The Chinese Economy, vol. 38, no. 2, March–April 2005, pp. 56–87.

ISSN 1097–1475 / 2005 $9.50 + 0.00.

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

Regional Distribution of Foreign Direct Investment in China

A Multivariate Data Analysis of Major Socioeconomic Variables

Abstract: *This study aims to identify the driving forces behind the regional distribution of inward FDI flows in China for the period 1998–2003. Based on the eleven selected socioeconomic variables and by adopting factor analysis, we accept the hypothesis that the overall socioeconomic environment in the administrative regions of China is a fundamental determinant of regional disparity in FDI. We use the complete linkage clustering technique to classify these regions into broader groups and then explore the similarities and dissimilarities between them. The findings from the study provide a yardstick for multinational firms regarding location decisions at the provincial level.*

According to the *IMF Balance of Payments Manual*,[1] 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,[2] 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 in which to establish their firms.

Tai-Yuen Hon is an assistant professor of economics, Che-Cheong Poon is an associate professor of economics, and Kai-Yin Woo is an assistant professor of economics, Hong Kong Shue Yan College; Wai Tsui Crescent, Braemar Hill Road, North Point, Hong Kong (tel.: 852-25707110; fax: 852-28068044; e-mail: ccpoon@hksyc.edu).

The material is based on work that was supported in part by grants from the Research and Staff Development Funds of Hong Kong Shue Yan College. The authors are grateful to Louis Cheng and Andrea Hope for helpful comments.

Based on a regional analysis of the major socioeconomic 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.

Between 1994 and 2003, China's economy grew on average by 8.1 percent (10.7 percent in nominal terms). The *IMF World Economic Outlook* 2004[3] forecast growth at 9 percent 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 has been 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 socioeconomic environment is one of the principal determinants of the regional distribution of inward FDI flows to China, we first select a set of variables based on the criteria of the availability of official statistics, significant regional differences, and the socioeconomic environment factors that are highly correlated with FDI. Then, we use the method of principal components factor analysis to construct a socioeconomic environment index for each of the thirty regions. A simple correlation analysis is conducted between the regional socioeconomic environment index and the regional inward FDI flows to provide evidence for our hypothesis.

The thirty-one administrative regions in China[4] (except for Taiwan, Hong Kong, and Macao) are grouped into three areas (eastern, central, and western) 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, Yunnan, Shaanxi, Gansu, and Qinghai), three autonomous regions (Tibet, Ningxia, and Xinjiang) located in the northwest and southwest of China, and one municipality (Chongqing).[5] The central area covers eight provinces (Shanxi, Jilin, Heilongjiang, Anhui, Jiangxi, Henan, Hubei, and Hunan) and one autonomous region (Inner Mongolia). Owing to the unavailability of FDI statistics in Tibet, we have to exclude Tibet from our list, leaving thirty regions in our sample.

Since it would be a time-consuming task for foreign entrepreneurs to select from a group of thirty a region in which to locate their firms, we have used a hierarchical clustering method to classify regions into broader regional groups (clusters) according to the similarities (internal cohesion) and dissimilarities (external isolation) in their socioeconomic environment. Last, based on the socioeconomic environment 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.

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 in attracting foreign direct investment. According to the *China Statistics Yearbook* 2004, China received direct inward investment totaling about 442.8 billion yuan in 2003, which equaled about 3.8 percent of gross domestic product (GDP). This represents a more than eighteen-fold increase of FDI in 1991 (23.2 billion yuan).[6] An in-depth study of the FDI in China enables us to point out several characteristics.

First, 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, the euro area,[7] and Japan entered China in larger numbers. In 2003, the FDI flows to China from these advanced countries were about 23 percent of total FDI flows to China (US$12.27 billion), up 127 percent over 1994;[8] however, Hong Kong, Taiwan, and Singapore still accounted for over 43 percent of total FDI in the same year.

Second, the contribution made to China by FDI has been to raise productivity rather than meet financial needs. Following the standard four-sector GDP determination model,[9] 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 equal nearly 47 percent of GDP,[10] 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.

Third, 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[11] range from 0.9945 to 0.9998). On the other hand, although the correlation among the FDI series is high, the decreasing value of correlation coefficients demonstrates that 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 that the rigid pattern of regional GDP cannot capture the changing pattern of regional FDI and that the correlation between two variables may be due to their common relation to other variables. Accord-

Table 1

Correlation Matrix of Regional GDP and FDI, 1998–2003

	GDP1998 (FDI1998)	GDP1999 (FDI1999)	GDP2000 (FDI2000)	GDP2001 (FDI2001)	GDP2002 (FDI2002)	GDP2003 (FDI2003)
GDP1998	1					
(FDI1998)	(1)					
GDP1999	0.9997	1				
(FDI1999)	(0.9951)	(1)				
GDP2000	0.9990	0.9996	1			
(FDI2000)	(0.9916)	(0.9908)	(1)			
GDP2001	0.9986	0.9993	0.9998	1		
(FDI2001)	(0.9880)	(0.9831)	(0.9948)	(1)		
GDP2002	0.9970	0.9979	0.9985	0.9992	1	
(FDI2002)	(0.9460)	(0.9363)	(0.9664)	(0.9717)	(1)	
GDP2003	0.9945	0.9957	0.9966	0.9976	0.9994	1
(FDI2003)	(0.8211)	(0.8065)	(0.8556)	(0.8763)	(0.9515)	(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, table 17-16, and *China Statistics Yearbook* 2004, tables 3-10, 18-2, and 18-16.

