0.69 <0.001	1.00	0.59 <0.001	0.97 <0.001
x_6	0.79 <0.001	0.84 <0.001	0.76 <0.001	0.74 <0.001	0.59 <0.001	1.00	0.63 <0.001
x_7	0.77 <0.001	0.64 <0.001	0.95 <0.001	0.78 <0.001	0.97 <0.001	0.63 <0.001	1.00

Table 3. Pearson partial correlation coefficient.

Category	x_5	x_6	x_7
x_2	0.84 <0.001	0.48 0.007	0.89 <0.001
x_4	0.40 0.03	0.30 0.11	0.53 0.003

When (x_1) manufacturing and (x_2) construction industries are known, the correlation coefficient between (x_3) wholesale and retail industries and (x_7) residential services, repairs, and other service industries reaches 0.886. From the correlation coefficient and partial correlation coefficient, we can see that there is indeed a significant correlation between (x_3) wholesale and retail industries and (x_7) residential services, repairs, and other service industries.

3.3. Cluster Analysis

Cluster analysis is the process of dividing classification objects into several categories according to certain rules, which are not predetermined but determined based on the characteristics of the data [4]. There is no need to make any assumptions about the number and structure of the categories. These objects in the same class tend to be similar to each other in a sense, while objects in different classes tend to be dissimilar [5].

Systematic clustering (or hierarchical clustering) is carried out through a series of successive mergers or divisions

There are two types of clustering and segmentation, which are suitable for situations where the number of samples n is not very large. The basic idea of the clustering system method is to start by treating n samples as one class and specifying the distance between samples and between classes. Then, the two closest classes are merged into a new class, and the distance between the new class and other classes is calculated; Repeat the merging of two nearest classes, reducing one class at a time until all samples are merged into one class [6].

The clustering steps of the segmentation system method are exactly opposite to those of the clustering system method. Starting from n samples forming a class, divide it into two subclasses that are as far apart as possible according to some optimal criterion, and then further divide each subclass into two classes using the same criterion. Choose the subclass with the best segmentation from them, and continue until all n samples belong to one class or adopt some stopping rule [7] [8].

1) Shortest distance method:

The distance between classes is defined as the distance between the closest samples of two classes, *i.e.*

$$D_{KL} = \min_{i \in G_K, j \in G_L} d_{ij} \tag{2}$$

This aggregation system method is called the shortest distance method or single connection method. The clustering steps are as follows:

a) Specify the distance between samples, calculate the distance matrix of n samples, where $D_{(0)}$ is a symmetric matrix.

b) Select the smallest element in $D_{(0)}$, set it as D_{KL}, and merge G_K and G_L into a new class, denote as G_M, which is $G_M = G_K \cup G_L$.

c) The recursive formula for calculating the distance between the new class G_M and any class G_j is

$$D_{MJ} = \min_{i\in G_M, j\in G_j} d_{ij} = \min\left\{\min_{i\in G_K, j\in G_j} d_{ij}, \min_{i\in G_L, j\in G_j} d_{ij}\right\} = \min\left\{D_{KJ}, D_{LJ}\right\}$$

In $D_{(0)}$, the row and column where G_K are located and G_L are merged into a new row and column. Corresponding G_M, the new distance value on this row and column is calculated using the recursive formula above, while the distance values on other rows and columns remain unchanged. This results in a new distance matrix, denoted as $D_{(i)}$.

d) Repeat the above two steps for $D_{(i)}$ to obtain $D_{(0)}$ for $D_{(2)}$, and continue until all elements are merged into one class.

