RIETI Discussion Paper Series 01-E-007

Development of Longitudinal Micro-Datasets and Policy Analysis for Japanese Industrial Sectors

MOTOHASHI Kazuyuki
RIETI

Research Institute of Economy, Trade & Industry, IAA
The Research Institute of Economy, Trade and Industry
http://www.rieti.go.jp/en/

RIETI Discussion Paper Series 01-E-007

Development of Longitudinal Micro-Datasets and Policy Analysis for Japanese Industrial Sectors

by Kazuyuki Motohashi[1]
Ministry of Economy, Trade and Industry (METI)
and
Research Institute of Economy, Trade and Industry (RIETI)

November 2001

[ABSTRACT]

This paper provides an overview of development of longitudinal micro-dataset and its use for policy analysis by METI's statistical survey data. First part of this paper presents methodology in compiling two datasets, establishment level data by Census of Manufacturing and firms level data by Basic Survey of Business Structure and Activities (BSBSA). This is followed by showing two examples in policy analysis based on these datasets, use of manufacturing census data for SME innovation policy formulation and BSBSA data for comparative study for technology, productivity and employment in France, Japan and the United States. This paper concludes with future perspectives in data development and institutional arrangements to further promote policy analysis based on longitudinal micro-dataset.

JEL Classification: D20, O33, O57
Keywords: Longitudinal Data, SME Policy, Productivity, International Comparison

[1] Executive Deputy Director, Research and Statistics Department, METI, and visitign fellow of Research Institute of Economy, Trade and Industry. e-mail: motohashi-kazuyuki@meti.go.jp. The views expressed here are those of the author's and do not necessarily reflect those of his organizations.

1. Introduction

Longitudinal Micro-Datasets, based on establishment or firm-level data from governmental statistics such as the Manufacturing Census, have been developed and actively used for policy analysis in the United States since the beginning of the 1990s. Recently, in several OECD countries, including Japan, there is also a growing awareness of the usefulness of such micro-datasets for analyzing various policy issues, such as technology and productivity and its impact on employment.

The OECD "Technology, Productivity and Job Creation" project is one of most important driving forces on micro-dataset activities in OECD countries. This research project on analyzing the relationships between technology, productivity, and employment in the OECD countries, centered on a comparison of economic conditions in Europe, where the unemployment rate had been steadily increasing since the mid-1980s, and the U.S. economy, which was characterized by a positive cycle of technological innovation, improvement in productivity, and expanded employment. OECD was asked to do this project by G7 Ministers Responsible for Labor Policy at the "G7 Employment Summit" meeting held in Detroit, USA in 1994.

In carrying out the project, it was increasingly understood that not only was it important to conduct analysis using industry-level data compiled by OECD, but that it was also necessary to do analysis using the micro-data at the establishment or firm level. For example, looking at R&D intensity indicators such as the ratio of sales to R&D expenditure, there is substantial variation at the firm level even within the same industry. For example, even in fields considered traditional manufacturing sectors, such as textiles, there are substantial numbers of high-tech firms (Cohen and Klepper, 1992). Also, in spite of the fact that aggregated statistics on the number of employees tend to decrease during times of economic decline, one can still see many offices that expand their number of employees (Davis and Haltiwanger, 1994). This kind of economic picture cannot be shown without analysis based on data at the establishment or firm level.

With active support by member countries, OECD has held experts' workshops[2] on analysis of technology, productivity and employment using micro-data, and has done follow-ups on the development of each country's data, as well as conducting various kinds of research projects at its Working Group on Industrial Statistics (OECD, 1996). In Japan, the Ministry of Economy, Trade and Industry (METI, former MITI) was actively involved in this project, compiled panel data using the establishment-level data from the Manufacturing Survey and the firm-level data from the Basic Survey on Business Structure and Activities (hereafter called BSBSA), and participated in various research projects coordinated by OECD, and I had played a central role in the project at METI, as

[2] For example, the conference on "The Effect of Technology and Innovation on Firm Performance and Employment" was co-organized with the US Commerce Department. The main papers presented at the

well as at OECD.

When I started research on micro-datasets in 1994, there were only a few examples of research in this area by academia (Matsuda (1999)), and particularly in the field of industrial statistics such as the Manufacturing Census, my work was a pioneering one. This situation was caused partly by the fact that, for the use of individual sheet data in government statistics, the confidentiality regulation under the Statistics Law made it difficult to utilize individual survey sheet data for general use. In addition, the National Statistics Office did not make great efforts to develop micro-longitudinal data or cross-linkage of different kinds of statistics at the micro level.

However, these conditions have changed greatly over the last five years. Along with the increase in awareness of the value of the micro-datasets in governmental departments and academia, due to the OECD projects, there has been a movement in statistics divisions such as METI's Research and Statistics Department to promote the use of individual sheet data. Specifically, it has compiled micro-longitudinal data by linking individual sheet data by the Manufacturing Census at the establishment level, and created groups comprised of academic researchers to build up cases of policy analysis, by studying the globalization of corporate activities, job creation and the impact on productivity of innovation, using the longitudinal datasets at the firm-level by the Basic Survey of Business Structure and Activities.

In this paper, taking these recent trends into consideration, I introduce the state of development of micro-longitudinal data based on individual sheet data of the Manufacturing Census and the BSBSA, as well as examples of policy analysis using each set of data. Firstly, in the next section, I give an overview of the characteristics, usefulness and methods of micro-longitudinal datasets, based on examples from OECD countries, particularly the U.S. Then, in Section 3, the methodology of compiling establishment-level micro-data based on the Manufacturing Census and its policy analysis is introduced, while Section 4 describes some examples of policy use of the BSBSA, which is firm-level data. Finally, this paper concludes with the future direction of the data development and research agenda.

2. Growing Demand for Micro-datasets for Policy Analysis

(1) Characterization of longitudinal micro-datasets

One of the important points related to a micro-dataset is its data unit, which could be a firm in the case of BSBSA, etc., an establishment in the case of Manufacturing Census, etc., or an individual employee in the case of worker-level wage statistics.

---

conference were published in *Economics of Innovation and New Technology*, No.7, 1998.

