Discovering the Factors Affecting the Location Selection of FDI in China*

Li Zhang**, Yujie Zhu, Ying Liu, Nan Zhou, and Guoqing Chen

School of Economics and Management, Tsinghua University, Beijing, 100084, China
zhangl34@em.tsinghua.edu.cn

Abstract. Since the late 1970s, Foreign Direct Investment (FDI) has played an important role in the economic development of China. However, the growth of FDI in China has an increasingly unbalanced development between the eastern and western provinces. In this chapter, the temporal association rule (TAR) mining method is applied in discovering the factors affecting the location selection of FDI. In the light of the data set at the provincial level in the period of 1984–2001, our analysis reveals that the factors, including incoming and saving, residential consumption, GDP, infrastructure, population, education and cumulative FDI amount, have significant impacts on attracting FDI, which is consistent with existing studies. But some of other factors such as means of transportation, the loan policy, and the degree of modernization of the eastern rural area have seldom been discussed, and may represent certain special and unique characteristics of China. Such findings may be helpful for the local governments to take proper measures for attracting more FDI.

Key words: Temporal Association Rule, Foreign Direct Investment, location selection of FDI

1 Introduction

By the end of December 2003, China had ushered in a total of US$501.5 billion in actual FDI, making it the second largest destination for FDI in the world, after the United States. Since the late 1970s, FDI has played an important role in the economic development of China. However, the unbalanced distribution of FDI across provinces within China has widened the gap in economic

* Partly supported by the National Natural Science Foundation of China (79925001/70231010), and the Bilateral Scientific & Technological Cooperation between China and the Flanders.

** Corresponding author.

Li Zhang et al.: *Discovering the Factors Affecting the Location Selection of FDI in China*, Studies in Computational Intelligence (SCI) **5**, 219–236 (2005)
www.springerlink.com

development between the eastern, central and western provinces by accelerating the development of the east provinces. The "By Region" column of Table 1 shows that from 1985 to 1998, the eastern region received a lion's share of the total FDI amount, more than 85 percent, while the central and the western regions together only received less than 15 percent [17]. The reasons why the eastern region has always been the key region of the direct investment of foreigners refer to excellent commercial geography, history, some open and preferential policies, etc.

Table 1. Distribution of FDI across region/provinces within China

	Distribution of FDI (%)					
	By Region			By Province		
Year	Eastern	Central	Western	Guangdong	Fujian	Others
1985	90.18	5.56	4.26	46	11	44
1986	87.95	7.28	4.76	50	6	44
1987	88.58	6.02	5.40	36	4	59
1988	87.00	5.94	7.06	43	5	52
1989	92.16	3.84	3.99	38	11	51
1990	93.87	3.87	2.26	46	9	45
1991	92.46	4.48	3.06	44	11	45
1992	91.30	6.82	1.89	34	13	53
1993	87.38	8.88	3.74	28	11	62
1994	87.83	7.85	4.31	28	11	60
1995	87.71	9.21	3.08	28	11	62
1996	88.04	9.52	2.45	28	10	62
1997	86.90	10.56	2.54	27	9	64
1998	88.04	9.86	2.10	30	10	61

Source: DRI CEIC Database and various issues of the Statistical Yearbook of China

The increasing gap has created some social and political problems and will deter the durative development of national economy. In order to decrease the differences among the regions, China's central government has adopted a series of measures including the policies encouraging FDI in the central and western regions. However, the situation of the Central and Western shows that the policy-makers need more information about how the region attracts more FDI to guide the investment inflow.

The studies on FDI have grown rapidly in resent years. There have been quite a few research efforts on FDI so far. The pioneer industrial organization theory was put forward to explain FDI by Hymer [12]. The well-known "OLI" theory [13] presents that the host country must possess three advantages including Ownership, Location and Internalization (OLI), which explain the

who, where and how of FDI respectively, if it tries to attract more FDI. The international trade theory was used to explain the allocation aspects of FDI in [2] and [11]. In our study, we focus on the location perspective, which is often used to explain why a multinational corporation would choose to invest in a particular host country. It can also be used to explain why foreign investors would choose to invest in a specific location within a particular host country [17].