ingly, a set of collinear socioeconomic variables (including GDP), which is significantly correlated with FDI, 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 thirty regions in 1998 and 2003 are shown in Table 2. First, FDI is highly concentrated in the eastern area (87.41 percent in 1998 and 86.69 percent of total FDI inflow in 2003), with a significant portion going to Shanghai, Jiangsu, Zhejiang, Shandong, and Guangdong (56.92 percent in 1998 and 65.85 percent of total FDI inflow in 2003). Second, although the aggregate FDI has increased by 16.87 percent from 1998 to 2003, the regional distribution of FDI has experienced a substantial change. The fact that thirteen regions recorded an increase of inward FDI flows and seventeen 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 percent, 247 percent, and 173 percent, re-

Table 2

Regional Distribution of FDI in China, 1998 and 2003

	1998				2003				
	Total FDI		Per capita FDI		Total FDI		Per capita FDI		
Region	Amount (bn yuan)	Rank	Amount (yuan)	Rank	Amount (bn yuan)	Rank	Amount (yuan)	Rank	Percent change in total FDI over the years
Eastern area	327.70 (87.47)				379.86 (86.69)				15.91
1. Beijing	17.95 (4.79)	7	1,649	3	18.14 (4.14)	8	1,587	2	1.05
2. Tianjin	17.50 (4.67)	8	1,939	2	12.70 (2.90)	11	1,377	3	–27.41
3. Hebei	11.83 (3.16)	9	181	11	7.98 (1.82)	13	118	14	–32.54
5. Inner Mongolia	0.75 (0.20)	25	32	24	0.73 (0.17)	24	31	25	–2.53
6. Liaoning	18.13 (4.84)	6	436	8	23.38 (5.34)	6	555	8	28.90
9. Shanghai	29.82 (7.95)	4	2,284	1	45.26 (10.33)	4	3,383	1	51.80
10. Jiangsu	54.91 (14.65)	2	764	7	87.44 (19.96)	1	1,179	4	59.25
11. Zhejiang	10.91 (2.91)	10	246	9	41.22 (9.41)	5	884	5	277.79
13. Fujian	34.87 (9.30)	3	1,100	5	21.51 (4.91)	7	616	7	–38.31
15. Shandong	18.24 (4.87)	5	207	10	49.80 (11.37)	3	547	9	173.05
19. Guangdong	99.51 (26.54)	1	1,400	4	64.75 (14.78)	2	818	6	–34.93
20. Guangxi	7.34 (1.96)	12	157	12	3.46 (0.79)	16	76	15	–52.78
21. Hainan	5.94 (1.58)	14	815	6	3.49 (0.80)	15	432	10	–41.28

Central area	35.85 (9.56)				48.27 (11.02)				34.64
4. Shanxi	2.02 (0.54)	23	69	20	1.77 (0.40)	22	54	20	–12.66
7. Jilin	3.39 (0.90)	19	129	14	1.58 (0.36)	23	58	19	–53.43
8. Heilongjiang	4.36 (1.16)	16	117	15	2.66 (0.61)	20	70	17	–38.88
12. Anhui	2.29 (0.61)	22	37	23	3.04 (0.69)	18	49	21	32.66
14. Jiangxi	3.85 (1.03)	17	93	18	13.34 (3.04)	9	315	11	246.61
16. Henan	5.10 (1.36)	15	55	21	4.46 (1.02)	14	48	22	–12.59
17. Hubei	8.06 (2.15)	11	137	13	12.99 (2.96)	10	217	12	61.21
18. Hunan	6.78 (1.81)	13	108	17	8.43 (1.92)	12	137	13	24.37
Western area	11.37 (3.03)				10.05 (2.29)				–11.61
22. Chongqing	3.57 (0.95)	18	117	16	2.16 (0.49)	21	69	18	–39.51
23. Sichuan	3.08 (0.82)	20	37	22	3.41 (1.78)	17	40	23	10.67
24. Guizhou	0.38 (0.10)	26	10	28	0.37 (0.08)	26	10	28	–0.34
25. Yunnan	1.21 (0.32)	24	29	25	0.69 (0.16)	25	16	27	–42.46
26. Shaanxi	2.48 (0.66)	21	69	19	2.75 (0.63)	19	74	16	10.57
27. Gansu	0.32 (0.09)	27	13	27	0.19 (0.04)	28	7	29	–39.40
28. Qinghai	0.00 (- -)	30	0	30	0.21 (0.05)	27	39	24	+
29. Ningxia	0.15 (0.04)	29	29	26	0.14 (0.03)	29	25	29	–6.11
30. Xinjiang	0.18 (0.05)	28	10	29	0.13 (0.03)	30	7	30	–29.23
Country total	374.92 (100)				438.18 (100)				16.87
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	—	—

(*continued*)

Table 2 (*continued*)

	1998				2003				
	Total FDI		Per capita FDI		Total FDI		Per capita FDI		
Region	Amount (bn yuan)	Rank	Amount (yuan)	Rank	Amount (bn yuan)	Rank	Amount (yuan)	Rank	Percent change in total FDI over the years
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	—	2,284.00	—	87.44	—	3,383.00	—	—

Sources: Calculated from *China Statistics Yearbook* 2000, 17-16, and *China Statistics Yearbook* 2004, 18-2 and 18-16.

Notes:

1. The FDI figures have been converted from US dollars to yuan by using the average exchange rate of 821.91 yuan and 827.7 yuan against US$100 in 1998 and 2003, respectively.
2. 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.
3. Figures in parentheses are percentages to total FDI.

spectively. 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 drops 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[12] toward 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.

