院學仁樹

SHUE YAN COLLEGE

香 港 樹 仁 學 院

The Relationship between Socio-Economic Environment and Regional Distribution of Foreign Direct Investment in China

Che-cheong Poon, Kai-Yin Woo, Tai-Yuen Hon

Department of Economics

Hong Kong Shue Yan College

December, 2004

經 濟 學 系

Working Paper Series

Economics Department

Hong Kong Shue Yan College

The Relationship between Socio-Economic Environment and Regional Distribution of Foreign Direct Investment in China[1]

By
Che-Cheong Poon, Kai-Yin Woo, Tai-Yuen Hon

I. Introduction

According to the IMF Balance of Payments Manual[2], foreign direct investment (FDI) statistics cover all directly and indirectly owned subsidiaries, associates and branches of multinational firms. With a stock of direct investment US$448 billion in 2002[3], China is probably the most attractive location for new business today. Once multinational firms determine to invest in China, they have to decide which region is the best business location to establish their firms. Based on a regional analysis of the major socio-economic variables, this article is intended to provide a decision-making tool for foreign multinational entrepreneurs on the destination of their direct investment in China at the provincial level.

Over the past ten years from 1994 to 2003, China's economy has grown on average by 8.1 percent (10.7 percent in nominal terms). The IMF Economic Outlook 2004[4] makes a forecast of 9 percent growth in 2004 and 7.5 percent in 2005. The growth rates forecasted are the highest among advanced economies, emerging markets and developing countries. This visible success in the economic transformation of China is brought about by twenty-five years of economic reforms since 1978. It is recognized that one of the key driving forces of this transformation is the progressive opening of China to the outside world through foreign direct investment.

To test the hypothesis that the regional socio-economic environment is one of the principal determinants of the regional distribution of inward FDI flows to China, we

---

[1] The revised version of this article titled "Regional Distribution of FDI in China: A Multivariate Data Analysis of Major Socio-Economic Variables" will be published in the March 2005 issue of Chinese Economy.

[2] IMF Balance of Payment Manual, 5th edition, paragraph 359 states "Direct investment is the category of international investment that reflects the objective of a resident entity in one economy obtaining a lasting interest in an enterprise resident in another country"; paragraph 362 states "Direct investment enterprises comprise those entities that are subsidiaries (a non-resident investor owns more than 50 percent), associates (an investor owns 50 percent or less) and branches (wholly or jointly owned unincorporated enterprises) either directly or indirectly owned by the direct investor".

[3] IMD World Competitiveness Yearbook 2004, pp 588. The value of China's direct investment stocks inward was US$447.89 billion in 2002 and ranked fifth in the world.

[4] Overview of the World Economic Outlook Projections, World Economic Outlook, September 2004, Table 1.1.

first select a set of variables based on the criteria of the availability of official statistics, significant regional differences, and the socio-economic environment factors that are highly correlated with FDI. Then, we use the method of principal components factor analysis to construct a socio-economic environment index for each of the 30 regions. A simple correlation analysis is conducted between the regional socio-economic environment index and the regional inward FDI flows provide evidence for our hypothesis.

The thirty-one administrative regions in China [5] (except for Taiwan, Hong Kong, and Macao) are grouped into three areas (eastern area, central area, and western area) according to their geographical locations. The eastern area comprises those regions along the eastern coast. It covers eight provinces (Hebei, Liaoning, Jiangsu, Zhejiang, Fujian, Shandong, Guangdong and Hainan), one autonomous region (Guangxi), and three municipalities (Beijing, Tianjin, and Shanghai). The western area covers six provinces (Sichuan, Guizhou, Yuanan, Shaanxi, Gansu, and Qinghai), three autonomous regions (Tibet, Ningxia and Xinjiang), and one municipality (Chongqing [6]) located in the Northwest and Southwest of China. The Central area covers eight provinces (Shanxi, Jilin, Heilongjiang, Anhui, Jiangxi, Henan, Hubei and Hunan) and one autonomous region (Inner Mongolia) situated between the eastern and western areas. Owing to the unavailability of FDI statistics in Tibet, we have to exclude Tibet from our list, leaving 30regions in our sample. The Map of China showing the location of provinces, autonomous regions and municipalities is attached at Figure 1.

Since the selection of a region to locate their firms from a group of 30 regions would be a time-consuming task for foreign entrepreneurs, we have used a hierarchical clustering method to classify regions into broader regional groups in different fusion levels according to the similarities and dissimilarities in their socio-economic environment. Based on the socio-economic environment index in 2003, we provide a prediction about the trend of inward regional FDI flows in 2004, and discuss the implications of our research results for multinational firms investing in China.

II. FDI in China

Since Deng Xiaoping's tour of the southern provinces in 1992 when he reaffirmed the commitment of the Chinese government to market-oriented reform and policies to open the economy, China has been successful to attract foreign direct investment. According to the China Statistical Yearbook 2004, China has received direct investment flows inward of about 442.8 billion Yuan in 2003 which was about 3.8 percent of GDP. This represents an increase of more than eighteen folds of FDI in

[5] "The Constitution of the People's Republic of China stipulates that the administrative areas in China are divided as: 1) The whole country is divided into provinces, autonomous regions and municipalities directly under the central government; 2) Provinces and autonomous regions are divided into autonomous prefectures, counties, autonomous counties and cities; 3) Autonomous prefectures are divided into counties, autonomous counties and cities; 4) Counties and autonomous counties are divided into townships, nationality townships and towns; 5) Municipalities and large cities are divided into districts and counties; 6) The state shall, when necessary, establish special administrative regions." Explanatory Notes on Main Statistical Indicators, Chinese Statistical Yearbook 2004, Chapter 1.

[6] Chongqing was separated from Sichuan and promoted to become a municipality in 1997.

1991 (23.2 billion Yuan)[7]. An in-depth study of the FDI in China enables us to point out the several characteristics.

Firstly, the main sources of FDI in China have historically been areas with a large Chinese population, but their importance declined somewhat in the past decade as enterprises from the United States, Euro area [8] and Japan entered China in larger numbers. In 2003, the FDI flows to China from these advanced countries were US$12.27 billion (about 23% of total FDI), up 127 percent over 1994 [9]; however, Hong Kong, Taiwan, and Singapore still accounted for over 43 percent of total FDI flows to China in the same year.

Secondly, the contribution made to China by FDI is to raise productivity rather than meeting financial needs. Following the standard four-sector Gross Domestic Product determination model[10], it is easy to derive an estimator for domestic savings, that is: S = I – (T-G) + (X-M). By using the 2003 statistics for Gross Capital Formation (I), Net export (X-M), and the Balance of Total Government Revenue and Expenditures (T-G), we calculate that China's domestic savings is nearly 47 percent of GDP[11], which is probably the highest in the world. With a 42 percent capital formation rate in the same year, from a financial point of view of Balance of Payments, China's high domestic savings rate should be able to finance the equally astounding domestic investment rate by itself. Hence, the role of foreign investment is not so much to contribute financially to the Balance of Payments, but to improve directly and indirectly the productivity of all domestic investment and, as a consequence, contribute to GDP growth.

Thirdly, a high and increasing association between FDI and GDP across regions in China demonstrates the economic significance of FDI to the economy of China in recent years. As shown in Table 1, the almost perfect correlation among the series of regional GDPs for the period 1998 – 2003 reveals the rigidity of regional GDP patterns (their Pearson correlation coefficients [12] range from 0.9945 to 0.9998). On

---

[7] Some of the FDI flows may be "round-tripping" from the mainland to take advantage of the preferential treatment of foreign investors in China. Prasad, Eswar (Ed.) (2004). "Hong Kong SAR: Meeting the Challenges of Integration with the Mainland." Occasional Paper 226, International Monetary Fund. p.4 footnotes 2.

[8] Euro area includes 12 countries: Germany, France, Italy, Spain, Netherlands, Belgium, Austria, Finland, Greece, Portugal, Ireland, and Luxembourg.

[9] Calculated from Table 16-15 and Table 18-15 of the China Statistical Yearbook 1996 and 2004 respectively.

[10] $Y^d = C + I + G + (X\text{-}M)$, $Y^s = C + S + T$; in equilibrium, $I + G + (X\text{-}M) = C + S + T$

[11] From Tables 3-13 and 8-1 of China Statistical Yearbook 2004, I = 51382.7 billion Yuan, (T-G) = 2934.7 billion Yuan, (X-M) = 2686.2 billion Yuan, GDP (by expenditure approach) = 121511.4; then, the calculated S is 57003.6 billion Yuan and the ratio of S to GDP is 0.4691.

[12] Pearson correlation coefficient is a measure of the degree of closeness of the linear relationship between two variables. The definitional formula for the correlation coefficient between X and Y is:

$$r = \frac{Cov(XY)}{\sqrt{Var(X) \cdot Var(Y)}}$$

the other hand, although the correlation among the FDI series is high, the decreasing value of correlation coefficients demonstrates the regional FDI distribution pattern has changed over the past five years. With a high correlation coefficient of 0.8863 between GDP and FDI in 2003, one should not hastily conclude that the regional distribution of FDI could be predicted in terms of the prior regional GDP figures. The reasons are the rigid pattern of regional GDP cannot capture the changing pattern of regional FDI and the correlation between two variables may be due to their common relation to other variables. Accordingly, a set of collinear socio-economic variables (including GDP) which is significantly correlated with FDI, and thus contributes more to the regional distribution of FDI than GDP alone does.

