Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC-ECCBSO-CBRT Conference on *"Uses of Central Balance Sheet Data Offices' information"*

Co-organised by the IFC, the European Committee of Central Balance Sheet Data Offices (ECCBSO) and the Central Bank of the Republic of Turkey (CBRT)

Özdere-İzmir, Turkey, 26 September 2016

Squaring the circle - providing annual account information for research in Germany

The annual accounts scientific database and its dissemination in the Deutsche Bundesbank[1]

Ulf von Kalckreuth, Deutsche Bundesbank

[1] This paper was prepared for the meeting. The views expressed are those of the author and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

Squaring the Circle – Providing Annual Accounts Information for Research in Germany

The Annual Accounts Scientific Database and its Dissemination in the Bundesbank

Paper presented at the IFC / ECCBSO / CBRT Conference on “Uses of Central Balance Sheet Data Offices’ information” in Ozdere-Izmir, 26 September 26 2016

Dr Ulf von Kalckreuth, Deutsche Bundesbank*

*Ulf von Kalckreuth is Head of Section and Deputy Head of Division and is responsible for corporate financial statements statistics at the Deutsche Bundesbank.
E-mail: ulf.von-kalckreuth@bundesbank.de. Address: Deutsche Bundesbank, DG Statistics, Wilhelm-Epstein-Strasse 14, 60431 Frankfurt am Main, Germany.

Squaring the Circle – Providing Annual Accounts Information for Research in Germany

The Annual Accounts Scientific Database and its Dissemination in the Deutsche Bundesbank

Paper presented at the IFC / ECCBSO / CBRT Conference on "Uses of Central Balance Sheet Data Offices' information" in Ozdere-Izmir, 26 September 2016

By Dr Ulf von Kalckreuth, Deutsche Bundesbank[1]

1. Data confidentiality is important in Germany

Working with statistics that involve granular data in Germany entails several additional layers of complexity compared to other European countries. In principle, German confidentiality and data privacy laws are fully compatible with EU standards. But since the ruling of the German Constitutional Court concerning the 1983 census in West Germany, the right to data protection has been considered a "basic right", ie a constitutionally protected individual right that directly binds the legislature, the executive and the judiciary. The Constitutional Court blocked the 1983 census to prevent the infringement of this right. The census was consequently delayed until 1987, after which no census took place for another 24 years, although the reunification of Germany in 1991 created a major need for new census information. As a result, courts and all levels of administration treat data privacy issues in a very principled way.

Data protection issues have repercussions for corporate financial statements statistics on many levels. In Germany, there is no balance sheet office that is officially tasked with processing and publishing financial statements.[2] With regard to the publication of financial statements, there are numerous exemptions for smaller companies. Thus, statistical information on smaller firms is difficult to obtain and firm-level data needs to be collected from various sources, involving diverse formats and levels of detail. There is no unique identifier for non-financial firms in Germany. Therefore, in order to use data from different sources, much work has to be done to preclude double entries, ie the same annual account entering the statistical database more than once. This and other aspects of data quality management become especially cumbersome when a significant part of the financial statements have to be processed in anonymised format for data protection reasons. And finally, it was not until 2016 that a law was passed allowing the use of the Federal Business Register for statistical purposes at the Bundesbank.

2. A database for statistical purposes

As is often the case, specific limitations call for specific solution strategies. Traditionally, annual accounts statistics uses financial statements data from the rating activities of the Deutsche

---

[1] Ulf von Kalckreuth is Head of Section and Deputy Head of Division and is responsible for corporate financial statements statistics at the Deutsche Bundesbank. Timm Körting is advancing this project. I thank him for indispensable and valuable input, also in the context of this presentation. E-mail: ulf.von-kalckreuth@bundesbank.de. Address: Deutsche Bundesbank, DG Statistics, Wilhelm-Epstein-Str. 14, 60431 Frankfurt am Main, Germany.

[2] The "Bundesanzeiger" is evolving in this direction, though.

Bundesbank. Before monetary union, the Deutsche Bundesbank bought commercial bills from commercial banks for refinancing purposes. Thus, for the year 1998, more than 60,000 balance sheets were available. After monetary union, the number of available annual accounts dropped sharply as a result of the new regime in refinancing. Thus, in 2005, a "pool" was created by merging balance sheets available from rating activities with those from external providers. Some of these are designated data providers selling identified annual accounts; others are private sector "pool partners" sharing anonymised balance sheet information. Today, the pool is the backbone of the statistical infrastructure on company finances at the Bundesbank.