Data Description

Many determinants of FDI have been identified in the economic literature (Coughlim and Segev 1999; Crum, Brigham, and Houston 2005, 97–99; Ng and Tuan 2003; Wang 2004). 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 in attracting FDI, our criteria for including elements in the list of socioeconomic variables that determine the attractiveness of FDI are based largely on the consideration of market size and factor productivity.

Market size refers to the extent to which a specific production output could be sold. At the macrolevel, the number of potential buyers in the market and the income of consumers are major determinants of market size. Among the socioeconomic variables restricted by the availability of data in China's official statistics, we choose per capita GDP,[13] per capita retail sales, average wage, and population density as proxies for market size. In addition, since FDI 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 chosen 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 matrices in 1998, 2002 and 2003 are provided in Tables 3, 4, and 5, and 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 the yuan against the US dollar.

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 aged six and over with education level at senior secondary school and higher (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.

The FDI Attraction Index

Having determined the set of socioeconomic variables to capture the abstract concept of regional attractiveness of FDI, it is necessary to go on to construct an index number which purports to measure that concept. To do this, we apply principal components analysis to summarize the eleven socioeconomic variables into m principal components[14] ($m < 11$), and then use the resulting factor score coefficients of

Table 3

Data Matrix for the Observed Values of the Eleven Regional Socioeconomic Variables in 1998

	Socioeconomic variable										
Region	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	1,197
2. Tianjin	14.81	6.51	10.24	799	65.75	94.5	6.33	22.76	138	0.44	1,183
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	5,854
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

(*continued*)

Table 3 (*continued*)

Region	Socioeconomic variable										
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
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. Guizhou	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 (1,000 yuan). *China Statistics Yearbook 1999*, 3–9. X_2: PRS = Per capita retail sales (1,000 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). *China Statistics Yearbook* 1999, 16–2. X_3: Average wage of staff and workers (1,000 yuan) in state-owned units, *China Statistics Yearbook* 1999, 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, 231; 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 percent; (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 US$100 = 827.91 yuan, 100 million yuan, calculated from *China Statistics Yearbook* 1999, 17–2 and17–11. X_6: Contribution of secondary and tertiary industries to GDP, *China Statistics Yearbook* 1999, 3–9. X_7: Per capita Total Investment in Fixed Assets, TIFA (1,000 yuan) = TIFA / GDPpop, calculated from *China Statistics Yearbook* 1999, 6–4. X_8 The percentage of population aged six and over with education level at senior secondary school and higher, calculated from *China Statistics Yearbook* 1999, 4–9. X_9: Per capita government expenditure for innovation enterprises (yuan) = GEIE / GDPpop, calculated from *China Statistics Yearbook* 1999, 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. *China Statistics Yearbook* 1999, 15–3. X_{11}: Resource density (10,000 yuan per sq. km.) is measured by GDP divided by land area.

Table 4

Data Matrix for the Observed Values of the Eleven Regional Socioeconomic Variables in 2002

	Socioeconomic variable										
Region	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	1,912
2. Tianjin	22.38	5.62	16.83	811	92.21	95.90	8.81	33.10	336	0.96	1,815
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	8,585
10. Jiangsu	14.39	3.21	13.89	720	57.99	89.47	4.67	16.97	71	0.83	1,036
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

(*continued*)

Table 4 (*continued*)

Region	Socioeconomic variable										
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
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. Guizhou	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 (1,000 yuan). *China Statistics Yearbook* 2003, 3–9. X_2: PRS = Per capita retail sales (1,000 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). *China Statistics Yearbook* 2003, 16–3. X_3: Average wage of staff and workers (1,000 yuan) in state-owned units, *China Statistics Yearbook* 2003, 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, 231; 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 percent; (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 US$100 = 27.7 yuan, 100 million yuan, calculated from *CSYB* 2003, 17–2 and 17–11. X_6: Contribution of secondary and tertiary industries to GDP, calculated from *China Statistics Yearbook* 2003, 3–9. X_7: Per capita Total Investment in Fixed Assets, TIFA, (1,000 yuan) = TIFA / GDPpop, calculated from *China Statistics Yearbook* 2003, 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 *China Statistics Yearbook* 2003, 4–9. X_9: Per capita government expenditure for innovation enterprises (yuan) = GEIE / GDPpop, calculated from *China Statistics Yearbook* 2003, 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. *China Statistics Yearbook* 2004, 15–3. X_{11}: Resource density (10,000 yuan per sq. km.) is measured by GDP divided by land area.

Table 5

Data Matrix for the Observed Values of the Eleven Regional Socioeconomic Variables in 2003

	Socioeconomic variable										
Region	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	2,180
2. Tianjin	26.53	10.00	19.35	816	101.55	96.40	11.27	33.94	318	0.97	2,166
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	2,124	146.34	98.55	18.68	45.72	1,155	1.42	9,922
10. Jiangsu	16.81	4.81	17.50	723	80.56	91.12	7.06	19.22	69	0.89	1,215
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

(*continued*)

Table 5 (*continued*)

