

Linkage of Ifo Survey and Balance-Sheet Data: The EBDC Business Expectations Panel & the EBDC Business Investment Panel

By Anja Hönig

1. Introduction

The Economics and Business Data Center (EBDC), founded as a cooperation of the University of Munich (LMU) and the Ifo Institute for Economic Research in 2008, has the aim of opening new fields of economic research by providing innovative datasets of German companies. These datasets contain part of the survey data of the Ifo Institute as well as external balance-sheet and financial statement data. More precisely, in a first step, micro data from the Ifo Business Survey and from the Ifo Investment Survey has been linked with balance-sheet data from the enterprise databases Amadeus and Hoppenstedt. As a result, we can offer two extensive datasets of German companies that allow for the simultaneous research of rather qualitative aspects contained in the Ifo micro data together with quantitative accounting or structural company information: the EBDC Business Expectations Panel (focus: cyclical economic factors/balance-sheet data) and the EBDC Business Investment Panel (focus: investment/balance-sheet data). Both company panels contain historical as well as current data and are updated regularly. In general, because of the high levels of confidentiality and data security the Ifo Institute ensures its panel members, the EBDC company panels can only be used for research purposes on the premises of the Ifo Institute. Furthermore, the data is only provided in an anonymized way and with a one-year time lag. Its use is subject to strict security precautions. The aim of this paper is to give an overview on the data sources, and to describe the scope of and access to the EBDC data. Furthermore, it describes some research projects taking advantage of these datasets up till now.

2. Data Sources in Short

To generate the EBDC datasets micro data from the Ifo Business Survey for manufacturing and the Ifo Investment Survey has been linked with external balance-sheet data from the enterprise databases Amadeus and Hoppenstedt.

Researchers are offered four regularly conducted standard enterprise surveys at the Ifo Institute: the *Ifo Business Survey* (KT), the *Ifo Investment Survey* (IT), the *Ifo Innovation Survey* (INNO) and the *Ifo World Economic Survey* (WES). The KT, which is conducted monthly, concentrates on enterprise-specific appraisals and expectations concerning business as well as market conditions and is the basis for the monthly published Ifo Business Climate Index which is of great interest in the public debate in Germany. The Ifo Business Survey is structured around four sectors: manufacturing (KT VG), wholesaling/retailing (KT HAN), construction (KT BAU) and service providers (KT DVDL). In each sector, the questionnaires contain standard monthly as well as periodically recurring special questions and relate to a product or product line (KT VG, KT DVDL) or to a sector or business field (KT HAN, KT BAU), respectively. The Ifo Investment Survey, on the other hand, focuses on investment behaviour/investment activities and is conducted twice a year – in spring and autumn. It refers to the company as a whole with each company being assigned to its branch of production.

Finally, researchers can take advantage of the Ifo Innovation Survey (annual) which asks for companies' innovation activities and the WES (quarterly) relating to the international outlook for economic activity. In the business surveys of the Ifo Institute, the participants are generally not requested to provide absolute or monetary figures, but to answer so-called "appraisal questions" on current business conditions, production constraints, the expected development of consumer demand, investment and innovation activities, etc., which implicitly mirror a company's actual and expected economic situation. More detailed information on the different surveys, i.e. the topics of questions posed and numerous studies based on the different surveys can be found in Becker/Wohlrahe (2008).

In addition to the Ifo survey data, there are two other sources for the EBDC datasets: the *Amadeus company database*, which contains business and financial information on more than 11 million, mainly non-quoted enterprises, from 41 countries in Europe¹, and the *Hoppenstedt database*, which gives detailed financial statement information on German companies. The key source for the Amadeus database is the MARKUS Database, containing business information of companies in the German Commercial Register with a bank credit index of a maximum of 499 (Creditreform Association) and the DAFNE database including annual accounts, investment data, etc. of all disclosing German firms (Creditreform Rating AG). Unlike the DAFNE database (raw data format), however, the data in Amadeus are in a homogeneous, standardised accounting format based on generalised national and/or international account-

