

## KombiFiD: A First Glimpse of the Bundesbank Data

By Elena Biewen\*

### Abstract

This contribution gives a first overview of the data in KombiFiD which refers to information on the balance sheets of firms (as collected in the Corporate Balance Sheet Statistics, USTAN) and also to information on foreign direct investment activities of German firms (as collected in Microdatabase Direct Investment, MiDi). Using descriptive statistics and a simple regression analysis we compare the results based on the original and the KombiFiD data. Although many statistics based on the KombiFiD sample are similar to the original, we observe partly large deviations, in particular regarding the MiDi sample. These differences may be attributable to the fact that large firms and firms from the manufacturing sector are over-represented in the KombiFiD sample.

*JEL Classification: D22, F23*

### 1. Introduction

Recently a unique cross-institutional dataset that provides a comprehensive information on German firms has been made available to the research community. This dataset combines several data provided by the Federal Statistical Office and Statistical Offices of the German Länder, Federal Employment Agency and Deutsche Bundesbank which were linked in the KombiFiD (combined firm data for Germany) project.<sup>1</sup> Now, for the first time research questions can be addressed in Germany that require the data of these three institutions. For example, studies that directly compare exporting firms and firms that undertake foreign direct investments, analysis of the qualification structure of the personnel in firms with FDI, and the effects of FDI on domestic labour market become feasible.

The data of the Deutsche Bundesbank in the KombiFiD project are Corporate Balance Sheet Statistics (USTAN) and Microdatabase Direct Investment

---

\* This paper represents the authors' personal opinions and does not necessarily reflect the views of the Deutsche Bundesbank.

<sup>1</sup> For more detailed information on KombiFiD project see Biewen, Gruhl, Gürke, Hethy-Maier, and Weiß in this issue.

(MiDi). Due to the fact that only part of the firms have agreed that data of the three institutions can be matched and as not all firms included in the dataset of the other project partners can be identified in the datasets of the Bundesbank the KombiFiD data referring to balance sheets and foreign direct investment is only available for a subsample of USTAN respectively MiDi.

The aim of this paper is to give a first overview of the quality of the Bundesbank data included in the KombiFiD sample. In doing so, we carry out some descriptive statistics and regression analysis on the basis of the original USTAN respectively MiDi data and replicate them using the KombiFiD data. As we do not have original data for the combined dataset, the study must be restricted to single datasets. Section 2 describes the study for the USTAN. In section 3 the results for the MiDi are presented. Section 4 concludes.

## 2. Corporate Balance Sheets (USTAN)

### 2.1 Descriptive Statistics

The Corporate Balance Sheets Statistics (USTAN) provides information on balance sheet items as well as the operating statement of non-financial firms which are collected by the Bundesbank in the context of refinancing transactions. The dataset is described in Stöss (2001) in more detail.

In the following analysis we exclude holding companies because their coverage in the KombiFiD data is not complete. As many holdings do not have employees liable to social security and are, therefore, not in the data of the German Federal Employment Agency<sup>2</sup>, they could not be linked in the KombiFiD project. The original data covers between 27,000 and 56,000 firms per year (Table 1). The KombiFiD data contains about 6–8 percent of firms from the original sample (between 2,134 and 3,257 firms annually).

*Table 1*

**Firm number in USTAN**

year	original	KombiFiD
2003	27,032	2,134
2004	39,449	2,625
2005	56,362	3,257
2006	45,764	3,041

*Data Base:* USTAN, authors' own calculations.

<sup>2</sup> The Establishment History Panel of the German Federal Employment Agency includes all firms with at least one employee liable to social security.

Starting with the distribution of legal forms (Table 2) we do not observe large differences between the original and KombiFiD sample. There are slightly more firms with the legal form AG/KGaA (joint stock company/partnership limited by shares) and slightly less “sole proprietorships, etc”.

Table 2

**Legal form: USTAN 2003–2006, in %**

legal form	original	KombiFiD
AG and KGaA	4.1	6.8
GmbH	63.6	63.2
Sole proprietorships, KG, OHG, OHG mbH, GmbH & Co. KG, AG & Co. KG, GbR, GbR mbH, e.V., e.G., a.G.	32.4	29.9
Branch or permanent establishment	0.01	0.03

*Data Base:* USTAN, authors’ own calculations.

