



EBDC Business Expectations Panel  
&  
EBDC Business Investment Panel

Linkage of  
Ifo Survey and Balance-Sheet Data

23.12.2010

Economics and Business Data Center

Ifo Institute for Economic Research  
Poschingerstr. 5  
81679 Munich

Anja Hönig, Heike Mittelmeier, Andreas Neudecker

## **Abstract**

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, aims at opening new fields of economic research by providing innovative datasets of German companies. These datasets contain both survey data of the Ifo Institute as well as external balance-sheet data from the company databases Amadeus and Hoppenstedt. Currently, researchers may use the EBDC Business Expectations Panel (focus: cyclical economic factors/ balance-sheet data) as well as the EBDC Business Investment Panel (focus: investment/ balance-sheet data) which will both be continually updated as regards time and content. 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 anonymised way and with a one-year time lag and its use is subject to strict security precautions. The aim of this paper is to give an overview of the data sources and to describe the scope of and access to the EBDC data. Furthermore, it also provides information on the matching method of probabilistic record linkage.

## 1. Introduction

The LMU Economics and Business Data Center (EBDC) was founded as a cooperation of the economics and business administration faculties of the University of Munich (LMU) and the Ifo Institute for Economic Research at the beginning of 2008. It receives funding from *LMUexcellent*, this being a project within the framework of the *Exzellenzinitiative* of the German federal and state governments for the promotion of science and research at German universities.

The Economics and Business Data Center has the aim of creating innovative company panels that link part of the Ifo survey data, i.e. the Ifo Business Survey and the Ifo Investment Survey, with different external company databases containing financial statement or corporate governance data. By this means, we want to supplement the more qualitative aspects (business expectations, assessments, etc.) contained in the Ifo micro databases with accounting and structural company information and thus facilitate innovative approaches for empirical economic research. Therefore, 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.<sup>1</sup> Moreover, the EBDC provides a suitable hard- and software technical infrastructure and offers support with regard to software-specific knowledge transfer. Additionally, within the framework of the Ifo Datapool, it offers access to the survey data of the Ifo Institute,<sup>2</sup> 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.

With regard to the new EBDC enterprise datasets, in a first step, we have linked the micro data of the Ifo Business Survey, as well as from the Ifo Investment Survey 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 together with quantitative business factors. Both company panels contain historical as well as current data and will be updated

---

<sup>1</sup> Accessible external databases at the EBDC include Bankscope, Thomson OneBanker, etc...

<sup>2</sup> The Ifo Institute and the EBDC also provide access to other external micro and macro data.

regularly. Currently, the EBDC Business Expectations Panel starts in 1980, whereas the EBDC Business Investment Panel includes data from 1987 onwards.

The aim of this paper is to give interested professors, guest researchers and doctoral students a general overview and to describe the sources, scope of and access to the EBDC data. Furthermore, it also provides information on the matching method of probabilistic record linkage. To this end, Section 2 describes the data sources, i.e. the Ifo Business Survey and the Ifo Investment Survey, as well as the company databases Amadeus and Hoppenstedt. An overview of the linkage method is provided in Section 3, while Section 4 explains the structure and components of the EBDC Business Expectations Panel and the EBDC Business Investment Panel. Finally, Section 5 describes the access to the data.

## **2. Data Sources**

To generate the EBDC datasets we have linked the micro data from the Ifo Business Survey (KT) and from the Ifo Investment Survey with external balance-sheet data from the enterprise databases Amadeus and Hoppenstedt. In this section, we briefly describe the individual data sources, while in Section 3, the matching technique of probabilistic record linkage is explained.