2) Center of gravity method:

The distance between classes is defined as the Euclidean distance between their centroids (mean). If the centroids of G_K and G_L are respectively $\bar{x}_k$ and $\bar{x}_L$, then the square distance between G_K and G_L is

$$D_{KL}^2 = d_{\bar{x}_K\bar{x}_L}^2 = \left(\bar{x}_K - \bar{x}_L\right)'\left(\bar{x}_K - \bar{x}_L\right)$$

This clustering method is called the centroid method. The focus of the merged G_K and G_L new classes G_M is

$$\bar{x}_M = \frac{n_K\bar{x}_K + n_L\bar{x}_L}{n_M}$$

It is the weighted average of $\bar{x}_k$ and $\bar{x}_L$, where $n_M = n_K + n_L$ is the number of G_M samples. The recursive formula for the center of gravity method is

$$D_{MJ}^2 = \frac{n_K}{n_M}D_{KJ}^2 + \frac{n_L}{n_M}D_{LJ}^2 - \frac{n_K n_L}{n_M^2}D_{KL}^2$$

3) Sum of squared deviations method:

The sum of squared Euclidean distances from each sample in a class to the class center of gravity is called the sum of squared deviations. If classes G_K and G_L are merged into a new class G_M, then the sum of squared deviations of G_K, G_L and G_M are respectively

$$\begin{aligned} W_K &= \sum_{i\in G_K}\left(x_i - \bar{x}_K\right)^{\mathrm{T}}\left(x_i - \bar{x}_K\right) \\ W_L &= \sum_{i\in G_L}\left(x_i - \bar{x}_L\right)^{\mathrm{T}}\left(x_i - \bar{x}_L\right) \\ W_M &= \sum_{i\in G_M}\left(x_i - \bar{x}_M\right)^{\mathrm{T}}\left(x_i - \bar{x}_M\right) \end{aligned} \tag{3}$$

For a fixed number of intra class samples, they reflect the degree of dispersion of each intra class sample. If categories G_K and G_L are relatively close together, the sum of squared deviations $W_M - W_K - W_L$ added after merging should be smaller; otherwise, it should be larger. So we define the square distance between G_K and G_L as

$$D_{KL}^2 = W_M - W_K - W_L$$

This clustering method is called the sum of squared deviations method or method.

4) Dynamic clustering method:

The basic idea is to select a batch of aggregation points or give an initial classification, so that the samples cluster towards the aggregation points according to certain principles, and continuously modify or iterate the aggregation points until the classification is more reasonable or the iteration is stable. The number of classes k needs to be specified first. A simple method for selecting initial condensation points is to randomly select samples, which can require at least a certain distance value to be separated between the condensation points. The dynamic clustering method can only be used for clustering samples, and cannot be used for clustering variables. The k-means method is one of the popular methods in dynamic clustering.

K-means method:

1) Select k samples as the initial condensation point, or divide all samples into k initial classes, and then use the mean of these k classes as the initial condensation point.

2) Classify all samples one by one, categorize each sample into the class closest to its condensation point, and update the condensation point of that class to its current mean until all samples are classified.

3) Repeat the steps until all samples can no longer be allocated.

3.4. Dynamic Clustering

From the initial seed information in Table 4, we can obtain that the center of Cluster 1 is the seven data in the first row of the table, and the other cluster centers can be extrapolated accordingly. From the clustering summary in Table 5,

Table 4. Information for initial seed.

x_1	x_2	x_3	x_4	x_5	x_6	x_7
3.27	2.30	1.36	2.31	1.34	2.61	1.57
−0.66	−0.80	−1.05	−1.15	−1.28	−0.90	−1.20
1.85	0.88	3.76	1.88	3.01	2.43	3.07
−0.17	0.58	0.07	0.87	−1.10	2.05	−0.86
−0.09	−0.23	−0.22	2.15	−0.27	−0.49	0.10

Table 5. Cluster summary information.

Frequent and continuous	Root mean square standard deviation	Maximum distance from seed to observation	Recent Clustering	The distance between cluster centroids
2	0.85	1.59	3	3.58
17	0.33	1.44	5	2.12
2	0.85	1.58	1	3.58
2	0.35	0.66	5	2.58
8	0.53	2.02	2	2.12

specific information for each class can be seen. For example, Cluster 1 contains a total of 2 observations, with a root mean square standard deviation of pseudo 0.85. The maximum distance from the seed of this cluster to the other observations is 1.59, and the closest class to Cluster 1 is Cluster 3. The distance between the two is the distance between the centroids of the clusters, which is 3.58. The remaining clusters can be analyzed similarly based on Cluster 1.

