Schmollers Jahrbuch 132 (2012), 379 – 392
Duncker & Humblot, Berlin

The Quality of the *KombiFiD*-Sample of Enterprises from Manufacturing Industries: Evidence from a Replication Study

By 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 manufacturing 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. For West Germany a comparison of the results based the original data and on the KombiFiD sample points to by and large highly similar results. Contrary to this the big picture is not in favour of the quality of the KombiFiD sample for East Germany where the KombiFiD sample is too small and differences between the results based on this sample and the original data are too large to suggest the use of the KombiFiD data in empirical investigations.

JEL Classification: C81

1. Motivation

Micro data at the level of the firm – the establishment (local production unit) or the enterprise (legal unit) – are an indispensable tool for empirical research in a wide range of economic fields including industrial economics, labour economics and international economics. In Germany most high-quality firm level data are collected in surveys conducted by the statistical offices. The German Federal Statistical Office and the statistical offices of the German federal states opened research data centres (described in detail in Zühlke et al., 2004) in 2001 and 2002. This started a new era for researchers working in empirical econom-

* 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 Lueneburg, 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 I am alone responsible for the quality study presented here and the conclusions drawn.

ics. Access to confidential data for firms that were collected in surveys performed by the statistical offices became easy by using these research data centres (RDC). The number and variety of data sets provided by the RDC increased steadily (see for an overview Kaiser / Wagner, 2008), and so did the use of it by researchers. The high potential of these data as a basis to generate new stylized facts, to motivate assumptions used in formal theoretical models, to test theoretical hypotheses econometrically, and to be used in policy consultation and evaluation is documented in a large and growing number of publications.[1]

From their start the RDC offered access to micro level panel data that linked information from various waves of a survey over time. These panel data enormously extended the research potential of data from official statistics by allowing dynamic analyses and control for unobserved heterogeneity via panel econometric methods. Compared to this first generation of firm panel data sets, a second generation of data sets which became available recently has an even higher research potential. These new data combine information for firms gathered in different surveys that could not be analyzed jointly before. Merging firm level data from different surveys to construct data sets that cover information on a wider range of variables than the ones collected in any of these surveys, one at a time, is the basic idea of the project *AFiD* which is in detail described in Malchin / Voshage (2009). *AFiD* is an acronym for the German *A*mtliche *Fi*rmendaten für *D*eutschland (official firm data for Germany). Merging of firm data from different sources of official statistics is legal according to § 13a BStatG (Bundesstatistikgesetz, or federal statistics law), and it is technically feasible because an identical firm identifier is used in the different surveys. Furthermore, it is legal to add firm level data from other sources (e.g. from data bases offered by commercial providers of firm level information) to the data from official statistics provided that these data are publicly available (see Wagner, 2010a).

The latest generation of firm panel data that includes information from various surveys conducted by the statistical offices over time and firm level information from external publicly available sources offers rich potential for empirical research. The information provided in these data sets, however, is still far from complete. To mention only two important points in these data sets there is no information available on foreign direct investment (FDI) of the firms and on the structure of the employees with regard to age, level of qualification etc. This information is available from data sets prepared by other data producers – FDI data at the firm level are offered by the German central bank and detailed information on the employees in a firm are offered by the Institute for Employment Research (IAB) of the Federal Employment Agency. Obviously merging firm level data from various producers would increase the value of the data enormously.

[1] For partial surveys, see Wagner (2006, 2008).

Linking confidential firm level information across the borders of the data producers, however, is difficult. 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 *Kombi*nierte *Fi*rmendaten für *D*eutschland, or combined firm level data for Germany) that is in detail described on the web (see 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 these three data producers in one data set. These matched data will then be made available for scientific research while strictly obeying the data protection law, i.e. without revealing micro level information to researchers outside the data producing agencies. In KombiFiD 54,960 firms were asked to agree in written form to merge firm level information kept inside the statistical offices, the IAB and the German central bank. 30,944 firms replied and 16,571 agreed. These 16,571 firms are in the *KombiFiD Agreement Sample*. This data set is used here, and the term KombiFiD sample is used for it.

