

# Research Log: Corporate Transparency

ZHEN YANG

May 2025

GitHub URL: <https://github.com/zhenyang-econ/rct25.git>

## 1 Identify Local Firms

### 1.1 Data Collection and Verification

In the first step, i used Google Map to identify GmbH and AG located in my area(13353).

In the second step, i accessed the company website and the official website to verify their existence and collected their information .

In addition, all information was cross-checked against the Unternehmensregister to ensure accuracy and legitimacy.

Table 1: Corporate Register Information of Local Firms

Company Name	Postal Code	Address	Website	Legal Disclosure
ABAP Straßenbau GmbH	13353	Westhafen straße 1	<a href="https://www.northdata.de/ABAP%20Stra%C3%9Fenbau%20GmbH,%20Berlin/Amtsgericht%20Charlottenburg%20(Berlin)%20HRB%2094429%20B">https://www.northdata.de/ABAP%20Stra%C3%9Fenbau%20GmbH,%20Berlin/Amtsgericht%20Charlottenburg%20(Berlin)%20HRB%2094429%20B</a>	available
BEHALA - Berliner Hafen- und Lagerhaus- gesellschaft mbH	13353	Westhafen straße 1	<a href="https://www.behala.de/">https://www.behala.de/</a>	available
Wilhelm Frese GmbH	13353	Genterstraße 18	<a href="https://www.frese-berlin.de/">https://www.frese-berlin.de/</a>	available
Drum Clean Recycling GmbH	13353	Westhafen straße 1	<a href="https://dcr-europe.com/">https://dcr-europe.com/</a>	
Berliner Werkstätten für Menschen mit Behinderung GmbH(BWB)	13353	Westhafen straße 4	<a href="https://www.bwb-gmbh.de/de/">https://www.bwb-gmbh.de/de/</a>	available

Note: "Legal Disclosure" refers to the availability of legal registration information.

## 2 Find Local Firms in Corporate Register

I collected corporate registration information and financial report filings from the *Unternehmensregister*. For each firm, I recorded their legal registration number (HRB/HRA), the type of available financial statements, and the number of filings.

I noticed that most firms had available financial reports starting from the fiscal year 2006. This is consistent with the implementation of the *EHUG*, requiring companies to file financial reports electronically through the *Bundesanzeiger* and *Unternehmensregister*.

Table 2: Corporate Register and Financial Report Information

Company Name	HRB/HRA	Type of Report	Fiscal Period	Number of Filings
ABAP Straßenbau GmbH	HRB 94429 B	Single entity	31.12.2006–31.12.2023	17(no report for 2021)
BEHALA - Berliner Hafen- und Lagerhaus- gesellschaft mbH	HRB 89951 B	single entity	31.12.2006–31.12.2023	20(one revise;one supplement)
Wilhelm Frese GmbH	HRB 72849 B	single entity	31.12.2006–31.12.2023	18
Drum Clean Recycling GmbH	HRB 180011 B	single entity	14.09.2016-12.31.2017	3(including founding balance sheet)
Berliner Werkstätten für Menschen mit Behinderung - Gesellschaft mit beschränkter Haftung (BWB)	HRB 4469 B	single entity	31.12.2006-31.12.2023	18

Table 3: Report Filing Date

Year	ABAP	BEHALA	Wilhelm Frese	Drum Clean	BWB	Notes
2006	08.02.2008	15.02.2008	16.01.2008	—	18.02.2008	BEHALA refiled (orig. 08.02.2012)
2007	09.02.2009	29.09.2008	07.08.2008	—	27.11.2008	
2008	09.02.2010	18.02.2010	21.10.2009	—	01.03.2010	
2009	20.12.2010	02.12.2010	16.12.2010	—	01.03.2011	
2010	23.03.2012	25.04.2012	18.11.2011	—	09.01.2012	
2011	03.01.2013	04.01.2013	25.09.2012	—	07.03.2013	BEHALA corrected (orig. 23.12.2016)
2012	07.03.2014	12.03.2014	05.12.2013	—	06.12.2013	
2013	05.01.2015	02.03.2015	29.09.2014	—	26.03.2015	
2014	10.02.2016	02.02.2016	08.10.2015	—	08.01.2016	
2015	31.08.2016	02.03.2017	22.12.2016	—	19.01.2017	
2016	13.12.2017	15.02.2018	07.09.2017	06.11.2018	28.11.2017	Drum Clean: founded 2016
2017	18.12.2018	11.02.2019	22.08.2018	07.02.2019	11.12.2018	Opening filed 03.01.2019
2018	02.01.2020	23.10.2019	22.01.2020	—	02.01.2020	ABAP miss- ing
2019	26.03.2021	20.01.2021	16.12.2020	—	05.01.2021	
2020	04.03.2022	14.02.2022	08.12.2021	—	13.09.2021	
2021	—	05.12.2022	14.11.2022	—	10.05.2023	
2022	16.04.2024	07.03.2024	08.02.2024	—	18.03.2024	
2023	24.04.2025	19.11.2024	10.01.2025	—	03.01.2025	

### 3 Prepare Orbis Data

I filtered the target firms in R. The resulting list of companies is shown in Table 4, generated via R (code provided below for reproducibility).

I observed that the Orbis database does not fully cover all available financial reports. Specifically, the number of firm-year observations in Orbis is fewer than the filings accessible via the *Unternehmensregister*.