3. Towards a database for scientific use

Since the inception of the pool, research activities at the Bundesbank and by outside researchers using firm-level data have become very important. The statistical database cannot be used directly for research purposes. Because the anonymous pool partner data are strictly confidential, they cannot be matched with external information such as ratings or direct investment activity.

Fortunately, during the last decade, the volume and share of data obtained from identified information from external information providers has increased dramatically. Together with the Bundesbank's data from rating activities, these provide a solid, stand-alone basis for research activities. In 2014, the Research Data and Service Centre (RDCS) was established as the Bundesbank's provider of micro data information for both analysis and (internal and external) research purposes. The RDCS provides a secure environment for the analysis of granular data. Within its confines, the balance sheet data collected from the Bundesbank's rating activities and from commercial data providers can be enriched using external information. The scientific data set is then anonymised and made accessible to researchers under close surveillance, making sure that no reidentification activity takes place. Everything is therefore in place: attractive granular information for researchers, a protocol to resolve confidentiality issues and the resources needed to make the data accessible in a safe way.

The database under construction is provisionally labelled *Ustan*$^{+}$, referring to an earlier and very successful scientific database called *Ustan* that was composed exclusively of data originating from rating activities; see Stöss (2001)[3] for a description and Chatelain et al. (2003)[4] for a usage example. From a technical point of view, Ustan^{+} will be realised by extracting data from the pool, which is composed of data from eight different providers.

4. Ustan^{+} at a glance

Ustan^{+} provides balance sheet and profit and loss account information from the non-consolidated financial statements of non-financial firms in Germany. It comprises three different data sources: the Bundesbank's refinancing operations and two commercial data providers, Bisnode and Creditreform. The data base is free of duplicates and the accounting information is provided in a unified format, with a common underlying set of definitions, making the data commensurable to the utmost extent.

[3] Elmar Stöss, Deutsche Bundesbank's Corporate Balance Sheet Statistics and Areas of Application, Schmollers Jahrbuch 121 (2001), pp 131-137.

[4] Jean-Bernard Chatelain, Ignacio Hernando, Andrea Generale, Ulf von Kalckreuth and Philip Vermeulen, New findings on firm investment and monetary transmission in the Euro area. Oxford Review of Economic Policy, Vol. 19(1), 2003, pp 1-11.

Per financial year, the data base encompasses up to around 90,000 observations for the years from 1997 onwards.

The scientific dataset is based on non-anonymized data with firms' names and addresses. Thus, RDCS staff is able to enhance the financial information, matching it with complementary data sources. The first version of the database is expected to be ready for use by summer 2017. Afterwards, subsequent refinements involving external information and weights will be carried out.

Tables 1 and 2 give an overview of the size and composition of *Ustan*$^{+}$ with regard to sectors and firm size for the fiscal year 2013. *Ustan*$^{+}$ is compared to the statistical database, the pool, and the business register run by the German National Statistical Office as a proxy for the universe of non-financial companies in Germany.[5] There is little difference between *Ustan*$^{+}$ and the pool in terms of sector and size composition, concerning both number of firms and aggregate sales. A comparison with German business register data reveals a clear underrepresentation of micro firms with sales of less than €2 million. In terms of the number of firms, the manufacturing sector is overrepresented and the service sector is underrepresented. The same holds true – to a lesser extent – with regard to sales aggregates. Thus, in order to extrapolate on the aggregate, informative weights are important.

Table 3 demonstrates the panel structure of the new scientific data base, showing the number of observations per fiscal year, the number of observations with at least one predecessor (needed to calculate first differences or growth rates) and the number of firms that are part of a balanced panel that has observations for each year in the period from 2008 to 2013. Longer contiguous strings of observations are needed for many of the more elaborate techniques in panel econometrics. Starting from 1997 with 55,000 firms, the number of observations increases to up to 90,000 observations per year. As many as 35,000 firms are part of a balanced panel from 2008 to 2013. The reduction is even less pronounced when using sliding cylindered samples.

It is expected that *Ustan*$^{+}$ will quickly become an important part of the data infrastructure for research on financial structures and activity of companies in Germany and Europe.