While the firm level data from the three data producing institutions 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[2] 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 manufacturing industries.

2. Exports and Firm Characteristics in German Manufacturing Industries

In a recent paper Wagner (2010b) reports results of a comprehensive empirical study on the links between firm characteristics and export performance. This study uses firm level data from the AFiD panel industry enterprises (see Malchin/Voshage, 2009) provided by the RDC of the statistical offices of the German federal states. All variables used in this study are available in the KombiFiD sample for the firms from manufacturing industries.[3] The basic idea ex-

[2] 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.

[3] This statement is not exactly true. The information on the number of employees in a firm used in Wagner (2010b) is computed from the number of employees reported in the

plored here is to replicate the study from Wagner (2010b) using the KombiFiD sample made of all firms that agreed to the matching of their data across the boundaries of the data producers and to compare the results to shed light on the question whether the KombiFiD data are a reliable basis for empirical investigations for manufacturing enterprises.[4]

Table 1 indicates that about one third of the enterprises that are covered by the original data set can be found in the KombiFiD sample. For East Germany, this results in a fairly small sample. Given that even today a separate analysis of enterprises from East and West Germany is necessary due to large differences in the economic performance in both parts of Germany this might cast doubts on the usefulness of the KombiFiD sample for analyses of manufacturing firms from East Germany.

Furthermore, Table 1 demonstrates that firms that never exported (over the four years investigated in the study) are underrepresented in the KombiFiD sample while firms that exported every year are oversampled. These differences are larger for East Germany than for West Germany.

Table 2 reports the distribution of characteristics of exporters and non-exporters from both parts of Germany in the original data set and in the KombiFiD sample in 2003, the first year of the study. The distributions are fairly similar[5] in the original data and in the KombiFiD sample. Note, however, that the average number of employees in the firms tends to be somewhat larger in the KombiFiD sample. This fits with the observation mentioned above that the share of firms that exported in every year (in no year) is larger (smaller) in the KombiFiD sample than in the original data.

---

monthly survey of establishments in manufacturing. This information is not available in the KombiFiD sample. The number of employees reported in the cost structure survey is used instead. The correlation between these two variables is +0.9954 in 2003, indicating that both variables are nearly identical. Although the correlation is nearly perfect all computations were performed here with the new variable for the number of employees for both the original data and the KombiFiD sample. This explains the small differences between some of the results based on the original data reported in Wagner (2010b) and in the replication study here.

4 The focus of this paper is on the quality of the data from the KombiFiD sample. Therefore, neither the theoretical background of the empirical models estimated nor the economic conclusions drawn from the results are discussed; see Wagner (2010b) for economic flesh to the bones.

5 Note that it is not possible to test for the equality of distribution across samples because it is illegal to pool data from the original sample and the KombiFiD sample.

Table 1

Export participation patterns 2003 – 2006

West Germany: Original data

Pattern	Frequency	Percent	Cumulated
0000	1,658	16.52	16.52
0001	99	0.99	17.50
0010	23	0.23	17.73
0011	56	0.56	18.29
0100	25	0.25	18.54
0101	10	0.10	18.64
0110	12	0.12	18.76
0111	80	0.80	19.56
1000	74	0.74	20.29
1001	11	0.11	20.40
1010	7	0.07	20.47
1011	19	0.19	20.66
1100	31	0.31	20.97
1101	12	0.12	21.09
1110	41	0.41	21.50
1111	7,880	78.50	100.00
Total	10,038	100.00	

West Germany: KombiFiD data – Agreement sample

Pattern	Frequency	Percent	Cumulated
0000	1,658	16.52	16.52
0001	99	0.99	17.50
0010	23	0.23	17.73
0011	56	0.56	18.29
0100	25	0.25	18.54
0101	10	0.10	18.64
0110	12	0.12	18.76
0111	80	0.80	19.56
1000	74	0.74	20.29
1001	11	0.11	20.40
1010	7	0.07	20.47
1011	19	0.19	20.66
1100	31	0.31	20.97
1101	12	0.12	21.09
1110	41	0.41	21.50
1111	7,880	78.50	100.00
Total	10,038	100.00	