Several prominent features of the distribution of inward FDI flows among the 30 regions in 1998 and 2003 are shown in Table 2. First, FDI is highly concentrated in the Eastern area (87.2% in 1998 and 86.5% of total FDI inflow in 2003) with a significant portion going to Jiangsu, Guangdong, Shangdong, Shanghai, and Zhejiang (56.9% in 1998 and 65.9% of total FDI inflow in 2003). Second, although the aggregate FDI has increased by 16.9% from 1998 to 2003, the regional distribution of FDI has experienced a substantial change. The fact that 13 regions recorded an increase of inward FDI flows and 17 regions recorded a decrease has induced a reshuffle of FDI ranking among the regions in China. Over the years, FDIs in Zhejiang, Jiangxi and Shandong have increased by 278%, 247% and 173% respectively. Third, while the measures of central tendency (the mean) and spread (the standard deviation) of the regional distribution of FDI in China between 1998 and 2003 indicate an insignificant change, the measures of the degree of asymmetry (skewness) and the level of peakedness (kurtosis) of these two distributions have changed markedly. The coefficient of skewness reduces from 3.14 in 1998 to 2.03 in 2003, revealing that more regions have attracted less than the average regional FDI and fewer regions have attracted more than average regional FDI over the five-year period. The coefficient of kurtosis dropps from 11.32 in 1998 to 3.82 in 2003 demonstrating that the frequency curve of regional FDI distribution has changed from a leptokurtic curve [13] towards a normal curve. This indicates that the similarity of FDI among the middle ranked regions has increased and the full range of the distribution has reduced over the years.

III. Data Description

Many determinants of FDI have been identified in the economic literature[14]. However, our investigation concentrates on those for which official statistics are available and relevant for the case of China at the provincial level. In light of the above analysis, while neglecting the financial environment as a significant factor to attract FDI, our criteria for including elements in the list of socio-economic variables which determine the attractiveness of FDI are based largely on the consideration of market size and factor productivity.

[13] A leptokurtic curve has a narrower central portion and higher tails than does the normal curve.

[14] For examples: Crum, Brigham, & Houston (2005); pp97-99, Wang (2004);Ng & Tuan (2003); and Coughlim & Segev (1999).

Market size refers to the extent to which a specific production output could be sold. At the macro level, the number of potential buyers in the market and the income of consumers are major determinants of market size. Among the socio-economic variables restricted by the availability of data in China's official statistics, we choose per capita GDP[15], per capita retail sales, average wage, and population density as proxies for market size. In addition, since FDIs from Hong Kong, Taiwan and Singapore tend to be export-oriented manufactured products, the degree of openness to international trade and the contribution of secondary and tertiary industries are included as variables under the category of market size.

Factor productivity refers to the extent to which a specific production factor contributes to production output under a given average production cost. At the macro level, a better infrastructure and human capital investment are major favorable indicators of factor productivity. Thus, we have choose per capita total investment in fixed assets, the percentage of population with education level at senior secondary school and higher, per capita government expenditure for innovation enterprises, total length of transportation routes, and overall resource productivity measured as a ratio of GDP to land area as proxies for factor productivity.

Because there are substantial differences among regions in China in terms of population and land area, in order to make meaningful comparisons, with the exception of those that are expressed in terms of percentages, all the selected variables are expressed either as per capita or per square kilometer. In addition, since the nature of this research is cross sectional and there is no need to take changes in the price level into account, we have used variables in nominal rather than in real terms. For the sake of statistical manipulation, these selected variables are expressed symbolically as X_i, where i = 1 to 11. The data matrixes in 1998, 2002 and 2003 are provided in Tables 3 4, and 5; the definitions of these variables are listed below:

X_1 Represents per capita GDP at current market prices estimated by production approach.

X_2 Represents per capita retail sales of consumer goods. It is calculated as the total retail sales divided by the number of mid-year population (the ratio of GDP to per capita GDP of the same year).

X_3 Represents average wage of staff and workers in State-owned units.

X_4 Represents population density (persons per sq km)

X_5 Represents the degree of openness to international trade. It is calculated as (X+M)/GDP; (X+M) is the total Import and Export value of commodities by places of destination or origin. The value has been changed to Yuan by using the average exchange rate of RMB against USD.

[15] Per capita GDP can also be a proxy for the overall productivity that is calculated as the ratio of GDP to total number of persons employed.

X_6 Represents the contribution of secondary and tertiary industries to GDP. It is calculated as the ratio of total gross output values produced by the secondary and tertiary industries to GDP.

X_7 Represents per capita total investment in fixed assets.

X_8 Represents the percentage of population with education level at senior secondary school and higher to the population aged 6 and over (sample survey data).

X_9 Represents per capita government expenditure for innovation enterprises.

X_{10} Represents total length of transport routes (railways, waterways and highways) per sq. km.

X_{11} Represents resource density, which is calculated as the ratio of GDP to land area.

IV The FDI Attraction Index

Having determined the socio-economic variables to capture an abstract concept of regional attractiveness of FDI it may be necessary to go on to construct an index which purports to measure that concept. To do this, we apply Principal Components Analysis to summarize the 11 socio-economic variables into m principal components [16] (m<11), and then use the resulting factor score coefficients of these principle components (preferably m=1, i.e. the first principal component) as weights to calculate the weighted average of the 11 socio-economic variables for a specific region to obtain its FDI attraction index.

1. Methodology

Since the 11 regional socio-economic variables are measured on different scales or on a common scale with substantial difference in magnitude, it is necessary to transform the 11 original variables on the same scale by standardizing them for the subsequent factor analysis. Suppose that each observed socio-economic variable X_i has a constant mean μ_i with a finite variance σ_i^2 over 30 regions in China. We transform X_i in $\mathbf{X}' = (X_1, X_2, \ldots, X_{11})$ to Z_i in the random vector $\mathbf{Z}' = (Z_1, Z_2, \ldots, Z_{11})$, where $Z_i = \frac{X_i - \mu_i}{\sigma_i}$ with a mean of zero and a standard deviation of one. In the

[16] Principal components are defined as optional linear combinations of the original variables extracting a maximum of variability and being uncorrelated. The first principal component (F_1) is that linear combination of the original variables accounts for as large a proportion of the total variance of these variables as possible; the second principal component is then required to account for as much of the remaining variance as possible, subject to being uncorrelated with the first principal component and so on with each successive component being uncorrelated with its predecessors and accounting for as much residual variance as possible. (Bartholomew 1897, pp 12)

factor analysis model, the standardized variable Z_i is expressed exactly as a linear combination of common factor scores of principal components $F_1, F_2, ..., F_m$ and one additional specific factor (or the error term) ε_i, it can be written as:

$$Z_i = \sum_{k=1}^{m} l_{ik} F_k + \varepsilon_i = l_{i1} F_1 + l_{i2} F_2 + ... + l_{im} F_m + \varepsilon_i .$$

$$i = 1,\ldots,11. \quad k = 1, \ldots, m \quad \ldots\ldots (1)$$

where l_{ik} is known as factor loading of the *i*th standardized variable Z_i on the k^th principal component F_k; m stands for the number of principal components extracted. The common factor score of the k^th principal component F_j and the specific factor ε_i are assumed to satisfy the following conditions:

$$\left.\begin{array}{l} E(F_k) = 0 \text{ and } Var(F_k) = 1, \ \forall \ k, Cov(F_k, F_l) = 0, \ \forall k \neq l. \\ E(\varepsilon_i) = 0 \text{ and } Var(\varepsilon_i) = \psi_i, \ \forall \ i, \quad Cov(\varepsilon_i, \varepsilon_p) = 0, \ \forall \ i \neq p. \\ Cov(\varepsilon_i, F_k) = 0; \forall \ i, k. \end{array}\right\} \ldots\ldots (2)$$

In this paper, the factor loadings and variances of the original standard normal variables (Z_i) are estimated using the principal components method.[17] The principal component solution of the factor model is expressed in terms of the eigenvalue-eigenvector pairs, denoted, $(\lambda_1, e_1), (\lambda_2, e_2), \ldots, (\lambda_{11}, e_{11})$, of the 11-by-11 variance-covariance matrix of **Z**, where $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_{11}$, $e_1' e_1 = e_2' e_2 = ... = e_{11}' e_{11}$. The estimated factor loading l_{ik} is obtained by $\sqrt{\lambda_j} e_{ik}$, where e_{ik} stands for the i^{th} element of the k^th eigenvector. Moreover, the contribution to the total variance of the standardized variables, $\sum_{i=1}^{11} Var(Z_i)$, explained by the k^th principal component factor score, is calculated by adding the squared estimates of factor loadings of all the standardized variables in **Z** under the k^th principal component, that is, $\sum_{i=1}^{11} l_{ik}^2 = l_{1k}^2 + l_{2k}^2 + ... + l_{11k}^2$ or $(\sqrt{\lambda_k} e_k)'(\sqrt{\lambda_k} e_k)$, which gives the k^th eigenvalue λ_k. The total standardized variance must be equal to 11; hence, $\frac{\lambda_k}{11}$ represents the proportion of total standardized variance attributable to the k^th common factor. Since the estimate of each consecutive eigenvalue is on the decrease, each corresponding factor score will account for less and less total standardized variance. Kaiser (1960) suggests that only the factor scores, which have eigenvalues of one or greater, should be extracted. He reasons that while the maximum amount of variance explained by one standardized variable is one, a common factor extracted is then required to explain at least as much as the equivalent of the variance of one standardized variable.

Furthermore, the portion of $Var(Z_i)$ explained by all the m principal components

[17] The principal components method is considered as the simplest and most widely used kind of factor analysis (Cramer 2003).

extracted is called the ith communality, denoted ϕ_i^2, which is equal to the sum of squares of the estimated loadings of Z_i on the m common factors given by $\sum_{k=1}^{m} l_{ij}^2 = l_{i1}^2 + l_{i2}^2 + ... + l_{im}^2$; hence, the higher the ith communality, the more the common factors can explain the variance of the ith standardized variable. The estimated specific variance ψ_i is simply equal to the variance of Z_i minus the estimated value of ϕ_i^2 [18].

