

## **The Quality of the *KombiFiD*-Sample of Business Services Enterprises: Evidence from a Replication Study**

By Alexander Vogel and Joachim Wagner\*

### **Abstract**

This study tests whether the KombiFiD sample can be regarded as a high quality data set for empirical research on enterprises from business services industries. It performs an empirical investigation using the original data in a first step and replicates exactly this investigation using the KombiFiD sample in a second step. We find that large business services firms are oversampled in the KombiFiD agreement sample which leads to a higher share of exporting business services firms compared to the original data. After controlling for firm size and industries results based on the original data and on the KombiFiD sample are highly similar for West German firms. Therefore, the KombiFiD sample can be regarded as a sound base for empirical studies on West German firms from business services industries. For East Germany, however, the number of business services firms seems to be too small for empirical analyses, at least in the field of firm's export participation.

*JEL Classification: C81*

### **1. Motivation**

Firm-level data are an important base for empirical studies in many fields of economics. During the last ten years in Germany the research data centres (RDC) of many data producing agencies provided researchers the possibility to perform empirical analyses based on firm-level data while keeping information on single firms strictly confidential. The number and variety of data sets offered by the RDC increases steadily (for an overview see Kaiser/Wagner, 2008).

---

\* This paper is part of the project *KombiFiD – Kombinierte Firmendaten für Deutschland* that is financially supported by the German Ministry for Education and Research (BMBF). It is a joint project of the Institute of Economics of Leuphana University Lüneburg, the research data centres of the German Federal Statistical Office and the statistical offices of the German federal states, the Institute of Employment Research of the Federal Employment Agency and the research department of the German Central Bank. While members of the KombiFiD-team from all institutions contributed to the construction of the data sets used in this paper we alone are responsible for the quality study presented here and the conclusions drawn.

The information provided in these data sets, however, is still far from complete. In particular, some important information (for example, on foreign direct investments of firms, or on the detailed composition of the employees by age and qualification levels) is included in data sets only that do not have information on other firm characteristics like exports, expenditures on research and development or profitability. The problem is that all these different data sets cannot be easily merged because they are provided by different data producing agencies. For example, data on foreign direct investments of firms are provided by the Deutsche Bundesbank, data on the detailed composition of the employees in a firm by age and qualification levels are provided by the Institute for Employment Research of the Federal Employment Agency, and data on exports, expenditures on research and development or profitability are provided by the Statistical Offices.

Linking confidential firm level information across the borders of the data producers is difficult in Germany for two reasons. Details aside, it is technically not easy (but not impossible either) and it is legal only if the firm agreed in written form. The basic idea of the project *KombiFiD* (an acronym that stands for *Kombinierte Firmendaten für Deutschland*, or combined firm level data for Germany) that is in detail described on the web (see [www.kombifid.de](http://www.kombifid.de)) is to ask a large sample of firms from all parts of the German economy to agree to match confidential micro data for these firms that are kept separately by the Statistical Offices, the Institute for Employment Research of the Federal Employment Agency and the Deutsche Bundesbank, in one data set. In the project 54,960 firms were asked to agree in written form to merge firm level information kept by the three data producing institutions. 30,944 firms replied and 16,571 agreed. These 16,571 firms are in the *KombiFiD Agreement Sample*. This data set is used in this paper, and the term *KombiFiD sample* is used for it.

While the original firm level data are high quality data that are either a census of the respective population of firms or a representative sample of this population the *KombiFiD sample* is the result of self-selection of firms into this data set because participation in *KombiFiD* was voluntary. A crucial question is whether the *KombiFiD sample* can be regarded as a high quality data set that can be used as a solid basis for empirical research. One way to shed light on this<sup>1</sup> is to perform an empirical investigation using data for all firms available from the respective data producer (the original data) in a first step and to replicate exactly this investigation using the *KombiFiD sample* in a second step. This is done in this paper using data for enterprises from business services industries.<sup>2</sup>

<sup>1</sup> An alternative way is to compare means and correlations of variables from the original data and the *KombiFiD sample*. Note that it is illegal to pool the original data and the *KombiFiD sample*. Therefore, a direct comparison of both data sets and an investigation of non-respondents or firms that refused to agree to merge their data are not feasible.

<sup>2</sup> See Wagner (2012) for a comparable study using data for enterprises from manufacturing industries.