The unit to be used depends on the subject of analysis. For example, to estimate the production function at the micro-level in order to conduct an analysis of productivity and employment, it is desirable to make use of data at the establishment level, i.e., the smallest unit of production activity. A factory conducts production activities, with capital and labor as its input factors. In the short term, since capital is a fixed production factor, the business operator must make adjustments in response to changes in demand either through an injection of labor or through a change in labor productivity. It is observed that this pattern of activity at the establishment is not symmetric, whether economic conditions are improving or worsening (Davis et al, 1996), and differs depending on the type of industry and the country involved (Doms et al, 1995). Moreover, in cases where the economic situation differs by region of a particular country, firm-level aggregated statistics may lack a clear understanding of the big picture of individual plant-level activities throughout the country. On the other hand, in order to analyze the more complicated firm-level activities such as investment in research and development and global strategies, including overseas investment, we need to utilize firm-level data. That is because this kind of firm-wide strategic decision may be made not by each individual establishment, but rather by the firm as a whole.

Ideally, researchers want to use micro-datasets with the appropriate level of data unit, depending on the research needs. However, in reality, our research is often constrained by the existing data. In Japan, the Manufacturing Census has a long history, and it is possible to arrange the time-series data over the long term. On the other hand, the BSBSA, which is a survey at the firm level, has a rather short history, having begun in 1991, and so the only existing data is that compiled after the collapse of the economic bubble. In other countries, since large-scale micro-datasets are compiled by using existing data and surveys, available datasets differ depending on the statistics system of each country. For example, in the U.S., the Longitudinal Research Dataset (LRD) of the Center for Economic Studies at the Commerce Department is an establishment-level dataset, since this is based on Manufacturing Census surveys at the establishment level. Other survey data, such as the Survey on Manufacturing Technologies, are linked to the LRD at the establishment level. On the other hand, in European countries such as the U.K. and France, various kinds of surveys such as the Manufacturing Census, etc. take place at the firm level, and panel data is also compiled at this level. The firm lists used for these kinds of surveys are based on administrative records such as Business Registers and taxation data, and this is one possible reason why surveys are carried out at the firm level in these countries.[3]

Furthermore, data at the establishment level and data at the firm level should be used in close connection with each other. Productivity at the establishment level is probably influenced by investment in R&D and/or information technology conducted by the firm as a whole while, at the same time, management performance at the corporate level may be affected by the regional distribution of the plants. However, linking the data of different statistical surveys at the micro-level is

3 For the details of longitudinal micro-datasets in OECD countries, refer to Motohashi (1998a)

in fact an extremely difficult challenge. In Japan's case, firm-level aggregation of Manufacturing Census establishment data has been conducted, but this kind of data is limited to establishments with at least 4 employees. Moreover, because the coding system for the surveyed firms is different for the Manufacturing Census and the BSBSA, it is impossible to mechanically match the two datasets at the firm level. These kinds of problems exist in other countries as well. The U.S. LRD is panel data at the establishment level, and it can be aggregated to the firm level. However, this aggregation can be applied only to manufacturing plants. In France, as mentioned earlier, complete firm-level data exists, which was compiled by using administrative information such as taxation data as a base, but it is said that the break-down into establishment data is inaccurate.

(2) Use of Microdata for Policy Analysis

A typical application of micro-datasets is for the creation new statistical indices, through types of statistical compilation other than those seen in official publications (such as classification by industry and by firm size). These new indices are not suitable for use in specific policies, but they could become powerful basic data to use in policy planning and evaluation.

In the U.S., the Gross Job Flow, which shows fluctuations in employment at an establishment, is a well-known example of statistical index based on LRD, the micro-dataset at the establishment level (Davis et al, 1996). To calculate Gross Job Flow, one should estimate the degree of changes in employment at the establishment level, based on the method of calculation shown below.

$$g_i = \frac{EMP_t - EMP_{t-1}}{1/2(EMP_t + EMP_{t-1})}$$

Furthermore, the following four indices are compiled, by aggregating the establishment-level figures above while stressing the number of employees of each establishment, in such a way that trends in individual establishments where employment has increased or decreased, are not lost in the total figures reached in the process of aggregation.

$$POS = \sum_{g>0} \left(\frac{EMP_i}{\sum EMP_i} \right) g_i$$

$$NEG = \sum_{g<0} \left(\frac{EMP_i}{\sum EMP_i} \right) |g_i|$$

$$SUM = POS + NEG$$
$$NET = POS - NEG$$

Table 1: Gross Job Flow at U.S. Manufacturing Plants

	POS	NEG	SUM	NET
1970-75	6.2	16.5	22.7	-10.3
1975-80	8.0	9.1	17.1	-1.1
1980-85	7.9	11.1	19.0	-3.2
1985-88	8.3	8.3	16.7	0.0

Source: Davis et al (1996)

As you can see in Table 1, showing the Gross Job Flow of US manufacturing plants, the gross changes in employment for each term (SUM) are considerably larger than the net changes in employment (NET). This clearly shows that fluctuations of employment at the establishment level are much greater than the macro-changes (NET) that we usually refer to. Also, it is noteworthy that, even in the period of 1970 – 1975 when the macro-economy was depressed after the first oil shock, there were some (if not many) establishments that actually increased their employment. The characteristic of the employment changes at the establishment level being rather large compared to the net employment changes, as seen here, has also been observed in other OECD countries (OECD, 1994).

By using firm-level panel data, the dynamics of a firm's growth and decline, which cannot be explained only by comparing the size distribution of employment between two periods, can be traced. For example, let us postulate that the number of employees classified in the SME category increases between two periods. However, we cannot deduct from this that employment of SMEs has increased over time. That is, it is possible that, as a result of corporate restructuring, firms that were originally classified in the large firm category at the beginning of the period of comparison, moved to the SME category at the end of the period of comparison, so that different groups of firms were compared. For this reason, we must observe how the number of employees changes over time, by tracking the same group of firms based on panel data. According to Baldwin (1995), in the case of Canadian manufacturing plants, the smaller the size of the firm, the higher its corporate size growth rate, and the lower the productivity level compared to large firms, but this gradually becomes higher over time. Plant-level productivity dynamics have been studied for U.S. manufacturing plants as well, and various kinds of methodology to shed light on this issue have been developed. (Baily, et al,1992, Baily et al., 1994).

Furthermore, micro-longitudinal datasets comprise a strong tool for the analysis of specific policy. To give one example from the U.S., concerning the evaluation of the Department of Commerce's Manufacturing Extension Program (MEP), a consulting service for small and medium enterprises, Jarmin (1996) linked the establishment-level data (LRD) to the MEP client list, to estimate the effect of MEP on plant-level employment and productivity growth. In this work, you can see the differences in performance factors such as productivity and growth of employment between establishments that received the MEP service and those that didn't. Since the MEP clients are supposed to be those establishments that have high growth potential to start with, the question of how to control this kind of selection bias becomes an important point. Even after this kind of bias has been eradicated through econometric techniques,[4] it is observed that a certain degree of policy benefits still exists.