Region	Socioeconomic variable X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
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. Guizhou	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 (1,000 yuan). *CSYB* 2004. 3–11. X_2: PRS = per capita retail sales (1,000 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). *CSYB* 2004, 17–3. X_3: Average wage of staff and workers (1,000 yuan) in state-owned units, *CSYB* 2004, 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, 231; figures for Sichuan and 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 percent; (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 US$100 = 827.7 yuan, 100 million yuan, calculated from *CSYB* 2004, 18–2 and 18–11. X_6: Contribution of secondary and tertiary industries to GDP, *CSYB* 2004, 3–11. X_7: Per capita Total Investment in Fixed Assets, TIFA (1,000 yuan) = TIFA / GDPpop, calculated from *CSYB* 2004, 6–4. X_8: The percentage of population with education level at senior secondary school and higher to the population aged six and over, calculated from *CSYB* 2004, 4–11. X_9: Per capita government expenditure for innovation enterprises (yuan) = GEIE / GDPpop, calculated from *CSYB* 2004, 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. *CSYB* 2004, 16–3. X_{11}: Resource density (10,000 yuan per sq. km.); it is measured by GDP divided by land area.

these principle components (preferably $m = 1$, that is, the first principal component) as weights to calculate the weighted average of the eleven socioeconomic variables for a specific region to obtain its FDI attraction index.

Methodology

Since the eleven regional socioeconomic variables are measured on different scales or on a common scale with substantial difference in magnitude, it is necessary to transform the eleven original variables on the same scale by standardizing them for the subsequent factor analysis. Suppose that each observed socioeconomic variable X_i has a constant mean μ_i with a finite variance σ_i^2 over thirty regions in China. We transform X_i in $X' = (X_1, X_2, \ldots, X_{11})$ to Z_i in the random vector $Z' = (Z_1, Z_2, \ldots, Z_{11})$, where:

$$Z_i = \frac{X_i - \mu_i}{\sigma_i}$$

with a mean of zero and a standard deviation of one. In the factor analysis model, the standardized variable Z_i is expressed exactly as a linear combination of common factor scores of principal components $F_1, F_2, \ldots, F_m$ and one additional specific factor (or the error term) e_i, it can be written as:

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

$$i = 1, \ldots, 11.\ \mathrm{k} = 1, \ldots, \mathrm{m} \qquad (1)$$

where l_{ik} is known as factor loading of the ith standardized variable Z_i on the kth principal component F_k; and m stands for the number of principal components extracted.

The common factor score of the kth principal component F_k and the specific factor ε_i are assumed to satisfy the following conditions:

$$\mathrm{E}(F_k) = 0 \text{ and } \mathrm{Var}(F_k) = 1,\ \forall\ \mathrm{k},\ \mathrm{Cov}(F_k, F_l) = 0,\ \forall\ k \neq l.$$

$$\mathrm{E}(\varepsilon_i) = 0 \text{ and } \mathrm{Var}(\varepsilon_i) = \Psi_i,\ \forall\ i,\ \mathrm{Cov}(\varepsilon_i, \varepsilon_p) = 0,\ \forall\ i \neq \mathrm{p}. \qquad (2)$$

$$\mathrm{Cov}(\varepsilon_i, F_k) = 0; \forall\ i, k.$$

In this paper, the factor loadings and variances of the original standard normal variables (Z_i) are estimated using the principal components method.[15] The principal component solution of the factor model is expressed in terms of the eigenvalue-eigenvector pairs, denoted (λ_1, e_1), $(\lambda_2, e_2), \ldots, (\lambda_{11}, e_{11})$ of the 11×11 variance-covariance matrix of Z, where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{11}$, $e'_1 e'_1 = e'_2 e_2 = \ldots = e'_{11} e_{11}$. The estimated factor loading l_{ik} is obtained by:

$$\sqrt{\lambda_j}\, e_{ik},$$

where e_{ik} stands for the ith element of the kth eigenvector. Moreover, the contribution to the total variance of the standardized variables,

$$\sum_{i=1}^{11} \mathrm{Var}(Z_i),$$

explained by the kth principal component factor score, is calculated by adding the squared estimates of factor loadings of all the standardized variables in Z under the kth principal component, that is,

$$\sum_{i=1}^{11} l_{ik}^2 = l_{1k}^2 + l_{2k}^2 ... + l_{11k}^2 \text{ or } (\sqrt{\lambda_k} e_k), (\sqrt{\lambda_k} e_k),$$

which gives the kth 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 kth 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. 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 equivalence of the variance of one standardized variable.

Furthermore, the portion of Var (Z_i) explained by all the m principal components extracted is called the ith communality, denoted

$$\phi_i^2,$$

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.[16]

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:

$\sqrt{\lambda_k}$.[17]

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:

$$(\sqrt{\lambda_1})^{-1} e_{1.k}, (\sqrt{\lambda_1})^{-1} e_{2.k}, \ldots, (\sqrt{\lambda_1})^{-1} e_{11k}. \quad (3)$$

Empirical Results

The principal components analysis is conducted on the data matrix of socioeconomic 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 percent of the total standardized variance in 1998, and about 78.6 percent and 78.5 percent in 2002 and 2003. Also, the estimated values of factor loadings can be used to measure the degree to which the socioeconomic 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 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 light of the above, we use the factor score coefficients of the first principal component, F_1, as the weighting system applied to the regional values of socioeconomic variables to obtain the common factor score of the first principal component (CFSFPC). It is sensible statistically to use the CFSFPC to represent the socioeconomic environment across the thirty regions in China. Taking Beijing (region code 1) and Shanghai (region code 9) as examples, their respective CFSFPC