3.5. Principal Component Analysis

Principal component analysis is a statistical analysis method that converts multiple variables into a few principal components (composite variables) through dimensionality reduction techniques. These principal components can reflect the vast majority of information about the original variables, and are typically represented as some linear combination of the original variables. In order to achieve the most efficient dimensionality reduction, the information contained in these principal components (in the sense of linear relationships) should not overlap with each other, that is, they should be required to be uncorrelated with each other. In short, principal component analysis is a statistical dimensionality reduction method that replaces a large number of correlated variables with a relatively small set of unrelated (composite) variables.

Let $x=\left(x_1,x_2,\cdots,x_p\right)^{\mathrm{T}}$ be a p-dimensional random vector and assume the existence of a second-order moment, Recorded as a $\mu=E\left(x\right)$, $\Sigma=V\left(x\right)$. Consider the following linear transformation

$$\begin{cases} y_1=a_{11}x_1+a_{21}x_2+\cdots+a_{p1}x_p=\boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{x} \\ y_2=a_{12}x_1+a_{22}x_2+\cdots+a_{p2}x_p=\boldsymbol{a}_2^{\mathrm{T}}\boldsymbol{x} \\ \vdots \\ y_p=a_{1p}x_1+a_{2p}x_2+\cdots+a_{pp}x_p=\boldsymbol{a}_p^{\mathrm{T}}\boldsymbol{x} \end{cases} \tag{4}$$

$y_1,y_2,\cdots,y_p$ has specific meanings in this chapter. We first attempt to use a composite variable to represent the original p variables. In order to make y_1 most representative in all linear combinations, we should maximize their variance to preserve the information about the variance and covariance structure of this group of variables as much as possible. Due to the presence of any constant k, have $V\left(ka_1^{\mathrm{T}}x\right)=k^2V\left(a_1^{\mathrm{T}}x\right)$, maximizing variance becomes meaningless if we don't restrict a_1. So we will limit a_1 to the unit vector, that is $\left\|a_1\right\|=1$, we hope to seek a vector a_1 under this constraint to reach the maximum $V\left(y_1\right)=a_1^{\mathrm{T}}\Sigma a_1$, is called the first principal component.

If the information content of the first principal component is not enough to represent the original p variables, it is necessary to consider using again. In order to ensure that the information contained does not overlap, it should be required to

$$Cov\left(y_1,y_2\right)=0$$

Find the second principal y_2 component using the same method.

The total variance of the principal component is

$$\sum_{i=1}^{p}\lambda_i=\sum_{i=1}^{p}\sigma_{ii} \quad \text{or} \quad \sum_{i=1}^{p}V\left(y_i\right)=\sum_{i=1}^{p}V\left(x_i\right)$$

and the proportion of the total variance belonging to the *i* principal component (or being explained) is

$$\lambda_i \Big/ \sum_{j=1}^{p}\lambda_j$$

refers to the contribution rate of the main component y_1. The greater the contribution rate of the first principal component, the stronger its ability to explain the original variable $x_1, x_2, \cdots, x_p$, while the $y_2, \cdots, y_p$ explanatory power of decreases in sequence. The purpose of principal component analysis is to reduce the number of variables, so it generally does not use all p principal components. Ignoring some principal components with smaller variances will not have a significant impact on the total variance. The sum of the contribution rates of m principal components

$$\sum_{i=1}^{m}\lambda_i \Big/ \sum_{i=1}^{p}\lambda_i .$$

is called the cumulative contribution rate of the main component $y_1, y_2, \cdots, y_m$, which indicates $y_1, y_2, \cdots, y_m$'s ability to explain $x_1, x_2, \cdots, x_p$. Usually, a smaller m (relative to *p*) is taken to achieve a higher cumulative contribution rate (such as 80% - 90%). At this point, $y_1, y_2, \cdots, y_m$ can be used as $x_1, x_2, \cdots, x_p$ to achieve the goal of dimensionality reduction, but the loss of information is not significant.