Existing studies have identified some factors that may attract FDI in one country. The studies on FDI in USA [1, 5, 7, 10, 15] accounted for most of the effort compared with that in other countries. The factors found include market potential, labor cost, transportation network, state expenditures, population size, and so on. In recent years, more and more attention has been paid to FDI in China. Chen [4] found such factors as market potential, labor cost, infrastructure (railway) and R&D. Borensztein et al. [6] identified the labor quality to have impact on attracting FDI. Zhang et al. [26] showed that GDP, GDP per capita, cumulative FDI amount, and transportation density were important factors affecting the location selection of FDI in China. Sun [24] identified the preferential policy, industrial structure, degree of openness, market and education as the determinants of FDI across the provinces within China. Based upon these findings, 7 potentially important factors associated with FDI distribution are summarized in Table 2.

In this chapter, we will investigate the factors that may affect the way regions across China attract FDI. First of all, we will examine whether the findings from existing studies are also applicable for the location selection of FDI in China. Secondly, we will explore whether there are factors that are only applicable in certain regions, which may reflect the national characteristics. Such a study is deemed meaningful for policy-makers.

Table 2. Summary of factors

Factor	Index
market potential	It means that the underlying market and the actual market scale, including the indexes such as GDP, GDP per capita, resident consumption, and incoming and saving
Infrastructure	The level of infrastructure is mainly represented by transportation and communication facilities
labor cost	wage
labor quality	education
Population	population size, population growth
degree of openness	It is often measured with the rate of Import/GDP
cumulative FDI amount	–
preferential policy	–
Finance	state expenditures, taxes

Instead of using the conventional approaches that apply empirical analysis or statistics methods based on models of selected variables/factors, our study attempts at employing a novel approach that is data-driven to identify the variables/factors that are associated. This approach, namely temporal association rule (TAR) mining, considers all possible combinations of factors in terms of their interconnections based on certain association measures, and is regarded particularly useful in the case where the data volume is huge primarily with a large number of variables/factors. It is worth indicating that this approach may be used in a supplementary manner to conventional approaches in that TAR identified factors, some of which may be new and interesting, could be further described in terms of their analytical relationships.

The chapter is organized as follows. Section 2 explains how the data is prepared. The TAR method is discussed in Sect. 3. In Sect. 4 and Sect. 5 the mining process with respective data and the mining results are discussed. Notably, for the sake of convenience, the terms variable, factor and index are used interchangeably in the text (otherwise indicated where necessary).

2 Data Preparation

The data set is available at the website of China Economic Information Network (www.cei.gov.cn) developed by China Economic Information Network Data Co. Ltd. whose holding company is State Information Center. The data provided by this website were collected from some authorities including the National Planning Commission, National Bureau of Statistics of China, and State Information Center. The data set over the years 1984–2001 for each of 31 provinces consists of ten sections in terms of the attributes, i.e., finance, foreign trade, GDP (including the primary industry, second industry, tertiary industry and architecture industry), investment in fixed assets, culture & education and sanitation, population, employment and labor cost, people's life, price index and the order of economic indexes for each province. Furthermore, some detail indexes are included in each section. The total number of the indexes is 372. Using the same data set, [26] examined the location selection of FDI in China by building a regression model based on certain selected factors.

Although the data source appears to be authoritative, the data may not be ready directly for use due to existence of certain noises, missing values, inconsistency and/or redundancy. Thus, data cleaning is necessary. Furthermore, data normalization and discretization are carried out in order to obtain appropriate formats for mining purposes.

2.1 Data Cleaning

The data cleaning process attempts to fill out missing values, smooth out noise, remove the redundancy, and correct inconsistencies in the data. First, for each of 31 provincial tables, each of 372 attributes and each of 18 records

Table 3. Steps and results of data cleaning

Step	Object	Operation	Before Operation	After Operation
Step 1	province	Removal of incomplete data at the provincial level	31	19
Step 2	attribute	Elimination of redundant attributes and the attributes with too many missing values	372	More than 200
Step 3	Record (year)	Filling out the missing value	–	–

are examined. Then, correction measures are taken on the deficient data. The steps and result of data cleaning are shown in Table 3.

Step 1 Removal of incomplete data at the provincial level. The whole data set includes 31 data tables. Note that Chongqing is excluded from the list because it was established only in 1997. Further, in some western provinces, such as Tibet, Xinjiang, Qinghai and Guizhou, the receipt of FDI was negligible and they are excluded from the list. Finally, the following 19 provinces are included: namely Anhui, Beijing, Fujian, Gansu, Guangdong, Hainan, Hebei, Heilongjiang, Hubei, Hunan, Jiangsu, Jiangxi, Jilin, NingXia, Shandong, Shannxi, Shanghai, Tianjin, and Yuan'nan. Some of these provinces are located along the Chinese coast, and others are situated in the Center or West of China.

