The Establishment File of the German Social Insurance Statistics

By Michael Fritsch and Udo Brixy

1. The German Social Insurance Statistics

The German Social Insurance Statistics („Statistik der sozialversicherungspflichtig Beschäftigten", SIS) was introduced in West Germany in 1973. In eastern Germany, it became compulsory from January 1991 on, shortly after unification. The SIS requires all public and private employers to report certain information about every employee who is subject to obligatory social insurance, i.e. health and unemployment insurance along with pension funds. The information collected can be transformed into an establishment file that provides longitudinal information about the establishments and their employees.

In this contribution we give an overview of the content of the SIS establishment file as well as of the possibilities and the problems involved in using this data for scientific analyses. We first describe the coverage and information contained in the SIS (section 2) and give a brief outline of the process by which this information is raised and prepared. Section 3 deals with data errors and reliability problems, particularly the definition of an establishment and the identification of start-ups and closures. Section 4 depicts the main topics of studies based on the data and describes the current possibilities of data access.

2. Coverage and informational content

The German SIS is derived from information that every employer is obliged to report about those employees who have to contribute to the social insurance system. The statistics therefore contain only data about this group of persons. Civil servants, army personnel, self-employed and unemployed persons are not included. Due to some changes of the criteria for being obliged to contribute to the social insurance scheme there was some variation in the definition of the observation unit over time. In manufacturing, the SIS currently covers more than 90 percent of all employed persons. In the service sector, this share is about 80 percent. It is relatively low in agriculture (less than 24 percent) and in the public sector (about 50 percent). The employers report the informa-

tion to the responsible health insurance company, which passes it on to the statutory pension insurance. The pension insurance transfers it to the Federal Employment Service (Bundesanstalt für Arbeit) where the data are collected and prepared. Each of these institutions may check the accuracy of the data and correct errors.

For every employee who is obliged to contribute to the social insurance system the SIS contains the following information:

- social insurance number;
- gender and date of birth;
- a number for the establishment where the person is employed;
- code number of the municipality (Gemeindkennziffer) in which the establishment is located;
- industry affiliation (since 1998 according to the five-digit NACE classification system);
- profession according to a classification of 334 categories;
- employment status (e.g. apprentice, unskilled, skilled, master craftsperson; full time employees only);
- the educational level attained and vocational training concluded;
- nationality and
- gross income (truncated at the income limit at which employees are free to opt out of the statutory health insurance and choose private health insurance).

For those establishments that belong to a firm with several establishments, no information about the respective firm is contained in the data.

3. Different forms of data preparation

The most frequently used form of the SIS to date is the *quarterly end of period files* (Vierteljahres-Stichtagsdateien). These files contain information about every employed person that was subject to compulsory social insurance on the last day of each quarter, i.e. as of March 31, June 30, September 30 and December 31. The quarterly end of period files are available about nine months after the respective date. It has become common practice in scientific analysis of these data to prefer the use of the quarterly file of June 30 for two reasons. First, there is no indication of any specific bias on that date. And second, information relating to the middle of a year appears to be particularly representative of that year. Another form of the SIS is the *history file* (Historikdatei), which contains all of the information gathered about a certain em-

ployee since the introduction of the statistical system. It makes it possible to
follow an employee's work history. Because the history file includes even
information that is reported very late and also accounts for late error correc-
tions, it is the most complete and reliable version of the data.

Although the original unit of observation of the SIS is the employee, the
data can be transformed into an *establishment file* by means of the establish-
ment number that is reported. In this procedure, all employees with the same
establishment number are allocated to this number. The current version of the
establishment file is generated from the quarterly end of period files that relate
to June 30 of each year. These data comprise information about all establish-
ments in the Federal Republic of Germany that report employees who were
obliged to contribute to the social insurance system on at least one of these
dates. Because the SIS does not contain any information about the respective
firm, the establishment file can not be further aggregated to that level. The
current West German file goes back as far as 1981 and does not cover the
earlier years because the respective quarterly files no longer exist. Currently, a
new version of the establishment file is being generated on the basis of the
history file. This new data set will provide information from the introduction
of the SIS in 1973[1] onwards and it will be more comprehensive and reliable
than the establishment file based on the quarterly end of period data.