### **2.1. Ifo Survey Data**

In general, there are 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. In contrast, the Ifo Investment Survey (semi-annual) focuses on investment behaviour, the Ifo Innovation Survey (annual) asks for companies' innovation activities and the WES (quarterly) relates 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, since experience has shown that a survey of

exact purchasing plans does not provide an accurate picture of reality. This is due to the fact that half of the purchase decisions of private households, as well as those of small and medium-sized enterprises, are spontaneous.<sup>3</sup> Therefore, “appraisal questions” such as the question on current business conditions or the demand situation have proven to be reliable since these implicitly mirror a company’s actual and expected situation. In addition, as enterprise decisions are influenced by judgements and other subjective factors, the variables also provide an outlook on the probable direction of economic activity, which is published monthly as the Ifo Business Climate. In this regard it is important that the Ifo approach is not limited to Germany, instead it is also supported by the European Commission and the OECD. Therefore, in January 2002 and as a result of the Europe-wide harmonisation of business surveys, there were also some changes in the Ifo panels. Specifically, there is no longer a distinction between survey month and questionnaire month. More detailed information on the KT sectors, the topics of questions posed and numerous studies based on this survey can be found in Becker and Wohlrabe (2008).

### **For illustration: Ifo Business Survey for manufacturing**

The Ifo Business Survey is structured around four sectors: manufacturing (KT VG), wholesaling/retailing (KT HAN), construction (KT BAU) and service providers (KT DL). In each sector, the questionnaires contain standard monthly as well as periodically recurring special questions with the questionnaires relating to a product or product line (KT VG, KT DL) or to a sector or business field (KT HAN, KT BUILDING), respectively. For this reason, large enterprises whose main sales consist of the production of several products frequently report more than one questionnaire per month. For example, the Ifo Business Survey for manufacturing focuses on questions as follows:

---

<sup>3</sup> See Goldrian (2004).

| <i>Standard questions:</i>                                                                                                               | <i>Special questions:</i>                                        |
|------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------|
| <i>Order stocks, capacity utilisation, constraints to production, competitive position (at home and abroad), employment, inventories</i> | <i>Profits, access to credit, innovations, special occasions</i> |

**Fig. 1: Contents Ifo Business Survey for Manufacturing**

In the manufacturing sector, participating companies are mostly presented with binary or ordinal scaled response categories<sup>4</sup> i.e., for each standard question there are only two (“1” yes, “2” no) or three different possible responses (“1” better, “2” the same, “3” worse) to choose from. Therefore, the company gives an assessment on its current (and anticipated) business situation and its response will also depend on the respective interpretation: e.g. the question on its “business situation” indeed refers to the economic conditions the company faces, but it is left to the firm to determine the basis on which to make this judgement.

In total, the Ifo Business Survey for manufacturing contains about 300 product groups which are defined such that they are maximally homogeneous in themselves. Specifically, in order to attain “authentic survey data”<sup>5</sup> consideration was given both to area representativeness (product variety) as well as to company representativeness (size, legal status, etc.). Moreover, since 1991 the KT VG has no structural breaks, so that there is an average of 3000 responses per month resulting in a return rate of 92%. Intense contacts with the companies maintain the panel size and composition at a representative level.

### **Ifo Investment Survey**

In the Ifo Investment Survey, companies of the manufacturing sector are requested to provide information on their investment activity. The survey is conducted twice a year, in spring and autumn, and refers to the company as a whole. Thus, each firm is assigned to the branch of its production focus. In general, there is no distinction between standard questions and special questions, however, there are different

<sup>4</sup> In some instances percentage specifications are required.

<sup>5</sup> See Goldrian (2004).

special subjects depending on the survey year, etc. which can be summarized as follows:

| <i>Regular questions:<sup>6</sup></i>                                                                                                                | <i>Questions until 2001:</i>                                                                          |
|------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------|
| <i>Total investment, investment plans this year/ last year/ next yea, aims and structure of investments, factors influencing investment activity</i> | <i>Development of production capacity, share of rented investment goods, financing of investments</i> |

**Fig. 2: Contents Ifo Investment Survey**

The Ifo Investment Survey informs about realized and planned investments and points out aims of investment activities, influencing factors, etc.. With regard to survey years and questions one has to distinguish between west and east: in the west, the Ifo Investment Survey is conducted since 1987 while in the east, the slightly modified questions started in 1992. In general, however, there are not only assessment questions in the Investment Survey but also questions on the exact values of realized or planned investments. For details, see the Ifo Investment Survey and the respective list of variables.