Continued next page

Table 1 (continued)

East Germany: Original data

Pattern	Frequency	Percent	Cumulated
0000	565	30.51	30.51
0001	45	2.43	32.94
0010	8	0.43	33.37
0011	26	1.40	34.77
0100	8	0.43	35.21
0101	7	0.38	35.58
0110	8	0.43	36.02
0111	28	1.51	37.53
1000	25	1.35	38.88
1001	5	0.27	39.15
1010	3	0.16	39.31
1011	12	0.65	39.96
1100	12	0.65	40.60
1101	6	0.32	40.93
1110	17	0.92	41.85
1111	1,077	58.15	100.00
Total	1,852	100.00	

East Germany: KombiFiD data – Agreement sample

Pattern	Frequency	Percent	Cumulated
0000	180	26.32	26.32
0001	11	1.61	27.92
0010	3	0.44	28.36
0011	8	1.17	29.53
0100	XXX	XXX	XXX
0101	XXX	XXX	XXX
0110	XXX	XXX	XXX
0111	5	0.73	31.14
1000	10	1.46	32.60
1001	XXX	XXX	XXX
1010	XXX	XXX	XXX
1011	5	0.73	33.63
1100	4	0.58	34.21
1101	XXX	XXX	XXX
1110	4	0.58	34.94
1111	445	65.06	100.00
Total	684	100.00	

Note: Frequency is the number of enterprises with a pattern. A zero indicates that an enterprise did not export in a year, a one indicates that it did export. A firm with the pattern 0000 did never export between 2003 and 2006, a firm with the pattern 0001 exported only in the last year (2006), etc. XXX indicates that there are between one and two cases; this number has to be treated as confidential.

Table 2

Distribution of characteristics of exporters and non-exporters in 2003

		Share of exports in total sales (percent)	Number of employees	Physical capital per employee (Euro)	Wage per employee per year (Euro)	Share of R&D expend. in total sales (percent)
West Germany						
Original data						
Exporters (N = 8,075)	mean	31.62	412.46	89,047.93	33,225.46	1.25
	sd	24.79	2,951.78	102,511.2	10,273.19	2.97
	p1	0.11	22	1,788.25	10,978.35	0
	p5	0.97	28	10,085.77	18,065.56	0
	p25	10.39	55	35,216.2	26,911.46	0
	p50	26.89	118	63,453.77	32,885.24	0
	p75	49.54	298	109,589.5	38,847.5	1.20
	p95	78.98	1,252	243,676.4	49,405.31	6.45
	p99	93.46	3,656	457,971.1	59,956.52	13.81
Non-exporters (N = 1,963)	mean	0	132.92	73,771.79	26,652.00	0.18
	sd	0	238.03	108,221.0	11,955.45	1.22
	p1	0	21	881.34	3,596.30	0
	p5	0	24	4,549.22	7,060.43	0
	p25	0	36	20,290.49	19,401.84	0
	p50	0	59	41,792.13	26,463.31	0
	p75	0	120	87,254.01	33,336.76	0
	p95	0	525	233,180.7	45,037.39	0.43
	p99	0	1,187	545,217.6	59,168.65	5.76
KombiFiD data – Agreement sample						
Exporters (N = 3,294)	mean	32.63	505.43	91,776.13	33,706.79	1.40
	sd	24.73	4,056.61	107,014	10,565.01	3.01
	p1	0.16	22	2,172.43	11,977.87	0
	p5	1.22	28	11,134.62	18,598.14	0
	p25	11.10	60	37,471.35	27,759.82	0
	p50	28.60	133	66,052.25	33,335.38	0
	p75	51.14	330	114,548.7	39,080.59	1.56
	p95	79.40	1,323	244,822.7	49,701.05	6.97
	p99	94.17	4,522	431,836.00	60,340.57	14.29
Non-exporters (N = 693)	mean	0	157.43	83,309.81	27,216.75	0.20
	sd	0	280.80	131,834.3	12,288.36	1.29
	p1	0	21	748	4,214.29	0
	p5	0	24	6,615.73	7,846.15	0
	p25	0	37	23,562.98	19,566.14	0
	p50	0	66	47,890.77	27,134.58	0
	p75	0	140	92,394.00	33,453.55	0
	p95	0	641	248,462.9	46,201.14	0.69
	p99	0	1,209	581,413.6	62,481.18	6.28

