Longitudinal data file compilation for historical analysis
Matsuda, Yoshiro

Longitudinal Data File Compilation
for Historical Analysis

Yoshiro Mat suda*

1. Introduction

Recent development of computer environment has made it possible to handle a large scale database. Thus, to compile a longitudinal data file based on the micro data for historical analysis becomes possible if the source of data remains. Our recent effort is to compile a historical database especially aiming to reconstruct a list data of factories and companies obtained from various sources and to integrate these list data into one Iongitudinal file linking the data of same factory in different time periods. (1)

Our focus of attention is to analyze the industrialization process of the Japanese economy since the Meiji Restoration in 1859 after the opening the Japanese market to the world. The basic economic policy of the government was how to catch up the standard of the Western World through introducing the Western technology, both, of hardware and software. The factory production system was one of the examples. (2)

The factory production system was characterized with the power machine and the new management system of assembling workers in one workshop. Since the 1980s, the Japanese government tried to grasp the actual situation of the factory system and left various list data of the factories. Our file uses these fragmental data to compile a longitudinal one between 1902 and 1909 for manufacturing factories. This period is extremely important to understand the situations of the results of the first industrial revolution. The Russo-Japanese War was in 1904-5 and the First World War was in 1914-18 and the period we analyze lies between these two phases.

2. Survey System and High Refusal Rations

The list of factory of 1902 was based on the survey of which the basic enumeration or reporting unit was the factory of more than ten employees. But the coverage of the list of factory of 1909 was expanded to more than

* Address all communications to Yoshiro Matsuda, Institute of Economic Research, Hitotsubashi University, Kunitachi, Tokyo, Japan.

five employees by the new Act for Factory Survey which required an census survey every five years. Except this census survey year, every province was to carry out survey on the factory of more than ten employees. The result was the erosion of the factory a little more than ten employees due to the result of factory managers against the survey in order to avoid the troublesome work of keeping records in detail. The growth rate of the factory more than ten employees marked a sharp decline for the next year of the census. The result is shown in Table 1. In 1910, after the survey year, with more than five employees, the number of factories declined 12.33% and in 1915 again it declined 1.48%.

Table 1:   Annual Growth of the Number of Factories more
than 10 Employee

Years	Factories more than 10 employee (A)	Growth Sate (%)	Factories •ore than 5 eiployee (B)	(A) X 100 (B) (%)
1905	9,776			
1906	10,361	5.98		
1907	10,938	5.56		
1908	11,390	4.13		
1909	15, 426	35.43	32,288	47.86
1910	13,523	-12.33		
1911	14,228	5.21		
1912	15,119	6.26		
1913	15,811	4.57		
1914	17,062	7.91	31,715	53.79
1915	16,809	-1.48		
Average		6.12		

The longitudinal file which linked the factory between 1902 and 1909 contains the name and owner of the factory and other attributes such as the location, the employees and the power machines, its products and the date of establishment. Matching procedures by computer are using mat-

ching keys of various attributes and the proper name of factories. The summary of the result is shown in Table 2. (3)

Totally 32,288 factories in 1909 are divided into two categories. One is those established before 1902 (19,557 factories) and others established after 1902 (12, 165 factories). However, among those established before, the rest of the factories are not found in 1902 file. Among them, 6,542 are more than ten employees in 1909, 10,253 are between five and nine employees in 1909. Thus, some of the former ones may be attributed to the erosion of the matching or refusal to the 1902 survey.

On the other hand, looking at the 7,818 factories in 1902 from the stand point of linkage with 1909 list, 200 establishments belonging to mining are excluded from the scope of the survey in 1909. The rest of the 7,618 is divided into two categories; one is those factories that have the corresponding name in 1909 (2,740 factories) and others are those that are not found in 1909 (4,878). The point of issue is whether the latter 4,878, which may be comparable to 6,542 factories in 1909, are really vanished until 1909 or are not identifiable in 1909 file due to the incompleteness of the matching technique.

It is clear that some part of the 1909 lists may be attributed to the erosion of factories, both, in 1902 and 1909. The rates of erosion should be different by the size of the employees of the factories. The rate of erosion of those with more than 21 employees is about 39.7% as a whole but that of more than 10 to 20 employees is about 58.6%. And so if the rate of erosion is the same in both employee size, at least 2,514, or 31,4% may be attributed to the refusal of the managers.

3. Matching with Company List

The company list of 1908/9 is also compiled from the two source data, one covered Southern District including Kyoto and Osaka and the other covered the Northern District including Tokyo and Yokohama. At that period no single list of company all over Japan existed. The factory lists contained the information of the owner. Some owners or owner companies held more than two factories. We integrated the names of the factories of the same owner into one name. Thus, we can match the company list and the derived list of the factory owners including companies. Among 8,556 companies, 1,270 companies are matched with the name of the company owing factories. The matching result is shown in Table 3.