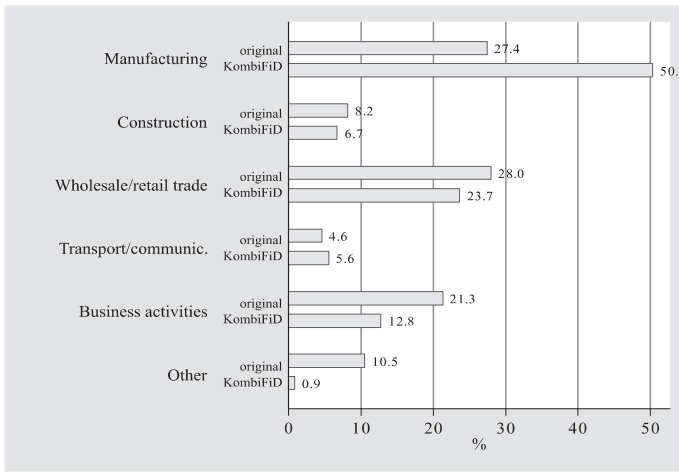
Figure 1 shows the distribution of firms across industrial sectors.<sup>3</sup> While every fourth firm in the original sample is from the manufacturing industry, in the KombiFiD sample this is the case for every other company. On the contrary, firms from the sectors “business activities” (including financial intermediation, property and housing, computer and related activities, research and development, other business-related services) and “other” are largely under-represented.

Figure 2 compares the distribution of employee numbers for both samples. First, we group the employee numbers into five larger categories (less than 50 employees; 50–249; 250–499; 500–999 and more than 1000). Almost 50 percent of the original USTAN sample are firms with less than 50 employees. In the KombiFiD small firms make up just 19 percent of the sample. Other size groups, on the contrary, are over-represented.

The fact that the KombiFiD sample consists of larger firms is reflected in the distribution of USTAN variables. Figure 3 shows for example the distribution of employment, balance sheet total, total sales, equity capital using kernel density estimations.<sup>4</sup> We observe a clear shift of the distributions to the right.

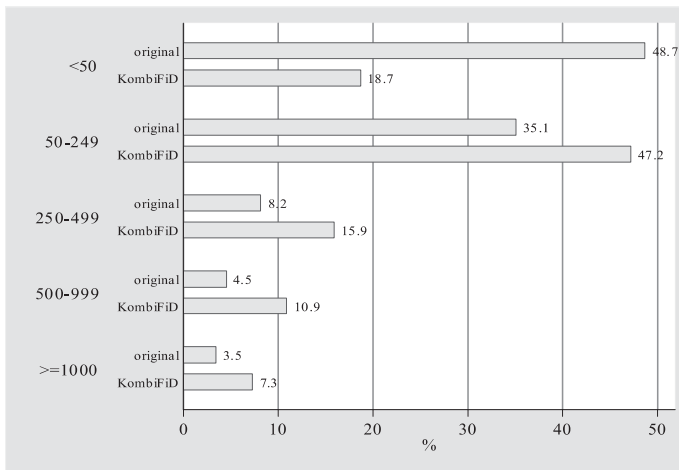
<sup>3</sup> In this figure pooled results are presented. The distributions for single years do not show large differences from the pooled results and are available from the author upon request. The same applies for other statistics presented in this paper.

<sup>4</sup> The distributions of other variables from the USTAN lead to similar conclusions and are available from the author upon request.



Data Base: USTAN, authors' own calculations.