[5] Due to lower thresholds concerning turnover and number of employees in the register, this is not entirely correct: in all of the data collections, a large number of very small firms is missing.

Number of Firms and Sales by Sector in Comparison (FY 2013)

	USTAN+		*Data Pool*		*Company Register*	
Sector	**Number of Firms**	**%**	**Number of Firms**	**%**	**Number of Firms**	**%**
Mining & Quarrying	255	0.3	373	0.3	2.279	0.1
Manufacturing	18,396	20.5	23,487	19.7	248,135	7.0
Energy & Water	3,643	4.1	4,669	3.9	74,273	2.1
Construction	11,320	12.6	13,820	11.6	389,557	11.0
Trade	21,102	23.5	27,747	23.3	655,102	18.6
Transportation	4,659	5.2	7,381	6.2	119,016	3.4
Information & Communication	3,708	4.1	4,929	4.1	130,027	3.7
Business-related Services	8,260	9.2	10,871	9.1	685,547	19.4
Other	18,496	20.6	25,773	21.6	1,223,844	34.7
Total	89,839	100.0	119,050	100.0	3,527,780	100.0

	USTAN+		*Data Pool*		*Company Register*	
Sector	**Sales € bn**	**%**	**Sales € bn**	**%**	**Sales € bn**	**%**
Mining & Quarrying	16	0.5	17	0.4	16	0.3
Manufacturing	1,385	38.1	1,566	37.9	1,988	33.9
Energy & Water	626	17.2	636	15.4	616	10.5
Construction	82	2.3	93	2.3	247	4.2
Trade	970	26.7	1,147	27.8	1,801	30.7
Transportation	130	3.6	150	3.6	264	4.5
Information & Communication	116	3.2	155	3.7	217	3.7
Business-related Services	87	2.4	100	2.4	371	6.3
Other	219	6.0	264	6.4	342	5.8
Total	3,632	100.0	4,127	100.0	5,861	100.0

Number of Firms and Sales by Size in Comparison (FY 2013)

Size Sales of …	*USTAN+* **Number of Firms**	**%**	*Data Pool* **Number of Firms**	**%**	*Business Register* **Number of Firms**	**%**
< 2 m€	38,967	43.4	54,501	45.8	3,326,856	94.3
2 to 10 m€	23,729	26.4	32,767	27.5	150,146	4.3
10 to 50 m€	18,224	20.3	21,969	18.5	38,879	1.1
> 50 m€	8,919	9.9	9,813	8.2	11,899	0.3
Total	89,839	100.0	119,050	100.0	3,527,780	100.0

Size Sales of …	*USTAN+* **Sales bn€**	**%**	*Data Pool* **Sales bn€**	**%**	*Business Register* **Sales bn€**	**%**
< 2 m€	27	0.7	37	0.9	694	11.8
2 to 10 m€	115	3.2	158	3.8	627	10.7
10 to 50 m€	417	11.5	494	12.0	810	13.8
> 50 m€	3,074	84.6	3,438	83.3	3,731	63.6
Total	3,632	100.0	4,127	100.0	5,861	100.0

Panel Structure: Number of Firms

Financial Year	*Unbalanced Panel*	*With Predecessor*	*Balanced Panel*
1997	55,146	.	.
1998	42,216	37,728	.
1999	36,765	31,748	.
2000	35,081	29,554	.
2001	36,038	28,028	.
2002	41,242	28,783	.
2003	51,458	33,566	.
2004	62,587	42,628	.
2005	72,198	49,275	.
2006	75,978	53,500	.
2007	70,268	51,673	.
2008	77,593	54,738	35,007
2009	83,129	61,130	35,007
2010	87,109	65,623	35,007
2011	89,941	68,780	35,007
2012	91,331	71,118	35,007
2013	89,674	71,096	35,007
2014	41,942	38,417	.

Irving Fisher Committee on
Central Bank Statistics

BANK FOR INTERNATIONAL SETTLEMENTS

IFC-ECCBSO-CBRT Conference on *"Uses of Central Balance Sheet Data Offices' information"*

Co-organised by the IFC, the European Committee of Central Balance Sheet Data Offices (ECCBSO) and the Central Bank of the Republic of Turkey (CBRT)

Özdere-İzmir, Turkey, 26 September 2016

Squaring the circle - providing annual account information for research in Germany

The annual accounts scientific database and its dissemination in the Deutsche Bundesbank[1]

Ulf von Kalckreuth, Deutsche Bundesbank

[1] This presentation was prepared for the meeting. The views expressed are those of the author and do not necessarily reflect the views of the BIS, the IFC or the central banks and other institutions represented at the meeting.