**2.2. Company Database Amadeus**

In addition to the Ifo survey data an the Hoppenstedt database, the Amadeus company database is the main source for the EBDC Business Expectations Panel and the EBDC Business Investment Panel. It is a product of the Bureau van Dijk Electronic Publishing GmbH (BvDEP), which is one of the leading European providers of global enterprise information. Amadeus contains business and financial information on more than 11 million, mainly non-quoted enterprises, from 41 countries in Europe. Currently, about 1 million German enterprises are registered.

For the company databases information of market-leading local institutions and well-known businesses of the respective countries are used. The financial closing data for

---

<sup>6</sup> Among others, until spring 2002, the question on factors influencing investment activities included “technical development” as well as “acceptance of new technologies”. From autumn 2002 onwards, however, these were subsumed in “technical factors”. For further changes, see the list of variables.

German businesses come from Creditreform or the Creditreform Rating AG, which belongs to the Creditreform Group. The Creditreform Group has provided bank credit information appertaining to customer-supplier relationships for more than 125 years and is the European market leader for bank credit information. The key source for the Amadeus database is the MARKUS Database, which contains 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), the data in Amadeus are in a homogeneous, standardised accounting format based on generalised national and/or international accounting 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.<sup>7</sup> In addition, with regard to currentness, Amadeus guarantees that the financial closing data are available in the database after 15 months at the latest.

### **2.3. Hoppenstedt Accounting Database**

The Hoppenstedt Accounting Database is a product of Hoppenstedt Business Information GmbH, which is one of the leading providers of business and industry information in Germany and is a part of the Hoppenstedt Group. Key business areas, in addition to the provision of company information, are the sales of postal addresses, credit and risk analyses as well as the publication of technical journals, so that – depending on customer status and rights of use – choices can be made from a great number of databases. For the company databases, information from external sources such as the Federal Official Gazette, the Commercial Register, the business press or annual reports, is used and if required also derived in direct dialog. According to Hoppenstedt all known changes are updated, evaluated and incorporated daily into the

---

<sup>7</sup> These, however, can be exported via the historical databases in the EBDC.

corresponding database, which is why the supplied business information is marked by currentness, quality and data depth.

The information for the EBDC Business Expectations Panel and the EBDC Business Investment Panel was taken from the Hoppenstedt Accounting Database,<sup>8</sup> currently containing 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, which are closely oriented on the respective original, were developed for HGB, IAS and US-GAAP.

### **3. Matching Method - Probabilistic Record Linkage**

To link the different data sources, recourse is taken to the companies' address information contained in each database. By this means, we can generate two (four) allocation tables (Ifo-Amadeus and Ifo-Hoppenstedt) and combine them in the EBDC Business Expectations Panel and the EBDC Business Investment Panel. In the following, we give a short overview of the record linkage between the addresses of the Ifo Business Survey and those of the Amadeus company database. This process is exactly the same in the case of the Hoppenstedt database or with regard to the linkage of the Ifo Investment Survey. The only difference is that, here, a so-called "gold standard" was additionally created to determine the match or non-match weights for every address variable. The 28,636 data records of the Ifo Business Survey (including construction, wholesaling/retailing, service providers, status: 03/2008) were linked with the 923,946 German companies in the Amadeus database (status: 01/2008 and 10/2008) by using the matching software MTB (Merge Toolbox) developed at the Center for Quantitative Methods and Survey Research of the University of Konstanz.

---

<sup>8</sup> At the time of data collection for the EBDC, ca. 120,000 enterprises were included in Hoppenstedt.