In addition, it is useful to compute the values of factor scores for further analysis of the FDI inflows in China. In the principal components analysis, the common factor score of the kth principal component F_j is given by $(e_{1.k}Z_1 + e_{2.k}Z_2 + ... + e_{11.k}Z_{11})$ divided by $\sqrt{\lambda_k}$.[19] In other words, the computation of F_k involves the linear combination of the standardized variables $Z_1, Z_2, ,..., Z_{11}$ with the respective factor score coefficients being equal to

$$(\sqrt{\lambda_1})^{-1} e_{1.k},\ (\sqrt{\lambda_1})^{-1} e_{2.k},\ \dots,\ (\sqrt{\lambda_1})^{-1} e_{11k}. \qquad \text{.................................} (3)$$

2. Empirical Results

The principal components analysis is conducted on the data matrix of socio-economic variables for 1998, 2002 and 2003. The first step is to estimate the pairs of eigenvalue and eigenvector from the variance-covariance matrix of **Z**. Based upon the criterion of Kaiser (1960), we only retain the first eigenvalue λ_1 associated with the first corresponding eigenvector e_1 and the first principal component factor score F_1 for the years under study. In other words, only one principal component (m = 1) is generated. We present the first principal component solution in Table 6. In Table 6, the first principal component factor score F_1 accounts for about 80.2% of the total standardized variance in 1998, and about 78.6% and 78.5% in 2002 and 2003. Also, the estimated values of factor loadings can be used to measure the degree of which the socio-economic variables are correlated with F_1. We find that during the years under consideration, GDP per capita (X_1) , retail sales per capita (X_2) and per capita investment (X_7) are the variables with the highest correlation coefficient with F_1. With m = 1, the estimated communalities, ϕ_i^2, are simply the squares of the respective factor loadings. The values of the factor score coefficients reported in Table 6 reflect the weights or relative importance of the individual standardized variables in the construction of F_1. Since the factor score coefficient of Z_i, given by $(\sqrt{\lambda_1})^{-1} e_{i1}$, is

---

[18] It is assumed that the number of principal components extracted is smaller than the number of the original variables under study; otherwise the factor model in equation (1) would become exact and the vector of specific factor would be a null vector (elements in the vector are all zero).

[19] See Chapter 8 in Johnson and Wichen (2002) for the detailed discussion of principal components.

equal to the respective factor loading $\sqrt{\lambda_1} e_{i1}$ divided by λ_1, the values of factor loadings are exactly the same as those of the factor score coefficients.

In the light of the above, it is sensible statistically to use the factor score coefficients of the first principal component as the weighting system applied to the regional values of socio-economic variables to obtain the common factor score of the first principal component (CFSFPC). It can be used to represent the socio-economic environment across the 30 regions in China, Taking Beijing (region code is 1) and Shanghai (region code is 9) as examples, their respective CFSFPC is calculated as the weighted average of the standardized values of the 11 socio-economic variables:

$$\left.\begin{array}{ll} \text{Beijing:} & \text{CFSFPC}_1 = w_1 Z_{1,1} + w_2 Z_{2,1} + w_3 Z_{3,1} + \ldots + w_{11} Z_{11,1} \\ \text{Shanghai:} & \text{CFSFPC}_9 = w_1 Z_{1,9} + w_2 Z_{2,9} + w_3 Z_{3,9} + \ldots + w_{11} Z_{11,9} \end{array}\right\} \quad \ldots\ldots\ldots\ldots (4)$$

Where: the weight (w_i) of Z_i stands for the factor score coefficient of the first principal component $(\frac{e_{i.1}}{\sqrt{\lambda_1}})$,

so that $w_1 = \frac{e_{1.1}}{\sqrt{\lambda_1}}$, $w_2 = \frac{e_{2.1}}{\sqrt{\lambda_1}}$,… , $w_{11} = \frac{e_{11.1}}{\sqrt{\lambda_1}}$.

Since the CFSFPCs for the 30 regions calculated based on standard normal variables that contain both positive and negative values, we have to convert the CFSFPC series to an index number for the sake of mathematical manipulation. In doing this, we first change the CFSFPC values to their corresponding probability values by using the Cumulative Standardized Normal Distribution Table, and then express the series in the form of an index number called socio-economic environment index (SEEI). The formula of the SEEI for the j^{th} region is:

$$\text{SEEI}_j = \frac{\Pr(-\infty < \text{CFSFPC}_j)}{\frac{\sum_{j=1}^{30} \Pr(-\infty < \text{CFSFPC}_j)}{30}} \times 100 \qquad j = 1,\ldots,30 \quad \ldots\ldots\ldots\ldots\ldots\ldots (5)$$

Table 7 shows the CFSFPC, the SEEI and the ranking of SEEI across the 30 regions in 1998, 2002 and 2003. The high correlation between the socio-economic environment index (SEEI) and per capita FDI (PCFDI) in consecutive years as shown in the lower portion of the Table implies that the regions with higher SEEI is expected to attract more PCFDI in the current and next few years. That is to say, we cannot reject the hypothesis that the regional socio-economic environment is one the principal determinants of the regional distribution of inward FDI flows in China. Hence, the SEEI can be thought of as the FDI attraction index.

Note however that the socio-economic environment index (SEEI) in a given year is largely a relative concept built on the variances of the socio-economic variables amongst regions in China in a specific year, and its absolute value is therefore not appropriate to be used as a basis for comparison over time.

V. Cluster Analysis

1. The Agglomerative hierarchical clustering technique

The socio-economic environment index (SEEI) provides us with an analytical tool through which we can determine not only the order of FDI attractiveness amongst regions in China but also their degree of difference. However, as indicated by the data matrix of the 11 selected variables, to identify the characteristics of the socio-economic environment of each of the regions in China is very difficult. So, it is desirable to partition the 30 regions in China into subgroups so that those in each particular group are more similar to each other before carrying on further investigation. We use the cluster analysis [20] method to complete this task.

The process of clustering begins by finding the closest pair of regions according to a particular distance measure of attributes and combine those regions with the nearest distance to form a cluster. The procedure continues one step at a time, linking pairs of regions, pairs of clusters, or a region with a cluster, by a linkage method until all the clusters are merged into a single cluster. This algorithm is known as the hierarchical clustering method. The results can be presented in dendrograms, which are known as hierarchical tree diagrams.

The attribute variables in our cluster analysis are the set of 11 regional socio-economic variables. We choose the squared Euclidean distance as the measure of dissimilarity among the attribute variables. This requires the computation of distance in variables on the same scale. Thus, we employ the standardized attribute variables for clustering analysis to avoid problems caused by scale differences. The squared Euclidean distance is the sum of the squared distance over all standardized attribute variables under consideration, which can place progressively greater weight on regions that are further dissimilar. Moreover, we use the complete linkage method for linking clusters in the hierarchical clustering algorithm. The results of the hierarchical clustering analysis for 1998, 2002 and 2003 are presented in the form of dendrograms and are attached in Figure 2.

2. Characteristics of distinct clusters

The dendrograms that appear in Figures 2, 3 and 4 are tree diagrams, which give a visualization of the hierarchical structure of the 30 regions in China in terms of their respective socio-economic environment. Taking the complete linkage dendrogram in 2003 as an example, before joining the regions together, each region is considered to be a single group at the first stage; at each of the second to the fourth stage in the agglomerative procedure, the number of regions is reduced. When the final (sixth) stage is reached, there is a single group containing all 30 regions.

At the fifth stage, the regions are classified into two area groups. One group contains three regions and the other contains 27 regions. It is obvious that all regions (Beijng,

[20] See Everitt & Dunn (1991) and Friedman (2004) for detailed description of the clustering method.

Tianjin and Shanghai) in the 3-region group are municipalities and are the earliest regions to be opened to the foreign world; and thus, the socio-economic environment in this group on average is the best in the country. The establishment of the Chongqing municipality in 1997 does not lead immediately to success in the overall socio-economic environment here despite it being given the leading role in the "Open up the West" campaign which started in January 2000. It therefore does not have the attributes to be classified in area group one.

At the fourth stage, the 30 regions are classified into the following four area groups:
Area group 1 : Shanghai
Area group 2 : Beijing, Tianjin
Area group 3 : Jiangsu, Zhejiang, Guangdong
Area group 4 : The other 24 regions

Shanghai itself forms a distinct group because it ranks first in a number of socio-economic variables, these include per capita GDP (X_1), population density (X_4), contribution of secondary and tertiary industries to GDP (X_6), percentage of population with education level at senior secondary school (X_8), per capita government expenditure for innovation enterprises (X_9), length of transport routes (X_{10}), resource density (X_{11}). It ranks second in the whole country in the following variables: per capita retail sales (X_2), average wage (X_3) and degree of openness to international trade (X_5), per capita total investment in fixed assets (X_7). In addition, the score of Shanghai in each of the above variables is only slightly lower than the score of the first ranking area. For X_2, Shanghai has a score of 16.6, which is only slightly lower than the score of Beijing of 16.78. For X_3, the score of Shanghai is 28.41, which is also only slightly lower than the score of Beijing of 28.46. For X_5, the score of Shanghai of 146.34 is also lower than that of Guangdong of 175.69. For X_7, the score of Shanghai of 18.68 is also slightly lower than that of Beijing of 18.99.

All regions in area group 3 are coastal provinces that are opened to the foreign world and undergo economic reforms earlier. Since their percentage of State-owned and State-holding enterprises to all enterprises in terms of gross output value are the lowest amongst the 30 regions in China[21], economic freedom in Jiangsu, Zhejiang and Guangdong should be the greatest amongst all the regions in China. In this area group, per capita GDP (X_1), average wage (X_3), per capita total investment in fixed assets (X_7), length of transport routes (X_{10}), resource density (X_{11}) ranks first amongst all the 22 provinces and 4 autonomous regions (Tibet is not included).

The majority of the twenty-four regions in area group 4 are situated in the interior parts of China. Their scores for the eleven socio-economic variables are the lowest in the country indicating the fact that geographical factors have had a substantial effect on the socio-economic environment of these regions.

If we compare the dendrograms in 1998 and 2003, we observe the following

---

[21] In 2002, the percentage of State-owned and State-holding enterprises with an annual sales income of over 5 million yuan in terms of gross output value at current prices of Jiangsu, Zhejiang and Guangdong is 22.8%, 13.6% and 19.3% respectively; in 2003, they are 19.0%, 13.1% and 18.4%. Since this statistics is not available in 1998, it is not included in the list of socio-economic variables. See China Statistical Yearbook 2003, Table 13-3; China Statistical Yearbook 2004, Table 14-2.

characteristics. Firstly, Shanghai remains in Area Group One in both the 1998 cluster and the 2003 cluster. This indicates that Shanghai has been able to stay in the leading position throughout the period; secondly, Tianjin has moved to area group 2 in 2003 from area group 3 in 1998 cluster because of the significant improvement in the degree of openness to international trade (X_5), investment in fixed assets (X_7) and length of transport routes (X_{10}); thirdly, Fujian, Shandong and Liaoning which belong to area group 3 in 1998 are classified into area group 4 in 2003 because their scores in government expenditure for innovation enterprises (X_9) are the lowest of members in Area Group 3 (with the exception of Guangdong). Finally, there is no noticeable change in the majority of regions with lower socio-economic scores during the period from 1998 to 2003.

