Schmollers Jahrbuch 120 (2000), 143 – 150
Duncker & Humblot, Berlin

European Data Watch

This section will offer descriptions as well as discussions of data sources that may be of interest to social scientists engaged in empirical research or teaching courses that include empirical investigations performed by students. The purpose is to describe the information in the data source, to give examples of questions tackled with the data and to tell how to access the data for research and teaching. We will start with data from German speaking countries that allow international comparative research. While most of the data will be at the micro level (individuals, households, or firms), more aggregate data and meta data (for regions, industries, or nations) will be included, too. Suggestions for data sources to be described in future columns (or comments on past columns) should be send to: Joachim Wagner, University of Lueneburg, Institute of Economics, Campus 4.210, 21332 Lueneburg, Germany, or e-mailed to ⟨wagner@uni-lueneburg.de⟩.

Firm Panel Data from German Official Statistics

By Joachim Wagner

Introduction

The past decade has seen a mushrooming growth of empirical studies in industrial economics, labor economics, international trade, and small business economics using large panel data sets at the firm (establishment or enterprise) level that were built from data collected regularly by the statistical offices. Firm panel data of this type are now available for most of the large industrial countries (Motohashi 1998), and their application lead to new and sometimes exciting findings (see Caves 1998 for a recent survey).

One problem with these micro data from official statistics lies in the strict rules of confidentiality that have to be followed in any application. For good reasons (not to be discussed here in any detail) the producers of these data are not allowed to give it away to external researchers because it is impossible to ensure anonymity for big firms. However, there are ways to allow scientists from universities and research institutes to use these confidential

micro data inside the statistical offices to produce results (e.g., to estimate parameters of an empirical model) that are not confidential at all. How this is done differs from country to country, and from agency to agency.

This paper describes the German case in six steps: (1) It gives an outline of the contents of the data collected regularly by the statistical offices of the federal states in establishments from the manufacturing sector. (2) It looks at the topics of selected empirical studies using these data. (3) It shows how researchers from outside the statistical offices use the data that are kept inside. (4) It introduces the German user group of firm level data from official statistics. (5) It points to the potential use of these data in teaching. (6) It ends with a peek into the future.

1. Information in the data

In Germany, the population of establishments in the manufacturing sector (including the mining one) is surveyed regularly by the statistical office of the respective federal state, the *Statistisches Landesamt*.[1] Details aside, all establishments with at least twenty employees (including the owner, and members of his family) or that are part of a multi-establishment enterprise of at least this size have to take part in a monthly survey, the *Monatsbericht für Betriebe im Bergbau und Verarbeitenden Gewerbe,* and in an annual survey of investment, the *Jährliche Investitionserhebung bei Betrieben des Bergbaus und des Verarbeitenden Gewerbes*. All other (smaller) establishments from mining and manufacturing are surveyed once a year in September in the *Erhebung für industrielle Kleinbetriebe im Bergbau und Verarbeitenden Gewerbe*. For convenience, let us label the first group of establishments large firms, and the second group small firms.

Using the unique identifier, the establishment number (*Betriebsnummer*), information for an establishment from various monthly and annual cross section surveys can be linked over time to form an (unbalanced) panel covering the population of large and small firms and combining information from the *large firms panel* and the *small firms panel* leads to a panel covering the population of all firms in mining and manufacturing, i.e. the *all firms panel*. There is no "panel attrition" and the data in these panels are (or at least, should be) reliable, because the units are mandated to take part in the surveys and to report all figures correctly.

[1] Note that these surveys are neither conducted by the federal statistical office, the *Statistisches Bundesamt,* nor are the micro data collected in the federal states transferred to the federal office.

As is often the case with data from official regular firm level surveys, information available in these surveys is not very comprehensive. To start with the *large firms panel,* we have the information listed in Box 1 for each establishment that was a member of the respective universe in at least one year since 1978.[2]

Box 1

Variable list for the *large firms panel*

Establishment number
Enterprise number
Location (county [*Kreis*])
Industry (4-digit identifier)
Number on monthly reports
Sum of hours worked by blue collar workers
Sum of gross wages
Sum of gross salaries
Total number of persons in the establishment (average of monthly reports)
Number of blue collar workers (average of monthly reports)
Sum of sales in Germany
Sum of sales outside Germany
Sum of investment in land with buildings
Sum of investment in land without buildings
Sum of investment in machinery
Sum of payments for rent and leasing
Sum of value of production

For *small firms* information is available for the following variables:

- Establishment number
- Enterprise number
- Location
- Industry
- Total number of persons in the establishment at the end of September
- Total sales in September
- Total sales in the year before the survey

[2] The exact content of the panel data base may vary from federal state to federal state. The description given here applies inter alia to Lower Saxony *(Niedersachsen),* the state I live in and where I am working with these panel data. For the questionnaires used in the surveys, see Methner (1992); a more detailed description of the data is given in Gerlach and Wagner (1997).