4. Data errors and reliability problems

4.1 The definition of an establishment

The definition of what constitutes an establishment is of crucial importance
for the usefulness of the establishment file. Generally, the establishment iden-
tifies a certain location, i.e. a plant or a workplace. There are, however, two
main deviations from this definition. First, if a firm has several plants in one
municipality which are affiliated to the same industry, the employees in these
plants may be subsumed under the same establishment number. And second, a
certain establishment may be split into several establishment numbers if this
seems appropriate to the management for any reason. This could be the case
particularly in relatively large establishments with different lines of operation.
The fact that the establishment number of the SIS does not relate to a certain
plant or location can be a problem particularly in the public sector, in non-
profit organizations and in industries where larger firms tend to have many
small branch offices such as in the retail trade, banking and the insurance
industry. However, such incorrect definitions of the establishment are rela-

---

[1] However, due to diverse errors during the implementation of the SIS the quality of
the data gathered in the first two years must be regarded as rather inadequate.

tively rare (well below ten percent of all cases) and large deviations of the number of employees as recorded in the SIS from the 'true' number are an exception (cf. Brixy and Fritsch 2002, 74 f.).

4.2 Identification of start-ups and closures

Under the assumption that new establishment-numbers in the German Employment Statistics may be regarded as indicating newly founded establishments and that disappearing numbers represent closures, the establishment file of the SIS provides the opportunity of analyzing establishment dynamics in the economy. In fact, the start-up of a new establishment with employees that have to contribute to compulsory social insurance will usually[2] lead to the inclusion of a new establishment number. And in the case of a closure, the number will normally disappear. However, there are other possible reasons for the appearance and disappearance of an establishment number in the statistics, such as

- the redefinition of an existing establishment, which can lead to either the cancellation of an established number and the introduction of a new one,

- ownership change, and

- in some rare cases the alteration of a firm's legal form.

Because establishments may exist for some considerable time before hiring an employee, start-ups may appear in the SIS 'too late'. Correspondingly, closures may disappear from the SIS prematurely if the establishment continues to exist without any employees.

One main problem in identifying start-ups and closures is establishments with 'perforated' employment, where information is interrupted for some time. In this case, it is unclear whether in the years with missing information the respective establishment

- has provided the data too late to be included in the quarterly end of period files which are currently the basis for the establishment file,

- had existed without an employee, or

- had been closed down and the entrepreneur has set up a new business after some time.[3]

---

[2] It will not lead to a new number if the new establishment belongs to a firm with an existing establishment of the same industry in that municipality and if the parent firm wants to subsume these establishments under one number.

[3] The establishment number is linked to the person of the owner. If an entrepreneur has gone out of business and then starts a new firm after some time, the old establishment number is used for the new business.

Because establishments with a perforated number tend to be rather small[4], one can assume that in many cases they have existed without an employee during the interruption.[5] In the current version of the establishment file, the length of the time period for which there is no information about an establishment in the data is used to separate continuing establishments from start-ups and closures.[6] Establishments with an interruption of less than three years are assumed to have continued without an employee during that time. If an interruption has lasted longer than three years, the disappearance of the establishment number is classified as a closure and re-emergence as a start-up. The range of possible misidentification associated with this criterion can be estimated in a rather simple way. According to the criterion, about 87 000 establishments that reappeared in the 1981–97 period are classified start-ups. If all of these establishments had continued business during the time of interruption, the number of start-ups would be overestimated by about four percent. If about half of these establishments had continued, the mistake would amount to about two percent.

There are a quite a lot of new establishment-numbers in the data that cover an untypically large number of employees. One might suspect these new numbers to be related to the reorganization of an existing firm or to the establishment of a larger branch plant. In order to single these cases out, the number of employees is used as a second criterion for identifying start-ups. New establishment-numbers are not classified as start-ups if more than 20 employees are reported in the first year. Therefore, a new establishment-number is classified as a start-up if it was not included in the establishment file during the three preceding years and does not report more than 20 employees on the date it enters the statistics. Because no additional size criterion is applied for the identification of a closure, classification as a closure is somewhat less restrictive than classification as a start-up. As a consequence, these definitions lead to an underestimation of net-entry, i.e. the number of start-ups minus the number of closures.