The service sector becomes more and more important in Germany (and in other advanced economies; see Jorgensen/Timmer (2011) and Eichengreen/Gupta (2011), and services are considered as an engine of growth (Statistisches Bundesamt, 2009). Micro-econometric studies on the performance of services firm, however, are still rare, not least due to the fact that suitable longitudinal firm level data became available only recently (see Vogel, 2009). Therefore, the KombiFiD sample for services firms can provide an important data base – if they can be considered to be high-quality data. This motivates the present paper.

## **2. Exporter Performance in the German Business Services Sector**

In a recent paper Vogel (2011) focuses on the relationship between exports and several performance characteristics in the German business services sector in order to determine whether export premia and self-selection into export markets exist in the business services sector. This study uses a dataset from the German business services statistics panel 2003–2007, which contains, among other things, information about the export activities, number of persons employed, total turnover, and average wage of more than 20,000 business services enterprises per year (see Vogel (2009) for more detailed information about the dataset). All variables used in this study are also available in the KombiFiD sample. The basic idea explored here is to replicate the study from Vogel (2011) to shed light on the question whether the KombiFiD data are a reliable basis for empirical investigations for business services firms. In doing so we first replicate the results of Vogel (2011) using the data of the German business services statistics panel for the period 2003 to 2006, the period of the KombiFiD dataset. In a second step we replicate the results using the KombiFiD agreement sample. More precisely we use a subset of the KombiFiD agreement sample that contains only firms with available information from official statistics as well as the Institute for Employment Research (IAB) of the Federal Employment Agency.<sup>3</sup>

### **2.1 Export Participation and Comparison of Exporters and Non-Exporters**

First, Table 1 indicates that about 10 percent of the East German firms and about 12 percent of the West German firms that are covered by the original data set can be found in the KombiFiD agreement sample. For East Germany, this results in a fairly small sample of less than 500 firms per year.

---

<sup>3</sup> The focus of this paper is on the quality of the data from the KombiFiD agreement sample. Therefore, neither the theoretical background of the empirical models estimated nor the economic conclusions drawn from the results are discussed. See Vogel (2011) for further reading on these topics.

Second, Table 1 provides information about the firm’s export activities measured by export intensity (defined as the percentage of exports in total turnover). In 2003 and 2006 the distribution of the export intensity was highly skewed – most of the exporters sold a relative small share of their total production abroad, and only a few firms exported a very high share. This picture is true for both parts of Germany and can be found in the original data as well as in the KombiFiD agreement sample. However, looking at the share of exporters in all enterprises it is obvious that the share of exporters in the KombiFiD agreement sample is higher than in the original data. Thus in 2006, the share of exporters in West Germany ranges from about 21 percent in the original data to 28 percent in the KombiFiD agreement sample. In East Germany we find an export participation of 14 percent based on the original data and an export participation of 21 percent based on the KombiFiD agreement sample in 2006.

Table 1  
Export participation of business services enterprises  
in West and East Germany 2003 and 2006

	2003		2006	
	Number of enterprises	Share of enterprises on all enterprises	Number of enterprises	Share of enterprises on all enterprises
West Germany				
Export intensity				
0%	1,525	77.29	1,341	71.56
> 0% and < 5%	239	12.11	263	14.03
≥ 5% and < 10%	64	3.24	77	4.11
≥ 10% and < 25%	72	3.65	88	4.70
≥ 25% and < 50%	49	2.48	56	2.99
≥ 50% and < 75%	X	X	25	1.33
≥ 75%	X	X	24	1.28
Total number of observations	1,973	100.00	1,874	100.00
East Germany				
Export intensity				
0%	411	83.54	365	78.50
> 0% and < 5%	35	7.11	49	9.25
≥ 5% and < 10%	X	X	X	2.75
≥ 10% and < 25%	X	X	X	X
≥ 25% and < 50%	X	X	X	X
≥ 50% and < 75%	X	X	X	X
≥ 75%	X	X	X	X
Total number of observations	492	100.00	469	100.00

**Original data**

	2003		2006	
	Number of enterprises	Share of enterprises on all enterprises	Number of enterprises	Share of enterprises on all enterprises
<b>West Germany</b>				
Export intensity				
0%	13,473	82.02	14,720	78.95
> 0% and < 5%	1,395	8.49	1,752	9.40
≥ 5% and < 10%	409	2.49	566	3.04
≥ 10% and < 25%	511	3.11	729	3.91
≥ 25% and < 50%	336	2.05	413	2.22
≥ 50% and < 75%	151	0.92	217	1.16
≥ 75%	151	0.92	247	1.32
Total number of observations	16,426	100.00	18,644	100.00
<b>East Germany</b>				
Export intensity				
0%	3,946	89.40	4,467	85.99
> 0% and < 5%	225	5.10	317	6.10
≥ 5% and < 10%	70	1.59	107	2.06
≥ 10% and < 25%	82	1.86	130	2.50
≥ 25% and < 50%	37	0.84	92	1.77
≥ 50% and < 75%	20	0.45	44	0.85
≥ 75%	34	0.77	38	0.73
Total number of observations	4,414	100.00	5,195	100.00

*Note:* Only enterprises with a sum of turnover and other operating income greater than or equal to €250,000 and with one or more employees are considered. All values are unweighted. Firms that belong to the 1st or 99th percentile of the wage, turnover profitability or value added distribution are excluded from all computations. X: values deleted due to confidentiality reasons.