Figure 1: Industrial sectors: USTAN 2003–2006

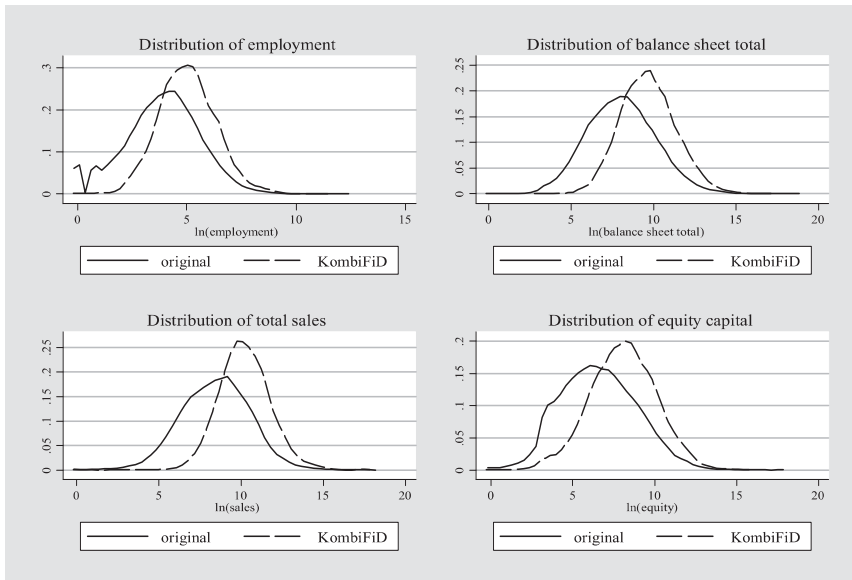


Data Base: USTAN, authors' own calculations.

Figure 2: Employment: USTAN 2003–2006

KombiFiD firms are on average larger, with a mean employee number of 421 versus 246 in the original sample (see Table 3). Their average balance sheet total is roughly twice as much, and total sales are even three times larger. Standard deviations are somewhat smaller in the KombiFiD sample which can be attributed to a smaller observation number. However, it is striking, that the dif-

ferences in the distributions are reduced if we calculate balance sheet ratios. Table 3 shows for example that differences in the equity ratio (calculated as the ratio equity/balance sheet total), current assets intensity (current assets/balance sheet total), returns on sales (profit before taxes/total sales), and leverage ratio (liabilities/equity) are cancelled out and means as well as percentiles generally remain very similar in both samples.



Data Base: USTAN, authors' own calculations.

Figure 3: Distribution of some USTAN variables: USTAN 2003–2006

## 2.2 Determinants of an Equity Ratio

Next, we estimate determinants of an equity ratio using OLS and panel techniques (fixed effects and random effects estimations) and compare the results on the basis of the original and KombiFiD samples. Specifically, we test how the return on sales, tangible assets ratio and firm size affect the equity ratio. We also control for the firm branch (two-digit) and legal form. The results are presented in Table 4. Although the coefficients are not exact the same in both samples, they are remarkably similar in most cases. The differences in coefficients are even smaller in the fixed and random effect estimations. We do not observe any sign changes: returns on sales and tangible assets ratio positively affect the equity ratio in both samples and all regressions. The effect of firm size remains always negative and suggests that larger companies tend to have lower equity ratios.

Table 3

**Distribution of some variables: USTAN 2003–2006**

		mean	sd	p5	p25	p50	p75	p95
Number of employees*	orig	246	2343.321	–	16	52	152	750
	KF	421	1896.392	19	63	146	352	1354
Balance sheet total	orig	44.435	831.721	0.111	0.710	2.955	12.187	103.383
	KF	85.321	520.051	1.079	4.837	14.767	45.630	285.152
Total sales	orig	47.345	617.358	0.064	0.950	4.476	18.291	120.030
	KF	139.688	1313.469	2.224	9.490	25.265	71.902	353.311
Labour productivity	orig	1.526	89.662	0.055	0.121	0.208	0.392	1.543
	KF	1.598	52.820	0.070	0.130	0.206	0.370	1.160
Equity ratio	orig	30.26	23.667	1.96	11.48	24.74	43.73	79.20
	KF	27.92	19.247	2.53	12.85	24.62	39.37	65.44
Current assets intensity	orig	63.45	29.793	6.31	41.55	70.68	89.48	99.46
	KF	64.25	24.867	13.51	48.03	67.87	84.92	97.37
Returns on sales	orig	3.59	24.413	–11.63	0.41	2.75	7.76	28.85
	Kf	3.40	13.672	–6.41	0.52	2.77	6.83	17.68
Leverage ratio	orig	207.54	147.551	15.84	85.37	177.55	309.23	494.00
	KF	230.57	140.310	18.19	42.37	116.88	327.29	499.44

\* Without observations equal 0 (not specified).