Step 2 Elimination of the attributes with too many missing values and of repeated and redundant attributes. Firstly, the attributes whose values are unbearably incomplete should be deleted. On the other hand, in fact, some indexes are repeated because of multiple data sources. In addition, some other indexes are redundant, for instance, "the number of doctors" have a linear relation with "the number of doctors per ten thousands". In the above-mentioned cases, the repeated or redundant attributes will be eliminated.

Step 3 Filling out the missing value in the records. The missing values are values that actually exist but have gone astray. Clearly where possible they should be looked for, but this may either be impossible or just too expensive and time-consuming. One of the solutions to the problem is to find a suitable value as the replacement of the missing value. Since most missing values are between 1999 and 2001 while economy developed steadily in China during the three years, it is reasonable that the missing values are estimated with the previous and next values of the missing value. The formulation of the filled-in value used is $a_n = (a_{n-1} + a_{n+1})/2$, where a_n, a_{n-1}, a_{n+1} represent the attribute values of the nth, $(n-1)$th and $(n+1)$th year, respectively. For

Table 4. TV cover rate, with missing value for 1995 and 2001

Year	92	93	94	95	96	97	99	00	01
Value	70	77	79	**80**	81	84	87	89	**91**

example, in Table 4, value of 1995 is $a_{95} = (a_{94} + a_{96})/2 = 80$, and the value of 2001 is $a_{01} = 2 * a_{00} - a_{99} = 91$.

2.2 Data Normalization and Discretization

With data normalization, attribute values are scaled so as to fall within a small specified range, such as -1.0 to 1.0. In the chapter, the values with initially large ranges are normalized to the increase percentage as:

$$IP_n = \frac{a_n - a_{n-1}}{a_{n-1}} \times 100\% \tag{1}$$

where IP_n is the increase percentage of the nth year, a_n and a_{n-1} represents the attribute values of the nth, and $(n-1)$th year, respectively. Another issue that should be considered in data preparation is discretization. For most mining methods, discretization of numeric attributes is helpful and sometimes necessary. For our data, it is rational that the data be divided into several intervals using equidistance subsection method. That is, with max(a) and min(a) representing the maximum and minimum values of an attribute, the interval [min(a), max(a)] is equally divided into some sub-intervals. And then these sub-intervals are replaced by some discrete numbers denoting different categories. As shown in Fig. 1, the attribute values are classified into four categories: $[\min(a), \min(a) + c]$ $[\min(a) + c\, \min(a) + 2c]$ $[\min(a) + 2c\, \min(a) + 3c]$ and $[\min(a) + 3c\, \max(a)]$ where c stands for the length of each sub-interval and is calculated as $c = (\max(a) - \min(a))/4$. These four sub-intervals are labeled categories $C1, C2, C3$ and $C4$.

3 The Method

Data mining includes several kinds of technologies such as association rule analysis, classification, clustering, sequential pattern etc. In the chapter, we focus on association rule mining since it has been applied in many fields and

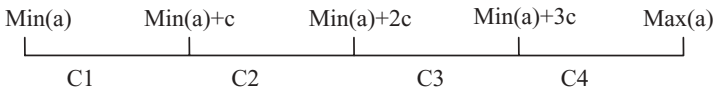

Fig. 1. Four categories

considered an important method for discovering associations among data [23]. Let $I = \{I_1, I_2, \ldots, I_m\}$ be a set of binary items, and D be a database of transactions. Each transaction t is represented as a binary vector, with $t[k] = 1$ if the item I_k occurred, and $t[k] = 0$ otherwise. Let X be a set of some items in I. We say that a transaction t satisfies X if for all items I_k in X, $t[k] = 1$. An association rule is an expression $X \Rightarrow Y$ meaning that if X occurs, then Y occurs at the same time, where X and Y are sets of items, $X \subset I, Y \subset I$, and $X \cap Y = \emptyset$.

Table 5. Number of beds in sanitary organization of a province during 1987 to 2001

Year	The Number of Beds	The Increase Percentage	Categorical Value
1987	10.11		
1988	10.45	0.034	4
1989	10.65	0.019	3
1990	10.78	0.012	2
1991	11.13	0.032	4
1992	11.45	0.029	4
1993	11.67	0.019	3
1994	11.94	0.023	3
1995	11.98	0.003	1
1996	12.08	0.008	1
1997	12.31	0.019	3
1998	12.35	0.003	1
1999	12.29	0.004	1
2000	12.39	0.008	1
2001	12.52	0.01	1

During recent years, many scholars have improved the conventional mining method and corresponding algorithm, namely Apriori algorithm [19], mainly in two directions: algorithmic efficiency [3, 16, 18, 23] and model extensions [18, 20]. However, most of these research efforts do not take into account the time dimension. On the other hand, time series data are common in many domains, such as business and economics in which temporal relationships are very important. In the research on discovering the factors affecting the location selection of FDI, some factors may not affect FDI immediately but in a period of time (e.g., years). Therefore, the approach to discovery of temporal association rules will be used.