Continued next page

Table 2 (continued)

		Share of exports in total sales (percent)	Number of employees	Physical capital per employee (Euro)	Wage per employee per year (Euro)	Share of R&D expend. in total sales (percent)
East Germany						
Original data						
Exporters (N = 1,157)	mean	25.78	201.27	147,509.9	24,350.37	1.63
	sd	24.87	542.68	244,110.3	8,401.32	4.74
	p1	0.03	22	2,549.01	8,212.23	0
	p5	0.35	25	14,161.62	13,208.97	0
	p25	5.04	44	44,917.57	18,921.51	0
	p50	17.83	83	89,609.07	23,149.81	0
	p75	40.43	173	159,656.4	28,643.04	1.04
	p95	77.43	634	460,087.3	39,819.79	9.35
	p99	96.51	2,070	980,751.9	49,494.31	21.08
Non-exporters (N = 695)	mean	0	126.07	127,668	20,765.25	0.28
	sd	0	363.74	434,151.9	10,022.81	1.26
	p1	0	20	1,913.78	3,613.04	0
	p5	0	23	8,814.00	9,609.26	0
	p25	0	35	27,986.15	15,207.94	0
	p50	0	56	57,968.74	19,332.38	0
	p75	0	107	126,857.1	24,283.16	0
	p95	0	371	402,719	35,590.24	1.48
	p99	0	1,107	865,491.4	46,278.13	7.37
KombiFiD data – Agreement sample						
Exporters (N = 471)	mean	27.55	213.17	140,523.8	24,814.68	1.86
	sd	25.04	585.12	160,156.9	7,989.62	5.22
	p1	0.03	22	3,182.56	10,476.07	0
	p5	0.41	25	16,578.36	14,737.18	0
	p25	6.42	43	45,626.19	19,481.00	0
	p50	20.13	83	94,566.66	23,436.62	0
	p75	44.13	180	154,116.1	29,022.13	1.47
	p95	75.95	649	435,317.9	38,684.29	9.62
	p99	94.71	2,918	908,272.3	53,445.36	23.80
Non-exporters (N = 213)	mean	0	122.96	107,715.5	20,979.83	0.54
	sd	0	321.94	152,463.4	8,442.35	1.88
	p1	20	20	2,414.22	4,230.05	0
	p5	0	23	7,205.02	9,281.76	0
	p25	0	33	29,708.26	15,331.03	0
	p50	0	54	57,968.74	19,453.51	0
	p75	0	97	112,981.4	24,793.55	0
	p95	0	377	399,402.9	39,997.57	4.25
	p99	0	1107	640,933.1	46,278.13	10.03

Note: sd is the standard deviation; p1 is the first percentile, etc. The mimima and maxima are confidential because they are information for single enterprises.

A comparison of the firm characteristics among exporters and non-exporters based on the original data and on the KombiFiD sample leads to identical conclusions when the distributions of characteristics in both groups of firms are compared using a Kolmogorov-Smirnov test for first-order stochastic dominance (see Table 3): Exporters are larger, employ more physical capital per employee, pay higher wages per employee and spend a higher share of total sales for R&D expenditures than non-exporters in both parts of Germany.