Note: According values are reported in mill. €. Balance sheet ratios are reported in %.

Data Base: USTAN, authors' own calculations.

Table 4

**Determinants of equity ratio: USTAN**

	OLS		RE		FE	
	original	KombiFiD	original	KombiFiD	original	KombiFiD
ln return on sales	0.1517 (0.000)	0.1258 (0.000)	0.0743 (0.000)	0.0639 (0.000)	0.0481 (0.000)	0.0420 (0.000)
ln tangible assets ratio	0.0843 (0.000)	0.1197 (0.000)	0.0788 (0.000)	0.1020 (0.000)	0.0632 (0.000)	0.0783 (0.000)
ln size	–0.0194 (0.000)	–0.0264 (0.010)	–0.0207 (0.000)	–0.0242 (0.092)	–0.0371 (0.000)	–0.0117 (0.736)
dummy 2004	0.0227 (0.054)	0.0410 (0.250)	0.0442 (0.000)	0.0455 (0.000)	0.0472 (0.000)	0.0472 (0.000)
dummy 2005	0.0380 (0.001)	0.0711 (0.037)	0.0891 (0.000)	0.0827 (0.000)	0.0955 (0.000)	0.0861 (0.000)
dummy 2006	0.0630 (0.000)	0.1131 (0.001)	0.1232 (0.000)	0.1142 (0.000)	0.1328 (0.000)	0.1169 (0.000)

constant	3.0372 (0.000)	3.0486 (0.000)	3.5405 (0.000)	3.2609 (0.000)	2.8353 (0.000)	2.7206 (0.000)
<i>N</i>	67,803	7,065	67,803	7,065	67,803	7,065
$R^2$ within			0.033	0.031	0.034	0.032
$R^2$ between			0.112	0.156	0.035	0.051
$R^2$ overall	0.122	0.159	0.112	0.150	0.032	0.042
Breusch-Pagan test			(0.000)	(0.000)		
Hausman test			(0.000)	(0.000)		

*Data Base:* USTAN, authors' own calculations.

However, there are some changes in the significance. The coefficient of firm size that is significant at 1% level becomes insignificant in fixed effects estimation and significant at 10% level using RE (in OLS at 5% level).  $R^2$  is slightly larger for the KombiFiD sample. The usual tests in panel models (Breusch-Pagan, Hausman) lead to the same implications in both samples.

### 3. Microdatabase Direct Investment (MiDi)

#### 3.1 Descriptive Statistics

The Microdatabase Direct Investment (MiDi) provides a comprehensive information on foreign direct investment activities of German firms abroad (outward FDI) as well as of foreign firms in Germany (inward FDI).<sup>5</sup> In the following analysis we consider the outward FDI only and as in the USTAN study we also exclude holding companies because of their incomplete coverage in the KombiFiD dataset. The sample on the basis of original data consists then of about 3,000 German investors per year (Table 5). About 11–13 percent of these firms are contained in the KombiFiD data. German investors held participating interests in 8,607 investment enterprises in 2003 with rising tendency to 9,231 enterprises in 2006. This increase is also reflected by the KombiFiD data which covers 13 percent of the affiliates.

The distribution of legal forms of investors (Table 6) is very similar in original and KombiFiD. The majority of investor firms have legal form GmbH (limited liability company) following by “sole proprietorships, etc.” and AG (joint-stock company)/KGaA (partnership limited by shares). However, the KombiFiD sample does not include any branch or permanent establishment. Nevertheless, since the branches make up only a very small part of the original sample, this fact can be neglected.

<sup>5</sup> The data is described in Lipponer (2003, 2009) in more detail.

Table 5

**Firm numbers in MiDi**

year	Number of investors		Number of direct investment enterprises	
	original	KombiFiD	original	KombiFiD
2003	3,052	329	8,607	1,074
2004	2,990	350	8,610	1,093
2005	2,978	383	8,800	1,164
2006	3,090	379	9,231	1,248

*Data Base:* MiDi, authors' own calculations.