Some researchers have extended the classical association rules mining with temporal semantics. However, most of temporal mining is not applied on the data but on the rules extracted from the data at various time points [14]. Reference [8] proposed an approach to extending the notion of a typical $X \Rightarrow Y$ to be a rule of the form $X \overset{T}{\Rightarrow} Y$, stating that if X occurs then Y will occur

Table 6. The reconstructed data set of one province (min_support = 0.3, min-confidence = 0.6)

year	Attri.1	Attri.2	……	Attri.N	FDI	**FDI-1**	**FDI-2**	**FDI-3**
1984	1	1		1	1	1	1	2
1985	1	1		1	1	1	2	3
1986	1	2		1	1	2	3	.....
1987	1	2		1	2	3	.....	2
1988	2	1		2	3	.....	2	1
……	……	……	……	……	……	2	1	3
1999	3	2		2	3	3	4	
2000	4	3		3	3	4		
2001	4	4		4	4			

within time T. They first form subsequences by sliding a window through the time series, and then cluster these subsequences by using a suitable measure of time series similarity. The discretized version of the time series is obtained by taking the cluster identifiers corresponding to the subsequence. Once the time-series is obtained, simple rule finding method is used to obtain rules from sequence. Chen & Yu extended the traditional association rules to temporal association rules with a delay time T [17], which is of the form

$$X \overset{T}{\Rightarrow} Y(T \geq 0) .$$

When $T = 0, X \overset{T}{\Rightarrow} Y$ degenerates to traditional association rule $X \Rightarrow Y$. When $T > 0$, it is a temporal association rule showing that if X occurs, Y will occur in the next T units of time. Concretely, after reconstructing the data set (as the shadow area shown in Table 6 of Sect. 4, in the case of $T = 1$), the number of the records in the reconstructed data set is $|D| - T$. Thus, the support degree (Dsupport) and confidence degree (Dconfidence) can be represented as follows [17].

$$\text{Dsupport}(\text{X} \overset{T}{\Rightarrow} Y) = ||X \underset{T}{\cup} Y||/(|D| - T)(T > 0)$$

where $||X \underset{T}{\cup} Y||$ is the number of occurrences in that X is followed by Y in delay time T.

$$\text{Dconfidence}(\text{X} \overset{T}{\Rightarrow} Y) = ||X \underset{T}{\cup} Y||/||X||'(T > 0)$$

where $\|X\|'$ is the number of occurrences of X in the reconstructed data set. Furthermore if we have $X \overset{T}{\Rightarrow} Y$ for any T value in an interval $[T_1, T_2]$, then we may use an extended form

$$X \overset{[T_1,T_2]}{\Rightarrow} Y(T_1 \leq T_2\ T_1\ T_2 > 0)$$

to reflect the association that occurs during the interval.

4 Mining Temporal Association Rules

4.1 Rules at The Provincial Level

For the purpose of discovering the factors affecting the location selection of FDI in a delay-time, a new data set is needed. Let us concentrate on the individual province, as shown in Table 6, where data for FDI-i ($i = 1, 2, 3$) are the FDI values with a delay of i years. Here FDI-i is regarded as a new attribute. Hence, with the reconstructed dataset and regarding each new tuple as a transaction, an association rule has a specific form:

$$\text{the trend of } X \overset{T}{\Rightarrow} \text{the trend of FDI } (T = 0, 1, 2, 3)$$

where the consequent of the rule is the definite item FDI, and X is a subset of the attribute set {Attri.1, Attri.2, ..., Attri.N, FDI-1, FDI-2, FDI-3}.

Using the Apriori algorithm on the reconstructed dataset, the mining process has discovered a good number of rules when $T = 2$ or $T = 3$, as listed in Table 7. Such results indicate that there exist some temporal relationships between the factors and FDI. For instance, the rule FDI - $1 \uparrow \overset{1}{\Rightarrow}$ FDI $\uparrow$ represents that the receipt of FDI in the previous year may have a positive effect on attracting more FDI in this year. Here and thenafter, symbol $\uparrow$ is used to stand for trend of increase.