VI. Implications for Multinational Firms' Location Decisions

With its continuing reduction in trade and financial barriers after WTO accession as well as advances in the communication networks, China has become an economic powerhouse in Asia and the focus of many foreign direct investments. The optimistic prediction made by researchers that the growth of FDI in China will reach US$100 billion annually during the Five-Year Plan period of 2006-10 [22] reflects the fact that investing in China is a golden chance for multinational firms from advanced countries. As we know, the motivation for multinational firms invest in foreign production facilities and related ventures in other countries is to enhance their investment return. To sustain growth opportunities, they have to reassess where it is best to produce their products regularly; also, after their home markets mature and competition becomes more intense, they have to expand their markets abroad. Thus, these eleven socio-economic variables, which are chosen to capture both market size and factor productivity, appear sensible in reflecting the attractiveness of different regions of China to multinational firms.

The socio-economic environment index (SEEI), which is derived from the first principal component in factor analysis, provides a useful tool for multinational entrepreneurs to rank their location decisions on a region-by-region basis regarding the socio-economic environment. The SEEI in 1998, 2002 and 2003 in Table 7 reveals that regions with a better socio-economic environment can attract more inward FDI flows. Furthermore, the high correlation between per capita FDI (PCFDI) and SEEI in the ensuing years as shown in the bottom portion of the Table enables us to say that SEEI is one of the most important leading indicators in location decisions.

For example, the SEEI Shanghai in 2003 was 223, 1% higher than that of Beijing, 29% higher than that of Guangdong, and 329% higher than that of the poorest province, Guizhou. With the best socio-economic environment in the country, Shanghai has strengthened its power to attract foreign investments and improved the quality of those investments by putting more emphasis on the development of modern manufacturing industries, modern servicing industries and new high-tech industries. In addition to its outstanding achievements in socio-economic development, taking

[22] "China to grab $100 billion annual FDI in 2006-10", Emerging Markets Economy, 1/2/2003. Available at Business Source Premier.

into account the role of Shanghai as the major financial center of China as well as its strategic geographical location, these enable us to believe that Shanghai will continue to be the most attractive business location for multinational firms.

On the other hand, there are 8 regions with their SEEI of less than 70 in 2003. They are Hainan and Guangxi from the eastern area; Sichuan, Guizhou, Yunnan, Gansu and Xinjiang and from the western area; and Inner Mongolia from the central area. The inferior economic development level and lower standards of living in these regions greatly restrict the expansion of their market size and, make it difficult for them to attract foreign investors. Moreover, the inferior conditions of the infrastructure facilities have constituted a barrier to socio-economic development in these regions, and have thus weakened their attractiveness to foreign capital. However, since these poor regions are characterized by a vast land area, rich mineral and forests resources and a sparse population, their marginal returns to investment are relatively higher than those of the regions in the developed eastern area. With the advantages brought about by the campaign to "Open Up the West" in 2000 [23] and the implementation of WTO commitments in 2006, these regions will probably become the target locations of some multinational firms.

As shown in Table 2, we note that the southern regions that contain the original four special economic zones (Shenzen, Zhuhai, Shantou, and Xiamen) have experienced a significant drop in FDI while there is an increase in northern regions. The FDI figures for Guangdong, Fujian and Hainan dropped 50.57 billion yuan (-36%) over the period of five years from 1998 to 2003; in contrast, the FDI figures for Shanghai, Jiangsu and Zhejiang soare by 78.28 billion yuan (+82%). This reveals that foreign investors in China are moving northward from the Pearl River Delta to the Yangtze River (Chang-jiang) Delta. FDI has begun to spread from the traditional investment base in the south to new regions because China has introduced new policies aiming at easing foreign investment restrictions and attracting more foreign investments to other parts of the country. For this reason, we can conclude that the signaling effect of prior FDI [24] is no longer an appropriate factor to determine location decisions.

VII. Concluding Remarks

Regional or location disparities in China have been one of the hot research topics for more than 10 years. Many models have been constructed (Hu, Wang & Hong, 1995; Poon, Hon and Woo, 1996; Coughlin & Segev, 1999; Ng & Tuan, 2003; Wang, 2004) to explain the pattern of disparities. Some models emphasize the role of government FDI promotion policy as the basic cause of regional disparity in inward FDI flows,

[23] A workshop to examine the causes, content and potential impact of the drive to "Open Up the West" with the emphasis on the provincial and local levels was hosted by the German Institute of Asian Affairs in Hamburg during 8-10 May 2003. Selected articles are published by The Chinese Quarterly (number 178) in June 2004.

[24] "Once there are a large enough number of foreign investors present in a certain area, it is a signal to other investors that conditions are apparently good or are good enough in that area to do business, and that will subsequently attract more foreign investors." Transcript of an Economic Forum: Foreign Direct Investment in China: What Do We Need To Know? International Money Fund, IMF Auditorium, Thursday, May 2, 2002. Available at http://www.imf.org/external/np/tr/2002/tr020502.htm.

while others emphasize economic size, labor productivity and coastal location, and still others cite the influence of the proximity of regions. This study aims to understand the driving forces behind the regional distribution of FDI and to provide a comprehensive image for multinational entrepreneurs regarding the overall investment environment in China. It analyzes the fixed effect of the socio-economic environment for individual regions, and determines whether or not the regions are attractive business locations as well as whether their attractiveness are changing over time. Since the socio-economic environment is an abstract concept that cannot be directly measured, we have collected information on variables likely to be indicators of the concept, and synthesized these indicators in the form of an index number to mirror the regional socio-economic environment in China.

The ambiguity of the definition of regions in China is an unavoidable analytical pitfall when dealing with the relation between the concepts of per capita (per city or per sq. km.) and aggregation. Regarding FDI, we believe that governmental authorities have a preference for total FDI figures rather than per capita FDI. However, multinational firms will find per capita FDI more relevant. Therefore, we use per capita FDI as the dependent variable in analyzing the location decisions. In addition, it is worth pointing out that while progress continues to be made in upgrading China's economic statistics, weaknesses in terms of their timeliness, accuracy and consistency in key areas including the national income statistics and the international direct investment flows have imposed an inevitable handicap to our analysis.

It should be stressed that socio-economic variables are strongly interrelated, and thus we should not attempt to use the multiple regression method for analysis [25]. The objective of constructing the socio-economic environment index (SEEI), which is based on the first principal component of the eleven socio-economic variables, is to demonstrate the ranking and magnitude of the socio-economic environment amongst the 30 regions in China. On the other hand, we use the complete linkage clustering technique to classify these regions into area groups in order to present the similarities and dissimilarities in their socio-economic environment. Taken together, the SEEI and the resulting clusters convey useful information to multinational entrepreneurs regarding the ranking of a specific region within the chosen area group in which the socio-economic environment is suitable for them to establish their direct investment enterprises. These results provide direction for foreign investors investing in China to make location decisions at the macro level.

[25] The regression coefficients may be interpreted as a measure of the change in the dependent variable when one unit increases in the corresponding independent variable and all other independent variables are held constant. Such an interpretation would no longer be valid in the presence of strong linear relation amongst the independent variables, simply because in such situation it is obviously impossible to change one variable whilst holding all other constant.

Table 1 Correlation Matrix of Regional GDP and FDI, 1998 to 2003

	GDP1998 (FDI1998)	GDP1999 (FDI1999)	GDP2000 (FDI2000)	GDP2001 (FDI2001)	GDP2002 (FDI2002)	GDP2003 (FDI2003)
GDP1998 (FDI1998)	1 (1)					
GDP1999 (FDI1999)	0.9997 (0.9951)	1 (1)				
GDP2000 (FDI2000)	0.9990 (0.9916)	0.9996 (0.9908)	1 (1)			
GDP2001 (FDI2001)	0.9986 (0.9880)	0.9993 (0.9831)	0.9998 (0.9948)	1 (1)		
GDP2002 (FDI2002)	0.9970 (0.9460)	0.9979 (0.9363)	0.9985 (0.9664)	0.9992 (0.9717)	1 (1)	
GDP2003 (FDI2003)	0.9945 (0.8211)	0.9957 (0.8065)	0.9966 (0.8556)	0.9976 (0.8763)	0.9994 (0.9515)	1 (1)
FDI1998	**0.7301**	0.7360	0.7455	0.7460	0.7449	0.7421
FDI1999	0.7241	**0.7293**	0.7387	0.7399	0.7403	0.7380
FDI2000	0.7739	0.7793	**0.7882**	0.7890	0.7889	0.7873
FDI2001	0.7766	0.7831	0.7922	**0.7930**	0.7931	0.7923
FDI2002	0.8245	0.8304	0.8369	0.8382	**0.8402**	0.8415
FDI2003	0.8391	0.8460	0.8504	0.8535	0.8603	**0.8663**

Source: Calculated from China Statistics Yearbook 1999, 2000, 2002 issues, Table 17-16, and from China Statistics Yearbook 2004, Tables 3-10, 18-2 and 18-16.

