Official Firm Data – New Options For Analyzing Firm Data On Micro Level

Malchin, Anja
State Statistical Institute Berlin-Brandenburg, Research Data Centre
Alt-Friedrichsfelde 60
10315 Berlin, Germany
E-mail: anja.malchin@statistik-bbb.de

Höninger, Julia
State Statistical Institute Berlin-Brandenburg, Research Data Centre
Alt-Friedrichsfelde 60
10315 Berlin, Germany
E-mail: julia.hoeninger@statistik-bbb.de

1 Introduction

To facilitate access to official microdata and to develop new options to analyze the great treasure of official statistics have been two of the main tasks of the research data centers (RDCs) of the German statistical offices since their founding about ten years ago. In this period microdata concerning social and economical issues, agricultural information as well as environmental and tax-related data from official statistics have been compiled and edited to create a huge data pool to the benefit of the scientific community. Microdata from official statistics cannot be made available for everybody because of the strict legal regulations about data transmission in Germany. According to the German Federal Statistics Law microdata have to be kept confidential, meaning that re-identification of respondents is not possible. Nevertheless under certain conditions the scientific community is granted a privileged access for scientific purposes (Zühlke et al. 2007).

While in the first years the work of the RDCs focused on compiling certain frequently requested statistics, over time the demands of the scientific community shifted and new challenges for the regional offices of the RDCs arose. Growing expectations especially in the field of research in economics led to further enlargement of the research potential of the micro data. Particularly since the amendment of the federal statistics law in the year 2005, which allows the combination of economic and environmental microdata from different statistical sources, new research possibilities could be offered (Malchin / Pohl 2007).

In the following the project AFiD will be presented. In this project several new combined data sets have been created. The research potential of these new firm- and enterprise level data will be briefly discussed, leading to the question if the demands of the scientific community for more detailed information have really been met. To assess this question, the RDCs of the German Statistical offices held a workshop in April 2010. It addressed scientists who had already worked with those new data sets and were willing to share their findings in public. Some examples of enterprise- and firm-related research and their results are to be found later on in this paper, followed by some concluding remarks.

2 Project AFiD (Official Firm Data in Germany)

In the first years of the RDCs a wide variety of data sets of units and enterprises[1] has been provided for

---

[1]   A unit or establishment is being defined as the smallest local unit, an enterprise being the smallest legal entity. When meaning data sets for establishments and/or enterprises, in the following they are referred to as firm data.

the scientific community. Data consisted mostly of cross-sectional data for individual branches. After the legal frame was changed, some pilot projects tested if the combination of micro data from different statistics was a promising way to gain more detailed information about the behavior of firms in Germany. Due to the positive results and noticing an ever growing interest for combined firm data by the scientific community, it was planned to set up standardized data sets during the course of the project AFiD[2] (Brandt et al. 2008).

In the course of the project AFiD several statistics have been matched, always trying to keep the level of investigation in the data. Also the sampling design (full survey or sample survey) had to be considered. Unique business numbers mostly stemming from the German business register served as direct identifiers when matching the different micro data to either form new cross-sectional data from different surveys covering similar groups of respondents or longitudinal panel data (Malchin / Voshage 2009). Ten panel data sets were created, which can further be supplemented by modules, always according to the specific research questions.

2.1 Agriculture

For the new panel A*griculture* data from the census of agriculture as well as the census of agricultural structure have been merged. Altogether three waves are included in the data set, giving researchers the opportunity to observe the reactions of Germany's agricultural firms to the changing circumstances like liberalization and opening of new of markets. A second version includes on top of the complete counts also two representative samples of agricultural holdings, each consisting at most of 100.000 agricultural holdings, while each of the agricultural censuses covers about 400.000 farms. The new panel data set *Agriculture* has considerably improved the information potential of the current agricultural statistics and is a great asset especially for historic agricultural research.

2.2 Industry

In the field of manufacturing two data sets were created. Units and enterprises in this sector have to respond to a bundle of alike surveys, which could be combined to two panels on the two investigation levels. The panel *Industrial Enterprises* covers information from the annual report, the investment survey and the cost structure survey of the years 2001 to 2008. It is suitable for analyses of structural changes in the German industrial sector over time. On the level of the local unit the panel *Industrial Units* covers information from the establishment panel (Konold 2007) combined with information from the quarterly production survey. Since it includes a broad variety of economic information as well as a long time span (1995 to 2008), this panel is a multifunctional data set for all research questions on the level of the single production unit. The information in both industrial data sets can be further enhanced by environmental modules or modules with information about energy use or earnings.