For example, rules Education $\uparrow \overset{2}{\Rightarrow}$ FDI $\uparrow$ and Education $\uparrow \overset{3}{\Rightarrow}$ FDI $\uparrow$ mean that the improvement of education has a prominently positive effect on the inflow of FDI in two or three years.

4.2 Rules at The Regional Level

In the above stage, TARs are obtained at the provincial level. This subsection discusses the rules discovered, which are based upon the rules extracted from the provincial data.

Let R_i be the rule set of the ith province, so the rule set of all provinces is expressed as follows:

$$R = \bigcup_{i=1}^{19} R_i$$

The support count of a rule at the regional level is defined as the number of the provinces in the region in which the rule is applicable. Obviously, the support degree of the rule is

$$\text{Region\_Dsup} = \frac{|R_x|}{|\text{Region}|} \times 100\%$$

Table 7. The number of the rules when $T = 0, 1, 2, 3$

No.	Province	The Number of the Rules with Delay Time T				
		$T = 0$	$T = 1$	$T = 2$	$T = 3$	Total
1	Anhui	4	5	22	17	48
2	Beijing	1	0	0	0	1
3	Fujian	20	12	43	62	137
4	Gansu	2	18	36	13	69
5	Guangdong	5	26	10	27	68
6	Hainan	7	7	20	15	49
7	Hebei	1	26	39	36	102
8	Heilongjiang	3	9	30	18	60
9	Hubei	6	27	49	31	113
10	Hunan	1	9	20	21	51
11	Jiangsu	5	19	1	10	35
12	Jiangxi	1	3	9	28	41
13	Jilin	1	8	24	33	66
14	NingXia	1	10	43	30	84
15	Shandong	4	11	30	22	67
16	Shannxi	9	12	34	30	85
17	Shanghai	5	21	34	35	95
18	Tianjin	1	21	24	29	75
19	Yuannan	0	2	0	57	29
20	Total	77	246	468	514	1305

Where |Region| represents the number of provinces belonging to the region/country, $|R_x|$ represents in the region/country the support count of the rule Rx. If Region_Dsup of one rule at the regional level is not less than the support threshold, we say the rule is applicable in the region/country and can be considered as one of the regional characteristics.

The mining result shows that no rule exists in the whole country, which may be due to the significant economic difference among provinces. Mining the rules in a particular region, in which the economic levels of the provinces are similar, may be interesting and meaningful. In China, the economy of the Eastern, the Central and the Western is at developed, developing and underdeveloped level, respectively. Furthermore, China's western development strategy is one of the most important projects at present. The research on the difference between the Eastern and the Western has attracted remarkable attention.

Let R be divided into three disconnect subsets

$$R_e = \bigcup_{i=1}^{11} R_i \quad R_m = \bigcup_{i=1}^{4} R_i \quad R_w = \bigcup_{i=1}^{4} R_i$$

where R_e is the rule set of the eleven eastern provinces, including Beijing, Tianjin, Hebei, Shandong, Shanghai, Jiangsu, Fujian, Hainan, Guangdong,

Heilongjiang, and Jilin; R_m is the rule set of the four central provinces, including Anhui, Hunan, Hubei, and Jiangxi; and R_w contains four provinces in the Western: Shannxi, Gansu, Ningxia, and Yunnan. If the support degree Region_ Dsup is equal to or larger than the threshold (e.g., min_sup = 0.6), the rule is deemed applicable in the region.

5 Results and Discussions

As shown in Table 2, the factors identified by existing studies are macroeconomic indexes, such as labor cost, labor quality, market potential and infrastructure. Our study considers these factors, along with some microcosmic variables, such as the number of the students in college, and the TV cover rate. In addition, these variables are classified into the following categories in Eastern, Middle, and Western area; income and saving, resident consumption, finance, infrastructure, education, population, GDP, investment in fixed asserts, income from tourism, and FDI, as summarized in the Appendix.