However, comparisons with the data of a yearly interview survey of some 4,000 firms which is based on a random sample of the SIS establishment file (Establishment Panel of the German Institute for Employment Research, IAB)

---

[4] 94.6 percent of all West German establishments that were characterized by perforated information in the 1981–97 period had no more than three employees in the year before the interruption occurred. 77.8 percent had only one employee. 96.1 percent of these establishments reported no more than three employees when they re-appeared in the statistics. 80.3 percent had only one employee.

[5] Interrupted information due to late notification does not occur if the history file is used for generating the establishment data.

[6] Taking the West-German data, we find that between 1981 and 1997 about 60% of all interruptions were just for one year. Only 14.5% of all interruptions related to a period longer than three consecutive years.

showed that these relatively crude assumptions lead to quite satisfactory results (Brixy 1999). Furthermore, comparisons of different data bases for identifying start-ups in Germany have revealed that errors and mis-specifications of start-ups were in no way more relevant in the SIS than in the other data sources (Fritsch et al. 2002). Moreover, the numbers of start-ups by year, industry and region that were identified from the different data sources tended to be quite closely correlated. So even if the information about start-ups generated from the establishment file of the SIS can not be regarded as the true number of newly founded establishments, it represents quite a good indicator of the actual number. There is currently no other data base available in Germany from which closures can be identified as comprehensively and reliably as from the establishment file of the SIS (cf. Fritsch et al. 2002). The level of errors and mis-specifications in identifying closures should be in about the same range as for start-ups, so the data can be regarded as an appropriate indicator of the 'true' number of closures. The higher data quality of the new version of the establishment file that is based on the history file of the SIS can be expected to permit a considerable reduction of errors in identifying closures and start-ups.

5. Main topics of studies based on the data and data access

Empirical studies based on the establishment file of the German SIS have focused on start-up processes, particularly regional variations of the number of newly founded firms[7], the effects of new firm formation on economic development[8], the performance of establishment cohorts[9] and job-turnover[10].

Direct access to the SIS is limited to the employees of the Federal Employment Service and the associated IAB. Currently, some aggregate information (e.g. the number of employees with certain qualifications in different industries or regions) are published on a regular basis. In the future, information from the establishment file of the SIS will be available from a research data centre that is planned to be set up at the IAB. An anonymized one percent random sample drawn from the history file that provides information on the number of employees and also about the respective establishments can be made available for scientific analysis (cf. Bender, Haas and Klose 2000). Information about the number of start-ups and closures per year, industry and region are provided by the Chair of Economic Policy at the Tech-

---

[7] Audretsch / Fritsch (1994, 1999), Brixy (1999a), Brixy / Grotz (2002), Fritsch (1992), Fritsch / Falck (2003), Lehmann, Mössinger (1996), Lehmann (1996a),

[8] Audretsch / Fritsch (2002), Fritsch (1996, 1997).

[9] Boeri / Cramer (1991), Brixy (1999a und b), Brixy / Kohaut (1999), Brixy / Grotz (2004).

[10] Cramer / Koller (1988) Lehmann (1996b), Brixy (1999a).

nical University of Freiberg and can be downloaded from the internet under http://www.wiwi.tu-freiberg.de/wipol/gründungsatlas/gründungs_indikatoren.

References

Audretsch, D. B./*Fritsch,* M. (1994): The Geography of Firm Births in Germany, Regional Studies 28, 359–365.

– (1999): The Industry Component of Regional New Firm Formation Processes, *Review of Industrial Organization* 15, 239–252.

– (2002): Growth Regimes over Time and Space, Regional Studies 36, 113–124.

Bender, S./*Haas,* A./*Klose,* Ch. (2000): The IAB Employment Subsample 1997–1995, Schmollers Jahrbuch 120, 649–662.