DEUTSCHE BUNDESBANK
EUROSYSTEM

Squaring the Circle – Providing Annual Accounts Information for Research in Germany

The Annual Accounts Scientific Database and its Dissemination in the Bundesbank

Ulf von Kalckreuth, Deutsche Bundesbank, DG Statistics

IFC / ECCBSO / CBRT Conference on “Uses of Central Balance Sheet Data Offices’ information” in Ozdere-Izmir, September 26th, 2016

Data confidentiality is important in Germany!

Confidentiality and data protection laws in Germany **in principle fully compatible with EU standards**

But: in Germany, right to data protection is considered a **fundamental individual right** by the powerful German Constitutional Court

- The Court blocked the 1983 census to make sure that this right was not infringed
- The census was carried out only in 1987 (Western Germany). After that, there was no census for 24 years!! (Reunification was in 1991).
- The first census after reunification was carried out only in 2011!

Courts, and – as a consequence – the administration treats data privacy issues in a rather principled way.

Doing statistics involving granular data is hard work in Germany!

Annual accounts information in Germany

Some consequences for annual accounts statistics

- No compulsory balance sheet central office that processes annual accounts (Though the “Bundesanzeiger” is evolving in this direction)
- Lots of exceptions from publication obligations for smaller firms
- **Statistical information for smaller firms hard to get**

Annual accounts information in Germany

Traditionally, annual accounts statistics uses data base from Bundesbank rating activities. Bundesbank was buying commercial bills from commercial banks for refinancing. In 1998, more than 60 000 balance sheets available.

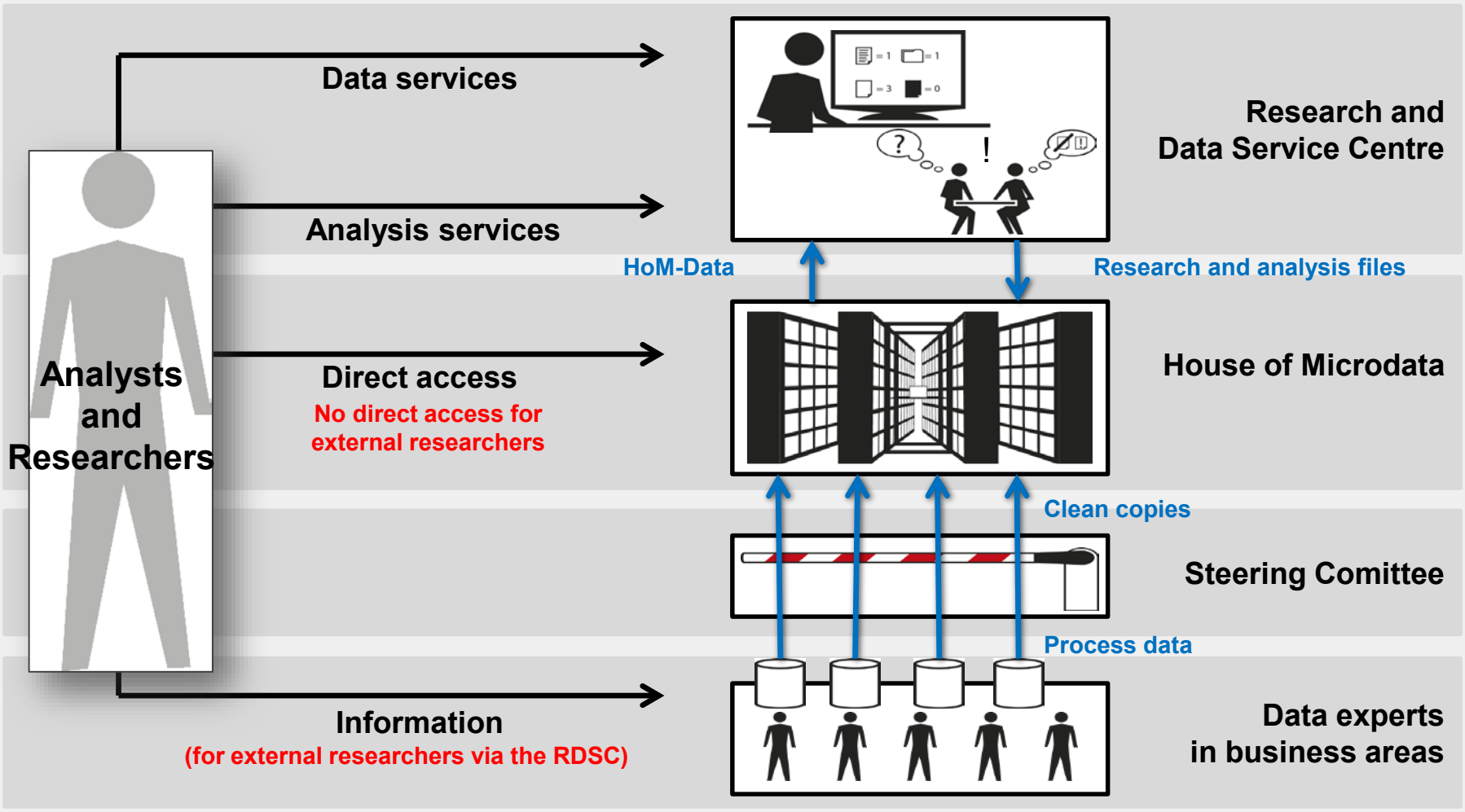
- After monetary union, number of available annual accounts drops sharply as a result of the new regime in refinancing
- In 2005, the **“Pool”** is created, from merging balance sheets available from rating activities with those from external providers.
- Some are designated data providers selling identified annual accounts, others are **“pool partners”**, sharing anonymised balance sheet information.
- **We must not disclose the accounts provided by the “pool partners”**