Table 6

**Legal form (investor): MiDi 2003–2006, in %**

legal form	original	KombiFiD
AG and KgaA	18.67	17.21
GmbH	56.42	56.84
Sole proprietorships, KG, OHG, OHG mbH, GmbH & Co. KG, AG & Co. KG, GbR, GbR mbH, e.V., e.G., a.G.	24.80	25.95
Branch or permanent establishment	0.11	–

*Data Base:* MiDi, authors' own calculations.

From 2005 on the MiDi reports the mode of entry for the direct investment enterprise. In 2005, about 4 percent of direct investment enterprises were reported for the first time because of the exceeding of the reporting limits (Table 7). 2 percent of direct investment enterprises were greenfield investments, and slightly less than 2 percent were the result of mergers, acquisitions or takeovers. The KombiFiD data gives us a very similar picture.

Table 7

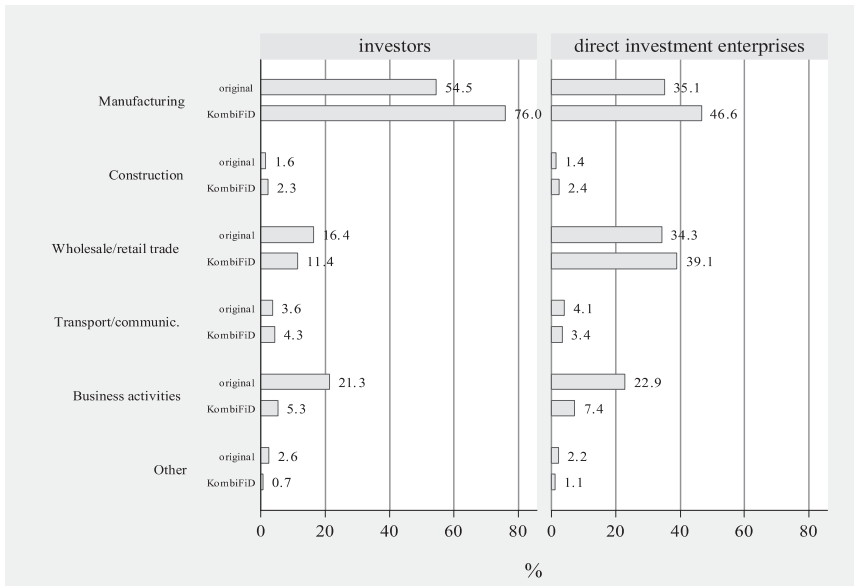
**Mode of entry of direct investment enterprises: MiDi 2003–2006, in %**

mode of entry	2005		2006	
	original	KombiFiD	original	KombiFiD
existing previously	91.87	92.35	90.38	91.75
new establishment/greenfield	2.17	1.63	2.36	1.84
merger/acquisition/takeover	1.88	1.80	2.39	1.36
exemption limits exceeded first time	4.08	4.21	4.87	5.05

*Data Base:* MiDi, authors' own calculations.



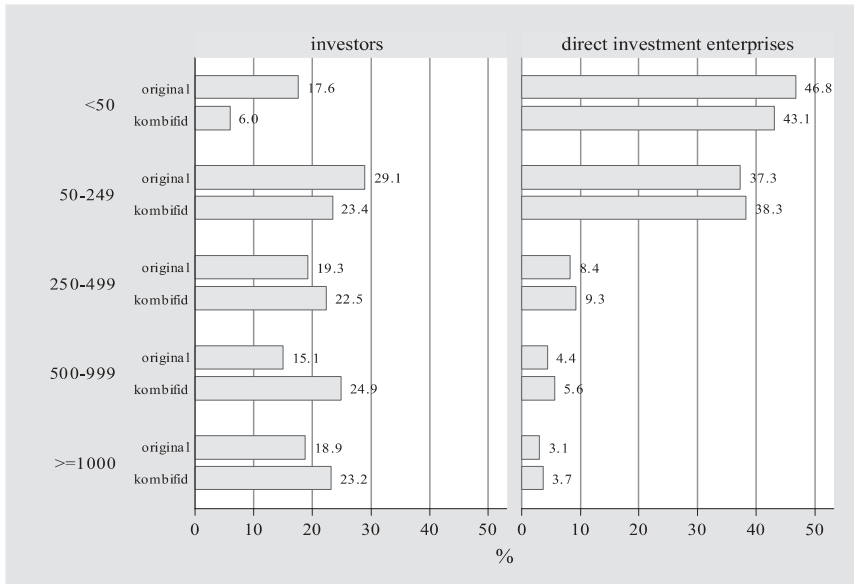
Figure 4 presents the distribution of German investors (graph on the left) and direct investment enterprises (graph on the right) across industrial sectors. The manufacturing sector makes up the largest share of investors in the original MiDi sample. It is also the most important sector in the KombiFiD sample, but it accounts for a much larger part of investors (76 percent in KombiFiD versus 55 percent in original). On the contrary, firms from the business activities sector are under-represented (5 in KombiFiD versus 21 percent in original). As for other industries, the differences between both samples are much less pronounced. In the case of direct investment enterprises, the differences between the original and KombiFiD sample are somewhat less distinct.