One reason for the higher share of exporters in the *KombiFiD* agreement sample can be found in the fact that the firms in the *KombiFiD* agreement sample are on average larger than the firms in the original dataset. Thus Table 2 shows that in 2006 East and West German exporters and non-exporters in the *KombiFiD* agreement sample are clearly larger in terms of the number of employees and turnover than the corresponding firms in the original dataset. However, concerning the comparison of exporting and non-exporting business services enterprises the results based on the original data and the *KombiFiD* agreement sample lead to the same conclusion: Exporting business services enterprises pay on average higher average wages and are on average also more productive (i.e. have a higher turnover and value added per employed person) than enterprises that serve only the domestic market. This is true for both parts of Germany.

Table 2

**Exporters vs. non-exporters in the West and East German  
business services sector 2006**

**Kombifid agreement sample**

	Non-exporters		Exporters	
	Mean	Standard Deviation	Mean	Standard Deviation
West Germany				
Number of Employees	144.1	372.8	108.7	301.0
Turnover (in € 1,000)	7,110.1	16,600.0	11,100.0	19,100.0
Average wage (in € 1,000)	30.8	18.8	41.1	18.0
Turnover per employee (in € 1,000)	89.0	108.0	129.8	118.3
Value added per employee (in € 1,000)	50.7	41.7	67.5	41.9
Number of observations	1,341		533	
East Germany				
Number of Employees	109.1	392.3	94.1	160.9
Turnover (in € 1,000)	4,127.0	9,799.6	11,600.0	36,000.0
Average wage (in € 1,000)	24.4	13.8	33.4	13.4
Turnover per employee (in € 1,000)	65.7	68.1	128.8	198.4
Value added per employee (in € 1,000)	40.3	30.2	63.2	55.9
Number of observations	365		104	

**Original data**

	Non-exporters		Exporters	
	Mean	Standard Deviation	Mean	Standard Deviation
West Germany				
Number of Employees	55.0	163.1	49.4	105.7
Turnover (in € 1,000)	3,012.5	8,241.1	5,948.7	11,100.0
Average wage (in € 1,000)	31.6	22.8	39.5	22.3
Turnover per employee (in € 1,000)	146.8	665.7	187.6	464.4
Value added per employee (in € 1,000)	69.2	255.6	82.1	216.4
Number of observations	14,720		3,924	
East Germany				
Number of Employees	44.3	123.2	41.4	82.5
Turnover (in € 1,000)	1,942.8	4,532.4	4,262.9	15,000.0
Average wage (in € 1,000)	24.1	16.3	31.7	15.7
Turnover per employee (in € 1,000)	104.3	224.9	125.0	149.3
Value added per employee (in € 1,000)	51.5	105.4	58.1	54.9
Number of observations	4,467		728	

*Note:* Only enterprises with a sum of turnover and other operating income greater than or equal to €250,000 and with one or more employees are considered. Reported are the unweighted mean, the unweighted standard deviation and the unweighted number of observations. Firms that belong to the 1st or 99th percentile of the wage, turnover profitability or value added distribution are excluded from all computations.

Note that these mean values give only an indication and overview of the differences between exporters and non-exporters without controlling for other firm characteristics like size and industries. Particularly in the heterogeneous business services sector it is important to control for industry effects. Therefore, a more thorough comparison between exporters and non-exporters is presented in the next section.

## 2.2 Exporter Premia

In this section we estimate so-called exporter premia that indicate the *ceteris paribus* differences of enterprise attributes between exporting and non-exporting enterprises, controlling for other characteristics of the enterprises. The results are presented in Table 3.