Given that information in the *all firms panel* is restricted to variables that are collected for both large and small firms, this describes the contents of the panel data set covering the entire population of establishments from mining and manufacturing, too. Note that both the *large firms panel* and the *all firms panel* are unbalanced panels due to firm entry and exit.

Efforts to build this type of firm panel data sets started in Germany in 1990 in the federal state of Lower Saxony, and it took quite some time until statistical offices and research teams in other federal states joined this endeavor. Beginning with Saxony in 1997, more and more states came on board, and at the beginning of the new millennium these firm panels are either in use or under construction in almost all states.[3]

2. A look at topics of selected studies with the data

Readers not familiar with the kind of data described in this note may wonder which questions can be investigated using these panels, given that information is extremely scarce in the *all firms panel* and not that rich in the *large firms panel*. I cannot discuss the numerous empirical investigations based on the panels here; interested readers may look at the comprehensive survey of the work with the data from Lower Saxony given by Gerlach and Wagner (1997). Instead I will do my best to whet the appetite of prospective users by enumerating selected topics:

- *Job turnover studies* decompose the net change in employment between two points in time in all firms from a region or industry, or a firm size class into gross flows due to entering, growing, shrinking and exiting firms. These studies point to an enormous amount of heterogeneity inside any group of firms – underneath the smooth path of aggregates there is a very active micro world! (Gerlach and Wagner 1993)
- *Firm size and firm growth* are investigated in studies testing whether Gibrat's Law holds, or whether small firms tend to grow faster than large firms, and whether growth rates tend to be linked positively or negatively over time. (Schmidt 1995)
- *The post-entry performance of cohorts of new firms* can be documented by looking at the fate of new firms in the years following their birthyear, and the long-run employment effect of entries can be computed taking account of the liability of newness on the one hand and the growth of surviving firms on the other hand. (Wagner 1994)

---

[3] For up-to-date information, see my homepage: http://www.uni-lueneburg.de/fb2/vwl/wifo/fidast.

- *Determinants of labor demand* can be studied using dynamic panel models (Breitung 1992).
- *Export entry and exit* can be investigated, and exporters and non-exporters can be compared to export starters over time, showing that the good go abroad but going abroad does not make better firms. (Bernard and Wagner 1997)

There are many more topics. Curious about how to get access to these data?

3. Data access

Establishment level data from official statistics are strictly confidential by law *(Statistikgeheimnis),* and the statistical offices are not allowed to give these data to outside researchers – unless there is either an agreement with each and every firm surveyed (and this means, never), or unless the data are manipulated so that the units cannot be identified without any disproportionate effort (and this means, more or less useless for most any serious research project due to the need to drop or disguise information on location, industry, size, sales etc.). However, there are ways and means that allow "outsiders" who are no "official statisticians" to use the micro data without giving them away. Details differ from federal state to federal state, but two models can be identified:[4]

- When we started the first of these projects in Hannover some ten years ago, the statistical office in Niedersachsen offered a rather simple way to access the micro data: send us a diskette with your program (written in Stata, SPSS, GAUSS, or LIMDEP); we will run it, check whether the results reveal any secret information, and send you the output file if this is not the case. Variants of this procedure are in use in some federal states now.
- When the project in Baden-Wuerttemberg was started in 1998, another way to access the micro data was chosen: to a member of the research team the status of a sworn employee (without any salary) in the statistical office was given. He can work with the micro data inside the office by himself the way all other employees do, and he has to present any results to the office for a privacy check before taking them out and publishing them. Several other statistical offices adopted this way of data access.

[4] For a more detailed description and a discussion of pros and cons see the papers presented at a workshop on the use of firm level data from official statistics published in Statistisches Bundesamt (1999).

Any details aside, access to the confidential firm panel data from official statistics needs a contract between the external user and the statistical office. Readers interested in working with these data should get in contact with a research group that is working with the data in the respective federal state.[5] This is easily done via the web (see footnote 3), and if you find out that there is no such group in your state, start one.

4. The FiDASt-network – a user group

Research teams working with firm panel data from official statistics or groups starting such a project are active in nearly every federal state. All of these teams are connected via a network labeled *FiDASt* (an acronym for *Fi*rmen-*D*aten aus der *A*mtlichen *St*atistik). Information about the regional teams is available from the FiDASt homepage (see footnote 3), and there is a discussion list.

From time to time members of FiDASt meet to discuss technical questions and research papers. The first workshop *FiDASt'99* took place in Hannover in October 1999, and revised versions of the papers presented are published in Schasse and Wagner (1999).