Table 2. Regional distribution of per capita FDI in China, 1998 & 2003

Region	1998				2003				% Change in total FDI
	Total FDI		Per capita FDI		Total FDI		Per capita FDI		
	Amount (Billion Yuan)	Rank	Amount (Yuan)	Rank	Amount (Billion Yuan)	Rank	Amount (Yuan)	Rank	
1. Beijing	17.95	7	1649	3	18.14	8	1587	2	1.05
2. Tianjin	17.50	8	1939	2	12.70	11	1377	3	-27.41
3. Hebei	11.83	9	181	11	7.98	13	118	14	-32.54
4. Shanxi	2.02	23	69	20	1.77	22	54	20	-12.66
5. Inner Mongolia	0.75	25	32	24	0.73	24	31	25	-2.53
6. Liaoning	18.13	6	436	8	23.38	6	555	8	28.90
7. Jilin	3.39	19	129	14	1.58	23	58	19	-53.43
8. Heilongjiang	4.36	16	117	15	2.66	20	70	17	-38.88
9. Shanghai	29.82	4	2284	1	45.26	4	3383	1	51.80
10. Jiangsu	54.91	2	764	7	87.44	1	1179	4	59.25
11. Zhejiang	10.91	10	246	9	41.22	5	884	5	277.79
12. Anhui	2.29	22	37	23	3.04	18	49	21	32.66
13. Fujian	34.87	3	1100	5	21.51	7	616	7	-38.31
14. Jiangxi	3.85	17	93	18	13.34	9	315	11	246.61
15. Shandong	18.24	5	207	10	49.80	3	547	9	173.05
16. Henan	5.10	15	55	21	4.46	14	48	22	-12.59
17. Hubei	8.06	11	137	13	12.99	10	217	12	61.21
18. Hunan	6.78	13	108	17	8.43	12	137	13	24.37
19. Guangdong	99.51	1	1400	4	64.75	2	818	6	-34.93
20. Guangxi	7.34	12	157	12	3.46	16	76	15	-52.78
21. Hainan	5.94	14	815	6	3.49	15	432	10	-41.28
22. Chongqing	3.57	18	117	16	2.16	21	69	18	-39.51
23. Sichuan	3.08	20	37	22	3.41	17	40	23	10.67
24. Quizhou	0.38	26	10	28	0.37	26	10	28	-0.34
25. Yunnan	1.21	24	29	25	0.69	25	16	27	-42.46
26. Shaanxi	2.48	21	69	19	2.75	19	74	16	10.57
27. Gansu	0.32	27	13	27	0.19	28	7	29	-39.40
28. Qinghai	0.00	30	0	30	0.21	27	39	24	+
29. Ningxia	0.15	29	29	26	0.14	29	25	29	-6.11
30. Xinjiang	0.18	28	10	29	0.13	30	7	30	-29.23
Mean	12.50	- -	408.97	- -	14.61	- -	427.93	- -	- -
Standard Deviation	20.51	- -	631.04	- -	21.76	- -	709.43	- -	- -
CV = SD/Mean	1.64	- -	1.54	- -	1.49	- -	1.66	- -	- -
Skewness	3.14	- -	1.86	- -	2.03	- -	2.88	- -	- -
Kurtosis	11.32	- -	2.48	- -	3.82	- -	10.00	- -	- -
Minimum	0.00	- -	0.00	- -	0.13	- -	7.00	- -	- -
Maximum	99.51	- -	2284.00	- -	87.44	- -	3383.00	- -	- -

Notes: i/ The FDI figures have been converted from USD to Yuan by using the average exchange rate of 821.91 and 827.7 RMB against 100 US dollar in 1998 and 2003 respectively.

ii/ Per capita FDI is calculated by dividing the total FDI by the number of mid-year population, which is borrowed from the quotient of dividing total GDP by per capita GDP.

Sources: China Statistical Yearbook 2000, 17-16, and China Statistical Yearbook 2004, 18-2 &18-16.

Table 3: Data Matrix for the Observed values of the 11 Regional Socio-economic Variables in 1998

Region	Socio-economic Variable										
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
1. Beijing	18.48	10.73	12.35	648	125.62	95.7	10.33	41.02	177	0.81	1197
2. Tianjin	14.81	6.51	10.24	799	65.75	94.5	6.33	22.76	138	0.44	1183
3. Hebei	6.53	2.04	6.59	348	8.22	81.5	2.44	13.93	28	0.32	227
4. Shanxi	5.04	1.72	5.94	203	5.75	87.1	1.43	13.52	10	0.33	102
5. Inner Mongolia	5.07	1.70	5.98	20	6.69	71.3	1.35	17.90	26	0.05	10
6. Liaoning	9.33	3.77	7.60	285	27.17	86.3	2.54	17.28	87	0.33	266
7. Jilin	5.92	2.58	6.81	141	8.78	72.4	1.64	21.00	31	0.21	83
8. Heilongjiang	7.54	2.53	6.54	83	5.90	83.7	2.05	17.14	46	0.12	62
9. Shanghai	28.25	11.27	13.75	2072	70.36	97.9	15.06	34.61	391	1.02	5854
10. Jiangsu	10.02	3.11	8.87	700	30.30	85.9	3.41	15.79	38	0.51	702
11. Zhejiang	11.25	4.31	10.48	436	24.66	87.3	4.06	13.83	44	0.49	490
12. Anhui	4.58	1.51	6.63	439	6.66	73.7	1.18	9.21	26	0.34	201
13. Fujian	10.37	3.53	8.68	265	42.66	81.7	3.28	11.84	23	0.44	274
14. Jiangxi	4.48	1.47	5.47	247	5.58	75.7	0.97	10.95	17	0.26	111
15. Shandong	8.12	2.41	7.47	563	19.22	83.1	2.19	10.74	35	0.43	457
16. Henan	4.71	1.62	6.20	554	3.30	75.4	1.39	11.90	15	0.36	261
17. Hubei	6.30	2.52	6.78	316	6.33	79.8	1.97	15.78	9	0.33	199
18. Hunan	4.95	1.74	6.80	306	4.59	74.2	1.23	12.46	11	0.34	152
19. Guangdong	11.14	4.57	11.29	399	135.73	87.3	3.72	15.42	70	0.59	445
20. Guangxi	4.08	1.57	6.24	198	10.48	69.8	1.20	8.21	18	0.24	81
21. Hainan	6.02	2.03	5.97	215	32.93	62.6	2.38	13.77	1	0.52	129
22. Chongqing	4.68	1.81	6.73	372	5.99	79.1	1.62	9.19	16	0.36	174
23. Sichuan	4.34	1.57	7.04	146	4.84	73.8	1.39	11.59	26	0.16	63
24. Quizhou	2.34	0.81	5.82	204	6.17	68.5	0.77	7.83	9	0.21	48
25. Yunnan	4.36	1.21	7.88	105	7.62	77.3	1.60	6.72	33	0.20	46
26. Shaanxi	3.83	1.44	6.26	175	12.30	79.5	1.44	13.90	14	0.22	67
27. Gansu	3.46	1.21	7.13	55	4.26	76.7	1.20	11.79	8	0.08	19
28. Qinghai	4.37	1.40	8.51	7	4.43	81.1	2.16	10.54	16	0.03	3
29. Ningxia	4.27	1.45	7.02	103	8.70	78.6	2.00	14.29	25	0.20	44
30. Xinjiang	6.23	1.83	7.17	11	11.31	74	2.87	17.67	14	0.02	7

Sources:

X_1 Per capita GDP at current market prices estimated by production approach (1000 yuan). CSYB1999. 3-9.

X_2 PRS = Per capita retail sales (1000 Yuan) = (Retail Sales / GDPpop), (GDPpop is the number of mid-year population which is the ratio of GDP to per capita GDP of the same year) CSYB1999, 16-2

X_3 Average wage of staff and workers (1000 Yuan) in State-owned units, CSYB1999, 5-18.

X_4 Population density (persons per sq km). It is calculated as GDPpop divided by Land area of the region. Data of Land area is from China Development Report 1995, P231; figures for Sichuan & Chongqing are obtained from http://hk.geocities.com/chinamap04

X_5 (X+M)/GDP is the measure of the degree of openness to international trade, expressed in %; (X+M) is the total Import and Export value of commodities by places of destination or origin. The value has been changed to yuan by using USD100 = RMB827.91 , 100 million yuan, calculated from CSYB1999, 17-2 &17-11.

X_6 Contribution of Secondary and tertiary Industries to GDP, CSYB1999, 3-9

X_7 Per capita Total Investment in Fixed Assets, TIFA, (1000 yuan) = TIFA/GDPpop, calculated from CSYB1999, 6-4

X_8 The percentage of population with education level at senior secondary school and higher to the population aged 6 and over, calculated from CSYB1999, 4-9.

X_9 Per capita government expenditure for innovation enterprises (yuan) = GEIE/GDPpop, calculated from CSYB1999, 8-14.

X_{10} Length of transport routes,TR, (railways, waterways and highways) per sq. km. It is calculated as the ratio of TR to land area of the respective region. CSYB1999, 15-3.

X_{11} Resource density (10,000 yuan per sq.km), it is measured by GDP divided by land area.

Table 4: Data Matrix for the Observed values of the 11 Regional Socio-economic Variables in 2002

Region	Socio-economic Variable										
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
1. Beijing	28.45	11.15	22.48	672	68.69	96.95	15.88	44.48	366	0.92	1912
2. Tianjin	22.38	5.62	16.83	811	92.21	95.90	8.81	33.10	336	0.96	1815
3. Hebei	9.12	2.06	10.72	358	9.23	84.37	3.01	16.30	16	0.36	326
4. Shanxi	6.15	1.58	9.89	210	14.76	90.20	2.48	17.30	5	0.40	129
5. Inner Mongolia	7.24	1.64	9.36	21	12.72	81.11	2.96	20.56	40	0.07	15
6. Liaoning	12.99	4.11	12.36	278	35.52	89.19	3.82	18.61	35	0.36	374
7. Jilin	8.33	3.11	10.39	144	15.01	80.14	3.10	23.66	43	0.25	120
8. Heilongjiang	10.18	2.70	9.51	84	9.99	88.49	2.74	19.57	26	0.16	85
9. Shanghai	40.65	12.10	24.77	2112	110.57	98.37	16.64	40.16	879	1.36	8585
10. Jiangsu	14.39	3.21	13.89	720	57.99	89.47	4.67	16.97	71	0.83	1036
11. Zhejiang	16.84	4.37	20.27	455	49.21	91.10	7.51	19.02	101	0.56	766
12. Anhui	5.82	1.32	9.53	438	9.75	78.35	1.75	10.03	24	0.54	256
13. Fujian	13.50	3.49	15.27	286	53.62	85.80	3.61	17.35	28	0.46	386
14. Jiangxi	5.83	1.24	8.95	252	6.74	78.13	2.11	14.39	14	0.41	147
15. Shandong	11.65	2.32	12.24	578	29.31	86.83	3.84	20.09	49	0.51	673
16. Henan	6.44	1.43	9.79	574	5.01	79.11	1.80	16.26	15	0.46	369
17. Hubei	8.32	2.37	9.62	312	7.54	85.79	2.68	16.08	14	0.52	268
18. Hunan	6.57	1.74	11.05	298	6.24	80.48	2.04	16.82	14	0.46	205
19. Guangdong	15.03	4.66	20.78	439	158.55	91.22	4.92	18.99	31	0.70	662
20. Guangxi	5.10	1.43	11.00	204	8.79	75.74	1.56	14.80	25	0.27	104
21. Hainan	7.80	1.64	7.60	226	24.57	62.10	2.91	18.20	1	0.63	178
22. Chongqing	6.35	1.55	11.78	379	8.49	83.98	2.90	13.73	16	0.42	240
23. Sichuan	5.78	1.31	11.48	149	7.57	78.92	2.25	14.19	26	0.22	86
24. Quizhou	3.15	0.75	10.63	214	6.85	76.30	1.68	11.06	19	0.27	67
25. Yunnan	5.18	1.03	12.00	109	8.63	78.92	1.89	8.36	24	0.43	57
26. Shaanxi	5.52	1.25	10.47	179	11.32	85.08	2.48	16.94	20	0.25	99
27. Gansu	4.49	1.12	11.48	57	7.40	81.54	2.04	14.66	24	0.10	26
28. Qinghai	6.43	1.34	15.89	7	5.69	86.84	4.38	12.11	14	0.04	5
29. Ningxia	5.80	1.16	11.75	110	12.42	83.95	4.00	17.60	58	0.24	64
30. Xinjiang	8.38	1.55	10.75	12	15.96	80.92	4.20	24.73	9	0.05	10