There are various other examples, including analysis of the benefits of R&D research consortia for Japanese firms (Branstetter and Sakakibara, 1997), firm-level impact of the unification of the EU market (Hildreth, 1996), and evaluation of the R&D tax credit system in the U.S. (Hall (1995)). To carry out analysis of specific policies, it is necessary to compile data for use in analysis, by such means as linking lists of users of the policies in question and data from surveys and questionnaires related to users, with the panel data of the Manufacturing Census, etc. As part of the administrative reforms of the Japanese Government, with the increasing importance of policy evaluation, this kind of quantitative analysis is a field on which we need to concentrate in the future.

3. Policy Analysis Using Japanese Manufacturing Census Data

(1) Development of Micro-Longitudinal Datasets by Manufacturing Census

The Manufacturing Census is a census survey for manufacturing establishments in Japan. It used to survey all establishments annually, but recently it is undertaken only in years that end in 0, 3, 5 or 8, while in other years there is a supplementary survey only of establishments with four or more employees. The survey consists of Survey A, for establishments with 30 or more employees, and the simpler Survey B, aimed at establishments with 29 or fewer employees. The total number of establishments covered is about 650,000, of which about 60,000 fall into the Survey A category.

Every year, the survey for establishments is conducted by using the identification number for each establishment, which is called the establishment code, so that the longitudinal data can be compiled based on this code. The establishment code is composed of ten digits - two digits for the prefecture, three for the city, town and village where the office is located, and five digits for each establishment located in the same city, town or village. It is possible to connect this data chronologically at the establishment level by tracing changes in each code. Concerning the codes for the geographical districts and local areas, changes can be seen when cities, towns or villages are amalgamated or abolished. Because the Management and Coordination Agency publishes the history of such changes, a list of alterations in the codes can be easily compiled. However, the problem lies in coping with changes in the codes for each establishment.

The Manufacturing Census survey is conducted through survey staff who have been appointed in each geographical district for on-site surveying, and the opening of new establishments or the closing of

[4] In Jarmin (1996), Heckman's two-step approach is applied. For details, refer to Maddala (1983).

existing ones are accurately reflected in the list of establishments in the survey. The problem is that about once every five years, the code number is revised to fill out missing numbers. To compile the long-term longitudinal datasets from 1970 on, one has to track the history of each establishment code, which has changed four times (at the times of the surveys in 1975, 1980, 1986 and 1991), but the matching table exists only for code changes in 1985 and 1990. Accordingly, for changes in the establishment codes made in 1975 and 1980, we need to find some other method to find a chronological connection at the establishment level.

One of the methods to conduct a chronological connection at the establishment level, without the help of a code-matching table, is, in the years before and after the changes in the office code took place, to select a survey item that is likely to remain unchanged at a particular establishment, and to match the establishments by using such establishment identifiers. For example, geographical data such as the area and district where the establishment is located, have probably not changed. Also, because inventory data exists for the year-beginning and the year-end, the year-end data for a particular year should match the year-beginning data for the following year. Moreover, the tangible fixed asset data exists only for the year-beginning, but by extracting the year-beginning data for the following term, using the amount of acquisition, scrap and depreciation in the case of physical assets, it is possible to match the data of the following year. Table 2 shows the items in the Manufacturing Census, Survey A that can be used for matching purposes.

Table 2: Survey Items Used for Chronological Linkage: Manufacturing Census, Survey A

Type of Data	Matching Key	Matching Method
Location code	Geog. district code City/town/village data Basic survey area code	The location of the establishment does not change over time.
Inventory data	Raw materials inventory Semi-product inventory Finished product inventory	The year-end inventory and the year-beginning inventory of the following year should have the same value.
Assets data	Land assets Machinery assets Construction assets	The year-beginning data + acquisition – scrap (- depreciation of physical assets) = the year-beginning data of the following year

Table 3 shows the results of the data linkage conducted on the basis of Survey A. The number of establishments in each year, the number of establishments that succeeded in connecting, and the number of deleted establishments are shown. As for deleted establishments, since they had exactly the same matching keys in the years before or after the timing of code change, meaning that discrimination between them was impossible, such establishments have been erased from the dataset.

Table 3: Results of Data Linkage: Manufacturing Census, Survey A

Establishment (yr.)	Establishment (yr.)	Linked establishments	Deleted establishments
57,455 (1974)	56,358 (1975)	31,068	1,499
54,203 (1979)	53,868 (1980)	35,982	1,692
57,626 (1985)	58,349 (1986)	41,586	505

Source: Motohashi, 1995

By full use of survey items in Survey A, the number of deleted establishments with the same matching keys is at the most about 3% of the total number, and it seems that a fixed result is being obtained. However, if we spread this method to Survey B, the number of offices involved will increase dramatically and, at the same time, because it is a simpler survey, the number of survey items where matching keys can be used will decrease, so we believe that chronological connection using this method would be extremely difficult.

In addition, it should be noted that there are problems associated with the panel data compiled by this method. According to Table 3, the number of linked establishments is considerably less than the number of offices in the original data. That is probably because during the period from the year before to the year after the code change, there were many establishments where the number of employees increased to 30 or more (or decreased to 29 or less). For example, if someone wanted to analyze the panel data at the two stages of 1970 and 1980, even if a certain establishment had 30 or more employees at both points in time, if the number of employees at either of the times of the linkages of data happened to be 29 or less, the office would not be included in the data set. As this kind of possibility of missing data increases every time data linkage takes place, for long-term chronology, the number of offices with a small number of employees, and the number of offices where there are major changes in the number of employees, are both under-estimated. We need to keep in mind the existence of this sampling bias when using this dataset.

(2) The Use of Manufacturing Census Panel Data for SME Policy Formulation

As part of the OECD project, the longitudinal datasets based on the Manufacturing Census Survey A were used to make an international comparison of employment fluctuations and productivity at the establishment level in Japan, France and the U.S. (For the results, see Doms et al, 1995; Motohashi, 1995, etc.) The 1986 establishment code converter was made available recently, and it became possible to compile longer-term panel data, including the Survey B data, going back to data from 1980. Accordingly, the Manufacturing Census panel data gradually came to be used for various kinds of analysis.[5] The Research and Statistics Department of MITI also compiled new statistics based on the

[5] For example, the MITI 1999 White Paper on SMEs featured plant-level dynamics of productivity and employment growth, by using Manufacturing Census panel data.

panel data to show the plant-level dynamics of the opening and closing of businesses. (MITI, 1997).