The results of our study reveal that some factors, such as income and saving, resident consumption, infrastructure, population, education and GDP, etc., had significant effects on the location selection of FDI in each region to different degrees, which are generally consistent with those of existing studies. Furthermore, by breadthwise summarizing and comparing these economic indexes, some of other factors such as means of transportation, the currency policy, and the degree of modernization of the eastern rural area have seldom been discussed previously, and may represent certain special and unique characteristics of China and also confirm the existence of gap between the Eastern and the Western. These factors and the corresponding rules are summarized in Table 8. Except for the last row, each index listed in the third column is the antecedent of the rule, which has a prominently positive effect on the location selection of FDI in the following three years. The last column is the region in which the rule can be applicable. For instance, in the Eastern, "Percentage of investment in fixed assets, all entities" has a positive effect on attracting FDI, which can be denoted by

$$\text{Percentage of investment in fixed assets, all entities} \uparrow \overset{[1,3]}{\Rightarrow} \text{FDI} \uparrow$$

Result 1 rules about investment in fixed asserts. The investment in fixed asserts refers to the economic activities of constructing and purchasing fixed asserts, which mainly calculates the actual investment in fixed asserts in one region and can reflect the capital scale. The investment in fixed asserts is a synthetical factor, involving the investment in transportation and communication facilities, farmland and water conservancy, power facilities and houses. All of these are indeed the most concerned part for foreign investors. With more investment in fixed asserts, the regions have a better investment environment and could attract more FDI. At present, most western provinces

Table 8. Factors and rules that could represent certain special characteristics of China

Result No.	Factors	Indexes (X) Rule: $X \uparrow \overset{[1,3]}{\Rightarrow}$ FDI $\uparrow$	Region
1	Investment in fixed asserts	Percentage of investment in fixed assets, all entities	Eastern
2	means of transportation	Added value index of transportation, storage, and postal and telecommunication services	Eastern
		Tons of cargo carried through railway	Central Western
		Tons of freight carried on waterways	Centra
		Tons of cargo carried through highways	Western
3	education	Enrollment at secondary schools	China
4	currency policy	Total amount of deposits in state-owned commercial banks	
		Total amount of loans from state-owned commercial banks	Central Western
		Total amount of loans to commercial enterprises from state-owned commercial banks	
		Total amount of loans to industrial enterprises from state-owned commercial banks	Western
5	the degree of modernization of the eastern rural area	Transaction amount in bazaar trade	
		Total retail sales on consumer products in rural areas	
		Total mechanical power owned in rural areas	Eastern
		Total retail sales on consumer products of collectively-owned businesses and others	Western
		Total retail sales of consumer products of individually-owned businesses	Central
6	self-reinforcing effect of FDI	$FDI \uparrow \overset{1}{\Rightarrow}$ FDI $\uparrow$	Eastern

have begun to seize the opportunities of implementing the west development strategy to strengthen the investment in fixed assets.

Result 2 rules about means of transportation. In each region, the factor infrastructure is always significant. But, note that the factor is materialized by tons of cargo carried through railway and highways for the Western, tons of cargo carried through railway and tons of freight carried on waterways for the Central, and added value index of transportation, storage, and postal and telecommunication services for the Eastern respectively. These rules are consistent with the fact that the road and railway are the main means of transportation in the western region while in the Central the water carriage is very often because all the four chosen provinces situating by the Yangtze River and Yellow River. Many means of transportation are used in the Eastern where possesses excellent geographical environment so that the synthetical factor "Added value index of transportation, storage, and postal and telecommunication services" emerged.

Result 3 rules about education. Education is a publicly accepted factor that could affect the location selection of FDI in the research of other countries, as well as of China. In the Appendix, we can find that the representation of Education only involves the indexes relating to secondary schools. The reason is that the investments of foreign companies are mainly clustered in the manufacturing and processing industries that do not require a high labor quality. However, FDI in China is coming into a new stage with China's entry into WTO. As shown in the appendix, in the Eastern and Central, the tertiary industry has begun affecting FDI. It is expected that Education will be more important in the future and foreign companies will employ more people that are better educated.

Result 4 rules about the currency policy. These rules show that governments play an important role in improving the investment environment of the Central and Western to attract FDI. In undeveloped regions, the Central and Western have to improve the investment environments by largely relying on the loan of national banks, while the Eastern as a developed region needs fewer support of this kind from governments.

Result 5 rules about the degree of modernization of the eastern rural area. Transaction amount in bazaar trade and Total retail sales on consumer products in rural areas in the Eastern shows that the east rural area has a relative large market potential and development from town to city. While both Total retail sales on consumer products of collectively-owned businesses and others in the Western, and Total retail sales of consumer products of individually-owned businesses in the Central could impact on the location selection of FDI. The above rules reflect the consumption market structure across regions in China, and meanwhile, such structure affects the distribution of FDI.

Result 6 rules about self-reinforcing effect of FDI. This is a public held rule in exiting studies. However, it is worth noticing that the rule holds only in

the Eastern. Observing the data values of FDI of each province, we find that the volume of FDI is very small in the western provinces and they can hardly produce the "scale effect". As the Central and the Western get more developed, and with more FDI attracted, it is expected that the self-reinforcing effect will emerge.