Table 3

Differences between exporters and non-exporters: Distributions of characteristics in 2003

	West Germany		East Germany	
	Original data	KombiFiD data – Agreement sample	Original data	KombiFiD data – Agreement sample
Number of employees				
K-S-Test H0: equality of distributions (*p*-value)	0.000	0.000	0.000	0.000
K-S-Test H0: differences favourable for non-exporters (*p*-value)	0.000	0.000	0.000	0.000
K-S-Test H0: differences favourable for exporters (*p*-value)	1.000	1.000	1.000	1.000
Physical capital per employee (Euro)				
K-S-Test H0: equality of distributions (*p*-value)	0.000	0.000	0.000	0.000
K-S-Test H0: differences favourable for non-exporters (*p*-value)	0.000	0.000	0.000	0.000
K-S-Test H0: differences favourable for exporters (*p*-value)	0.898	0.817	0.994	0.994
Wage per employee per year (Euro)				
K-S-Test H0: equality of distributions (*p*-value)	0.000	0.000	0.000	0.000
K-S-Test H0: differences favourable for non-exporters (*p*-value)	0.000	0.000	0.000	0.000
K-S-Test H0: differences favourable for exporters (*p*-value)	0.969	0.967	0.997	0.920
Share of R&D expenditures in total sales (percent)				
K-S-Test H0: equality of distributions (*p*-value)	0.000	0.000	0.000	0.000
K-S-Test H0: differences favourable for non-exporters (*p*-value)	0.000	0.000	0.000	0.000
K-S-Test H0: differences favourable for exporters (*p*-value)	1.000	1.000	1.000	1.000

Note: K-S-Test is the Kolmogorov-Smirnov test for first-order stochastic dominance.

By and large, empirical models for the determinants of participation of firms in the export market lead to identical results when estimated with the original data or with the KombiFiD sample. The one exception is the estimated coefficient for the physical capital per employee. As can be seen from Table 4 based on the original data the physical capital intensity is unrelated to the probability of exporting for firms from West Germany, while the estimated coefficient is statistically highly significant when the KombiFiD sample is used. Note, however, that the physical capital per employee is not computed from information collected in the survey by the statistical office directly. Instead it is estimated from information on the amount of depreciation reported by the firms, information on the composition of the capital stock into buildings and equipments at the industry level and information on the economic lives of buildings and equipments for the economy as a whole (for details see Wagner 2010c). Therefore, the quality of the capital intensity variable itself is doubtful, and the different results based on the original data and the KombiFiD sample should not be viewed as a convincing argument against the quality of the KombiFiD sample.

Table 4

Determinants of export participation: Probit-estimates

		West Germany		East Germany	
		Original data	KombiFiD data – Agreement sample	Original data	KombiFiD data – Agreement sample
Number of employees	β	0.00015	0.00015	0.00044	0.00068
	p	0.000	0.000	0.000	0.003
Number of employees (squared)	β	-9.27e-10	-9.17e-10	-5.62e-8	-7.74e-8
	p	0.000	0.000	0.001	0.004
Physical capital per employee (Euro)	β	-4.32e-8	-1.53e-7	1.70e-9	1.19e-7
	p	0.387	0.010	0.979	0.612
Wage per employee per year (Euro)	β	5.86e-6	5.13e-6	7.06e-6	1.48e-5
	p	0.000	0.000	0.001	0.002
Share of R&D expend. in total sales (%)	β	0.012	0.015	0.039	0.017
	p	0.009	0.024	0.000	0.091
Number of cases		9,410	3,294	1,597	448

Note: β is the estimated marginal effect at the mean of the independent variable; p is the prob-value. All models include a full set of 4digit industry-dummies plus a constant. The models are estimated for data from 2005.

Table 5 demonstrates that similar conclusions hold for empirical models for the share of exports in total sales. Results are rather similar for both the original data and the KombiFiD sample except for the estimated coefficients for physical capital per employee. The estimated coefficient for physical capital intensity

is statistically highly significant when the original data are used, but not significant at any conventional error level when the KombiFiD data are used.