2.3 Services

The survey in the services sector has been implemented in Germany for the first time in the year 2000. The new AFiD panel *Services* covers the years 2003 to 2007. This is on one hand due to qualitative considerations (especially in the first years) and on the other hand due to the sampling procedure which only allows comparative research for the years mentioned above. For the reporting year 2008 the sample as well as the sampling design have changed greatly. The data set includes information on the enterprise level and is suitable e.g. for analyses of growth over time or export behavior. More information about the variables included and the research potential of this survey can be found in Vogel 2009.

---

[2] AFiD is the German acronym for "Amtliche Firmendaten in Deutschland" ("Official Firm Data in Germany").

2.4 Energy

For analyses in the field of Electricity, gas and water supply two panel data sets have been created. The panel *Energy Enterprises* combines information from the cost structure survey and the investment survey, both being census surveys with a cut-off limit, available for the years 2003 to 2007. There is also a longitudinal data set *Energy Units* constructed by integrating the monthly report and the investment survey for establishments in energy and water supply, both surveys having the same group of respondents on the level of the local unit. With both data sets the effects of the changing global and national conditions on the national units and enterprises can be analyzed and give useful input for political strategies for an efficient energy policy and the according strategic measures.

More detailed information to all existing panel data (including the above not mentioned panels *Trade*, *Hotel and Restaurant Industry* and *Business Register*) can be found in Gramlich et al. (2009). The creation of the new AFiD panels (by linkage of cross-sectional data and/ or by combination of microdata from various branches over time) led to an enormous gain of information compared to the separate cross-sectional data sets. Research projects as discussed in the following chapter could not have been realized having only access to the traditional data.

3 Research potential of the new data sets

The RDCs of the German statistical offices have established these new AFiD-panels as a reaction to frequent requests from the scientific community for merged data sets. The demand for such kind of detailed micro data was obvious, but have the new options to analyze firm data on a micro level really expanded the knowledge about firm's competitiveness and entrepreneurship? Is the result of a complex and time-consuming procedure of merging data sets really meeting the demand of the scientific community? To assess the benefits of the newly created data sets in May 2010 the RDC of the Statistical offices of the German Länder held a workshop for the scientific community already using these new longitudinal data and asked to share their results. More than 20 abstracts were submitted as a reaction to the call for paper. After a selection procedure conducted by a committee consisting of both members of science and official statistics 11 projects were chosen to be presented at the workshop in Berlin. The selected papers represented the topics industry, services, agriculture and energy and water. Additionally projects were presented for which the AFiD-data had been merged to external data sources for further enhancement of the information potential.

3.1 Industry

Wondering why in Germany almost 40 percent of all manufacturing enterprises with more than 20 employees are one-product firms, one research question was if and in which way product differentiation affects the profitability of firms. Both positive and negative relations are possible to think of. Having analyzed the AFiD-Panel *Industrial Units* Braakmann and Wagner (2009) were able to show that firm performance in Germany significantly decreases with an increase in product differentiation, when controlled for observed and unobserved heterogeneity. Obviously concentrations on a core markets pays well for the firms.

Schiersch and Schmidt-Ehmcke (2011) tested indicators often used by the Monopoly Commission to measure the degree of competition in certain branches. All applied indicators (e.g. Herfindahl-Hirschmann-Index, Lerner-Index) have some sort of shortcomings, so the researchers of the DIW Berlin tried to find out the quality of a different indicator: the Boone-indicator. With access to the panel *Industrial Units* the indicator was tested before and after known breakdowns of cartels in the cement industry, the power-cable industry and the industry of ready-mixed concrete. If the Boone indicator was a good quality criterion the competition in the market should considerably increase after the break-up of the cartels. But in spite of the given theoretical advantages of the Boone-indicator compared to the traditional indicators, it showed an unexpected decrease in competition intensity. For this reason empirical usability can be doubted. Considering

the research status quo the Lerner-Index (Price-Cost-Margin) seems to be more reliable. Only new research or the breaking-up of a new cartel would affirm or disprove the theoretical disadvantages of the Lerner-Index.