Table 6

Principal Component Solution of the Factor Model

	1998				2002				2003			
Z_i	Factor loadings of F_1 $(\sqrt{\lambda_1}e_{i1})$	Factor score coeffi-cients of F_1 $(e_{i1}/\sqrt{\lambda_1})$	Com-munality (ϕ_i^2)	Specific variance (ψ_i)	Factor loadings of F_1 $\sqrt{\lambda_1}e_{i1}$	Factor score coeffi-cients of F_1 $(e_{i1}/\sqrt{\lambda_1})$	Com-munality (ϕ_i^2)	Specific variance (ψ_i)	Factor loadings of F_1 $(\sqrt{\lambda_1}e_{i1})$	Factor score coeffi-cients of F_1 $(e_{i1}/\sqrt{\lambda_1})$	Com-munality (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) 8.826		Percent variance (λ_1/11) 80.24 percent		Eigenvalue (λ_1) 8.641		Percent variance (λ_1/11) 78.55 percent		Eigenvalue (λ_1) 8.634		Percent variance (λ_1/11) 78.49 percent	

Notes: F_1 stands for the first common factor score; percent variance stands for the percentage of total standardized variance attributable to the first common factor.

is calculated as the weighted average of the standardized values of the eleven socioeconomic variables:

$$\text{Beijing: } CFSFPC_1 = w_1 Z_{1,1} + w_2 Z_{2,1} + w_3 Z_{3,1} + \ldots + w_{11} Z_{11,1}$$

$$\text{Shanghai: } CFSFPC_9 = w_1 Z_{1,9} + w_2 Z_{2,9} + w_3 Z_{3,9} + \ldots + w_{11} Z_{11,9} \quad (4)$$

Where: the weight (w_i) of Z_i stands for the factor score coefficient of the first principal component:

$$\left(\frac{e_{i.1}}{\sqrt{\lambda_1}}\right),$$

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

Since the CFSFPCs for the thirty regions are 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. To do this, we convert the CFSFPC values from the standard normal scale, $Z \sim N(\mu_z = 0, \sigma_z = 1)$, to a normal random variable, $Y \sim N(\mu_y = 100, \sigma_y = 50)$ by using the transformation formula

$$Z_j = \frac{Y_j - \mu_y}{\sigma_y}.$$

This derived variable is called socioeconomic environment index (SEEI). The formula of the SEEI for the jth region is:

$$SEEI_j = 50 \times CFSFPC_j + 100 \qquad j = 1, \ldots, 30 \quad (5)$$

Table 7 shows the CFSFPC, the SEEI, and the ranking of SEEI across the thirty regions in 1998, 2002, and 2003. The high correlation between the socioeconomic environment index (SEEI) and per capita FDI (PCFDI) in consecutive years (0.888 to 0.976) as shown in the lower portion of Table 7 implies that the regions with higher SEEI are expected to attract more PCFDI in the current and next few years. That is to say, we cannot reject the hypothesis that the regional socioeconomic environment is one of 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 SEEI in a given year is largely a relative concept built on the variances of the socioeconomic variables across the thirty regions in a specific year, and its absolute value is therefore not appropriate for use as a basis for comparison over time.

Table 7

The Socioeconomic 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	213	2	2.215	211	2
2. Tianjin	1.305	165	3	1.524	176	3	1.503	175	3
3. Hebei	–0.211	89	10	–0.297	85	11	–0.289	86	11
4. Shanxi	–0.35	83	14	–0.346	83	13	–0.347	83	12
5. Inner Mongolia	–0.623	69	27	–0.541	73	23	–0.526	74	24
6. Liaoning	0.198	110	7	0.050	103	9	0.059	103	9
7. Jilin	–0.342	83	13	–0.302	85	12	–0.372	81	14
8. Heilongjiang	–0.275	86	12	–0.349	83	14	–0.413	79	19
9. Shanghai	3.893	295	1	3.897	295	1	3.913	296	1
10. Jiangsu	0.420	121	6	0.471	124	6	0.581	129	6
11. Zhejiang	0.492	125	5	0.655	133	5	0.740	137	5
12. Anhui	–0.450	78	20	–0.506	75	22	–0.417	79	20
13. Fujian	0.159	108	8	0.116	106	7	0.061	103	8
14. Jiangxi	–0.581	71	26	–0.573	71	25	–0.456	77	21
15. Shandong	0.013	101	9	0.086	104	8	0.093	105	7
16. Henan	–0.379	81	16	–0.400	80	16	–0.393	80	16
17. Hubei	–0.235	88	11	–0.285	86	10	–0.277	86	10
18. Hunan	–0.436	78	18	–0.405	80	18	–0.402	80	17
19. Guangdong	0.901	145	4	0.840	142	4	0.757	138	4

20. Guangxi	–0.649	68	28	–0.617	69	27	–0.656	67	27
21. Hainan	–0.400	80	17	–0.569	72	24	–0.538	73	25
22. Chongqing	–0.378	81	15	–0.354	82	15	–0.354	82	13
23. Sichuan	–0.564	72	25	–0.585	71	26	–0.599	70	26
24. Guizhou	–0.815	59	30	–0.726	64	30	–0.725	64	30
25. Yunnan	–0.529	74	24	–0.617	69	28	–0.674	66	28
26. Shaanxi	–0.487	76	22	–0.479	76	21	–0.410	80	18
27. Gansu	–0.653	67	29	–0.652	67	29	–0.676	66	29
28. Qinghai	–0.517	74	23	–0.455	77	20	–0.489	76	22
29. Ningxia	–0.443	78	19	–0.404	80	17	–0.387	81	15
30. Xinjiang	–0.471	76	21	–0.430	79	19	–0.517	74	23
1998 PCFDI	—	(0.900)	—	—	—	—	—	—	—
1999 PCFDI	—	(0.888)	—	—	—	—	—	—	—
2002 PCFDI	—	(0.939)	—	—	(0.938)	—	—	—	—
2003 PCFDI	—	(0.966)	—	—	(0.970)	—	—	(0.976)	—

Notes:
1. Figures in parenthesis are Pearson correlation coefficients.
2. PCFDI represents per capita foreign direct investment.
3. CFSFPC represents the common factor score of the first principal component.