¹ Amadeus is a product of the Bureau van Dijk Electronic Publishing GmbH (BvDEP), one of the leading European providers of global enterprise information. For more information on the database, see the BvDEP webpage.

ing rules. Every enterprise report consists of a total of 23 accounting items, 25 positions of the financial statements, 20 key finance figures and numerous descriptive information such as industry codes, partnership structures, stocks and stock price information. For the enterprises in the EBDC panels, more than 50 positions have been selected, but initially not including partnership, stock and stock price information. The Hoppenstedt Accounting Database², which is the second source of financial information, currently contains more than 2.7 million closing statements from more than 1 million German enterprises in the areas of manufacturing, distribution, services, insurance and banks. Almost all final statements published since 2005 are registered here and historical information for large firms even dates back, in part, to 1987. The collected data on accounts and financial statements of individual companies are accessible in varying levels of detail (norm accounting: maximal available positions according to the respective accounting regulations; abridged accounting: ca. 90; short accounting: ca. 30 positions). Moreover, to account for the different types of final statements, separate accounting schemes closely oriented on the respective original were developed for HGB, IAS and US-GAAP.

3. Construction of the EBDC Datasets

To link the different data sources, we used the company specific address information contained in each database to generate two allocation tables (Ifo-Amadeus and Ifo-Hoppenstedt). These were subsequently combined to construct the EBDC Business Expectations Panel and, on the other hand, the EBDC Business Investment Panel. Linkage is done by using the matching software MTB (Merge Toolbox) developed at the Center for Quantitative Methods and Survey Research of the University of Konstanz. MTB is especially useful when there is no unambiguous key, e.g. a company code number, in the datasets. The method used is called Probabilistic Record Linkage and takes recourse on the theory of Newcombe et al. (1959), formalised by Fellegi/Sunter (1969). The idea here is to look for similar name/address data and link them by determining probabilities for the degree of agreement (so-called “similarity”) between the variables.

With regard to the EBDC Business Expectations Panel, by this means, the number of linked companies between the Amadeus database and the Ifo Business Survey amounted to 8,915 firms or, 3,819 for the sub-group of manufacturing. In comparison, the linkage of the Hoppenstedt firm database with the Ifo Business Survey resulted in a total of 4,377 enterprises, of which 2,454 companies are allocated to manufacturing. Both linkages were subsequently

² It is a product of the Hoppenstedt Business Information GmbH, one of the leading providers of business and industry information in Germany and which is part of the Hoppenstedt Group.

combined to get an allocation table containing, for each company, the key variable from the Ifo KT VG (Ifo-ID) as well as the keys of the respective accounting database (Amadeus-ID and/or Hoppenstedt-ID). These keys are necessary to append the really interesting business and financial data.

In the end, it amounts to 8,881 allocations (IDs) in the *EBDC Business Expectations Panel* and to 4,656 companies in the *EBDC Business Investment Panel*. In general, if balance-sheet information was available both from Amadeus and Hoppenstedt, preference was given to the latter as the Hoppenstedt database provides a more extensive variable selection and deliverance.

4. Structure of the EBDC Company Panels

The EBDC Business Expectations Panel and the EBDC Business Investment Panel are similar in their structure: Identification is possible through a company-specific ID, the respective year of observation and further dataset-specific time variables (like, e.g., the survey month). In general, when available, the EBDC company panels contain information from individual instead of corporate group accounts, however, the accounting schemes from the original databases were not taken over. Instead, we developed a new EBDC scheme which integrates both Amadeus as well as Hoppenstedt variables and abstracts from the existing differences of the two databases. Specifically, the EBDC accounting scheme is based on the accounting and earnings-statement structure of the German Commercial Code (HGB) and in part also contains variables according to total or turnover cost procedures. For a detailed and correspondingly structured overview of the available accounting and earnings-statement variables, recourse can be taken to the list of variables of the respective panel (both available on the webpage of the EBDC) which also contains a detailed description of the questions posed, their time horizon, the survey frequency, etc.