From the eigenvalue Table 6 of the correlation matrix, it can be seen that the principal component corresponding to eigenvalue 5.53 can solve 79.06% of the variation; the principal component corresponding to the eigenvalues 0.78114 can explain 11.16% of the variation, and the first two principal components can explain 90.22% of the variation.

From the feature vector Table 7, specific information about each linear combination of principal components can be obtained. For example, the first principal component is:

Table 6. Correlation matrix eigenvalues.

Eigenvalues	difference	Ratio	accumulate
5.53	4.75	0.79	0.79
0.78	0.50	0.11	0.90
0.28	0.05	0.04	0.94
0.24	0.12	0.03	0.98
0.12	0.08	0.02	0.99
0.03	0.03	0	1.00
0	0	0	1.00

Table 7. Eigenvector values table.

Category	Print	Print	Print	Print	Print	Print	Print
x_1	0.38	0.28	−0.44	−0.34	−0.64	0.15	0.19
x_2	0.36	0.50	−0.28	−0.17	0.63	−0.34	−0.01
x_3	0.40	−0.31	0.08	0.37	−0.19	−0.65	0.38
x_4	0.37	0.07	0.79	−0.47	0	0.08	0.10
x_5	0.39	−0.49	−0.22	0.03	0.38	0.55	0.36
x_6	0.36	0.42	0.19	0.70	−0.09	0.35	−0.20
x_7	0.39	−0.40	−0.12	−0.13	−0.06	−0.10	−0.80

$$\text{Print}1 = 0.382x_1 + 0.362x_2 + 0.396x_3 + 0.372x_4 + 0.374x_5 + 0.364x_6 + 0.394x_7$$
$$\text{Print}2 = 0.277x_1 + 0.450x_2 - 0.310x_3 + 0.066x_4 - 0.486x_5 + 0.417x_6 - 0.403x_7$$

From Table 7, it can be seen that the average load distribution of the first principal component print1 on each variable is 0.36 - 0.40, indicating that the pro-portion of the first principal component in these seven industries is equivalent. Therefore, the first principal component is interpreted as the regional comprehensive employment component, while the second principal component print has a relatively large load in (x_1) manufacturing, (x_2) construction, (x_4) transportation, warehousing, and postal industries, (x_6) leasing and commercial services, which is positive, The (x_3) wholesale and retail industry, (x_5) accommodation and catering industry, (x_7) resident services, repair and other service industries have negative loadings, therefore the second principal component is interpreted as a regional employment bias component.

4. Problem Conclusion and Suggestions

4.1. Method Introduction

Combined with these analysis methods, we know that Jiangsu, Guangdong and Shanghai, which are economically developed cities, provide more employment opportunities for the public, while in some remote cities, such as Xinjiang, Xizang, Qinghai, Inner Mongolia and other regions, the slow development and weak economic to fewer employment opportunities. The reasons for this phenomenon are as follows:

There is irrationality in the government's market access policies for private and individual enterprises, and China's "super national treatment" relative to individual and private enterprises has also caused foreign investment to monopolize certain industries.

With the acceleration of urbanization, especially the concentrated construction of large residential communities, there will be a large number of people living in urban areas. On the one hand, this population has a strong demand for employment, and on the other hand, there is a lack of labor-intensive and service-oriented employment positions in the vicinity of large urban communities

that match their scale, exacerbating the spatial mismatch and mismatch of employment.

Most individual and private enterprises mainly serve the secondary industry, such as construction and creative industries, while some enterprises also serve the tertiary industry, such as wholesale and retail, accommodation and catering industries. Relatively backward services are far from keeping up with the wishes of consumers, causing some enterprises to experience unfavorable business operations.

The development level of private individual enterprises is greatly affected by economic development among regions, and it is similar to economic development to a certain extent. The economic phenomenon of regional economic imbalance objectively exists and will not be eliminated in a short period of time. Overly obvious regional economic differences will affect the speed and quality of regional economic development. The western region has backward economic development, scarce resources, and too few employment opportunities.