Cluster Analysis

The Agglomerative Hierarchical Clustering Technique

The socioeconomic 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 relative magnitude. However, as indicated by the intercorrelation matrix of the eleven selected variables, to identify the characteristics of the socioeconomic environment of each of the regions in China could be very cumbersome and confusing. Therefore, it is desirable to partition the thirty regions in China into a smaller number of homogeneous clusters before carrying on further investigation. The most widely used method to complete this task is use of distance-based hierarchical clustering algorithms. Everitt and Dunn (1991) and Friedman and Meulman (2004) provide comprehensive reviews of the clustering techniques.

The process of clustering begins by finding the closest pair of regions using a particular interobject distance measure of attributes and combining those regions with the nearest distance to form a cluster. At each stage in the procedure, the number of clusters is reduced by one by joining or fusing the two clusters considered the closest to each other. The procedure continues one step at a time until all the clusters are merged into a single cluster.

The attribute variables in our cluster analysis are the eleven regional socioeconomic variables. We employ the standardized values of these attribute variables for clustering analysis to avoid problems caused by scale differences. We choose the squared Euclidean distance as the measure of dissimilarity among the attribute variables. 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. One useful representation of the hierarchy is a binary tree diagram called a *dendrogram*, and it can be obtained by using common statistical computer packages such as SPSS and Minitab. We can visualize this representation to assess the degree of clustering in the data set and manually choose a particular partition of the objects into clusters.

Characteristics of Distinct Clusters

Taking the SPSS output of the complete linkage dendrogram based on the set of socioeconomic variables in 2003 as an example, before joining or fusing the regions together, each region is considered to be a single group at the first fusion stage; at each of the second to the fifth fusion stages in the agglomerative procedure, the number of regions is reduced. When the final (sixth) fusion stage is reached, there is a single group containing all thirty regions. At the fifth fusion stage, the thirty regions are classified into two clusters: one cluster contains three regions and the other contains twenty-seven regions. It is obvious that all regions in the three-region cluster are municipalities (Beijing, Tianjin, and Shanghai) and were the earliest regions to be opened to the foreign world; and thus, the socioeconomic environment in this cluster on average is the best in the country. The establishment of the Chongqing municipality in 1997 did not lead immediately to success in the overall socioeconomic 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 the three-region cluster. At the fourth fusion stage, the thirty regions are classified into three clusters: Shanghai alone in one cluster, Beijing and Tianjin occupying another cluster, and the remaining twenty-seven regions belonging to the third cluster. When moving down one stage, the thirty regions are partitioned into four clusters: Cluster 1 contains Shanghai; Cluster 2 contains Beijing and Tianjin; Cluster 3 contains Jiangsu, Zhejiang, and Guangdong; and Cluster 4 contains the other twenty-four regions. Finally, we can identify eight clusters at the second fusion stage.

Although it is not easy to determine which fusion stage should be chosen in the dendrogram to identify naturally occurring groups of regions defined in terms of their socioeconomic environment, the interpretability of the resulting cluster structure will often be useful to justify the choice. With this in view, we choose the third fusion stage of the complete linkage dendrogram to partition the thirty regions in China into four relevant clusters based on the socioeconomic dataset over the years under consideration. The four clusters of regions at the third fusion stage together with the corresponding SEEI and PCFDI series in 1998, 2002, and 2003 are listed in Table 8.

In 2003, Shanghai itself forms a distinct cluster because it ranks first in a number of socioeconomic variables, including 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}), and 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), degree of openness to international trade (X_5), and 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 regions in Cluster 2. For X_2, Shanghai has a score of 16.6, which is only slightly lower than Beijing's score of 16.78. For X_3, the score of Shanghai is 28.41, which is also only slightly lower than Beijing's score of 28.46. For X_5, Shanghai's score of 146.34 is also lower than Guangdong's score of 175.69. For X_7, Shanghai's score of 18.68 is also slightly lower than Beijing's score of 18.99. Judging from their outstanding socioeconomic environment as measured by the SEEI, we call the regions in Cluster 1 and Cluster 2 the *matured* regions and *advanced* regions, respectively.

The three coastal provinces, Jiangsu, Zhejiang, and Guangdong, in Cluster 3

Table 8

Member Regions of the Four Socioeconomic Clusters with SEEI and PCFDI in 1998, 2002, and 2003

Cluster		1998		2002		2003	
	Region	SEEI	PCFDI	SEEI	PCFDI	SEEI	PCFDI
Cluster 1	Shanghai (E)	295	2,284	295	2,657	296	3,383
(mature regions)							
Cluster 2	Beijing (E)	221	1,649	213	1,264	211	1,587
(advanced regions)	Tianjin	—	176	1,429		175	1,377
Cluster 3	Fujian (E)	108	1,100	—	—	—	—
(developed regions)	Guangdong (E)	145	1,400	142	1,201	138	818
	Jiangsu (E)	121	764	124	1,142	129	1,179
	Liaoning (E)	110	436	—	—	—	—
	Shandong (E)	101	207	—	—	—	—
	Tianjin (E)	165	1,939	—	—	—	—
	Zhejiang (E)	125	246	133	550	137	884
Cluster 4	Fujian (E)	—	—	106	916	103	616
(developing /	Hainan (E)	80	815	72	553	73	432
underdeveloped regions)	Hebei (E)	89	181	85	96	86	118
	Inner Mongolia (E)	69	32	73	60	74	31
	Liaoning (E)	—	—	103	696	103	555
	Shandong (E)	—	—	104	342	105	547
	Guangxi (E)	68	157	69	72	67	76
	Anhui (C)	78	37	75	52	79	49
	Heilongjiang (C)	86	117	83	77	79	70