Overall the results based on the *KombiFiD* agreement sample and the original data are rather similar. The results of the pooled regression show for East and West Germany notable positive export premia concerning average wage, turnover and productivity (in terms of the turnover per employee and the value added per employee). These export premia are economically large and statistically highly significant when estimated with the original data as well as with the *KombiFiD* agreement sample. After controlling for unobserved heterogeneity by including fixed enterprise effects, the analyses based on the original data show that statistically significant differences in turnover are still present in both parts of Germany, even though on a much lower scale. For all other characteristics, there are no significant differences between exporters and non-exporters. Based on the *KombiFiD* agreement sample also concerning the turnover no significant differences occur.

When we look at the size of the coefficients based on the pooled regression it is obvious that in West Germany the coefficients do not differ more than eight percentage points between the two datasets. For East Germany the picture is different. Here, the export premia concerning the value added per employee is almost 15 percentage points higher when estimated with the *KombiFiD* agreement sample. Concerning the turnover and the turnover per employee the export premia is even more than 20 percentage points higher compared to the original data.

## 2.3 Pre-Entry Premia of Export Starters

The exporter premia reported in section 2.2 do not provide any information about the causality between exporting and the performance variables under consideration. This section tests whether the exporter premia reflect self-selection effects by analysing the differences between export starters and firms that continue to serve the national market only, several years before the export starters begin to export. Table 4 presents the pre-entry premia of enterprises that began to export in 2006 for two years before starting to export, one year before starting to export and at the starting year.

Table 3  
Export premia of business services enterprises  
in West and East Germany (2003–2006)

Kombifid agreement sample

	Estimation of (the log of) enterprise characteristics on export status and controls in $t$		Number of observations
	pooled regression	fixed effects model	
West Germany			
Average wage	29.39**	−1.03	7,652
Turnover	68.63**	1.37	7,652
Turnover per employee	47.89**	−0.37	7,652
Value added per employee	33.79**	−1.95	7,632
East Germany			
Average wage	30.02**	−0.60	1,889
Turnover	80.89**	2.31	1,889
Turnover per employee	61.70**	0.91	1,889
Value added per employee	37.67**	−0.67	1,881

Original data

	Estimation of (the log of) enterprise characteristics on export status and controls in $t$		Number of observations
	pooled regression	fixed effects model	
West Germany			
Average wage	27.93**	−0.12	69,679
Turnover	76.20**	2.78**	69,679
Turnover per employee	44.28**	0.20	69,679
Value added per employee	28.70**	−0.85	69,222
East Germany			
Average wage	27.25**	−0.50	18,867
Turnover	59.49**	4.10**	18,867
Turnover per employee	39.05**	1.86	18,867
Value added per employee	23.71**	−2.19	18,724

*Note:* The estimated regression coefficients and the levels of significance (+ indicates significance at the 10% level, \* at the 5% level, and \*\* at the 1% level, based on cluster robust standard errors) are presented for estimations of the log turnover, the log average wage, the log turnover per employed persons and the log value added per employed persons on the export status at  $t$ . Model 1 controls for a full set of interaction terms of year and economic activity (2-digit) dummies, the number of employed persons and its squared value. Model 2 also controls for fixed enterprise effects. To facilitate the interpretation, the estimated coefficient for the export dummy on the logarithmic variables has been transformed by  $100(\exp(\beta) - 1)$ . The transformation shows the average percentage difference of the respective variables (*ceteris paribus*) between exporters and non-exporters. Firms that belong to the 1st or 99th percentile of the wage, turnover profitability or value added distribution are excluded from all computations.



Table 4

**Self-selection into export markets of business services enterprises  
in West and East Germany 2006**

**Kombifid agreement sample**

	OLS estimation of the logarithmised characteristics on export start in $t = 2006$ and controls in $t, t - 1$ and $t - 2$			number of	
	Two years before starting ( $t - 2$ )	One year before starting( $t - 1$ )	In the starting year ( $t$ )	non- exporters	export starters
West Germany					
Turnover	44.51**	45.34**	53.27**	1,120	88
Average wage	27.77**	18.87**	26.37**	1,120	88
Turnover per employee	47.04**	40.36**	51.93**	1,120	88
Value added per employee	28.03**	20.14**	23.41*	1,118	87
East Germany					
Turnover	39.04+	40.48+	43.44+	306	21
Average wage	8.20	9.40	11.44	306	21
Turnover per employee	20.87	24.15	23.67	306	21
Value added per employee	13.52	13.10	6.43	304	21