Data Base: MiDi, authors' own calculations.

Figure 4: Industrial sectors: MiDi 2003–2006

Figure 5 shows that there are not significant differences for the distribution of employee numbers in direct investment enterprises (graph on the right). In the case of German investors (graph on the left) the smaller size group is under-represented, e.g. firms with less than 50 employees make up just 6 percent of the KombiFiD sample compared to 18 percent in the original data. On the other hand, larger investors (more than 250 employees) are over-represented in the KombiFiD sample.



Data Base: MiDi, authors' own calculations.

Figure 5: Employment: MiDi 2003–2006

Table 8 reports the ten most important FDI markets. In a pooled original sample sales to the US account for about 22 percent of overall FDI, following by the Netherlands with 11% and the UK with 10%. Using the KombiFiD sample we find somewhat different results. The most top countries from the original sample are among the top partners in the KombiFiD sample, but their order is changed. The Netherlands is ranked first with 22% of FDI shares. The US accounts for just 12% of overall FDI. Luxembourg and Sweden do not appear in the KombiFiD and are replaced by Czech Republic and Poland. However, the top 10 countries make up 69% of total FDI in the original sample and slightly more (about 74%) in the KombiFiD.

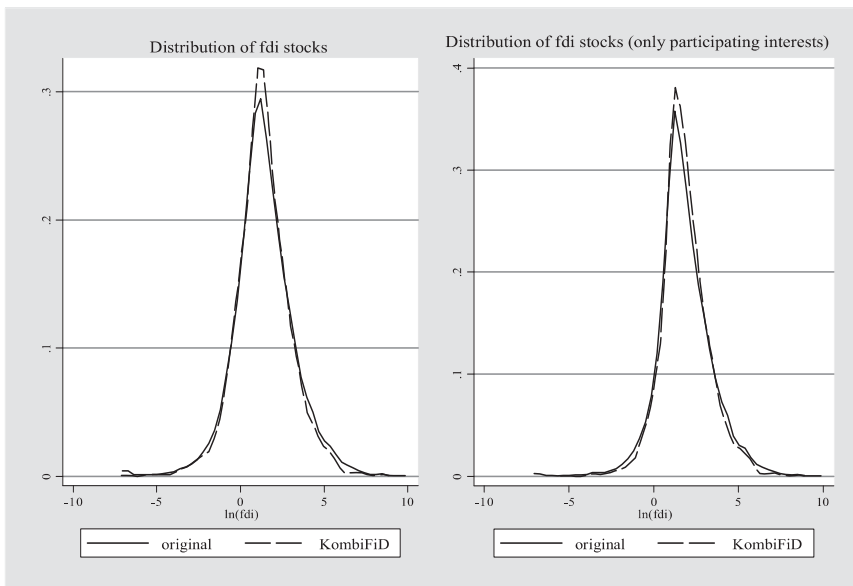
In the next step we compare distributions of some selected variables (direct FDI stocks, employment, total sales, balance sheet total, labour productivity) from the MiDi using the kernel density functions. Surprisingly, the distributions of FDI stocks in Figure 6 look very similar with minor differences, in particular in the middle of the distribution. This may reflect the fact that foreign direct investment activities are mainly driven by big companies. Therefore, the lack of small firms in the sample does not have major effects on the distribution of FDI. The graph on the right presents those FDI that include participating interests only (i.e. without loans).