Sources:

X_1 Per capita GDP at current market prices estimated by production approach (1000 yuan). CSYB2003. 3-9.

X_2 PRS = Per capita retail sales (1000 Yuan) = Retail Sales / GDPpop, (GDPpop is the number of mid-year population which is the ratio of GDP to per capita GDP of the same year), CSYB2003, 16-3

X_3 Average wage of staff and workers (1000 Yuan) in State-owned units, CSYB2003, 5-28.

X_4 Population density (persons per sq km). It is calculated as GDPpop divided by Land area of the region. Data of Land area is from China Development Report 1995, P231; figures for Sichuan & Chongqing are obtained from http://hk.geocities.com/chinamap04.

X_5 (X+M)/GDP is the measure of the degree of openness to international trade, expressed in %; (X+M) is the total Import and Export value of commodities by places of destination or origin. The value has been changed to yuan by using USD100 = RMB827.7 , 100 million yuan, calculated from CSYB2003, 17-2 &17-11.

X_6 Contribution of Secondary and tertiary Industries to GDP, calculated from CSYB2003, 3-9

X_7 Per capita Total Investment in Fixed Assets, TIFA, (1000 yuan) = TIFA/GDPpop, calculated from CSYB2003, 6-4

X_8 The percentage of population with education level at senior secondary school and higher to the population aged 6 and over, calculated from CSYB2003, 4-9.

X_9 Per capita government expenditure for innovation enterprises (yuan) = GEIE/GDPpop, calculated from CSYB2003, 8-22.

X_{10} Length of transport routes, TR, (railways, waterways and highways) per sq. km. It is calculated as the ratio of TR to land area of the respective region. CSYB2004, 15-3.

X_{11} Resource density (10,000 yuan per sq.km), it is measured by GDP divided by land area.

Table 5: Data Matrix for the Observed values of the 11 Regional Socio-economic Variables in 2003

Region	Socio-economic Variable										
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
1. Beijing	32.06	16.78	28.46	680	70.80	97.39	18.99	45.63	402	0.93	2180
2. Tianjin	26.53	10.00	19.35	816	101.55	96.40	11.27	33.94	318	0.97	2166
3. Hebei	10.51	3.23	11.78	360	11.29	85.01	3.67	20.05	15	0.37	378
4. Shanxi	7.44	2.21	11.21	211	17.45	91.24	3.33	18.08	21	0.43	157
5. Inner Mongolia	8.97	3.03	11.93	20	12.42	80.46	4.90	19.24	62	0.07	18
6. Liaoning	14.26	5.54	13.60	289	41.18	89.74	4.93	24.55	37	0.37	411
7. Jilin	9.34	4.11	11.12	144	22.08	80.70	3.59	23.59	39	0.26	135
8. Heilongjiang	11.62	3.61	11.03	84	11.61	88.70	3.06	18.93	33	0.17	97
9. Shanghai	46.72	16.60	28.41	2124	146.34	98.55	18.68	45.72	1155	1.42	9922
10. Jiangsu	16.81	4.81	17.50	723	80.56	91.12	7.06	19.22	69	0.89	1215
11. Zhejiang	20.15	6.77	27.29	458	58.43	92.25	10.17	19.93	105	0.56	923
12. Anhui	6.46	2.16	11.22	441	11.82	81.50	2.31	16.62	24	0.55	285
13. Fujian	14.98	4.98	16.46	288	61.00	86.70	4.28	18.11	41	0.49	431
14. Jiangxi	6.68	2.18	10.92	254	8.65	80.22	3.07	22.49	16	0.41	170
15. Shandong	13.66	4.32	13.98	581	32.89	88.09	5.84	19.05	51	0.51	794
16. Henan	7.57	2.61	11.40	558	6.56	82.41	2.43	15.24	18	0.47	422
17. Hubei	9.01	3.93	11.81	322	8.90	85.22	3.02	19.28	21	0.53	291
18. Hunan	7.55	2.96	12.60	290	8.38	80.89	2.59	19.32	17	0.47	219
19. Guangdong	17.21	7.08	22.94	445	175.69	91.97	6.08	18.52	28	0.70	766
20. Guangxi	5.97	1.87	12.33	194	9.75	76.15	2.01	15.80	28	0.28	116
21. Hainan	8.32	2.38	10.31	238	23.58	62.99	3.47	22.18	2	0.63	198
22. Chongqing	7.21	2.68	13.59	381	9.41	85.00	3.72	14.72	23	0.44	274
23. Sichuan	6.42	2.46	13.92	150	8.77	79.32	2.75	14.42	27	0.22	96
24. Quizhou	3.60	1.22	11.39	214	9.48	78.00	1.99	13.96	15	0.29	77
25. Yunnan	5.66	1.80	13.47	111	9.13	79.60	2.30	6.71	41	0.43	63
26. Shaanxi	6.48	2.31	11.83	180	12.27	86.66	3.24	23.18	25	0.26	117
27. Gansu	5.02	1.83	12.93	57	8.20	81.86	2.39	16.89	25	0.10	29
28. Qinghai	7.28	1.91	16.69	7	7.27	88.17	4.77	16.15	20	0.04	5
29. Ningxia	6.69	2.10	13.72	111	15.99	85.60	5.52	18.40	69	0.25	74
30. Xinjiang	9.70	2.18	13.20	12	21.40	78.01	5.03	22.36	14	0.05	11

Sources:

X_1 Per capita GDP at current market prices estimated by production approach (1000 yuan). CSYB2004. 3-11.

X_2 PRS = Per capita retail sales (1000 Yuan) = (Retail Sales / GDPpop), (GDPpop is the number of mid-year population which is the ratio of GDP to per capita GDP of the same year) CSYB2004, 17-3

X_3 Average wage of staff and workers (1000 Yuan) in State-owned units, CSYB2004, 5-28.

X_4 Population density (persons per sq km). It is calculated as GDPpop divided by Land area of the region. Data of Land area is from China Development Report 1995, P231; figures for Sichuan & Chongqing are obtained from http://hk.geocities.com/chinamap04.

X_5 (X+M)/GDP is the measure of the degree of openness to international trade, expressed in %; (X+M) is the total Import and Export value of commodities by places of destination or origin. The value has been changed to yuan by using USD100 = RMB827.7 , 100 million yuan, calculated from CSYB2004, 18-2 &18-11.

X_6 Contribution of Secondary and tertiary Industries to GDP, CSYB2004, 3-11

X_7 Per capita Total Investment in Fixed Assets, TIFA, (1000 yuan) = TIFA/GDPpop, calculated from CSYB2004, 6-4

X_8 The percentage of population with education level at senior secondary school and higher to the population aged 6 and over, calculated from CSYB2004, 4-11.

X_9 Per capita government expenditure for innovation enterprises (yuan) = GEIE/GDPpop, calculated from CSYB2004, 8-15.

X_{10} Length of transport routes, TR, (railways, waterways and highways) per sq. km. It is calculated as the ratio of TR to land area of the respective region. CSYB2004, 16-3.

X_{11} Resource density (10,000 yuan per sq.km), it is measured by GDP divided by land area.

Table 6 Principal component solution of the factor model

Z_i	1998				2002				2003			
	Factor loadings of F_1 ($\sqrt{\lambda_1} e_{i1}$)	Factor score coefficients of F_1 ($e_{i1}/\sqrt{\lambda_1}$)	Communality (ϕ_i^2)	Specific Variance (ψ_i)	Factor loadings of F_1 ($\sqrt{\lambda_1} e_{i1}$)	Factor score coefficients of F_1 ($e_{i1}/\sqrt{\lambda_1}$)	Communality (ϕ_i^2)	Specific Variance (ψ_i)	Factor loadings of F_1 ($\sqrt{\lambda_1} e_{i1}$)	Factor score coefficients of F_1 ($e_{i1}/\sqrt{\lambda_1}$)	Communality (ϕ_i^2)	Specific Variance (ψ_i)
Z_1	0.987	0.112	0.974	0.026	0.989	0.114	0.977	0.023	0.988	0.114	0.975	0.025
Z_2	0.978	0.111	0.956	0.044	0.966	0.112	0.933	0.067	0.961	0.111	0.924	0.076
Z_3	0.909	0.103	0.826	0.174	0.873	0.101	0.762	0.238	0.863	0.100	0.744	0.256
Z_4	0.877	0.099	0.769	0.231	0.887	0.103	0.787	0.213	0.901	0.104	0.812	0.188
Z_5	0.759	0.086	0.576	0.424	0.783	0.091	0.613	0.387	0.806	0.093	0.649	0.351
Z_6	0.797	0.090	0.636	0.364	0.733	0.085	0.538	0.462	0.721	0.083	0.519	0.481
Z_7	0.980	0.111	0.960	0.040	0.953	0.110	0.908	0.092	0.944	0.109	0.891	0.109
Z_8	0.836	0.095	0.698	0.302	0.849	0.098	0.720	0.280	0.855	0.099	0.731	0.269
Z_9	0.947	0.107	0.898	0.102	0.934	0.108	0.872	0.128	0.911	0.106	0.831	0.169
Z_{10}	0.866	0.098	0.750	0.250	0.851	0.099	0.725	0.275	0.863	0.100	0.745	0.255
Z_{11}	0.885	0.100	0.783	0.217	0.897	0.104	0.805	0.195	0.902	0.104	0.813	0.187
	Eigenvalue (λ_1)		% Variance ($\lambda_1/11$)		Eigenvalue (λ_1)		% Variance ($\lambda_1/11$)		Eigenvalue (λ_1)		% Variance ($\lambda_1/11$)	
	8.826		80.24%		8.641		78.55%		8.634		78.49%	

Notes:

i/ F_1 stands for the first common factor score.

ii/ % Variance stands for the percentage of total standardized variance attributable to the first common factor.