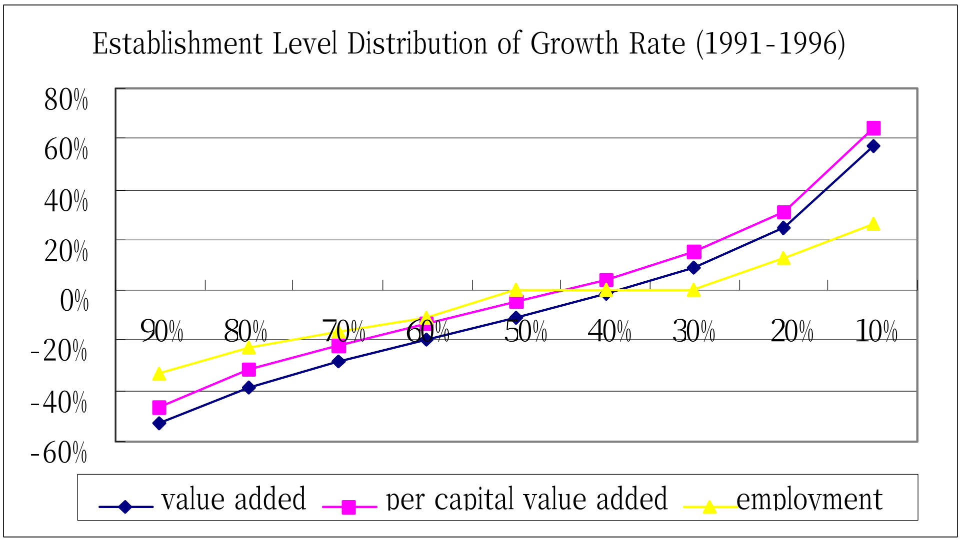
Following is an example that was used for designing the policy of the Law on Supporting Business Innovation of Small and Medium Enterprises. This Law was enacted in 1999, replacing two previous laws, the Law on Promoting the Modernization of Small and Medium Enterprises and the Law on Facilitating Small and Medium Enterprise Entry Into New Fields, which had served as the backbone of Japan’s SME promotion policies from 1963 onwards. The new law dramatically changed the basic idea of SME promotion policy, in that it facilitates the business innovation of each individual enterprise, instead of protecting weak SMEs from competition, which is the concept behind traditional SME policies. Under the new law, there are schemes whereby the manager of a small or medium enterprise can, either individually or as part of a group, create a business plan for a new project, and can also, by clarifying the goal of business performance measured by productivity growth in that plan, receive special measures such as subsidies, tax relief and low-interest loans.

In order to design the structure of business performance goals measured by productivity growth, the Manufacturing Census panel data was used. This work is not only for determining the actual cut-off point of productivity growth rates, such as a certain percentage within five years, but also for designing the structure of the performance goal itself, such as whether or not the goal needs to be changed according to the type of industry, and whether or not a special rate should be set for weak SMEs, for example an SME experiencing a decrease in its market share. The methodology of analysis used here is as follows.[6]

Graph 1 is a starting point, plotting the growth rates of establishments in the order of each 10% unit starting at the top, after calculating the order in the growth rate for all establishments, in terms of amount of value added, etc. for the five-year period from 1991 to 1996. For example, the graph indicates that, concerning the growth rate for the amount of value added, the establishments in the top 10% achieved growth rates of approximately 60% during the 1991 – 1996 period, those in the top 20% achieved about 30% growth rate, and so on. By compiling these kinds of graphs, the distribution of establishment-level growth rates of the value added, per capita value added and employment can be understood.

[6] Annual longitudinal datasets from 1986 to 1996 for all establishments with 4 or more employees were used for analysis. The total sample number of the data is about 190,000.

Graph 1: Distribution of Establishment Level Productivity Growth: 1991-1996



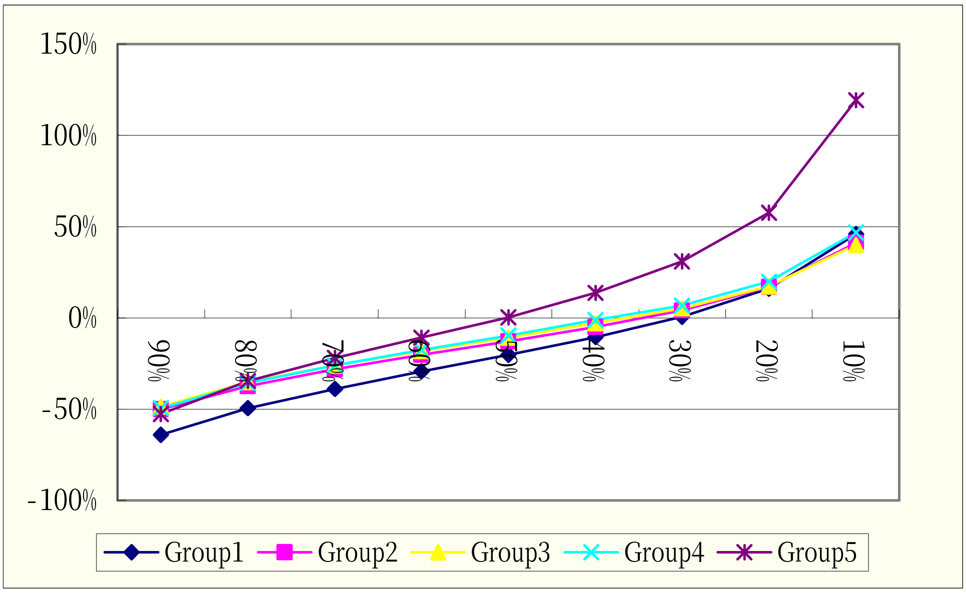
The same results were analyzed by size class, based on the number of employees, by comparing the time period and the type of industry. A summary of the observations made from these graphs follows:

- Looking at the growth rate from 1991 to 1996, about 30% of offices achieved a 15% growth rate for both the amount of value added and the labor productivity growth (per capita value added).
- If we compare the graphs for different periods, the growth rates for amount of value added, etc. show a downward trend from the 1986 to 1991 period to the 1991 to 1996 period, so that the goals obtained from the most recent results of 1991-1996 are the most severe.
- For the growth rates for the amount of value added, etc., there is greater differential between establishments within the same industry than between the different industries.