**Original data**

	OLS estimation of the logarithmised characteristics on export start in $t = 2006$ and controls in $t, t - 1$ and $t - 2$			number of	
	Two years before starting ( $t - 2$ )	One year before starting( $t - 1$ )	In the starting year ( $t$ )	non- exporters	export starters
West Germany					
Turnover	43.48**	45.89**	52.19**	9,171	600
Average wage	22.22**	23.26**	21.79**	9,171	600
Turnover per employee	35.90**	38.30**	39.40**	9,171	600
Value added per employee	21.63**	22.56**	24.61**	9,126	597
East Germany					
Turnover	32.91**	36.67**	45.02**	2,690	129
Average wage	19.78**	17.57**	18.25**	2,690	129
Turnover per employee	23.96**	23.68**	28.65**	2,690	129
Value added per employee	19.26**	19.61**	5.53+	2,676	129

*Note:* The estimated regression coefficients and the levels of significance (+ indicates significance at the 10% level, \* at the 5% level, and \*\* at the 1% level, based on robust standard errors) are presented from OLS estimations of the log turnover, log average wage, log turnover per employed persons and log value added per employed persons at  $t - 2$ ,  $t - 1$  and  $t$ . The control vector contains a full set of economic activity (2-digit) dummies, the number of employed persons and its squared value. To facilitate the interpretation, the estimated coefficient for the export starter dummy on the log variables has been transformed by  $100(\exp(\beta) - 1)$ . The transformation shows the average percentage difference in the respective variables at  $t - 2$ ,  $t - 1$  and  $t$  between enterprises that begin exporting at  $t$  and enterprises that do not start to export. Firms that belong to the 1st or 99th percentile of the wage, turnover profitability or value added distribution are excluded from all computations.

For West Germany the results do not differ much between the original data and the KombiFiD agreement sample. Based on both datasets prospective exporters in West Germany are economically and statistically significant larger (in terms of turnover), more productive and pay economically and statistically significant higher average wages than non-starters, even in the periods before the prospective exporters begin to export.

For East Germany the picture is different. The pre-export premia of the average wage and the productivity variables are no longer statistically significant and the pre-export turnover premia is only significant at the 10 percent level when the KombiFiD agreement sample is used. In contrast all pre-export premia are statistically highly significant when the original dataset is used. The missing significance of the pre-export premia may be because of the small number (21 firms) of East German business services enterprises that began to export.

Beside the look at the significance of the pre-export premia it has to be mentioned that the size of the coefficients differs in some cases a lot between the two datasets. Thus, we find for example in West Germany a pre-export premia of the the turnover per employee in  $t - 2$  that is eleven percentage points higher when estimated with the KombiFiD agreement sample. In East Germany we find an eleven percentage point lower pre-export premia of the average wage in  $t - 2$  when the KombiFiD agreement sample is used.

### 3. Conclusion

This study tests whether the KombiFiD sample can be regarded as a high quality data set for empirical research on enterprises from business services industries. It performs an empirical investigation using the original data in a first step and replicates exactly this investigation using the KombiFiD sample in a second step. We find that large business services firms are oversampled in the KombiFiD agreement sample which leads to a higher share of exporting business services firms compared to the original data. After controlling for firm size and industries results based on the original data and on the KombiFiD sample are highly similar for West German firms. Therefore, the KombiFiD sample can be regarded as a sound base for empirical studies on West German firms from business services industries. For East Germany, however, the number of business services firms seems to be too small for empirical analyses, at least in the field of firm's export participation.

## References

- Eichengreen, B./Gupta, P.* (2011): The two waves of service-sector growth, Oxford Economic Papers (forthcoming; advanced access published December 19, 2011).
- Jorgenson, D. W./Timmer, M. P.* (2011): Structural Change in Advanced Nations: A New Set of Stylised Facts, *Scandinavian Journal of Economics* 113 (1), 1–29.
- Kaiser, U./Wagner, J.* (2008): Neue Möglichkeiten zur Nutzung vertraulicher amtlicher Personen- und Firmendaten, *Perspektiven der Wirtschaftspolitik* 9 (3): 329–349.
- Statistisches Bundesamt* (2009): Der Dienstleistungssektor – Wirtschaftsmotor in Deutschland. Ausgewählte Ergebnisse von 2003 bis 2008, Wiesbaden.
- Vogel, A.* (2009): The German Business Services Statistics Panel 2003–2007, *Schmollers Jahrbuch/Journal of Applied Social Science Studies* 129 (4), 515–522.
- Vogel, A.* (2011): Exporter Performance in the German Business Services Sector, *The Service Industries Journal* 31 (7), 1015–1031.
- Wagner, J.* (2012): The Quality of the KombiFiD-Sample of Enterprises from Manufacturing Industries: Evidence from a Replication Study, *Schmollers Jahrbuch/Journal of Applied Social Science Studies* 132 (3), 379–392.