5. Teachers' corner

Readers who teach applied courses where students are engaged in performing their own empirical investigations are aware of the difficulties to get hold of "real life" micro data for use by students. While public use micro data (and even panel data) for individuals and households are easily available, firm level data for use by students are scarce, and panel data are even more scarce.

Given that the firm panel data from official statistics are strictly confidential, how can one expect that they can be used by students in a PC lab? One cannot, of course, and using them via the routes described in section 3 for this purpose is impossible, too. However, there are ways to work around this. Statistical offices release small artificial samples of the firm panels to research teams to allow team members to test their programs in their offices before submitting them for runs using the original panels. The test data are

[5] Remember that the establishment level data from each federal state are collected by and kept inside the statistical office of this federal state. Furthermore, users of the data should be members of a university or a non-profit research institute located in the federal state.

manipulated to prevent identification of the firms by changing establishment, enterprise, location and industry identifiers to nonsense numbers, but the figures reported for the other variables included are the original figures. The test data, therefore, are true firm panel data, and they can be used for empirical exercises. To give an example, I use a sample for 100 establishments that form my test data from the *all firms panel* of Lower Saxony in my course on empirical small business economics to let students test the validity of Gibrat's Law. Readers interested in using such test data in their own courses should contact a FiDASt team in their federal state.

6. A look ahead

Where do we go from here? Given that research teams working with firm panel data from official statistics or preparing to do so in the near future are active in nearly every federal state, but not all over Germany, the next step must be to fill the gaps. If we succeed in this, we can think about a way to link the panels across federal states, and to make these linked panels available in all federal states. Only then will we be able to investigate topics related to relocation across state borders, to do research based on enterprise data (made from data of all establishments belonging to an enterprise), and to perform comparisons across regions in one place. Readers sharing this vision are invited to help in making it reality.

References

Bernard, Andrew B. / *Wagner,* Joachim (1997): Exports and Success in German Manufacturing, Weltwirtschaftliches Archiv *133,* 134 – 157.

Breitung, Joerg (1992): Dynamische Modelle für die Paneldatenanalyse, Frankfurt am Main: HAAG + HERCHEN.

Caves, Richard E. (1998): Industrial Organization and New Findings on the Turnover and Mobility of Firms, Journal of Economic Literature *XXXVI,* 1947 – 1982.

Gerlach, Knut / *Wagner,* Joachim (1993): Gross and Net Employment Flows in Manufacturing Industries, Zeitschrift für Wirtschafts- und Sozialwissenschaften *113,* 17 – 28.

– (1997), Analysen zur Nachfrageseite des Arbeitsmarktes mit Betriebspaneldaten aus Erhebungen der amtlichen Industriestatistik – Ein Überblick über Ansätze und Ergebnisse für niedersächsische Industriebetriebe, in: Jürgen Kühl, Manfred Lahner und Joachim Wagner (Hrsg.), Die Nachfrageseite des Arbeitsmarktes – Ergebnisse aus Analysen mit deutschen Firmenpaneldaten (Beiträge zur Arbeitsmarkt- und Berufsforschung BeitrAB 204), Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung der Bundesanstalt für Arbeit, 11 – 82.

Methner, Eckhard (1992): Das Erhebungsprogramm der amtlichen Statistik im Bereich des Produzierenden Gewerbes, in: Rainer Ertel und Joachim Wagner (Hrsg.), Produzieren in Niedersachsen – Empirische Untersuchungen mit Betriebsdaten (NIW-Vortragsreihe Band 6), Hannover: Niedersächsisches Institut für Wirtschaftsforschung, 13 – 29.

Motohashi, Kazuyuki (1998): Institutional Arrangements for Access to Confidential Microlevel Data in OECD Countries, mimeo, Paris: OECD.

Schasse, Ulrich / *Wagner,* Joachim (Ed.) (1999): Entwicklung von Arbeitsplätzen, Exporten und Produktivität im interregionalen Vergleich – Empirische Untersuchungen mit Betriebspaneldaten (NIW-Vortragsreihe, Band 13), Hannover: Niedersächsisches Institut für Wirtschaftsforschung.

Schmidt, Elke Maria (1995), Betriebsgrösse, Beschäftigtenentwicklung und Entlohnung (Studien zur Arbeitsmarktforschung, Band 9), Frankfurt / New York: Campus.

Statistisches Bundesamt (Ed.) (1999), Möglichkeiten einer wissenschaftlichen Nutzung von Unternehmensdaten aus der amtlichen Statistik (Spektrum der Bundesstatistik Band 14), Stuttgart: Metzler-Poeschel.

Wagner, Joachim (1994), The Post-Entry Performance of New Small Firms in German Manufacturing Industries, Journal of Industrial Economics *XLII,* 141 – 154.