Towards a data set for scientific use

Since then

- Research activities in the Bundesbank and from outside researchers using Bundesbank data become very important
- Data pool not directly useable due to confidentiality and the missing possibility to match external information (such as rating or direct investment activity)
- Volume of data from outside information providers increases dramatically
- In 2014 Research Data and Service Centre (RDCS) founded as a Bundesbank provider of micro data information for analysis and (internal and external) research
- **Will provide safe environment for data analysis**

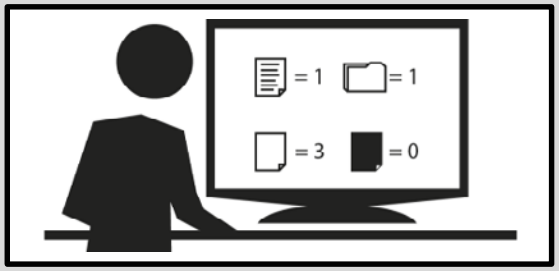
House of Microdata and RDCS – Value added for analysts and researchers



Research Data and Service Centre (RDSC)
Services for internal and external users

Data services

- Research and analysis files
- Documentation
- Tabulation
- Matching and merging
- Reference data administration



Analysis services

- Advice
- Clarification of access rights

Internal users

- Guidance
- Ad-hoc evaluations on analysis files

External Users

- Processing of applications
- Dissemination of data, making available on-site
- Output control and clearing

Towards a data set for scientific use

USTAN+

- Combines the Bundesbank rating data with the information from 2 commercial data providers
- Name refers to the old information data base USTAN that has gained a high reputation
- **Realised as extraction from data pool (itself consisting of 8 data sources)**

USTAN+ --- Basics

Overview

- Non-consolidated financial statements of non-financial firms in Germany
- 3 different data sources: Bundesbank's refinancing operations and two commercial data providers (Bisnode and Bureau van Dijks DAFNE database)
- Free of duplicates and a common structure
- Up to around 90.000 observations per year
- From 1997 onwards
- Non-anonymized data, i.e. containing firms' names and addresses
- Matching with complimentary data sources possible
- Data base expected to be usable in early spring 2017
- Afterwards refinements: matching, weights.