We became aware that, in the most severe period, the business performance goal which about 30% of firms (establishments) can achieve is 15% growth rate of value added or labor productivity over a five-year period. In addition, we saw that there is no need to change the business performance goal according to the type of industry. Furthermore, concerning whether or not a special rate for weak SMEs should be set, Graph 2 clearly refutes such an idea. Graph 2 looks at the distribution of the growth rate of value added, etc. by the group of business performance in the previous period. More specifically, all establishments are classified into five groups by the growth rate of the value added during the 1986 – 1991 period, with Group 1 being the group of firms which show the best performance, and Group 5 being the group for the worst firms. After classifying the establishments in this way, a graph was compiled for each group individually, according to the growth rate of the amount of value added in the succeeding period, i.e., the period of 1991 – 1996. According to Graph 2, establishments classified as Group 5, which showed the worst figures for the value added in the past, later achieved a higher growth rate of value added, implying that there is no necessity to consider a special rate of business performance goal for these firms, which may be considered weak SMEs.

Through this example, it should be noted that designing the scheme of new SME promotion policy could be done based on the objective facts, that is, the plant-level distribution of productivity growth, instead of on subjective judgement.

Graph 2: Distribution of the Value Added Growth by Group Based on Past Performance



4. Policy Analysis Using Firm-Level Survey Data (BSBSA)

(1) Compilation of Micro-Longitudinal Data by the BSBSA

The Basic Survey of Business Structure and Activities (BSBSA) is a comprehensive survey of manufacturers, wholesalers and retailers which have 50 or more employees and capital of at least ¥30 million. It is a relatively new survey, having begun in 1991. The BSBSA has been regarded as the basis of various firm-level surveys, such as one aiming to track the foreign activities of firms, short-term economic trends, etc. Since the data of other firm-level surveys is recommended to be linked with the BSBSA, the firm code system used in the BSBSA is carefully managed, and making panel data is relatively easy. When the survey began in 1991, it was expected that it would be conducted every three years, but since the second survey in 1994, it has been conducted annually. The latest data is for 1998, so there is usable data for five points in time.

Since the firm-level survey called Manufacturing Census, Survey C, which can be regarded as the predecessor of the BSBSA, was conducted in 1987 and 1989, we can utilize this earlier survey for compiling long-term panel data. The Manufacturing Census, Survey C is a complete enumerative survey of all manufacturers which have 50 or more employees and capital of ¥10 million or more, so it includes some manufacturers covered by the BSBSA. The list of companies in the first BSBSA Survey used the list of company names in this survey for reference, so tables corresponding to the firm codes for the 1989 Survey C and the BSBSA exist. However, between the 1987 survey and the

1989 one, the firm codes were changed, and there is no information about this code change.

As will be described later in more detail, the France-Japan-U.S. Comparative Analysis Project at OECD conducted analysis of technology, productivity and employment trends at the firm level from the 1980s to the 1990s. For that purpose, since connecting with the 1987 Manufacturing Census Survey C data is very important, we carried out the data linkage in the same manner as mentioned in the previous section. Because the survey is not conducted every year, unlike the Manufacturing Census Surveys A and B, we cannot use inventory and physical assets as matching keys. Therefore, data linkage was done by using the address of the company's headquarters and its industry code, and we compiled panel data on almost 4,000 firms after conducting data sorting.

(2) Policy Analysis of BSBSA Panel Data: Comparative Study of Technology, Productivity and Employment in France, Japan and the U.S.

Because it is easy to make panel data from the BSBSA, as the data accumulates, the firm-level data of this survey is becoming more widely used. In addition, at the Research and Statistics Department, groups of scholars which study how to utilize the panel data have been organized, and we are now compiling a stock of analysis cases on policy issues such as firms' foreign activities, firm-level changes of employment, and corporate diversification (MITI, 1999). In addition, the author has conducted cross-section analysis by using 1991 data such as studies on corporate performance in R&D and productivity (Motohashi, 1998b), and on corporate IT investment, organization and productivity (Motohashi, 1996).

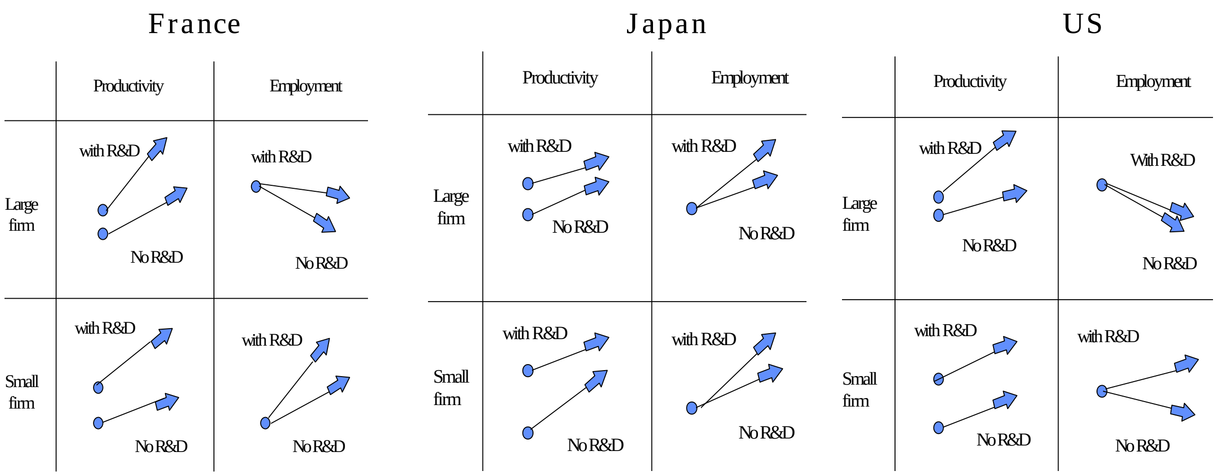
Here the author would like to introduce one of these cases, the International Comparative Study on technology, productivity and employment, which is a joint study with researchers from Japan, France and the U.S. as an OECD project. This is an advanced version of the former analytic project (Doms et al, 1995), which compares productivity and changes in employment in France, Japan and the U.S. by using establishment-level datasets. Since the study introduced here is extending the scope of analysis to innovation activities, we decided to use firm-level data where R&D related variables are available.