5. EBDC Business Expectations Panel

The EBDC Business Expectations Panel is a combination of monthly (Ifo) and annual (balance-sheet) data covering the period from 1980 to the end of 2009. Both balance-sheet and Ifo information in the same year is available for 8,318 enterprises. As the Ifo Business Survey is conducted monthly, each month contains the survey results from the Ifo KT resulting in up to 12 reports per product/company in each year. The survey information is then followed by the accounting information from the company databases. Sorted according to their function, the following variables are included in the columns of the EBDC Business Expectations Panel: identification variables, balance sheet and profit/loss account informations as well as Ifo variables referring to the questions on economic conditions. In addition to ID, year and month, there are

further variables of identification like, for example, the industry code, company size, German federal state, stock exchange quotation and legal form.³ Like the Ifo Business Survey for manufacturing, the Ifo Innovation Survey (INNO) refers to the products a company produces. Moreover, the participants constitute a fraction of those of the Ifo Business Survey, making the Innovation survey a suitable add-on for the EBDC Business Expectations Panel. The questions in the INNO focus on innovation activities, aims of innovative developments and factors influencing the innovation process. Therefore, it may be of special interest for various research projects. In general, the firms are asked to provide answers for product and process innovations with special annual questions addressing different topics. Another add-on to the EBDC Business Expectations Panel is the Special Question on Innovation which is posed once a year in December in the frame of the Ifo Business Survey. As regards content, it refers to the market perspectives of a specific product (growing, stagnating, decreasing) and its status of innovation / phase of development.

6. EBDC Business Investment Panel

Due to the semi-annual rhythm of the Ifo Investment Survey, the structure of the EBDC Business Investment Panel is slightly different. Instead of a monthly time indicator, there is a variable called “survey_base” that indicates if observations are from the spring survey, the autumn survey or from a balance-sheet. Currently, the dataset contains data from 1987 to 2008/2009 and is sorted by “EBDC_ID”, “year” and “survey_base”. The format in Stata is “long” such that each observation can be identified through these variables. Other firm identifiers are industry codes, federal state information, legal form, etc.⁴ Again, listed in the different columns, are the financial statement variables from the firm databases and the variables of the Ifo Investment Survey which refer to investment activities, obstacles to investment, changes in the planned production program, etc. In total, the EBDC Business Investment Panel contains 4,656 allocations (IDs) with about 21,000 balance-sheets and more than 40,000 spring and autumn observations, respectively. For 12,200 balance-sheets there is also information from the Investment Survey in the same year. A detailed evaluation of the EBDC Business Investment Panel including the list of variables can be obtained at the EBDC.

³ Due to anonymization, the federal state information for very large firms (> 10.000 employees) is deleted.

⁴ To guarantee the anonymity of firms, federal state information for very large companies (> 10.000 employees) was deleted.

7. Research Questions

There are, of course, lots of different research questions that can be tackled with the data and which cannot be mentioned here extensively. Some work in progress, however, deals with the question on how taxes, captured by the cost of capital, influence the investment activities of firms or how federal and local taxes in Germany impact capital structure choices. Büttner/Fuest (2010) analyze if the corporate income tax acts like an automatic stabilizer in the business cycle and show that there is a stabilization effect which tends to increase during cyclical downturns. Further studies include innovation activities and their reaction on capital market constraints, and it also interest, which parameters can be used to identify credit constrained firms. Rottman/Wollmershäuser (2010) analyze if there has actually been a credit crunch during the recent economic crisis and other authors test if trade credits have somehow compensated for constraints with regard to bank loans. Finally, some researchers try to use the information from the firm's balance sheets and from the Innovation survey to study the impact of taxation on innovation and growth opportunities of firms or to find out, where firms issue debt (see Eidenmüller/Engert/Hornuf, 2010). In general, not only the Ifo Business Survey but also the Ifo Investment and the Ifo Innovation Survey have been the subject of numerous studies and research projects. These are described in detail by Becker/Wohlrabe (2008), with some new ones being the recently published work by Seiler (2010), Bachmann/Elstner/Sims (2010), or Fidrmuc/Hainz (2009). In general, if you are interested in doing some project with the EBDC data, please feel free to contact the team. In short, the access modalities are the following.