USTAN+ --- Structure (1)

Number of firms by sector in comparison (FY 2013)

	USTAN+		*Data Pool*		*Company Register*	
Sector	**Number of Firms**	**%**	**Number of Firms**	**%**	**Number of Firms**	**%**
Mining & Quarrying	255	0.3	373	0.3	2.279	0.1
Manufacturing	18,396	20.5	23,487	19.7	248,135	7.0
Energy & Water	3,643	4.1	4,669	3.9	74,273	2.1
Construction	11,320	12.6	13,820	11.6	389,557	11.0
Trade	21,102	23.5	27,747	23.3	655,102	18.6
Transportation	4,659	5.2	7,381	6.2	119,016	3.4
Information & Communication	3,708	4.1	4,929	4.1	130,027	3.7
Business-related Services	8,260	9.2	10,871	9.1	685,547	19.4
Other	18,496	20.6	25,773	21.6	1,223,844	34.7
Total	89,839	100.0	119,050	100.0	3,527,780	100.0

USTAN+ --- Structure (2)

Sales of firms by sector in comparison (FY 2013)

	USTAN+		*Data Pool*		*Company Register*	
Sector	**Sales € bn**	**%**	**Sales € bn**	**%**	**Sales € bn**	**%**
Mining & Quarrying	16	0.5	17	0.4	16	0.3
Manufacturing	1,385	38.1	1,566	37.9	1,988	33.9
Energy & Water	626	17.2	636	15.4	616	10.5
Construction	82	2.3	93	2.3	247	4.2
Trade	970	26.7	1,147	27.8	1,801	30.7
Transportation	130	3.6	150	3.6	264	4.5
Information & Communication	116	3.2	155	3.7	217	3.7
Business-related Services	87	2.4	100	2.4	371	6.3
Other	219	6.0	264	6.4	342	5.8
Total	3,632	100.0	4,127	100.0	5,861	100.0

USTAN+ --- Structure (3)

Number of firms by size in comparison (FY 2013)

Size	*USTAN+*		*Data Pool*		*Business Register*	
Sales of …	**Number of Firms**	**%**	**Number of Firms**	**%**	**Number of Firms**	**%**
< 2 m€	38,967	43.4	54,501	45.8	3,326,856	94.3
2 to 10 m€	23,729	26.4	32,767	27.5	150,146	4.3
10 to 50 m€	18,224	20.3	21,969	18.5	38,879	1.1
> 50 m€	8,919	9.9	9,813	8.2	11,899	0.3
Total	89,839	100.0	119,050	100.0	3,527,780	100.0

USTAN+ --- Structure (4)

Sales of firms by size in comparison (FY 2013)

Size Sales of …	*USTAN+*		*Data Pool*		*Business Register*	
	Sales bn€	**%**	**Sales bn€**	**%**	**Sales bn€**	**%**
< 2 m€	27	0.7	37	0.9	694	11.8
2 to 10 m€	115	3.2	158	3.8	627	10.7
10 to 50 m€	417	11.5	494	12.0	810	13.8
> 50 m€	3,074	84.6	3,438	83.3	3,731	63.6
Total	3,632	100.0	4,127	100.0	5,861	100.0

USTAN+ --- Structure (5)

- **Little difference between USTAN+ and Data Pool** in terms of sector and size composition, concerning both number of firms and aggregate sales
- Comparison with German Company Register data (a proxy to the full population) reveals **underrepresentation of micro firms** with sales of less than 2 m€
- In terms of number of firms, an **overrepresentation of manufacturing sector** at the expense of **underrepresentation of the service sector**
- The same holds true – to a lesser extent – in terms of sales aggregates

USTAN+ --- Panel Structure

Number of Firms

Financial Year	*Unbalanced Panel*	*With Predecessor*	*Balanced Panel*
1997	55,146	.	.
1998	42,216	37,728	.
1999	36,765	31,748	.
2000	35,081	29,554	.
2001	36,038	28,028	.
2002	41,242	28,783	.
2003	51,458	33,566	.
2004	62,587	42,628	.
2005	72,198	49,275	.
2006	75,978	53,500	.
2007	70,268	51,673	.
2008	77,593	54,738	35,007
2009	83,129	61,130	35,007
2010	87,109	65,623	35,007
2011	89,941	68,780	35,007
2012	91,331	71,118	35,007
2013	89,674	71,096	35,007
2014	41,942	38,417	.

USTAN+ --- Panel Structure

- Starting from 1997 with 55.000 firms, the number of observations increases to up to 90.000 observations per year.
- Mainly due to the growing volume of the DAFNE database – particularly since 2006
- Number of firms drops to 35.000 when using a balanced panel from 2008 to 2013.
- Reduction is less pronounced when using sliding cylindered samples

THANK YOU!