MTB is especially useful when there is no unambiguous key, such as a company code number, in the dataset as it allows the allocation of different datasets containing large amounts of data with different postal-address presentations. Specifically, using the method of Probabilistic Record Linkage which is based on the theory of Newcombe et al. (1959) and which was formalised by Fellegi and Sunter (1969), similar name/address data can be linked by determining probabilities for the degree of agreement (so-called “similarity”) between the variables. This “similarity” is calculated from the quotient of the probability that the variable  $x$  from both datasets is identified as concurrent for similar companies ( $M$  probability) and the probability that the variable  $x$  from both datasets is evaluated as concurrent for non-similar companies ( $u$  probability). Ideally, the quotient amounts to  $1/0$ . Since deviations between the different address variables are not equally important, we use the logarithm of the  $M$  (= match weight) or the  $u$  probability (= non-match weight) of the respective variable to weight in case of a variable match or non-match. In sum, these (positive/negative) variable weights make up the overall “quality”, i.e. the degree of similarity of each address linkage.

As these parameters (weights) of the Probabilistic Record Linkage have to be gained empirically, an ensured partial set from the amount of data records is initially formed. This data subset is called a “gold standard”, whereby, in general, a gold standard refers to a linkage that allows for an unambiguous allocation of the data records of two databases. Thus, this first record linkage is performed using telephone, fax numbers and e-mail addresses. Since, however, these variables were not filled in all data records and since the systematics of telephone and fax number entries differed as well, a linkage of all the data was not possible in this way. Nevertheless, by this means, we received 11,225 matches meaning that about 40 percent of the Ifo entries could be directly assigned to an enterprise in Amadeus. However, a closer look revealed that also data records actually not belonging together were identified as matches. Frequently, this is the case if there are several subordinate units within a company which have the same central telephone and fax numbers or e-mail addresses – for example administrative units, holding and management companies. Therefore, several of the linked gold-standard pairs were not true matches in the literal sense. They were only true matches in the sense of their affiliation to a larger enterprise group.

Within these linked data records, we then computed weights for the individual address variables. To this end, the number of correct/false concurrences was compared with the total population of comparisons. In addition, in order not to distort the results, the weights were determined for different preparation variants of the variables. By this means, we could avoid giving too positive assessment of the results due to frequently occurring name sequences or to give too negative assessments due to different spellings of the same name: the variant of the match variables with the best differentiation is the one having the greatest number of true agreements (M probability) without displaying an increased number of wrong agreements (u probability). As already mentioned, the gold standard also contained matches that could not be designated as such in the true sense. Therefore, the correctness of the determined weights was again controlled by hand by means of a random sample, comprising 2000 data records from the gold standard. From the resulting similarity weights, which were allocated to the variables before the actual MTB run, we could then calculate the quality. If positive matches were determined for all variables when linking companies' addresses from both databases, the quality resulted as the maximum total weight from the sum of the individual concurrence weights. For example:  $10.6 + 8.6 + 4.9 + 12.9 = 37.2$ . On the other hand, if MTB identified one of the variables as a non-match, the (negative) non-match weight was used in the equation, so that the overall quality of the respective address linkage decreased accordingly.

The comparison of the variables from the Ifo and the Amadeus databases was carried out by means of a string similarity function, comparing N grams of length 2 (= bi-gram with blanks before and after all strings) from the respective variables by placing a raster of the length 2 over the string. Differences in the variable characteristics were weighted linearly according to bi-gram similarity, with MTB giving a lower evaluation to cases where there is a high agreement with a shorter name than in the case of a high agreement with a longer name. In addition, the determination of a Jaro factor for all variables (= weight adjustment) to the value "2" led to a faster allocation of the full agreement weight in the case of high concurrence and thus resulted in a better differentiation of the matches from the non-matches. Finally, for evaluating the MTB run, a threshold value for the quality variable has to be defined. In doing so, however, recourse was not taken to the program evaluation, which as of a certain quality identifies the linkage as a match. Instead, a higher value was applied from the very

start and, once again, we conducted a manual control for a larger quality area. As before, the idea here was to avoid the error of a false positive: pairs do not qualify as true matches only because they are above a threshold value (e.g., a place in a name can increase similarity or there can be the wrong legal form but the same name). On the other hand, some true matches, which were too strongly devalued due to lacking information (e.g. same name, street empty), can still be found below the defined threshold value in certain blocks.