There are strict regulations for the use of individual survey sheet data of government statistics, not only in Japan, but also in France and the U.S., and we cannot take such data out of the country. Therefore, in this project, we conducted the study by designing comparative methods of aggregation and gathering the aggregated data without any confidentiality concern at the OECD office, where I carried out comparative analysis. We asked the National Statistics Office of France (INSEE), the Research and Statistics Department of Japan's Ministry of International Trade and Industry, and the Center for Economic Studies of the U.S. Commerce Department, to cooperate in providing the data.

For the data used in the analysis, the Japanese data is firm-level panel data in which the Manufacturing Census Survey C and the BSBSA are connected, while the French data is panel data based on Manufacturing Census centered on individual firms, and linked to R&D survey data. For the U.S. data, the LRD, with the establishment as the unit of measurement, is used, and after summing up the data for each firm, the firm-based R&D survey data conducted by the NSF, were linked. However, only manufacturing plant data can be used for firm-level aggregation. Consequently, for the firms which have non-manufacturing establishments, employees and sales amounts are undervalued. This should be taken into account when comparing U.S. data with Japanese and French data.[7]

Based on this dataset, we conducted a study to see if there was a difference in firm performance in terms of productivity and employment between firms which are active in R&D investment and those which are not. Figure 3 indicates in pictorial form the growth of labor productivity and employment of both R&D-oriented and non-R&D firms, which are divided into large firms (more than 99 employees) and small firms (fewer than 100 employees).

Figure 3: Dynamism of Firm-Level Productivity and Employment



In Figure 3, the first point indicates the base year in the 1980s, and the arrow indicates the direction of change at the second point in the 1990s.[8] For example, if we look at the level of labor productivity in the base year, in all countries, the R&D-oriented firms are located above the non-R&D-oriented ones. That is, the R&D firms are relatively higher in productivity than non-R&D firms. In addition, it can be shown that this difference in productivity is widening in France and the U.S., while in Japan it has narrowed from the 1980s to the 1990s.

As for employment trends, while large enterprises in France and the U.S. seem to be downsizing their workforce, small and medium size enterprises seems to be expanding their size. In Japan, regardless

[7] For details of data in the three countries, refer to Appendix 2.
[8] Due to the data constraints, the period of comparison differs by country; France: 1985-91, Japan: 1987-94 and

of a firm's size, it seems to be expanding, a notable difference from the other two countries. And comparing the R&D-oriented and non-oriented firms, we can see better employment performance in the R&D-oriented firms than in the non-oriented ones in all countries. This implies that technology is strongly related to employment. The OECD Technology, Productivity and Job Creation Project aims to analyze the role of technology by comparing the structural unemployment issue in Europe with the healthy U.S. economy. The observations from this comparative study support the importance of technology and innovation policy for employment, not only in the U.S., but also in Europe.

Furthermore, we analyzed the relationship between technology, productivity and employment by grouping the enterprises based on the growth rate of labor productivity and employment. Specifically, we categorize all firms into the following four groups:

Group 1: Firms with rising labor productivity and rising employment
Group 2: Firms with rising labor productivity and falling employment
Group 3: Firms with falling labor productivity and falling employment
Group 4: Firms with falling labor productivity and rising employment

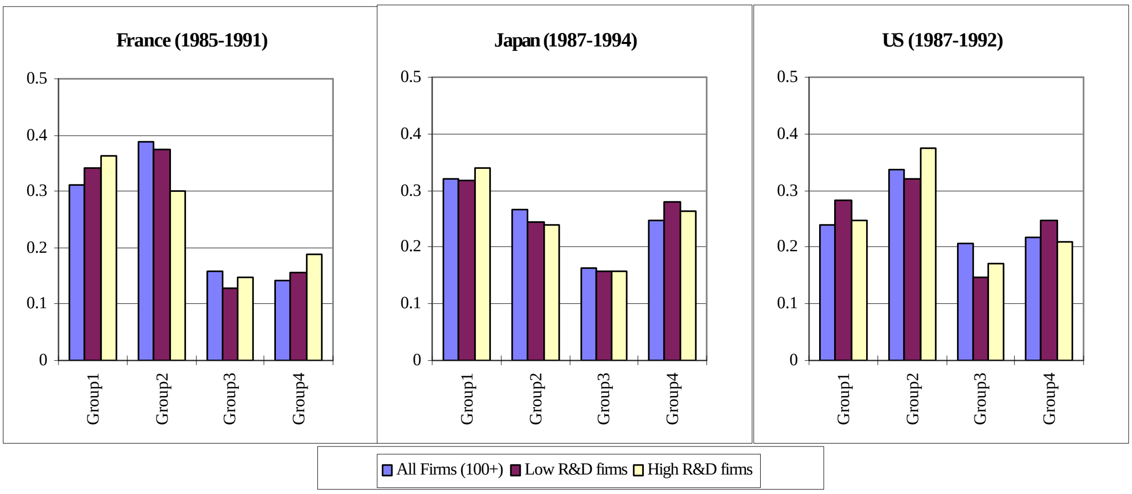
Both Group1 and Group 2 firms contribute to the productivity growth in the macro-economy, but the key difference between the two groups is that the share of Group1 firms in the economy is expanding, while Group 2's share is shrinking. Therefore, it is important to increase the share of Group1 firms in the total economy for long-term productivity growth. Figure 4 indicates the shares based on the number of firms in Groups 1 to 4. All enterprises are divided into 3 categories based on the ratio of R&D to sales for all firms, low R&D firms and high R&D firms, and then the share of each group in each R&D category is calculated. As for the change of shares in Groups 1 and 2 in France and Japan, the higher the R&D ratio, the higher the share of Group 1 and the lower the share of Group 2. This finding indicates that the share of Group1 firms is related to R&D intensity, which supports the importance of technology and innovation policy for long-term productivity growth of the whole economy.

In the U.S., this finding is the same as above if comparing all firms with low R&D firms, but the share of Group2 is much greater in high R&D firms. This might be because the U.S. economy during the analysis period (1987 to 1992) was in a so-called jobless recovery, and many large firms were trying to downsize their scale in order to strengthen their long-term corporate competitiveness. In this sense, Group 2's high share in high R&D groups might be a temporary phenomenon.

---

the U.S.: 1987-92.