Table 5

Determinants of the share of exports in total sales: Fractional logit estimates

		West Germany – 2003		West Germany – 2004	
		Original data	KombiFiD data – Agreement sample	Original data	KombiFiD data – Agreement sample
Number of employees	β	0.000083	0.000064	0.000081	0.000061
	p	0.000	0.003	0.000	0.004
Number of employees (squared)	β	-5.39e-10	-3.92e-10	-5.15e-10	-3.57e-10
	p	0.000	0.004	0.000	0.006
Physical capital per employee (Euro)	β	6.56e-7	3.18e-7	6.95e-7	2.76e-7
	p	0.001	0.251	0.000	0.274
Wage per employee per year (Euro)	β	0.000030	0.000027	0.000032	0.000037
	p	0.000	0.000	0.000	0.000
Share of R&D expend. in total sales (%)	β	0.057	0.058	0.052	0.035
	p	0.000	0.000	0.000	0.083
Number of cases		10,038	3,987	10,038	3,987

Note: β is the estimated regression coefficient; p is the prob-value. All models include a full set of 4digit industry-dummy variables plus a constant. The model cannot be estimated for East Germany; Stata reports that the variance matrix is non-symmetric or highly singular.

Using pooled data for the years 2003 to 2006 a decomposition of the overall variation of firm characteristics into variation between firms and variation within firms over time shows similar results for both data sets.

When fixed enterprise effects are included in the empirical models for export participation and the share of exports in total sales to control for time invariant unobserved firm specific characteristics (like the quality and the strategy of management) results for West Germany do not differ much between the two data sets (see Table 7). The only exception is the different sign for the coefficient estimated for the share of R&D expenditures in total sales in the two samples – this coefficient, however, is not statistically different from zero at a conventional error level in both cases. For East Germany, however, we get a completely different picture. Firm size is no longer statistically significant when the KombiFiD sample is used, and the wage per employee (the proxy variable for human capital intensity) is positively related to export participation at an error level of five percent while the estimated coefficient is insignificant with $p = 0.217$ when the model is estimated with the original data.

Table 6

Decomposition of overall variation into between and within variation

		West Germany		East Germany	
		Original data	KombiFiD data – Agreement sample	Original data	KombiFiD data – Agreement sample
Exporter (Dummy; 1 = yes)	overall	0.39	0.37	0.48	0.46
	between	0.38	0.36	0.46	0.44
	within	0.10	0.10	0.15	0.13
Share of exports in total sales	overall	26.10	26.17	23.87	24.64
	between	25.67	25.76	23.18	24.0
	within	4.72	4.64	5.71	5.57
Number of employees	overall	2,645.94	3,705.38	467.40	516.61
	between	2,642.66	3,700.91	465.37	515.09
	within	133.69	188.87	44.51	43.13
Physical capital per employee (Euro)	overall	105,417	115,620	259,887	155,722
	between	99,080	107,179	249,033	150,270
	within	36,009	43,394	74,489	41,146
Wage per employee per year (Euro)	overall	11,202	11,043	9,482	8,865
	between	10,668	10,436	9,161	8,557
	within	3,419	3,614	2,454	2,333
Share of R&D expend. in total sales (%)	overall	2.82	2.89	3.63	4.21
	between	2.60	2.73	3.33	3.81
	within	1.08	0.94	1.44	1.78
Number of observations		40,152	15,948	7,408	2,736
Number of firms		10,038	3,987	1,852	684

The bottom line, then, is that a comparison of the estimations using the original data and the KombiFiD sample points to highly similar results for West Germany (with the exception of the results related to the physical capital intensity). Contrary to this the big picture is not in favour of the quality of the KombiFiD sample for East Germany where the KombiFiD sample is too small and differences between the results based on this sample and the original data are too large to suggest the use of the KombiFiD data in empirical investigations.