6 Conclusions

Different from most of existing studies, this chapter has used a novel approach, namely temporal association rules mining, to discovering the factors that affect attracting FDI in China. The results have revealed factors that are consistent with those obtained in existing studies, along with certain other factors that may only pertain to the context of China.

Appendix

	Income & Saving
Eastern	Balance of saving deposit of town resident in year-end
	Average annual per capita net income of rural families
Middle	Balance of saving deposit of town resident in year-end
Western	Balance of saving deposit of town resident in year-end
	Average annual per capita disposable income
	Index of government purchasing prices for agriculture products
	Resident Consumption
Eastern	Consumption level of total residents
	The final consumption
	Average annual per capita consumptive expenditure of town residents
	Total retail sales on consumer products in rural areas
	Transaction amount in bazaar trade
Middle	Resident consumption
	Total retail sales of consumer products of individually-owned businesses
Western	Consumption level in total residents
	Total retail sales on consumer products of collectively-owned businesses and others
	Average annual per capital expenditure on food consumption of rural households

	Finance
Eastern	Tax revenue of local government, all items
Middle	Tax revenue of local government, all items
	Local Financial Expenditure
	Local financial assistance to agricultural and non-profit units
	Total amount of deposits in state-owned commercial banks
	Total amount of loans from state-owned commercial banks
Western	Local financial expenditures
	Local financial assistance to agricultural and non-profit units
	Total amount of deposits in state-owned commercial banks
	Total amount of loans from state-owned commercial banks
	Total amount of loans to commercial / industrial enterprises from state-owned commercial banks
	Infrastructure
Eastern	Added value index of transportation, storage, and postal and telecommunication services
	Total transactions of postal and telecommunication services
	Total number of passenger cars owned
	Total number of telephone connections owned
	Medical facility
	Electronic power consumption in rural areas
	Total mechanical power owned in rural areas
Middle	Total transactions of postal and telecommunication services
	Total number of passenger cars owned
	Tons of cargo carried through Railway
	Tons of freight carried on waterways
	Medical facility
Western	Total number of passenger cars owned
	Tons of cargo carried through railway
	Tons of cargo carried through highways
	Total number of passenger carried
	Education
Eastern	Enrollment and full-time teachers at secondary schools (excluding professional schools)
	Full-time classroom teachers at secondary schools
	Number of art communities
	Number of journals published
Middle	Enrollment at secondary schools (excluding professional schools)
	Enrollment at secondary schools
	Full-time classroom teachers at secondary schools
	Number of journals published
	School attendance rate
Western	Enrollment at secondary schools
	Number of books published

	Population
Eastern	Rural population
	Urban population
	Death rate
Middle	Year-end Permanent Population
Western	Death rate
	Income from tourism
Middle	Foreign exchange income from international tourism
	Investment in fixed assets
Eastern	Percentage of investment in fixed assets, all entities
	Investment in fixed assets in state-owned enterprises
	GDP
Eastern	Primary Industry:
	Gross output index of Agriculture, Forestry, Animal Husbandry, Fishery
	Gross output index of agriculture
	Gross output index of forestry
	Gross output index of fishery
	Second Industry:
	Gross output of medium-sized industrial enterprises
	Gross output of tobacco, iron, and cement products
	Tertiary Industry:
	Added value index of wholesale, retail trade and food services
Middle	Primary Industry:
	Gross output index of Agriculture, Forestry, Animal Husbandry, Fishery
	Second Industry:
	Gross output of industry
	Gross output of yarn, salt, and steel
	Tertiary Industry:
	Added value index of wholesale, retail trade and food services
Western	Primary Industry:
	Gross output index of Agriculture, Forestry, Animal Husbandry, Fishery
	Second Industry:
	Gross output of large industrial enterprises
	Gross output of industry
	Gross output of salt, coal, and sugar
	Electronic power generation
	FDI
Eastern	The self-reinforcing of FDI