Graph 4: Upsizing and Downsizing by R&D Class



5. Conclusion

As was mentioned earlier, the government census survey is conducted with the aim of showing the economic situation at the time the survey was conducted, and it is not used very much as panel data. However, with the increasing awareness of the importance of micro-longitudinal datasets as a tool of policy analysis, national statistics have started to improve the user environment for longitudinal micro-data analysis, such as the development of the establishment code converter at the Manufacturing Census. The Research and Statistics Department of METI is expected to proceed further in this direction, and the following are some concrete tasks which should be undertaken for the further development of micro-longitudinal datasets in Japan.

(1) Development of the long-term chronological longitudinal dataset of the Manufacturing Census
As in Section 3, the establishment code of the Manufacturing Census survey has been revised every 5 years, but the corresponding table of the revised codes is available only at the time of the code changes in 1986 and 1991. In order to make a long-term chronological dataset, the corresponding tables for the code change of 1975 and 1980 should be compiled.

(2) Further linkage of METI's firm-level survey data with the BSBSA
The statistical survey at the firm level conducted by METI is encouraged to use the firm list of BSBSA. Therefore, it is easy to conduct a firm-level link between METI's firm-level surveys and the data of the BSBSA. Such data linkage of various statistical sources is important not only for improving user utilities, but also for reducing the burden of the person surveyed in responding to statistical surveys by printing available information in advance.

(3) Linkage of establishment-level Manufacturing Census data and firm-level BSBSA data

The usefulness of datasets will improve dramatically if the establishment-level data of the Manufacturing Census and the firm-level data of the BSBSA can be linked each other. Few countries have achieved this and it is definitely a difficult task. However, it should be tried based on the company name file of both sets of statistics. In addition, since the Census for Establishments and Firms by the Management and Coordination Agency in 1999 was conducted simultaneously with the Census of Commerce by METI, establishment-level linkage could be done easily. Therefore, the linkage between the Census of Commerce and the BSBSA, via the Census for Establishments and Firms by MCA, should be tried as well.

(4) Establishment of a Micro-Longitudinal Dataset Center at the National Statistical Office

Under the current statistical system of Japan, we need permission from the Management and Coordination Agency in order to use micro-data from government statistical surveys. For METI's data, the Research and Statistics Department arranges the application process for this permission, and each data division inside the department also covers data gathering, compiling statistics and consulting with users. To respond to the increasing demand for the micro-data of Census surveys, the author recommends setting up a special section to cope with activities related to micro-longitudinal datasets.

The Center for Economic Studies (CES) in the U.S. Commerce Department provides a good example of such a data-centered approach. CES is in charge of maintaining the LRD based on the U.S. manufacturing census, linking with other statistics and planning new databases. CES is a research center as well, where about 10 Ph.D. economists carry out joint research with the data users. This is an effective way of not only maintaining the confidentiality of private data, but also of conducting useful policy analysis jointly with academics. It is not easy to establish such a system in a short time in Japan, but quantitative analysis in the process of policy planning and evaluation is becoming increasingly important, and a micro-longitudinal dataset has been found to be a strong tool to back up objective decision- making for policy makers. Hence, more resources should be devoted to developing and operating micro-longitudinal datasets in Japan.

Bibliography

Baily, Bartelsman and Haltiwanger (1994), “Downsizing and Productivity Growth: Myth or Reality?”, CES Discussion Paper 94-4, May 1994

Baily, Campbell and Hulten (1992), “The Distribution of Productivity in Manufacturing Plants”, *Brookings Papers on Economic Activities: Microeconomics*, 1992

Baldwin, J. (1995), *The Dynamics of Industrial Competition*, Cambridge University Press

Branstetter and Sakakibara (1997), “Japanese Research Consortia: A Microeconomic Analysis of Industrial Policy”, a paper presented at AEA meeting in New Orleans, 1997 January

Cohen, W and S. Klepper (1992), The Anatomy of Industry R&D Distribution, *American Economic Review*, vol. 82, no. 4, pp. 773 - 799

Davis and Haltiwanger (1992), Job Creation, Job Destruction and Worker Reallocation, *Quarterly Journal of Economics*, vol. 107, no. 3.

Davis, Haltiwanger and Schuh (1996), *Job Creation and Destruction*, MIT Press, USA

Doms, Jensen, Kramarz, Motohashi and Nocke (1995), “A Micro Economic Comparison of the Manufacturing Sectors in France, Japan and the United States”, a paper presented at the NBER Summer Institute, Productivity Workshop, July 1995, Cambridge,, USA

Hall, B. (1995), “Effectiveness of Research and Experimentation Tax Credits: Critical Literature Review and Research Design”, report to Office of Technology Assessment, Congress of the United States

Hildreth (1996), “UK Manufacturing Performance and the Single European Market”, a paper presented at EEA meeting of Eurostat, Luxembourg, April 1996

Jarmin, R. (1996), “Measuring the Impact of Manufacturing Extension Partnership”, a paper presented for the International Conference on Comparative Analysis of Enterprise Data, June 1996, Helsinki

Maddala (1983), *Limited Dependent and Qualitative Variables in Econometrics*, Econometrics Society Monographs, Cambridge University Press

Matsuda (1999), Micro-data no egaku Shakaikeizaizo, (in Japanese) Nihon Hyoronsha

MITI (1999), Research on Globalization and Job Creation of Japanese Firms (in Japanese), a report by Industrial Research Institute

MITI (1997), Research on New Field Entry of Japanese Manufacturing Establishments (in Japanese), a report by Industrial Research Institute

Motohashi, K.(1998a), “Institutional Arrangements for Access to Confidential Micro-level Data in OECD Countries”, STI Working Paper 98-3, OECD Paris

Motohashi, K.(1998b), “Innovation Strategy and Business Performance of Japanese Manufacturing Firms”, *Economics of Innovation and New Technology*, vol.7, pp.27-52

Motohashi, K.(1996), Use of Information Networks, Organizational Changes and Productivity: Firm Level Evidence in Japan, a paper presented at the NBER Summer Institute, Productivity Workshop, July 1996, Cambridge,, USA

Motohashi, K (1995), Analysis of Technology, Productivity and Employment, using Individual Sheet Data of Census Statistics (in Japanese), Keizai Tokei Kenkyu, MITI Statistics Association, vol. 23, no. 3

OECD (1996), *Technology, Productivity and Job Creation, vol.2 : Analytical Report,* OECD, Paris

OECD (1994), *Employment Outlook*, OECD, Paris

Annex 1. **Firm Level Panel Data by Manufacturing Census C and BSBSA**

[Primary Datasets]