Boeri, T./*Cramer,* U. (1991): Betriebliche Wachstumsprozesse – Eine statistische Analyse mit der Beschäftigtenstatistik 1977–1987. in Mitteilungen aus der Arbeitsmarkt- und Berufsforschung 24, 1, 70–80.

Brixy, U. (1999a): Die Rolle von Betriebsgründungen für die Arbeitsplatzdynamik. Beiträge zur Arbeitsmarkt- und Berufsforschung 230, Nürnberg.

– (1999b): Gründe und Hintergründe des Erfolgs neuer Betriebe in Ostdeutschland, in: Wiedemann, E. et al. (eds.): Die arbeitsmarkt- und beschäftigungspolitischen Herausforderungen in Ostdeutschland. Beiträge zur Arbeitsmarkt- und Berufsforschung 223, Nürnberg.

Brixy, U./*Fritsch,* M. (2002): Die Betriebsdatei der Beschäftigtenstatistik der Bundesanstalt für Arbeit, in: Michael Fritsch and Reinhold Grotz (eds.), 55–77.

Brixy, U./*Grotz,* R. (2002): Räumliche Differenzierungen von Betriebsgründungsintensität und Überlebenschancen in Westdeutschland 1983–1997, in: Raumforschung und Raumordnung 60 (2), 100–122.

– (2002): Räumlicher Vergleich der Datengrundlagen zum Gründungs- und Stilllegungsgeschehen in Deutschland, in: Michael Fritsch and Reinhold Grotz (eds.), 165–198.

– (2004): Entry-Rates, the Share of Surviving Business and Employment Growth: Differences between Western an Eastern Germany since Reunification, in: Michael Dowling, Jürgen Schmude and Dodo zu Knyphausen-Aufsess (eds.): Advances in Interdisciplinary European Entrepreneurship Research, Münster: LIT, 143–152.

Brixy, U./*Kohaut,* S. (1999): Employment Growth Determinants in New Firms in Eastern Germany. Small Business Economics, 13, 155–170.

Cramer, U./*Koller,* M. (1988): Gewinne und Verluste von Arbeitsplätzen in Betrieben – der „Job-Turnover“-Ansatz, in: Mitteilungen aus der Arbeitsmarkt- und Berufsforschung 21 (3), 361–377.

Fritsch, M. (1992): Regional Differences in New Firm Formation: Evidence from West Germany, Regional Studies 25, 233–241.

– (1996): Turbulence and Growth in West-Germany: A Comparison of Evidence by Regions and Industries, *Review of Industrial Organization* 11, 231 – 251.

– (1997): New Firms and Regional Employment Change, *Small Business Economics* 9, 437 – 448.

Fritsch, M. / Falck, O. (2003): New Firm Formation by Industry over Space and Time: A Multi-Level Analysis, Discussion Paper 322, German Institute for Economic Research (DIW Berlin).

Fritsch, M. / Grotz, R. (eds.) (2002): Das Gründungsgeschehen in Deutschland – Darstellung und Vergleich der Datenquellen, Heidelberg: Physica.

Fritsch, M. / Niese, M. (2002): Vergleich auf gesamtwirtschaftlicher und sektoraler Ebene, in: Michael Fritsch and Reinhold Grotz (eds.*),* 141 – 164.

Fritsch, M. et al. (2002): Zusammenfassender Vergleich der Datenquellen zum Gründungsgeschehen in Deutschland, in: Michael Fritsch and Reinhold Grotz (eds.*),* 199 – 214.

Lehmann, U. (1996a): Dynamik und Beschäftigungsentwicklung der ostdeutschen Betriebe 1991 – 1995. MittAB 29 (4), 580 – 588.

– (1996b): Flexibilität durch Job-Turnover und Betriebsgründungen.in: Flexibilität von Betrieben in Deutschland. Beiträge zur Arbeitsmarkt- und Berufsforschung. 200, 105 – 118.

Lehmann, U. / Mössinger, W. (1996): Regionale Arbeitsplatzdynamik in den neuen Ländern. In: Informationen zur Raumentwicklung, H.1, 29 – 38.