### **Results:**

With this procedure, 28,636 enterprise postal addresses from the Ifo Business Survey were compared with 923,046 enterprise postal addresses from the **Amadeus company database**. This led to a total of 103,377,535 pair comparisons, whereby in the subsequent manual control 9,472 were classified as matches.<sup>9</sup> Of these, 4,073 matches belong to the **Ifo Business Survey for manufacturing**, 1,703 to the **Ifo Business Survey of construction**, 1,859 to the **Ifo Business survey of wholesaling/retailing** and 1,837 matches belong to the **Ifo business survey of service providers**. Furthermore it can occur that one company in Amadeus is allocated to two or more Ifo entries. The number of linked companies thus amounts to 8,915. In comparison, the linkage of the **Hoppenstedt firm database** with the **Ifo Business Survey** resulted in a total of 4,811 matches or 4,377 enterprises.

If both linkages are subsequently combined, we get an allocation table containing, for each company, the key variable from the Ifo KT (Ifo-ID) as well as the key of the respective accounting database (Amadeus-ID and/or Hoppenstedt-ID). In this regard, companies of the Ifo panel can only be found in Amadeus, only in Hoppenstedt or in both enterprise databases. Therefore, there is a total of 10.699 allocations (IDs) in the EBDC Business Expectations Panel which can be identified via the new variable EBDC-ID. Figure 3 illustrates the generation of the EBDC Business Expectations Panel.<sup>10</sup>

---

<sup>9</sup>These results are primarily attributable to the differing currentness of the address databases. Amadeus companies that have not reported for more than 5 years are deleted; the Ifo panel contains all the addresses of the past 20 years, even those that have already been cancelled.

<sup>10</sup> At the time of construction, the Business Expectations Panel referred to the years 1980- 2009. However, as it is updated regularly, it always extends to the present minus one year.