- Manufacturing Census, Survey C[9] (MSC) : complete enumeration with a cut-off point of 50 employees and ¥10 million capital[10] for manufacturing firms (# of firms is **19,702 for 1987 and 21,271 for 1989**)
- Basic Survey of Business Structure and Activity (BSBSA): complete enumeration with a cut-off point of 50 employees and Y¥30 million capital[11] for manufacturing and commerce firms (# of firms is 23,776 and **# of manufacturers is 13,688**)

[Linking Procedure]

There is a firm level matching table between MSC of 1989 and BSBSA, but there is not between MSC of 1987 and that of 1989. Therefore, that linkage is done by using matching keys as follows;

- Location code of a headquarters (prefecture, city and region codes)
- 3-digit industrial classification

The linkage has been done as follows;

- For each of datasets in 1987 and 89, firms with identical keys are omitted. In this process, the number of firms goes down from 19,702 to 17,641 in 1987 and from 21,271 to 19,046 in 1989.
- Matched by keys. Data cleaning of matched sample is done by using the criteria that the employment change between both periods is from -50% to 100%.[12] (# of samples : 11,815)
- This matched dataset of 1987 and 1989 is linked with 1991 BSBSA by a firm ID conversion table. The same data cleaning is conducted, and the number of samples becomes 5,656. [13]

[Industrial Classification]

3-digit special classification system for BSBSA (77 industries in manufacturing)

[Available Variables]

- Basic financial statements (inputs and outputs, net book value capital etc.)
- The number of employees by occupation (blue collar, white collar, R&D person etc.)
- R&D expenditure and advertising expenditure
- Exports and imports

---

[9] Survey C is a special survey at the firm level, and is had been done only in 1987 and 1989. Regular establishment level surveys (Survey A and B) have been conducted every year.

[10] As of 31/12/'87 and 31/12/'89.

[11] As of 31/3/'92

[12] By this data cleaning, wrongly matched samples as well as firms with major restructuring are omitted.

[13] In MSC, R&D expenditure is surveyed only for firms with no less than 100 employees. The sample

- Detail technology variables only in 1991 (patent, technology licensing, use of information networks etc.)

[Deflators]

Deflators for output and input at the 3-digit industry classification of BSBSA are compiled, based on MITI's input-output tables. MITI's IO tables are available every year at more than 400 classification levels, and deflators in 1987, 1989 and 1991 are derived from aggregated constant and current values of inputs and outputs at a BSBSA classification level.

[Definition of R&D variables]

1987 & 89 data and 91 data come from different surveys, and the definition of R&D expenditure is different, as the following notes on questionnaires show.

Manufacturing Census Survey C (1987 & 1989 data): *All expenses for R&D activities, such as wages, materials, investments and utilities. Outsourced R&D is also included*.

Basic Survey of Business Structure and Activity (1991 data): As well as the above descriptions, *for manufacturing firms, not only pure research activities, but also development associated with products, production processes, and technical improvement activities are also included.*

As for the R&D employees, both statistics apply the head count method, instead of that of full-time equivalent (FTE). Although the definition of R&D employees is the same for all three periods, the data are not comparable between the two kinds of surveys. For the 1991 data, if there are any independent establishments in addition to a firm's headquarters, all employees of such establishments are classified by the firm's classification. For example, all employees in manufacturing plants are counted as manufacturing workers, even though some of them are R&D employees. The following table summarizes the statistics of firms with and without separated manufacturing plants.

	# of Firms	R_D/EMP87	R_D/EMP89	R_D/EMP91
all firms	3465	6.70	7.13	4.98
without plant	1605	4.29	4.92	4.18
with plant	1860	7.53	7.90	5.26

size of data with R&D expenditure is 3860.

Annex 2: Datasets for France-Japan-US Comparative Study

	FRANCE	JAPAN	US
Name of Database	SUSE - systeme unifie de statistiques d'entreprises (based on EAE)	Japanese R&D Panel	US Longitudinal Research Database (LRD)
Unit of Analysis	Firm (SIRENE)	Firm	Establishment, with firm identifier in every year.
Sectors	All private and public firms.	Manufacturing	Manufacturing
Years of Coverage	1984-1992, but older data are available from other datasets	1987, 89, 91 and 94	1963, 1967, 1972-1992.
Sample Characteristics	Complete every year. Small firms, under 20 employees may be under represented (small firms participating in the Forfait or BNC tax systems are not included)	All firms with 50+ employment Some cut-pff points on total capital is also applied. (20,000 firms), but R&D data are available for only 100+ firms	Census every 5 years (approx. 350,000 estabs.), probability sample in the intervening years (approx. 55,000 estabs.)
Industry Classification	NAP-600, some ISIC link	3-digit Japanese SIC	4 digit ISIC revision 2
Employment Data			
Level of Employment	annual mean number of employees	December 31st employment	March 12th employment
Salaries	Total annual salaries, and total labor costs by skill level- see below	Total annual salaries (including bonuses and other benefits, such as housing allowances.)	Total annual salaries. Also, supplemental labor costs.
Production Data			
Sales	Shipments as measured by freight on board prices	Shipments as measured by freight on board prices	Shipments as measured by freight on board prices
Value Added	Firm level value added (total sales - all cost incurred + wage + depreciation)	Firm level value added (total sales - all cost incurred + wage + depreciation)	Defintion of value added = sales - change in inventories - cost of purchased materials - cost of energy - cost of contract work + value of receipts of contract work performed.
Capital	Book value of machinery, equipment and buildings, as well as estimated one by perpetual inventory method	Book value of machinery, equipment and buildings	Book value of machinery, equipment, and buildings. 1972-1985, ASM establishments only. 1987 and 1992 for all non-admisnistraive records.
R&D Data	Firm level R&D survey data can be linked. Data are available from 1974 to 92	R&D expenditure is surveyed	NSF R&D survey data can be lined at the firm level. (LRD data are aggregated into firm level)
R&D Expenditure	Activity base R&D expenditure as well as self-financed R&D. The breakdown into basic, applied R&D and development can be done	Finance Base R&D expenditure, Breakdown into activity base one and self-financed one can be done for 1991 data	Activity base R&D expenditure as well as self-financed R&D. The breakdown into basic, applied R&D and development can be done
Number of Scientist	# in full time equivalent	# in head count	# in full time equivalent
Other comments	Innovation survey data can be linked at the firm level	Detail technology variables, such as patent, technology licencing and use of information network, are avaialble in 1991 and 94	Survey on Advanced Manufacturing Technology can be linked at the establishment level