Table 7 The socio-economic environment index (SEEI) in 1998, 2002 and 2003

	1998			2002			2003		
Region	CFSFPC	SEEI	Rank	CFSFPC	SEEI	Rank	CFSFPC	SEEI	Rank
1. Beijing	2.412	221	2	2.258	220	2	2.215	220	2
2. Tianjin	1.305	201	3	1.524	209	3	1.503	208	3
3. Hebei	-0.211	93	10	-0.297	85	11	-0.289	86	11
4. Shanxi	-0.35	81	14	-0.346	81	13	-0.347	81	12
5. Inner Mongolia	-0.623	59	27	-0.541	65	23	-0.526	67	24
6. Liaoning	0.198	129	7	0.050	116	9	0.059	117	9
7. Jilin	-0.342	81	13	-0.302	85	12	-0.372	79	14
8. Heilongjiang	-0.275	87	12	-0.349	81	14	-0.413	76	19
9. Shanghai	3.893	223	1	3.897	223	1	3.913	223	1
10. Jiangsu	0.420	148	6	0.471	152	6	0.581	160	6
11. Zhejiang	0.492	153	5	0.655	166	5	0.740	171	5
12. Anhui	-0.450	73	20	-0.506	68	22	-0.417	75	20
13. Fujian	0.159	125	8	0.116	122	7	0.061	117	8
14. Jiangxi	-0.581	62	26	-0.573	63	25	-0.456	72	21
15. Shandong	0.013	112	9	0.086	119	8	0.093	120	7
16. Henan	-0.379	78	16	-0.400	77	16	-0.393	77	16
17. Hubei	-0.235	91	11	-0.285	86	10	-0.277	87	10
18. Hunan	-0.436	74	18	-0.405	76	18	-0.402	76	17
19. Guangdong	0.901	182	4	0.840	178	4	0.757	173	4
20. Guangxi	-0.649	57	28	-0.617	60	27	-0.656	57	27
21. Hainan	-0.400	77	17	-0.569	63	24	-0.538	66	25
22. Chongqing	-0.378	78	15	-0.354	80	15	-0.354	80	13
23. Sichuan	-0.564	64	25	-0.585	62	26	-0.599	61	26
24. Guizhou	-0.815	46	30	-0.726	52	30	-0.725	52	30
25. Yunnan	-0.529	66	24	-0.617	60	28	-0.674	56	28
26. Shaanxi	-0.487	70	22	-0.479	70	21	-0.410	76	18
27. Gansu	-0.653	57	29	-0.652	57	29	-0.676	55	29
28. Qinghai	-0.517	67	23	-0.455	72	20	-0.489	70	22
29. Ningxia	-0.443	73	19	-0.404	76	17	-0.387	78	15
30. Xinjiang	-0.471	71	21	-0.430	74	19	-0.517	67	23
1998PCFDI	- -	(0.905)	- -	- -	- -	- -	- -	- -	- -
1999PCFDI	- -	(0.914)	- -	- -	- -	- -	- -	- -	- -
2002PCFDI	- -	(0.911)	- -	- -	(0.898)	- -	- -	- -	- -
2003PCFDI	- -	(0.880)	- -	- -	(0.877)	- -	- -	(0.882)	- -

Notes:

i/ Figures in parenthesis are Pearson correlation coefficients.

ii/ PCFDI represents per capita Foreign Direct Investment.

iii/ CFSFPC represent the common factor score of the first principal component.

Figure 1 Map of China

Source: http://www.chinapage.com/map/peopledailymap.html

Figure 2 Complete linkage dendrograms for the data matrix of the 10 socio-economic Variables in 2003

* * * * * * H I E R A R C H I C A L C L U S T E R A N A L Y S I S * * * *

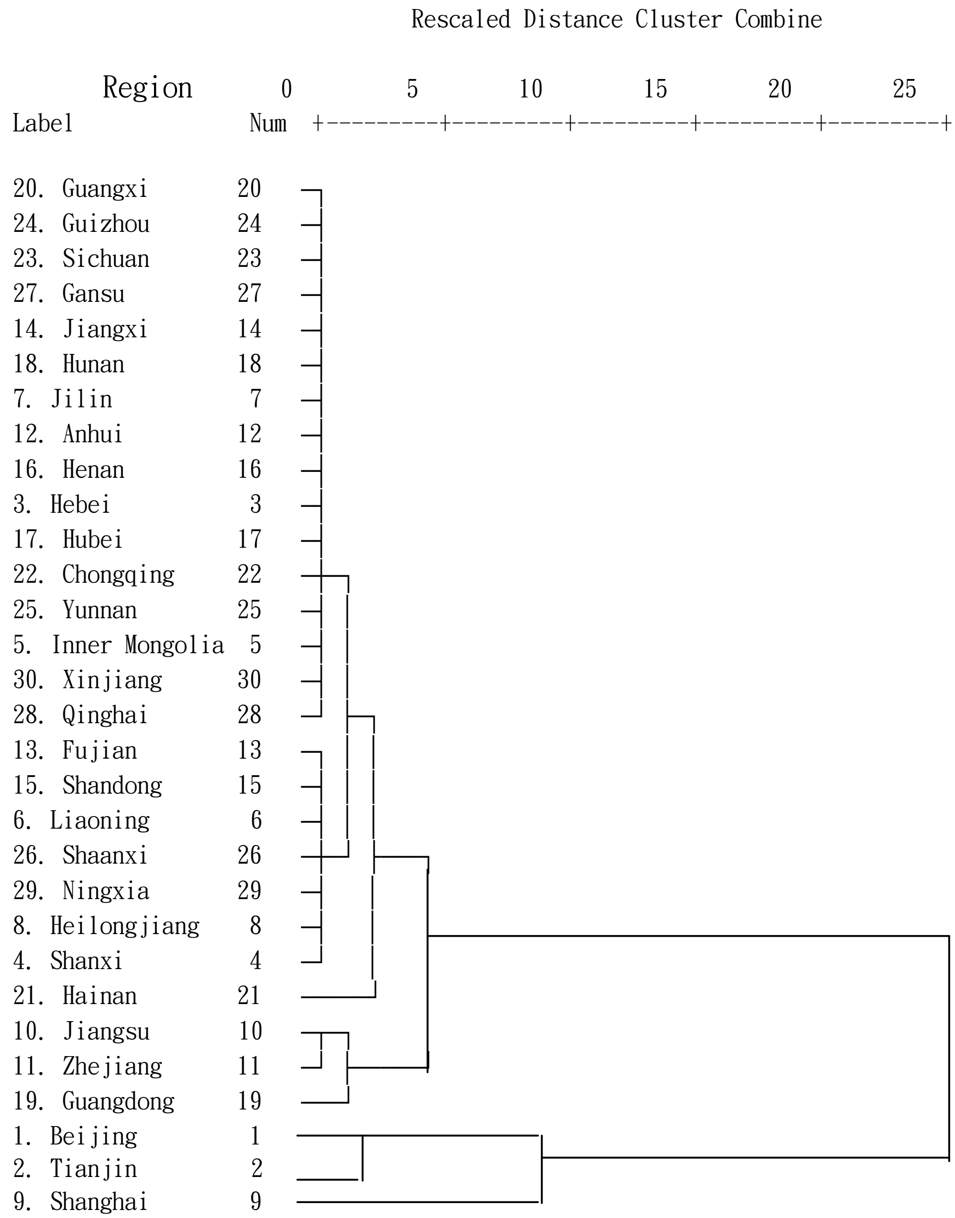


Figure 3 Complete linkage dendrograms for the data matrix of the 10 socio-economic Variables in 2002