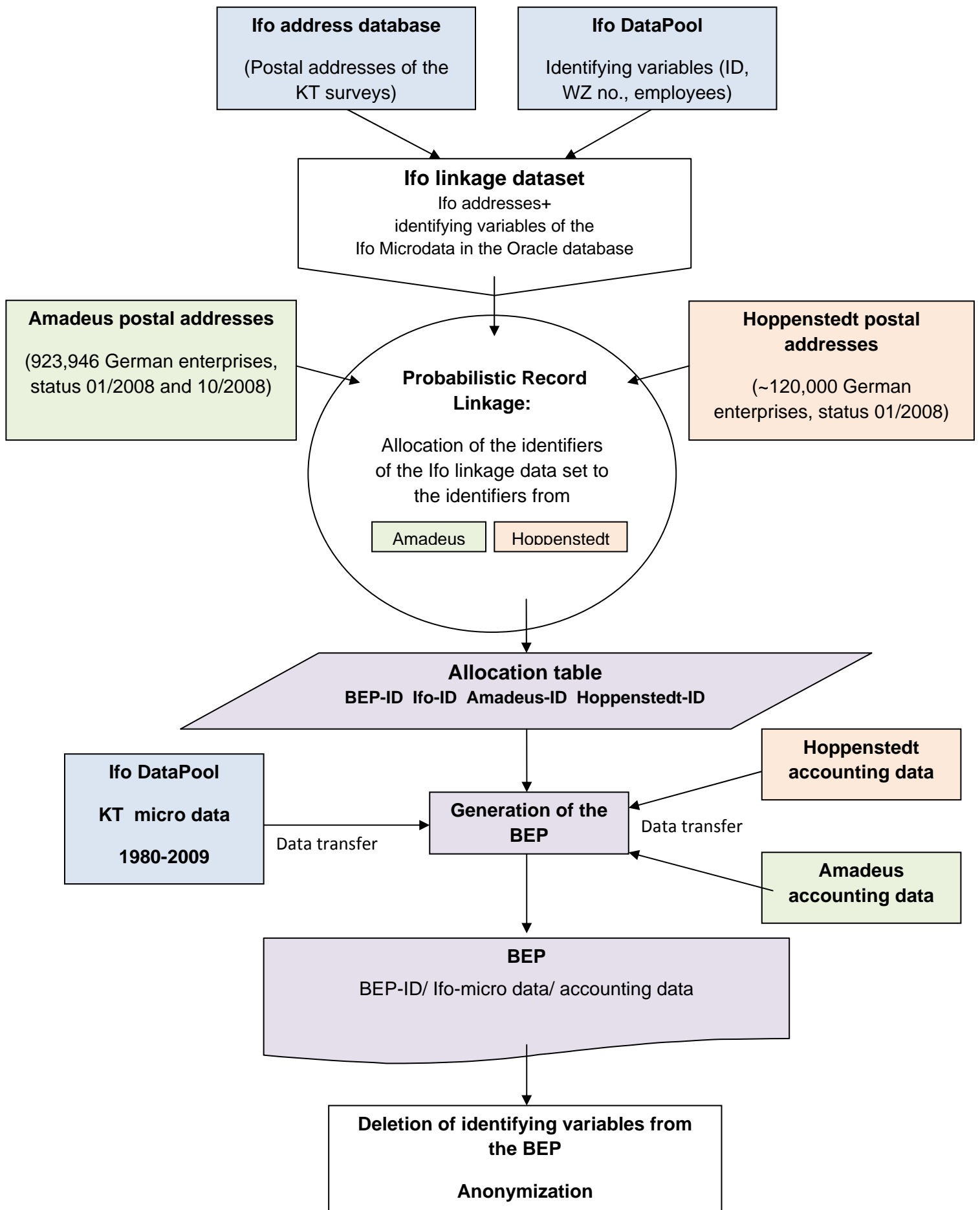


Fig. 3: Generation of the EBDC Business Expectations Panel (status 2010)

Considering the record linkage between the **Ifo Investment Survey** and the company database **Amadeus** 6,154 matches could be identified, while in the case of **Hoppenstedt** there were 3,101 fitting addresses. These values, however, are inclusive of firms that do not belong to the manufacturing but to the construction sector (IT BAU), which will be integrated later. Therefore, in total, there are 4,656 companies in the EBDC Business Investment Panel (BIP). Due to the fact that there is no distinction between products in the Ifo Investment Survey, each Ifo address belongs to one entry in Amadeus or Hoppenstedt. Thus, identification is again via the EBDC-ID. 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.<sup>11</sup>

#### **4. Structure of the EBDC Company Panels**

With regard to their structure, the EBDC Business Expectations Panel and the EBDC Business Investment Panel are very similar. Identification is possible with a company-specific BEP/BIP-ID, the respective year of observation and further dataset-specific time variables (see below).

In general, when available, the EBDC company panels contain information from individual instead of corporate group accounts,<sup>12</sup> however, we did not take over the accounting schemes from the original databases. 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.<sup>13</sup> 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.<sup>14</sup> For a detailed and correspondingly structured overview of the available accounting and earnings-statement variables, see the

---

<sup>11</sup> In general, accounting information from Amadeus is in thousand euros and rounded off; furthermore individual positions are often aggregated. In contrast, Hoppenstedt displays raw data with an accounting-structure being closely oriented to the HGB.

<sup>12</sup> The respective type of reported account is indicated by the variable "reporting\_basis". In this context, limited financial data means that the accounting information was not published but requested individually.

<sup>13</sup> The conversion scheme employed to transfer the initial variables into the newly generated EBDC accounting variables can be viewed at the EBDC. Furthermore, there is also a detailed standard accounting scheme available that can be used for orientation.

<sup>14</sup> Upon request and in special cases, the EBDC panels can also be made available with the initial accounting variables from Amadeus and Hoppenstedt.

respective list of variables of the EBDC Business Expectations Panel and the EBDC Business Investment Panel.