Table 8

**Top 10 countries: MiDi 2003–2006**

rank	original			KombiFiD		
	country	sum FDI	% of total	country	sum FDI	% of total
1	USA	264,780.1	21.57	Netherlands	25,078.0	21.85
2	Netherlands	132,437.6	10.79	USA	13,574.6	11.83
3	UK	126,214.0	10.28	Belgium	13,180.5	11.48
4	France	68,139.6	5.55	UK	7,051.2	6.14
5	Luxembourg	66,846.7	5.45	Austria	6,461.2	5.63
6	Austria	46,703.3	3.81	France	5,406.7	4.71
7	Belgium	39,013.2	3.18	Czech Republic	4,387.3	3.82
8	Sweden	34,966.0	2.85	Italy	3,498.0	3.05
9	Italy	34,410.1	2.80	Poland	3,120.9	2.72
10	Spain	34,246.5	2.79	Spain	3,074.1	2.68

Data Base: MiDi, authors' own calculations. Values are reported in mill. €.



Data Base: MiDi, authors' own calculations.

Figure 6: Distribution of FDI variables: MiDi 2003–2006

Regarding other MiDi variables we observe that in the case of direct investment enterprises the original distributional form is approximated with the KombiFiD sample.<sup>6</sup> However, the distributions of the variables for investor firms show more differences. Generally they are shifted to the right indicating that investors in the KombiFiD sample are larger with larger sales and balance sheet total.

### 3.2 Determinants of FDI

In this section we estimate a simple linear model of determinants of foreign direct investment activities in order to examine the differences between the original and KombiFiD data. The choice of the determinants of FDI stems from Buch/Lipponer (2004). As this study uses information from other datasets, e.g. Dafne and Hoppenstedt, that is not in the KombiFiD, we restrict our analysis on the determinants available in the KombiFiD data. The dependent variable is the logarithms of direct FDI stocks. As determinants of FDI we test total assets of the investor, gross domestic product (gdp), geographical distance (measured as distance between the capitals), dummies for common border (equal 1 if a partner country is a neighbour of Germany) and common language (German), and an EU dummy (equal 1 if a partner country is a member of European Union). The continuous variables are entered in logarithmic form in the model. We also control for year and sectoral effects. Since some of the variables do not vary over time and would be cancelled out by FE (fixed effects) estimation, we only apply OLS and RE (random effects) estimations.

The results are reported in Table 9. As expected the regression based on the original data shows that larger firms undertake more outward FDI. German investors tend to hold affiliates in countries with larger GDP. The distance has a negative impact on FDI stocks, which can be attributed to the increase of transaction costs. Common border coefficient has an expected positive sign since foreign activities with neighboring countries involve e.g. lower transportation costs. The coefficient of the EU dummy is also as expected positive. Surprisingly, the effect of common language is negative in the OLS estimation. However, in the RE regression it becomes positive but insignificant.<sup>7</sup> If we compare these results with the KombiFiD sample, some tendencies can be made out. Highly significant regressors remain significant at 1% level, have the same coefficient signs as in the original data, but their effects are on general larger what may reflect the importance of large firms in the KombiFiD data. Highly insignificant coefficients maintain insignificance almost at a similar level. However, there are still some irregularities in the results: regressors that are signifi-

<sup>6</sup> For reasons of space we do not present here the distributions, but they are available from the author upon request.

<sup>7</sup> Buch/Lipponer (2004) point out that multicollinearity between common border and common language and EU membership may be a problem.

cant at higher levels may change into insignificant using the KombiFiD sample (e.g. common border, EU membership, common language in the OLS). Similar to USTAN  $R^2$  is somewhat larger in the KombiFiD sample. The Breusch-Pagan test for individual effects yields the same results in both samples.