Henan (C)	81	55	80	35	80	48
Hubei (C)	88	137	86	203	86	217
Hunan (C)	78	108	80	118	80	137
Jiangxi (C)	71	93	71	213	77	315
Jilin (C)	83	129	85	75	81	58
Shanxi (C)	83	69	83	53	83	54
Chongqing (W)	81	117	82	52	82	69
Gansu (W)	67	13	67	20	66	7
Guizhou (W)	59	10	64	8	64	10
Ningxia (W)	78	29	80	32	81	25
Qinghai (W)	74	0	77	74	76	39
Shaanxi (W)	76	69	76	81	80	74
Sichuan (W)	72	37	71	55	70	40
Xinjiang (W)	76	10	79	9	74	7
Yunnan (W)	74	29	69	21	66	16

Notes:
1. SEEI is the socioeconomic environment index.
2. PCFDI is per capita foreign direct investment in yuan.
3. The capital letters E, C, and W in parentheses after the region name indicate that the region belongs to the eastern, central, or western areas of China.

are open to the foreign world and underwent economic reforms earlier. In this cluster, per capita GDP (X_1), average wage (X_3), per capita total investment in fixed assets (X_7), length of transport routes (X_{10}), and resource density (X_{11}) rank first amongst all the twenty-two provinces and four autonomous regions. Owing to the successful implementation of preferential economic development policies, the percentage of state-owned and state-holding enterprises to all enterprises in terms of gross output value[18] in this cluster is the lowest, relative to other clusters. And thus, regions in this cluster are believed to have the greatest degree of economic freedom in China. Taking into account their stage of socioeconomic development, we call the regions in this cluster the *developed* regions.

All the regions in the central and western areas, together with the seven eastern regions with mediocre socioeconomic conditions, are included in Cluster 4. Since the twenty-four regions in this cluster represent about 95 percent of the total land area and 80 percent of total population in China, their geographical and demographic factors should have a substantial effect on the socioeconomic environment of these regions. Considering their inferior socioeconomic condition, we call these regions the *developing* or *underdeveloped* regions.

If we compare the cluster structure in 1998 and 2003 shown in Table 8, we observe the following characteristics. First, Shanghai remains in the cluster of matured regions (Cluster 1), indicating that it has been able to stay in the leading position throughout the period; second, Tianjin has moved to Cluster 2 in 2003 from Cluster 3 in 1998 because of its 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}); third, Fujian, Shandong, and Liaoning, which belong to Cluster 3 in 1998, are classified into Cluster 4 in 2003 because their socioeconomic developments are stagnant during the period. Finally, there is no noticeable change in the central and western regions during the period from 1998 to 2003.

Implications for Multinational Firms' Location Decisions

With its continuing reduction in trade and financial barriers since the World Trade Organization (WTO) accession, as well as advances in communication networks, China has become an economic powerhouse in Asia and the focus of much foreign direct investment. 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[19] 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 investing 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 socioeconomic 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 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 socioeconomic environment. The SEEI in 1998, 2002, and 2003 in Table 7 reveals that regions with a better socioeconomic environment can attract more inward FDI flows. Furthermore, the high correlation between 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 296, 40 percent higher than that of Beijing, 114 percent higher than that of Guangdong, and 363 percent higher than that of the poorest province, Guizhou. With the best socioeconomic environment in the country, Shanghai has strengthened its power to attract foreign investment and improved the quality of that investment by putting more emphasis on the development of modern manufacturing and servicing industries, and new high-tech industries. Taking into account the role of Shanghai as the major financial center of China, as well as its strategic geographical location and its outstanding achievements in socioeconomic development, allows us to believe that Shanghai will continue to be the most attractive business location for multinational firms.

On the other hand, there are eight regions in the cluster of developing or underdeveloped regions (Cluster 4) with their SEEI less than 75 in 2003. They are: Hainan and Guangxi from the eastern area; Sichuan, Guizhou, Yunnan, Gansu, and Xinjiang 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 socioeconomic 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 forest 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[20] 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 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 percent) over the period of five years from 1998 to 2003; in contrast, the FDI figures for Shanghai, Jiangsu, and Zhejiang soared by 78.28 billion yuan (+82 percent). This reveals that foreign investors in China are moving northward from the Pearl River Delta to the Yangzi River (Changjiang) Delta. FDI has begun to spread from the traditional investment base in the south to new regions because China has introduced new policies aimed at easing foreign investment restrictions and attracting more foreign investment to other parts of the country. For this reason, we can conclude that the signaling effect of prior FDI[21] is no longer an appropriate factor in determining location decisions.

Concluding Remarks

Regional or locational disparities in China have been one of the hot research topics for more than ten years. Many models have been constructed (Hu, Wang, and Hong 1995; Poon, Hon, and Woo 1996; Coughlin and Segev 1999; Ng and 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, 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 socioeconomic environment for individual regions, and determines whether the regions are attractive business locations, as well as whether their attractiveness is changing over time. Since the socioeconomic 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 socioeconomic 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 national income statistics and international direct investment flows, have imposed an inevitable handicap to our analysis.