3.2 Services

Export behavior of firms in the industrial sector has been thoroughly investigated over the last years (e.g. see Wagner 2010). Since the structure survey in the services sector is quite a young statistic, only a few analyses to firm's behavior have been conducted yet. But for quite some time also the service sector has become a great asset in the national economy and similar to manufacturing the international export of services has increased considerably. Following comprehensive studies of the structural change in the service sector Eickelpasch and Vogel (2010) used the AFiD-panel *Services* to find answers to the questions why some enterprises start exporting while others do not, which factors are determining the export decision of service sector firms or if the same determinants found out for manufacturing firms are important in the service sector. Having included no control for unobserved heterogeneity the results from the studies in the manufacturing industry could be reproduced for the service sector: The size of the enterprise, a high share of human capital and a high level of productivity characterized the firms more likely to export. But with control for unobserved heterogeneity those clear result seem to blur. Human capital and the productivity level rather seem to be correlated to the unobserved time-constant firm characteristics. The result that only the size of the enterprise has a significantly positive influence on export performance is quite surprising.

3.3 Agriculture

Also in the field of agriculture complex analyses were conducted using the new AFiD-data. One project chosen to be presented at the workshop in Berlin addressed the determinants of agricultural businesses closing down, including firm characteristics like size, region and line of business.

In the other presented project possible determinants of the ongoing structural change in the agricultural sector were illustrated. A considerable number of research projects have already been trying to find explanations for the observed changes, taking in account various firm characteristics, but with inconsistent results. Röder and Kilian (2011) not only took advantage of the new AFiD-panel *Agriculture* of the years 1999 to 2007, but were also able to examine demographic and environmental variables from other data sources to shape so-called "marginal regions" to take a closer look at. They found out that above all other characteristics firm size, value added and the special line of business in keeping of animals are the parameters which determine whether the number of agricultural enterprises in a region is changing. Strikingly indeed are the big regional differences with respect to the speed of the structural changes as well as the differing relevance of the determinants across regions.

3.4 Energy and water

As mentioned before, the new AFiD-panels *Industrial Enterprises* and *Industrial Units* can be further enlarged by several modules, increasing the information potential of the official microdata even more. One of the modules contains detailed information about energy use, leading to a study analyzing almost 80.000 German industrial enterprises covering a period of 12 years to describe the determinants of changes in energy intensity, carbon intensity and CO_2 emissions in the German industry (Petrick et al. 2011). The study resumes that the increase in economic activities over the years did not lead to an equivalent increase in emissions and that companies' energy efficiency rather depends on firms' behavior in the business cycle than on the use of certain technologies.

The industrial sector is also a main user of water resources and uses water as a production factor, as water is being used in almost every production step of manufactured goods. Embedded in the ongoing project GLOWA-Danube researchers from Munich Ifo Institute of Economic Research presented their analyses partly conducted with the AFiD-panel *Industrial Units* in combination with the module *Water*

supply and waste water disposal. In this project the effects of climate change on industry as well as the effectiveness of different adaption strategies are being analyzed (Jeßberger et al. 2011). After the framework had already been tested with macro data, the new micro data sets gave the opportunity to perform more detailed analyses and simulations, which amongst others allowed a quantification of the effects of policy measures. Interesting is that the anticipated effect of climate change is outstanding the effect of any socio-economic scenario.

3.5 Combination of AFiD and external data

One frequent request by the scientific community was the wish for the combination of the official microdata with data from other sources. These external data only might be merged with official microdata when coming from a publicly available source. Even results from self-conducted analyses must not be combined with official data in Germany, except they are made public on the internet or in a paper. The range of external variables already been used by RDC-projects spreads from a simple unemployment rate over prize indices to complex data from commercial data providers.

Data of the latter kind was analyzed in a project trying to find out the economic consequences of one-third co-determination in German supervisory boards. The data for the project originated from the AFiD-Panel *Services* in combination with data from the Hoppenstedt Database. This data contains information about co-determination in German companies with more than 200 employees or a certain sales volume per year. In combination with the information from the survey in the services sector the performance of enterprises with and without a supervisory board was compared. The existence of co-determination could theoretically either lead to increasing or decreasing productivity and profitability. Having conducted various empirical investigations Boneberg (2010) showed that - against all fears of supporters of the property right theory - there is no negative effect of co-determination on the productivity and profitability of an enterprise. The results of the study imply that the discussion about co-determination regulations in Germany is quite overrated.

For more detailed information about the contents, the statistics included and the research potential of all AFiD-data see Wagner 2009.

4 Concluding remarks

The variety of already conducted research projects using the newly created AFiD-data shows that there is a need for highly informative data on the micro level. To understand economic structures and developments it is important to base political decisions on profound knowledge about regional, economic and time-related contexts. Through the establishment of this flexible data management concept an extensive treasure of data about economic units and enterprises in Germany was opened to the scientific community by official statistics. One of the frequently expressed demands of the scientific community has been met already: With the new data now enterprises changing their line of business can be looked at closely to find out the determinants of their economic decisions.