Table 9

**Determinants of FDI: MiDi**

	OLS		RE	
	original	KombiFiD	original	KombiFiD
ln total assets	0.3583 (0.000)	0.4815 (0.000)	0.2987 (0.000)	0.4215 (0.000)
ln gdp	0.1408 (0.000)	0.1345 (0.000)	0.1856 (0.000)	0.1946 (0.000)
ln distance	-0.0753 (0.000)	-0.1605 (0.000)	-0.0726 (0.004)	-0.1679 (0.000)
common border	0.0521 (0.032)	0.0197 (0.739)	0.1007 (0.013)	0.0693 (0.436)
common language	-0.0630 (0.021)	-0.0265 (0.698)	0.0049 (0.930)	-0.0419 (0.755)
eu member	0.0428 (0.072)	-0.0740 (0.209)	0.1150 (0.005)	-0.0203 (0.812)
dummy 2004	0.0139 (0.534)	0.0302 (0.568)	0.0197 (0.124)	0.0398 (0.194)
dummy 2005	0.0649 (0.004)	0.0880 (0.096)	0.0791 (0.001)	0.1022 (0.006)
dummy 2006	0.0748 (0.001)	0.0798 (0.122)	0.1074 (0.000)	0.1126 (0.004)
constant	-3.3609 (0.000)	-3.0929 (0.000)	-4.2051 (0.000)	-4.3308 (0.000)
$N$	29,337	4,018	29,337	4,018
$R^2$ within			0.054	0.077
$R^2$ between			0.236	0.343
$R^2$ overall	0.287	0.326	0.275	0.302
Breusch-Pagan test			(0.000)	(0.000)

Data Base: MiDi, authors' own calculations.

#### 4. Summary

Using simple descriptive statistics as well as regression analysis we got a first impression of the quality of the Bundesbank data provided in the KombiFiD project. Some descriptive statistics on the basis of KombiFiD data approx-

imate the original statistics. However, it is striking that USTAN firms as well as German investors in the MiDi are larger than firms in the original sample. This is reflected in the right-shifted distributions of the USTAN and MiDi variables for German investors. However, the bias in the USTAN is strongly reduced by calculating balance sheet ratios. It is also much less pronounced in the distributions of MiDi variables for direct investment enterprises. The reason for the latter is probably the fact that the Bundesbank data was linked with the data of other institutions by comparing of the addresses of the reporting units, i.e. German investors in the case of outward FDI. If an investor firm could be linked, it appears in the KombiFiD data with all its direct investment enterprises. Therefore, a larger bias is to be expected on the side of investors and not the affiliates. Furthermore, we also found an over-representation of firms from manufacturing industrial sectors.

By and large, for the USTAN the regression analysis (OLS, fixed and random effects estimations) led to more similar results using the original and KombiFiD sample. The size of coefficients in the KombiFiD was close to the original, and all coefficient signs remained the same. Nevertheless, some changes in the significance of the coefficients could be observed. For the MiDi we observed more divergences. The effects of determinants of FDI stocks, that were highly significant or insignificant, were larger in the KombiFiD but did not change coefficient signs and significance level. Some irregularities arised for coefficients that were significant at higher levels, but this may be attributable to multicollinearity problems in the model.

This contribution indicated a larger number of big firms and firms from manufacturing in the KombiFiD sample. Therefore, it would be interesting to restrict the analysis just on the manufacturing sector. Furthermore, the study presented here does not use any weights and compares the samples as they are. Thus, next steps in testing the data quality must take into consideration the construction of appropriate weights. It is possible that their use in the KombiFiD sample could improve the results.

## References

- Buch, C./Lipponer, A.* (2004): FDI versus Cross-border Financial Services: The Globalisation of German Banks, Discussion Paper, Series 1: Economic Studies, No. 05/2004, Economic Research Centre, Deutsche Bundesbank.
- Lipponer, A.* (2003): Deutsche Bundesbank's FDI Micro Database, Schmollers Jahrbuch/ Journal of Applied Social Science Studies 123, 593–600.
- Lipponer, A.* (2009): Microdatabase Direct Investment – MiDi a brief guide. Deutsche Bundesbank. [http://www.bundesbank.de/download/vf/fdi/vfz\\_mikrodaten\\_guide.pdf](http://www.bundesbank.de/download/vf/fdi/vfz_mikrodaten_guide.pdf).
- Stöss, E.* (2001): Deutsche Bundesbank's Corporate Balance Sheet Statistics and Areas of Application, Schmollers Jahrbuch/ Journal of Applied Social Science Studies 121, 131–137.