4.2. Suggestions

Party committees and relevant departments at all levels should effectively strengthen the leadership and specific guidance of party building work for individuals and private economic organizations, and incorporate the strengthening of the construction of private enterprises and individual businesses into the overall plan for the construction of grassroots party organizations. We need to strengthen supervision and inspection, timely study and solve practical problems in our work, closely cooperate, and form an integrated force to promote the construction of the Party in private enterprises and individual businesses.

Relax the threshold, optimize the environment, further eliminate industry and regional barriers, and fully leverage the flexibility advantages of individual and private economies in the labor market. The individual and private economy has become the most competitive sector in the market, with obvious advantages in resource allocation efficiency and labor market flexibility. It can be more agile in responding to the impact of economic adjustments, continue to deepen administrative approval system reform, commercial system reform, create a more fair competition market environment, and promote more sufficient free flow and effective allocation of capital and labor in the individual and private economy, It helps individual and private enterprises to better leverage their advantages, create and stabilize employment through efficiency improvement, and reduce frictional and structural unemployment.

Appropriately increase support for labor-intensive individual private economy. The labor-intensive industry will still be an important sector for absorbing employment for a long period of time in the future. For some individual private enterprises with strong employment absorption capacity, such as the creative industry, residential service industry, accommodation and catering industry, we should increase support, give appropriate policy tilt, effectively control and re-

duce labor costs and business burden, and encourage them to create stable employment positions.

Continue to improve and gradually expand tax and fee preferential policies. The series of policies currently introduced by the country to encourage employment and entrepreneurship, and promote the development of the private economy have played a positive role. We can continue to improve tax, social insurance, training, administrative fees and other preferential measures, simplify procedures, lower thresholds, expand the scope of policy application, effectively reduce the operating burden of individual and private enterprises, encourage innovation, protect intellectual property rights, and improve market competitiveness, Promote employment creation and employment stability.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Wang, R.L. (2020) Based on Multivariate Statistical Analysis of Regional Private Enterprises and Individual Employment Study. *Modern Marketing*, No. 7, 193-195.

[2] Luo, Z. (2011) The Effect of Trade Liberalization on Employment: Comes from the Data of the Tertiary Industry and Individual Private Enterprises. *Economic Review*, No. 5, 79-85+97.

[3] Chen, B.P. (2012) Employment Evaluation Model of Private Enterprises Based on Grey Correlation Method. *Computer and Information Technology*, **20**, 19-21.

[4] Mirosław, K., Peter, N., Waldemar, R., *et al.* (2024) Spatio-Temporal Principal Component Analysis. *Spatial Economic Analysis*, **19**, 8-29. http://dx.doi.org/10.1080/17421772.2023.2237532

[5] Sihai, Z. and Li, D. (2009) A Note on Optimal Weights and Variable Selections for Multivariate-Survival Data. *Science in China Series A*, *Mathematics*, *Physics*, *Astronomy*, **52**, 1131-1133. https://doi.org/10.1007/s11425-009-0134-3

[6] LeBreton, J.M. and Tonidandel, S. (2008) Multivariate Relative Importance: Extending Relative Weight Analysis to Multivariate Criterion Spaces. *The Journal of Applied Psychology*, **93**, 329-345. https://doi.org/10.1037/0021-9010.93.2.329

[7] De Gooijer, J.G. (2023) On Portmanteau-Type Tests for Nonlinear Multivariate Time Series. *Journal of Multivariate Analysis*, **195**, 105157. https://doi.org/10.1016/j.jmva.2023.105157

[8] Wang, Z.W., Katsaros, D., Wang, J.L., *et al.* (2023) Machine Learning-Based Cluster Analysis of Immune Cell Subtypes and Breast Cancer Survival. *Scientific Reports*, **13**, Article Number: 18962. https://doi.org/10.1038/s41598-023-45932-4