It should be stressed that socioeconomic variables are strongly interrelated, and thus we should not attempt to use the multiple regression method for analysis.[22] The objective of constructing the SEEI, which is based on the first principal component of the eleven socioeconomic variables, is to demonstrate the ranking and magnitude of the socioeconomic environment amongst the thirty 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 socioeconomic 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 socioeconomic 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 macrolevel.

Notes

1. IMF *Balance of Payment Manual*, 5th ed., para. 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." Para. 362 states: "Direct investment enterprises comprise those entities that are subsidiaries (a nonresident 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."

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

3. "Overview of the World Economic Outlook Projections," *World Economic Outlook* (September 2004), table 1.1.

4. "The Constitution of the People's Republic of China stipulates that administratively: 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; and 6) the state shall, when necessary, establish special administrative regions." "Explanatory Notes on Main Statistical Indicators," *China Statistics Yearbook* 2004, chap. 1. For the official classification of the thirty-one administrative regions into eastern, central, and western areas, see the "Explanatory Notes on Main Statistical Indicators," *China Statistics Yearbook* 1997, chap. 10.

5. Chongqing was separated from Sichuan and promoted to become a municipality in 1997.

6. Some of the FDI flows may be "round-tripping" from the mainland to take advantage of the preferential treatment of foreign investors in China. See Prasad 2004, 4, note 2.

7. The euro area includes twelve countries: Germany, France, Italy, Spain, Netherlands, Belgium, Austria, Finland, Greece, Portugal, Ireland, and Luxembourg.

8. Calculated from table 16-15 and table 18-15 of *China Statistics Yearbook* (1996 and 2004, respectively).

9. $Y^d = C + I + G + (X - M)$, $Y^s = C + S + T$; in equilibrium, $I + G + (X - M) = C + S + T$.

10. From tables 3-13 and 8-1 of *China Statistics Yearbook* 2004, $I = 5.14$ trillion yuan, $(T - G) = 2.93$ trillion yuan, $(X - M) = 2.69$ billion yuan, GDP (by expenditure approach) = 121511.4; then, the calculated S is 5.7 trillion yuan and the ratio of S to GDP is 0.4691.

11. The 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:

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

12. A leptokurtic curve has a narrower central portion and higher tails than does a normal curve.

13. 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.

14. 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 1987, 12).

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

16. 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).

17. See Johnson and Wichern 2002, chap. 8, for the detailed discussion of principal components.

18. In 2002, the percentages 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 are 22.8 percent, 13.6 percent, and 19.3 percent, respectively; in 2003, they are 19.0 percent, 13.1 percent, and 18.4 percent. Since these statistics are not available for 1998, they are not included in the list of socioeconomic variables. See *China Statistics Yearbook* 2003, table 13-3; *China Statistics Yearbook* 2004, table 14-2.

19. "China to Grab $100 Billion Annual FDI in 2006–10," *Emerging Markets Economy*, January 2, 2003.

20. 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, May 8–10, 2003. Selected articles were published in *China Quarterly,* no. 178 (June 2004).

21. "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, May 2, 2002, available at www.imf.org/external/np/tr/2002/tr020502.htm.

22. 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 among the independent variables, simply because in such a situation it is obviously impossible to change one variable while holding all others constant.

References

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

China Statistics Yearbook, various years. Beijing: China Statistical Publishing House.

Coughlim, C.C., and E. Segev. 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.* Maidenhead, UK: Open University Press.
Crum, R.L., E.F. Brigham, and J.F. Houston. 2005. *Fundamentals of International Finance*. N.p.: Thomson South-Western, 1st ed.
Everitt, B.S., and G. Dunn. 1991. *Applied Multivariate Data Analysis*. London: Edward Arnold.
Friedman, J.H., and J.J. Meulman. 2004. "Clustering Objects on Subsets of Attributes." *Journal of the Royal Statistical Society*, Series B, Statistical Methodology, vol. 66, part 4: 815–49.
Garelli, Stephane. 2004. *IMD World Competitiveness Yearbook.*
Hu, A.G., S.G. Wang, and S.K. Hong. 1995. *Report on China's Regional Disparities* (in Chinese). Shenyang: Liaoning People's Publishing House.
International Monetary Fund (IMF). 2004. *World Economic Outlook* (September 2004).
————. 1996. *Balance of Payment Manual* (5th ed.).
Johnson, R.A., and D.W. Wichern. 2002. *Applied Multivariate Statistical Analysis*. 5th ed. Englewood Cliffs, NJ: Prentice Hall.
Kaiser, H.F. 1960. "The Application of Electronic Computers to Factor Analysis." *Educational and Psychological Measurement* 20: 141–51.
Ng, L.F.Y., and C. Tuan. 2003. "Location Decisions of Manufacturing FDI in China: Implication of China's WTO Accession." *Journal of Asian Economics* 14, no. 1: 51–72.
Poon, C.C., T.Y. Hon, and K.Y. Woo. 1996. "A Cluster Analysis of Regional Economic Disparities in China" (in Chinese). Economics Working Paper Series 97012, Economics Department, Hong Kong Shue Yan College.
Prasad, Eswar, ed. 2004. "Hong Kong SAR: Meeting the Challenges of Integration with the Mainland." Occasional Paper 226, International Monetary Fund.
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: 116–25.