8. Access & Service

Like the EBDC Business Expectations Panel, the EBDC Business Investment Panel is updated regularly with a one-year time-lag due to anonymity reasons. Updating is done both with respect to time and base, as further companies that have recently been added to the respective address databases will be integrated in the EBDC panels as well. The EBDC sees itself as a service provider that supports research projects of professors, visiting researchers and doctoral students by providing, among other things, the EBDC Business Expectations Panel as well as the EBDC Business Investment Panel. In general, research projects must be non-commercial, high-level projects in economics that can be empirically analyzed using the EBDC data. Due to the high confidentiality and the obligation to maintain the secrecy of survey results as well as panel member identity, the EBDC company panels can only be used on the premises of the EBDC and are made available with a time lag. We provide a computer without access to the Internet, a printer or other external storage media and which can only be used in the presence of an EBDC staff member.

This person will ensure, on completion of the researcher's stay, that the anonymised data cannot be used to identify individual firms and that no inferences can be made regarding the panel composition. Moreover, after this examination has been successfully carried out, he/she will send the results in a Stata format. Access to the EBDC datasets can be applied for using a form at the Ifo Website.⁵ In addition, a short description of the research project and accompanying information as to scheduling must be submitted. Upon request the EBDC will send by e-mail a test package containing an anonymized EBDC test panel in Stata-format as well as the documentation on the respective original dataset. The EBDC expressly supports empirical research projects and is thus free of charge. Access to the EBDC data only depends on the availability of workplaces.

References

- Bachmann, R./Elstner, S./Sims, E.* (2010): Uncertainty and Economic Activity: Evidence from Business Survey Data, NBER Working Paper 16143, 2010.
- Becker, S. O./Wohlrabe, K.* (2008): Micro Data at the Ifo Institute for Economic Research: The "Ifo Business Survey", Usage and Access, *Schmollers Jahrbuch – Journal of Applied Social Science Studies* 128(2), 307–319.
- Büttner, T./Fuest, C.* (2010): The Role of the Corporate Income Tax as an Automatic Stabilizer, *International Tax and Public Finance* 17, 686–698.
- Eidenmüller, H./Engert, A./Hornuf, L.* (2010): Where Do Firms Issue Debt? – An Empirical Analysis of Issuer Location and Regulatory Competition in Europe, ECGI – Finance Working Paper No. 292/2010.
- Fellegi, I. P./Sunter, A. B.* (1969): A Theory for Record Linkage, *Journal of the American Statistical Association* 64, 1183–1210.
- Fidrmuc, J./Hainz, C.* (2009): Financial Integration and Banking Regulation: Cross-Border Lending in Germany, "Paolo Baffi" Centre Research Paper Series No. 2009–46.
- Newcombe, H. B./Kennedy, J. M./Axford, S. J./James, A. P.* (1959): Automatic Linkage of Vital Records, *Science* 130, 954–59.
- Rottmann, H./Wollmershäuser, T.* (2010): A Micro Data Approach to the Identification of Credit Crunches, CESifo Working Paper No. 3159.
- Seiler, C.* (2010): Dynamic Modelling of Nonresponse in Business Surveys, Ifo Working Paper No. 93.

Economics and Business Data Center

The tasks of the EBDC also include the procurement and administration of data sources for research and teaching, the central provision, updating and documentation of external databases, as well as the acquisition of corresponding support tools. Moreover,

⁵ http://www.cesifo-group.de/portal/page/portal/ifoContent/N/data/EBDC_Container/EBDC_Angebot_Container/Formal_obligation_EBDC_panels.pdf

the EBDC provides a suitable hard- and software technical infrastructure and offers support with regard to software-specific knowledge transfer. Within the framework of the Ifo Datapool, it offers access to the survey data of the Ifo Institute, which has conducted regular surveys throughout Germany since 1949. Thus, interested researchers and students working on empirical projects may profit from synergy and efficiency effects.