Table 7

Determinants of export participation and the share of exports in total sales: Fixed effects panel estimates

		Export participation (Fixed effects logit)		Share of exports in total sales (Fractional probit panel)	
		Original data	KombiFiD data – Agreement sample	Original data	KombiFiD data – Agreement sample
West Germany					
Number of employees	β	0.0033	0.0059	9.16e-6	1.40e-5
	p	0.097	0.095	0.488	0.455
Number of employees (squared)	β	-1.51e-7	-5.72e-7	-4.47e-11	-6.26e-11
	p	0.727	0.354	0.364	0.363
Physical capital per employee (Euro)	β	-2.62e-6	-4.92e-6	-1.86e-8	-3.56e-8
	p	0.031	0.027	0.347	0.142
Wage per employee per year (Euro)	β	0.000018	0.000013	9.82e-7	9.80e-7
	p	0.094	0.094	0.002	0.019
Share of R&D expend. in total sales (%)	β	-0.015	0.164	-0.00047	-0.00085
	p	0.792	0.330	0.806	0.537
Number of observations		2,000	740	40,152	15,948
Number of firms		500	185	10,038	3,987
East Germany					
Number of employees	β	0.026	0.061	0.00057	0.00044
	p	0.001	0.169	0.000	0.073
Number of employees (squared)	β	-0.000014	0.000021	-3.28e-8	2.31e-8
	p	0.002	0.866	0.009	0.837
Physical capital per employee (Euro)	β	-2.17e-7	6.92e-7	-7.90e-8	-1.13e-8
	p	0.544	0.857	0.212	0.902
Wage per employee per year (Euro)	β	0.000035	0.00016	3.35e-6	6.87e-6
	p	0.217	0.027	0.034	0.003
Share of R&D expend. in total sales (%)	β	-0.013	-0.110	-0.0037	-0.0038
	p	0.879	0.447	0.049	0.111
Number of observations		840	236	7,408	2,736
Number of firms		210	59	1,852	684

Note: β is the estimated regression coefficient; p is the prob-value. All models include a full set of year-dummies; the fractional probit panel models include a full set of mean values of the exogenous variables plus a constant, too.

3. Conclusion

This paper shows that results from empirical investigations for enterprises from West German manufacturing industries based on the KombiFiD sample are very similar to results computed with the original data. Therefore, the KombiFiD sample can be regarded as a sound data base for empirical studies on West German firms from manufacturing industries. Further research is needed to investigate whether this holds for other data from other parts of the economy (service industries, trade, buildings and construction), too.

References

Kaiser, U. / *Wagner,* J. (2008): Neue Möglichkeiten zur Nutzung vertraulicher amtlicher Personen- und Firmendaten, Perspektiven der Wirtschaftspolitik 9 (3), 329 – 349.

Malchin, A. / *Voshage*, R. (2009): Official Firm Data for Germany, Schmollers Jahrbuch / Journal of Applied Social Science Studies 129 (3), 501 – 513.

Wagner, J. (2006): Politikrelevante Folgerungen aus Analysen mit Firmendaten der amtlichen Statistik, Schmollers Jahrbuch / Journal of Applied Social Science Studies 126 (3), 359 – 374.

Wagner, J. (2008): Die Forschungspotenziale der Betriebspaneldaten des Monatsberichts im Verarbeitenden Gewerbe, AStA – Wirtschafts- und Sozialstatistisches Archiv 2 (2), 209 – 221.

Wagner, J. (2010a): The Research Potential of New Types of Enterprise Data based on Surveys from Official Statistics in Germany, Schmollers Jahrbuch / Journal of Applied Social Science Studies 130 (1), 133 – 142.

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

Wagner, J. (2010c): Estimated Capital Stock Values for German Manufacturing Enterprises Covered by the Cost Structure Surveys, Schmollers Jahrbuch / Journal of Applied Social Science Studies 130 (3), 403 – 408.

Zühlke, S. / *Zwick*, M. / *Scharnhorst*, S. / *Wende*, Th. (2004): The reserach data centres of the Federal Statistical Office and the statistical offices of the Länder, Schmollers Jahrbuch / Journal of Applied Social Science Studies 129 (4), 501 – 513.