Nevertheless the common impression received by the staff of the RDCs is that the scientific community is always evolving and requesting still more detailed information, more merged data to take advantage of the research potential of different statistics and of course more waves in the data to conduct long-time-research. To verify this impression and maybe to tackle even more specific research questions, at the end of the AFiD-workshop in Berlin the participants engaged in a discussion about their ideas for further enlargement of the information potential of official data for scientific purposes. Quite a lot of interesting and groundbreaking ideas were expressed, being primarily the wish for merged data from statistics which allow international comparisons, data about patents-assignments or foreign trade information merged to the already existing official micro data in Germany as well as the combination of official statistics and geographic information.

Since the RDCs of the German statistical offices have to strictly follow the legal regulations, some of the expressed suggestions are not realizable, but of course the full potential of the already existing data treasure ought to be exploited. If possible and corresponding with legal framework, the staff of the RDCs is glad to satisfy the wishes of the scientific community for a broader basis of data with an even higher data content. Suggestions for further enhancement are always welcome.

REFERENCES

Boneberg, F. (2010): The Economic Consequences of one-third Co-Determination in German Supervisory Boards: First Evidence for the Service Sector from a New Source of Enterprise Data. University of Lüneburg, Working Paper Series in Economics No. 177.

Braakmann, N. / Wagner, J. (2009): Product Differentiation and Profitability in German Manufacturing Firms. University of Lüneburg, Working Paper Series in Economics No. 115.

Brandt, M. / Oberschachtsiek, D. / Pohl, R. (2008): Neue Datenangebote in den Forschungsdatenzentren. Betriebs- und Unternehmensdaten im Längsschnitt, AStA Wirtschafts- und Sozialstatistisches Archiv, 3, 193-207.

Eickelpasch, A. / Vogel, A. (2010): Determinants of the Export Behaviour of German Business Services Companies. The Service Industries Journal, First published on: 28 September 2010 (iFirst), 1-14.

Gramlich, T. / Bachteler, T. / Schimpl-Neimanns, B. / Schnell, R. (2009): Panelerhebungen der amtlichen Statistik als Datenquellen für die Sozialwissenschaften. RatSWD Working Paper Series No. 132.

Jeßberger, C. / Sindram, M. / Zimmer, M. (2011): Global Warming Induced Water-Cycle Changes and Industrial Production – A Scenario Analysis for the Upper Danube River Basin. Journal of Economics and Statistics 231(3), 415-439 (forthcoming).

Konold, M. (2007): New possibilities for economic research through integration of establishment-level panel data of German official statistics, Schmollers Jahrbuch / Journal of Applied Social Science Studies 127 (2), 321-334.

Malchin, A. / Pohl, R. (2007): Firmendaten der amtlichen Statistik ‒ Datenzugang und neue Entwicklungen im Forschungsdatenzentrum. Vierteljahrsheft des DIW Berlin 3/2007, 8-16.

Petrick, S. / Rehdanz, K. / Wagner, U. (2011): Energy Use Patterns in German Industry: Evidence from Plant-level Data. Journal of Economics and Statistics 231(3), 379-414 (forthcoming).

Röder, N. / Kilian, S. (2011): Which Parameters Determine the Development of Farm Numbers in Germany? Journal of Economics and Statistics 231(3), 358-378 (forthcoming).

Schiersch, A. / Schmidt-Ehmcke, J. (2011): Is the Boone-Indicator Applicable? Evidence from a Combined Data Set of German Manufacturing Enterprises. Journal of Economics and Statistics 231(3), 336-357 (forthcoming).

Vogel, A. (2009): The German Business Services Statistics Panel 2003 to 2007. Schmollers Jahrbuch / Journal of Applied Social Science Studies 129 (3), 515-522.

Wagner, J. (2010): Exports and Firm Characteristics in German Manufacturing Industries. University of Lüneburg, Working Paper Series in Economics No. 188.

Wagner, J. (2009): The Research Potential of New Types of Enterprise Data based on Surveys from Official Statistics in Germany. Schmollers Jahrbuch / Journal of Applied Social Science Studies 130, 133-142.

Zühlke, S. / Christians, H. / Cramer, K. (2007): Das Forschungsdatenzentrum der Statistischen Landesämter – eine Serviceeinrichtung für die Wissenschaft. AStA Wirtschafts- und Sozialstatistisches Archiv 3-4, 169-178.