References

1. Bagchi-Sen, S., Wheeler, J. (1989), A spatial and temporal model of foreign direct investment in the United States. Economic Geography 65 (2), 113–129. 221
2. Bhagwati, J.N., Srinavasan, T.N. (1983), Lectures in International Trade. MIT Press, Cambridge, MA. 221
3. Brin S., Motwani R., Ullman J., Tsur S. (1997), Dynamic Itemset Counting and Implication Rules for Market Basket Data. *In Proc. Of the 1997 ACM-SIGMOD Int'l conf on the Management of Data*, 255–264. 225
4. Chen, Chien-Hsun (1996), Regional determinants of foreign direct investment in Mainland China. Journal of Economic Studies 23:18–30. 221
5. Coughlin, C.C., Terza, J.V., Arromdee, V. (1991), State characteristics and the location of foreign direct investment within the United States. Review of Economics and Statistics 73 (4), 675–683. 221
6. Borensztein, E., De Gregorio, J., Lee. J.-W. (1997), How does foreign direct investment affect economic growth? Journal of International Economics 45, 115–135. 221
7. Friedman, J., Gerlowski, D.A., Silberman, J. (1996), Foreign direct investment: the factors affecting the location of foreign branch plants in the United States. Global Finance Journal 7 (2), 209–222. 221
8. Gautam Das, King-Ip Lin, Heikki Mannila, Gopal Renganathan, Padhraic Smyth (1998), Rule discovery from time series. In *Proceedings of the international conference on KDD and Data Mining (KDD-98)*. 225
9. Gautam Das, King-Ip Lin, Heikki Mannila, Gopal Renganathan, Padhraic Smyth Graham, E., Krugman, P. (1991), Foreign Direct Investment in the United States, 2nd ed. Institute for International Economics, Washington, DC.
10. Grossman, G.M., Helpman, E. (1991), Innovation and Growth in the Global Economy. MIT Press, Cambridge, MA. 221
11. Hymer, S. (1960), The International Operations of National Firms: A Study of Direct Foreign Investments. MIT Press, Cambridge, MA. 221
12. Dunning, J. (1977), Trade, Location of Economic Activity and the Multinational Enterprise: A Search for an Eclectic Approach, in B. Ohlin, P.O. Hesselborn and P.M. Wijkman (eds.), *The International Allocation of Economic Activity*, Macmillan, London, pp. 395–418. 220
13. John F. Roddick, Myra Spiliopoulou (2002), A survey of temporal knowledge discovery paradigms and methods. IEEE transactions on knowledge and data engineering, vol. 14, No. 4, July/August. 220
14. Lipsey, R. (1993), Foreign direct investment in the United States: changes over three decades. In: Froot, Kenneth (Ed.), Foreign Direct Investment. University of Chicago Press, Chicago, pp. 113–170. 225
15. Park J.S., Chen M., Yu P.S. (1995), An Effective Hash Based Algorithm For Mining Association Rules. *In Proc. of SIGMOD Intl. Conf. on Management of Data*, San Jose, CA, May 1995. 175–186. 221
16. Qian Sun, Wilson Tong, Qiao Yu (2002), Determination of foreign direct investment across China, *Journal of International Money and Finance*, 21 (2002) 79–11. 225
17. R. Agrawal, R. Srikant (1994), Fast Algorithms for Mining Association Rules, *Proceedings of VLDB Conference*, Santiago, Chile, Sept. 1994. Expanded version available as IBM Research Report RJ9839, June. 220, 221, 226

18. R. Agrawal, T. Imielinski, A. Swami (1993), Mining association rules between sets of items in large database. In *Proc. of the ACM SIGMOD Conference on Management of Data*, pp. 207–216, Washington, D.C., May. 225
19. R. Srikant, R. Agrawal (1995), Mining Generalized Association Rules, *Proceedings of the 21st VLDB Conference* Zurich, Swizerland. 225
20. R. Srikant, R. Agrawal (1996), Mining Quantitative Association Rules in Large Relational Tables, *SIGMOD'96 6/96* Montreal, Canada. 225
21. Rastogi R., Shim K. (1998). Mining Optimized Association Rules with Categorical and Numerical Attributes, *In Proc. of the 14th Int'l Conf. on Data Engineering*, 503–512.
22. Savasere, E. Omiecinski, S. Navathe (1995), An Efficient Algorithm for Mining Association Rules in Large Databases, *Proceedings of the VLDB Conference*, Zurich, Switzerland, September.
23. Sun J. (2002), the Analysis of the factors of FDI location selection in China, China Economic Quarterly, Vol. 1, No. 3, Apr. 225
24. Yu. W., Chen G. (2002), Mining delayed association rules based on temporal data, *computer application and research*, November. 221
25. Zhang L., Gong Y.C. (2002), Determinants of foreign direct investment across provinces in China. The research report of Tianjin University (social science version), Vol. 4, Issue 2: 117–121, June.
26. Zhu Y.J., Empirical Analysis the factor affecting regional selection of FDI in China, International Economic Cooperation, 2003. 11. 221, 222