### **EBDC Business Expectations Panel**

For the EBDC Business Expectations Panel we get 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, however, only for 8,318 enterprises (status: 2010). For this reason, in the EBDC Business Expectations Panel, the BEP-ID consists of three components: the variable “company\_id”, a running company number, the variable “questionnaire\_id”, indicating the number of questionnaires per company and the variable “sector\_id”, which counts sectors per questionnaire. The questionnaire\_id in the Ifo Business Survey stands for different business units. In KT VG questionnaire\_id counts the number of products, in KT DL it enumerates the service division, in KT Han a new questionnaire\_id appears per product group, in KT Bau per division of construction. Several sectors may be requested on one questionnaire.

The dataset is sorted according to BEP-ID, year and month and is presented in a long format, i.e. every response is identifiable via these three variables. Each month contains the survey results from the Ifo KT resulting in up to 12 monthly reports per product/ company in each year. What follows is the accounting information in a constructed month “99”. This structure was chosen as it allows an easy dataset-handling and various allocations between months and years.

Sorted according to their function, the following variables are included in the columns of the EBDC Business Expectations Panel: identification variables, accounting and GUV as well as selected Ifo variables. In addition to BEP-ID, year and month, the identification variables contain information such as the industry code, company size, German federal state, stock exchange quotation and legal form.<sup>15</sup> The list of variables gives a general overview of the available information and sums up important features of the variables (e.g. questions of the individual surveys, survey frequency, etc.).<sup>16</sup> The

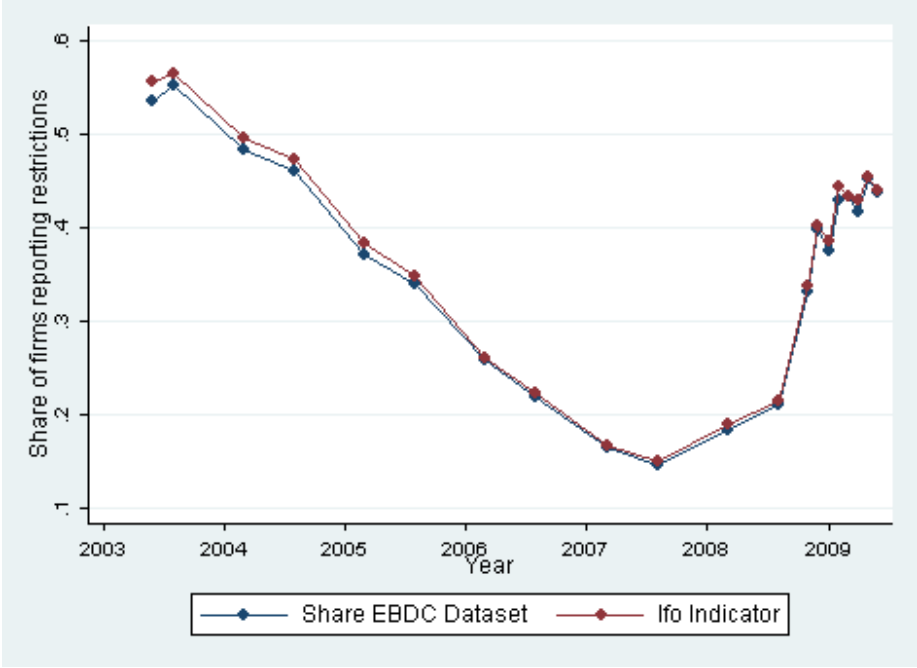
---

<sup>15</sup> Due to anonymization, the federal state information for very large firms (> 10.000 employees) was deleted.

<sup>16</sup> It may be in demand in the EBDC. A BEP panel with complete variables of the Ifo Business Survey can also be obtained.

criterion of their inclusion in the BEP panel was to occur in at least 2 of the 4 surveys of the Ifo KT.