Table 2:   Transition Matrix of Factories between 1902 and 1909

		1902 (!)	1909 m		**Hatching** toy
Excluded from the survey		200 (2.6)			
Not found in 1909 list		4,878 (62.4)			
Found in both lists latched ones)	PASS I	1,074 (13.7)	1,074 (3.3)		type of company naie of company naie of factory prefectural code
	PASS II	1,584 (20.3)	1,584 (4.9)	2,740 (8.5)	only owner
	PASS III	82 (1.0)	82 (0.3)		only naie of company
Established before 1902	More than 10 emplo- yee		6,542 (20.3)	16,817 (52.1)	
	5-9 employee		10,253 (31.7)		28,982 (89.8)
	•iscella- neous		22 (0.1)		
Established after 1902	More than 10 emplo- yee		6,016 (18,6)	12,165 (37.7)	
	5-9 employee		6,149 (19.0)		
Residual			566 (1.8)		
Total		7,818	32,288		

Table 3: Linkage of the 0»pany List and the Factories owned by Coipany

Company List
of *ma*

Company Mare obtained from
r^ptory u?t *of* m_____

Company List of *ma*		Company Mare obtained from r^ptory u?t *of* m	
Jointstock Companies 2,486			
Unlimited Partnership 1,671	7,286		
Limited Partnership 3,088			
Pseudo-Company 41			
Jointstock Companies 657		Jointstock Companies 657	
Unlimited Partnership 236	1,270	Unlimited Partnership 236	1,279
Limited Partnership 377		Limited Partnership 377	
		Jointstock Companies 144	
		Unlimited Partnership 200	
		Limited Partnership 450	954
		Other Type of Company 200	
		Pseudo-Company	2,269
		Individual Owner's Factories	26,850
Total	8,556	Total	31,343

4. Conclusions

By this new longitudinal file we can analyse the function of the company system introduced to gather capital for financing the manufacturing sector and the size distribution of firms by the number of the employees and the power of engines. Before our compilation of the data file very few summary tables were published without multi-classification tables by the government of the day. The statistical report on the factory survey of 1909 contained the tables of two way classification only such as the employee size vs industry or prefecture and industry vs prefecture.

The crucial classification key is the number of employees. By combining the company lists which contained the capital information with the factory owner lists, derived from the factory lists. Thus, we can reclassify the factory by the capital, employees and power of the engine. In order, the effectiveness between two variables is computed in Table 4. It indicates that the number of employees is no good criterion of the factory size because of the substitutability between capital assets and labour. Capital assets may be reflected in the nominal capital or the horse power of the power engine. Thus, our new data file makes it possible to classify factories in more than three ways.

Notes

1. The exact matching to compile a longitudinal file has been extensively experimented at the Census Bureau of US Department of Commerce. See US Department of Commerce [1980].
2. Matsuda [1988].
3. The matching procedures are as follows: PASS I uses i) type of company such as joint stock company etc., name of the company in character strings, name of the factory in character strings and the location such as prefectural code, PASS II uses only owner's name, and PASS III uses only name of the company. PASS II and III may contains some erroneous matching.

Table 4: Size Criteria of Manufacturing Companies

	Effective Sample Size			Adjusted Correlation Coefficient		
	EMP vs HP	CAP vs HP	EMP vs CAP	EMP vs HP	CAP vs HP	EMP vs CAP
Whole industry	716	711	1,264	.2400	.5158	.2319
10. Manufacturing	2	2	11	.7402	.9804	.0227
11. Textile	265	264	396	.5603	.6276	.4361
12. Food	130	130	233	.1564	.3730	.2019
13. Machinery	42	42	103	.3757	.5523	.6088
14. Chemicals	27	27	72	.4196	.4530	.2887
15. Agriculture	76	76	105	.2625	.2572	.1268
16. Ceramics	41	41	72	.3260	.4828	.3002
17. Pulp and Papers	29	29	35	.2377	.5879	.3917
18. Miscellaneous	32	32	94	.2168	.5302	.2319

(Notes)   EMP: Employee
           CAP: Nominal Capital
           HP :   Horse power of power engines

References

MATSUDA, Yoshiro. 1988. Manufacturing and Corporate Firm System in Meiji Japan, in: K. Uno and S. Shishido eds. Statistical Data Bank Systems, Elsevier Science Publishers B.V., 1988

US Department of Commerce, Office of Federal Statistical Policy and Standards, 1980, Report on Exact and Statistical Matching Techniques