******H I E R A R C H I C A L C L U S T E R A N A L Y S I S****

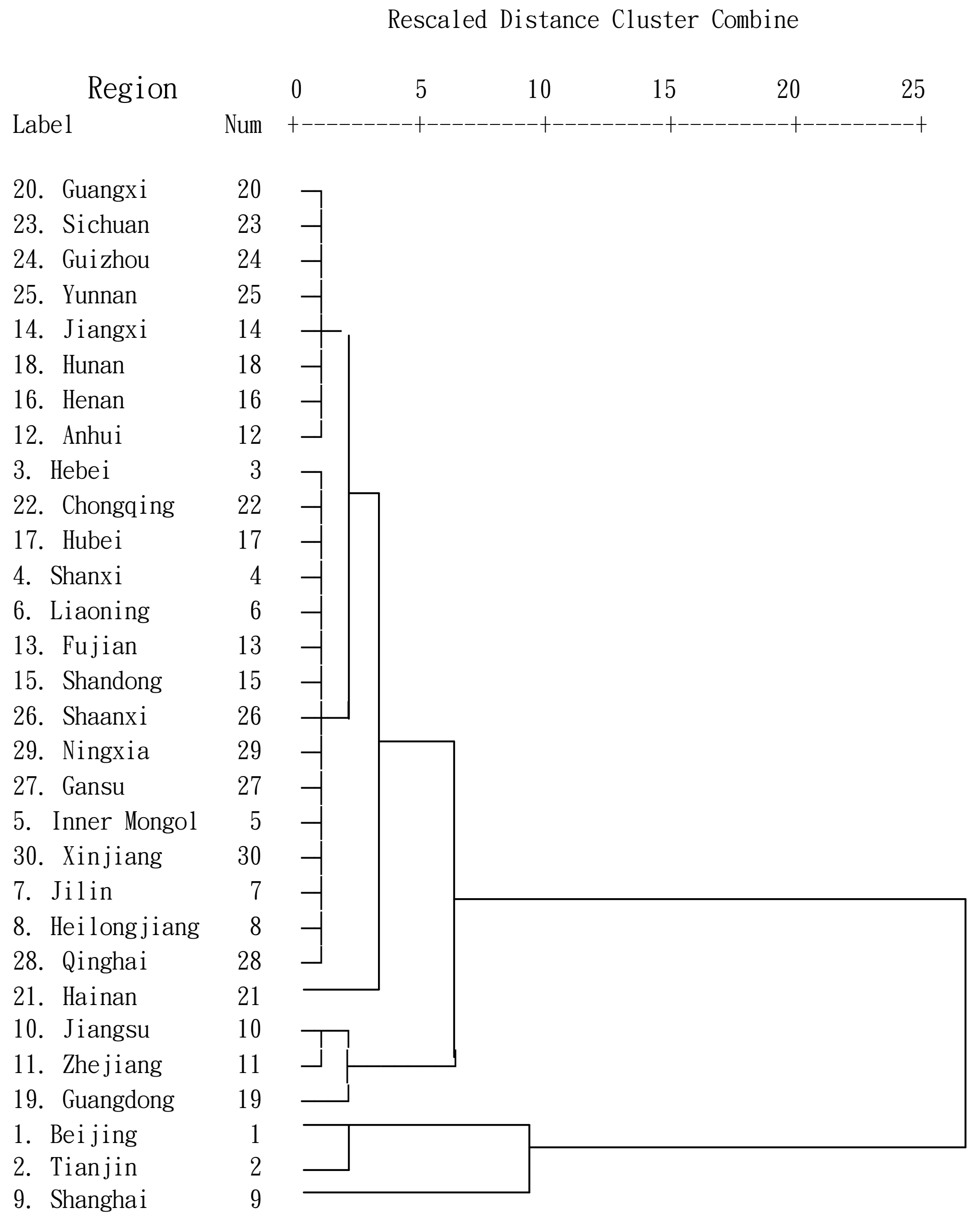
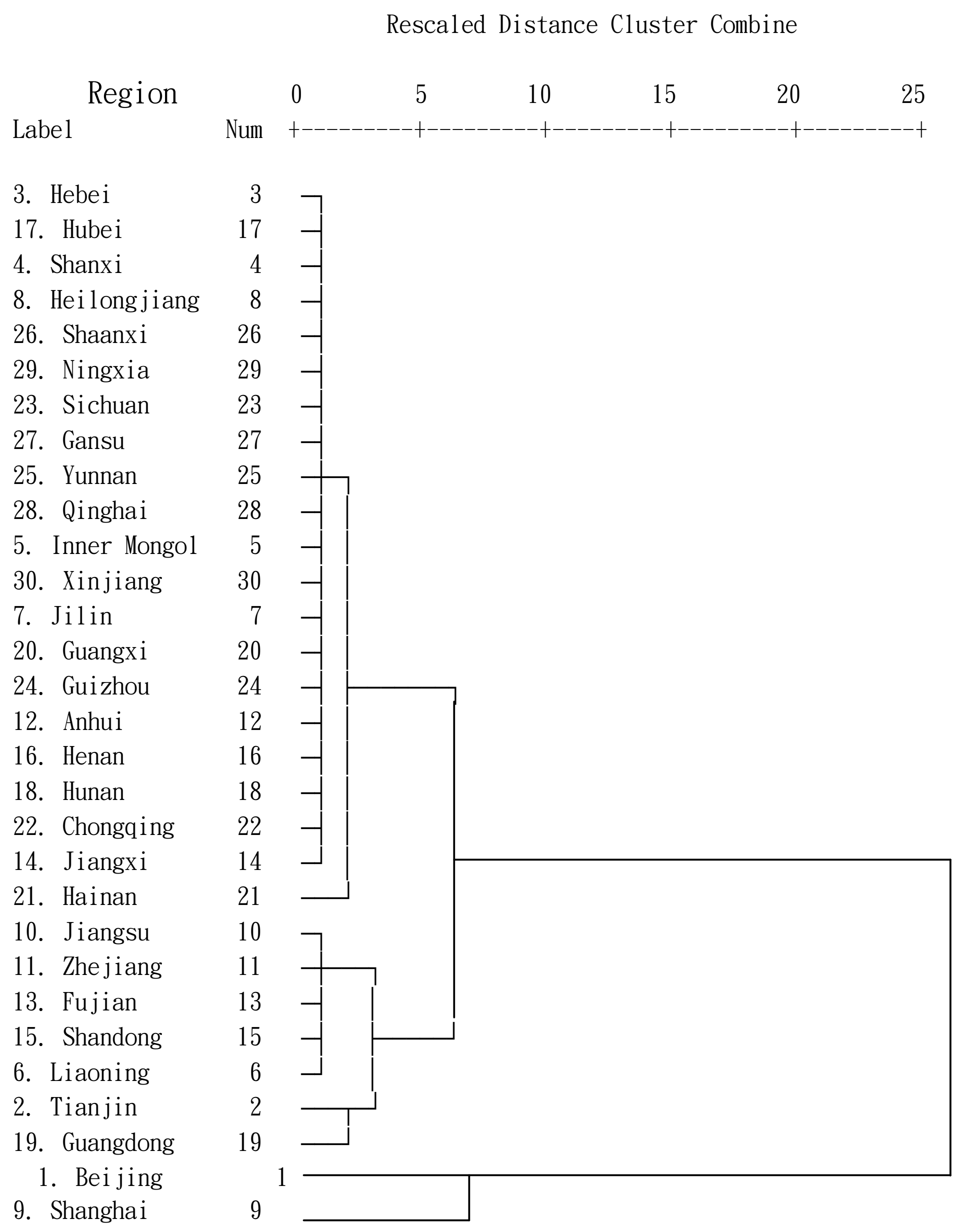


Figure 4: Complete linkage dendrograms for the data matrix of the 10 socio-economic Variables in 1998

* * * * * * H I E R A R C H I C A L   C L U S T E R   A N A L Y S I S * * * *



REFERENCE

Bartholomew, D. J. (1987). Latent Variable Models and Factor Analysis, Oxford University Press, New York.

Barnett, Steven (2004). "China's Growth and Integration into the World Economy: Prospects and Challenges." Occasional Paper 232, International Monetary Fund.

China Statistical Yearbook, various issues, China Statistical Publishing House.

Coughlim, C. C. & Segev, E. (1999). Foreign Direct Investment in China: A Spatial Econometric Study, Working Paper 1999-001A, Federal Reserve Bank of ST. Louis.

Cramer, D. (2003), Advanced Quantitative Data Analysis, Open University Press.

Crum, R. L., Brigham, E . F. and Houston, J. F. (2005). Fundamentals of International Finance, Thomson South-Western, 1st Edition.

Everitt, B.S. and Dunn, G. (1991). Applied Multivariate Data Analysis, Edward Arnold, London.

Friedman, J. H. and Meulman, J. J. (2004). Clustering objects on subsets of attributes, Journal of the Royal Statistical Society, Series B, Statistical Methodology, Vol. 66, Part 4, pp. 815~849.

Hu, A. G., Wang, S. G., and Hong, S. K. (1995). Report on China's Regional Disparities (in Chinese), Liaoning People PublishingHouse.

IMF (2004). World Economic Outlook September 2004

IMF Balance of Payment Manual (1996), 5th edition

IMD World Competitiveness Yearbook 2004

Johnson, R.A. and Wichern, D. W., (2002), Applied Multivariate Statistical Analysis, 5th edn., Prentice Hall.

Javorcik, B.S.(2004). "Does Foreign Direct Investment Increase the Productivity of Domestic Firm? In Search of Spillovers Through Backward Linkages." The American Economic Review (June), 94 (3).

Kaiser, H. F., (1960), The application of electronic computers to factor analysis. Educational and Psychological Measurement, vol. 20, pp.141-151.

Ng, L. F. Y and Tuan, C. (2003). Location decisions of manufacturing FDI in China: Implication of China's WTO accession, Journal of Asian Economics, 14(1), 51~72.

Poon, C. C., Hon, T. Y. and Woo, K. Y. (1997). A Cluster Analysis of Regional Economic

Disparities in China (in Chinese), Economics Working Paper Series 97012, Economics Department, Hong Kong Shue Yan College, 1996.

Prasad, Eswar (Ed.) (2004). "Hong Kong SAR: Meeting the Challenges of Integration with the Mainland." Occasional Paper 226, International Monetary Fund.

Rousson, V. and Gasser, Th. (2004). Simple component analysis, Journal of the Royal Statistical Society, Series C, Applied Statistics, Blackwell, Vol. 53, Part 4, pp. 539-555.

Wang, Jian.(2004). The Determinants of Regional Distribution of Foreign Direct Investment (in Chinese), Economic Science, Ministry of Education of the PRC, No. 5, Serial No. 104, pp. 116-125.

Important Note

The working paper series is a series of occasional papers funded by the Research and Staff Development Committee. The objective of the series is to arouse intellectual curiosity and encourage research activities. The expected readership will include colleagues within Hong Kong Shue Yan College, as well as academics and professionals in Hong Kong and beyond.

All opinions, information and/or statements made in the papers are exclusively those of the authors. Hong Kong Shue Yan College and its officers, employees and agents are not responsible, in whatsoever manner and capacity, for any loss and/or damage suffered by any reader or readers of these papers.

Please address any comments and further inquiries to:

Dr. Shu-kam Lee
Working Paper Coordinator
Department of Economics
Hong Kong Shue Yan College
10 Wai Tsui Crescent
Braemar Hill Road
North Point
Hong Kong
Fax: 28068044
Tel: 25707110
Email: sklee@hksyc.edu