In total, there are 8,881 firms with 49,692 accounting results linked to 1,135,300 Ifo responses for the whole period. A more detailed and up-to-date evaluation of the dataset (descriptive statistics, filling of selected variables, etc.) is available for duplication at the EBDC. Concerning representativeness of the dataset one can, however, take an immediate look at figure 4.<sup>17</sup> Here, the share of firms in the EBDC panel (only KT VG) reporting credit constraints is illustrated and compared to the Ifo Credit Constraint Indicator (Kredithürde) which covers the whole manufacturing sector. As can be seen, the much smaller EBDC Business Expectations Panel can in fact be called representative with regard to the Ifo Business Survey for manufacturing.



**Fig. 4: Comparison EBDC Panel and Ifo Credit Constraint Indicator**

### 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, we

<sup>17</sup> Source: EBDC, Ifo Institute

have the variable “survey\_base” that indicates if the observations are from the spring (1) survey, from the autumn (2) survey or from a balance-sheet (99).

Currently, the dataset contains data from 1987 to 2008/ 2009 and is sorted by “BIP\_ID” (company\_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, legal form, etc..<sup>18</sup> Again, what follows in the different columns are accounting variables and variables of the Ifo Investment Survey. The whole list for the EBDC Business Investment Panel contains more detailed information on the available parameters, their frequency, time horizon, etc..

In total, there are 4,656 companies, 21,485 balance-sheets and more than 40,000 spring and autumn observations, respectively. However, for only 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 can be obtained at the EBDC.

### **Additional modules**

Like the Ifo Business Survey for manufacturing, the aforementioned **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. Identification is via the same EBDC\_ID/ year/ month (with a constructed month “98”) combination what makes it possible to easily append the information on innovative activities. 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. To append it, we constructed the month “97”. As regards content,

---

<sup>18</sup> To guarantee the anonymity of firms, federal state information for very large companies (> 10.000 employees) was deleted.

it refers to the market perspectives of a specific product (growing, stagnating, decreasing) and its status of innovation/ phase of development.

## **Expansions**

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. Another step is to include the remaining information from the Ifo Investment Survey, the sector of construction (IT BAU) in the EBDC Business Investment Panel. In this regard, however, it must be taken into consideration that the KT surveys in construction and wholesale/retail are not directed to various product groups but to individual construction and distribution sectors. What is another option: to merge the two existing EBDC panels into an overall dataset or to link them with other external data such as supervisory board or proprietor structures.

## **5. Access**

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 do not allow the identification of 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.<sup>19</sup> 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.

## 7. Bibliography

Becker, S. O. and K. Wohlrabe (2008), Micro Data at the Ifo Institute for Economic Research: The "Ifo Business Survey", Usage and Access. *Journal of Applied Social Science Studies (Schmollers Jahrbuch)*, 128(2), 307-319.

Fellegi, I. P. and A. B. Sunter (1969), "A Theory for Record Linkage", *Journal of the American Statistical Association*, 64, 1183–1210.

Goldrian, G. (2004), *Handbuch der umfragebasierten Konjunkturforschung. Ifo Beiträge zur Wirtschaftsforschung 15*, Ifo Institute for Economic Research, Munich.

Newcombe, H. B., J. M. Kennedy, S. J. Axford and A. P. James (1959), "Automatic Linkage of Vital Records", *Science* 130, 954–59.

Oppenländer, K. H. and G. Poser, eds. (1989), *Handbuch der Ifo-Umfragen: 40 Jahre Unternehmensbefragungen des Ifo-Instituts für Wirtschaftsforschung*, Duncker & Humblot, Munich and Berlin.

---

<sup>19</sup> [http://www.cesifo-group.de/portal/page/portal/ifoContent/N/data/EBDC\\_Container/EBDC\\_Angebot\\_Container/EBDC\\_Vertrag.pdf](http://www.cesifo-group.de/portal/page/portal/ifoContent/N/data/EBDC_Container/EBDC_Angebot_Container/EBDC_Vertrag.pdf)