Table 4: Orbis Company Match

Company Name	Found in Orbis	Fiscal period in Orbis	Number of Reports
ABAP Straßenbau GmbH	Yes	2011-2020	10
BEHALA GmbH	Yes	2013-2022	10
Wilhelm Frese GmbH	Yes	2013-2022	10
Drum Clean Recycling GmbH	Yes	2017	1
BWB GmbH	Yes	2013-2022	10

```
## Load the dataset
data <- readRDS("~/Desktop/second_semester/research_on_transparency/rct25-main/data/generated/
  orbis_panel_berlin.rds")
View(data)
str(data)

## Finding companies in postcode:13353; and my five companies
install.packages("kableExtra")
library(dplyr)
library(knitr)
library(kableExtra)

filtered_fivecompanies <- data %>%
  filter(name_native %in% c("ABAP_Straenbau_GmbH",
    "BEHALA_Berliner_Hafen_und_Lagerhausgesellschaft_umbH",
    "Wilhelm_Frese_GmbH",
    "Drum_Clean_Recycling_GmbH",
    "Berliner_Werksttten_frf_Menschen_mit_Behinderung_Gesellschaft_mit_
      beschrnkter_Haftung_BWB"))
View(filtered_fivecompanies)

company_table <- filtered_fivecompanies %>%
  select(name_native, year)

kable(company_table,
  caption = "Key_Characteristics_of_the_Five_Selected_Companies",
  booktabs = TRUE, format = "latex") %>%
  kable_styling(latex_options = c("striped", "hold_position"))
```

Listing 1: Filter and display R code

## 4 Largest Firm in 13353: Mobility Trader Holding GmbH

I filtered firms located in postal code 13353 from Orbis database. I then sorted these firms by total assets for the fiscal year ending in calendar year 2021, in descending order. I found out the largest total assets company in 13353 is Mobility Trader Holding GmbH.

All filtering and sorting operations were conducted using R. The corresponding code snippet is shown in List 2 for reproducibility.

Table 5: Mobility Trader Holding GmbH

2021 Yrs	Orbis	Unternehmensregister	Notes
Total assets	378,113,923	378,113,923.11	Matched
Book value of equity	373,218,405	373,218,405.10	Matched
Net Sales	not published	not available	

```
##sort and find the top five largest total assets candidates for analysis

yers2021companies13353 <- data %>%
  filter(postcode == 13353 & year == 2021)

assetstop1 <- yers2021companies13353 %>%
  arrange(desc(toas)) %>%
  slice(1:1)

View(assetstop1)

##book of equity:use SHFD(capi+osfd). total equity:comparable across the countries or CAPI: more

cols_to_keep <- c("name_native", "shfd", "toas", "turn")
mobilitytrader <- assetstop1 %>%
  select(any_of(cols_to_keep))
View(mobilitytrader)
```

Listing 2: R code for the largest total asset Firm in Orbis

## 5 Compare Firms in 13353 to the Berlin Firm Population

### 5.1 should i fixed the postcode effect or firm effect after fixing the time effect? do we need a constant intercept in our model?

Refer to Breuer and Dehaan (2024) "Using and Interpreting Fixed Effects Models" [1]

Fixing firm effect theoretically is a good approach to avoid the firm-level correlated omitted variables. But in the fixed effects mechanism, the dependent variables will demean the sample in the same firm, then it only consider the within group variation. But in our data, the firm total assets and equity ratio may not fluctuate over time, so the fixed firm effect will greatly decrease the variation in Y, making them flat over time. Also, if we fixed firm effect, dummy variables of 13353 explicitly demean it self, dummy variables will end up being zero in firms located 13353; i think it is also multicollinearity.

if we control postcode effect, it will cause serious multicollinearity. because we already include the dummy variable of the 13353. if we fix postcode effect, it will absorb all the variation in dummy variables, then decrease the test power. We will greatly decrease the X variation, making it useless in the model

We need to include the intercept in our model to absorb some unobserved effect caused by confounders. In addition we need to include fixed effect, because the time-fixed effect will greatly control the economic changes and other time-related problem.

### 5.2 should i cluster postcode standard errors.

Refer to Abadie et al. (2023) "When Should You Adjust Standard Errors for Clustering?" [2].

They offer a new framework on whether we should use clustered standard errors depending on the sampling process and the assignment process. In clustered treatment assignment, clustering becomes relevant. We should cluster the standard errors at the level of assignment.

In my research design, the treatment variable is an indicator for postcode 13353. It means my assignment is based on the postcode level, and all firms in the two groups share the same treatment within their specific group.

Regarding sampling, I include all firms within postcode 13353, which corresponds to clustered sampling. Given the clustered structure in both the assignment and the sampling processes, I should cluster the standard errors at the postcode level.

### 5.3 should i scale the Total Assets in my research.

In this setting, we run two separate regressions using two dependent variables *Total Assets* and *Equity Ratio*. Although the different scale of two variables will not affect the significant differences test in this problem. But in valid OLS, it relies on the normality assumption in the error term. if not, it will ruin our test efficiency. So I scale the total assets here. In my histogram, we can easily observe that the raw data is seriously right-skewed, after the log transformation, it follows normal distribution approximately. So i use the log transformation here

### 5.4 should i handle the unusual values in equity ratio

From the summary table, we observe that the equity ratio exhibits unusual values. Specifically, the 1st to 99th percentile range lies between -30 and 1.

Considering negative equity ratios, since firms experiencing severe financial distress often have negative equity, it is reasonable to retain these observations. Although negative equity ratios are economically plausible, the extreme imbalance between total assets and equity raises concerns.

To mitigate the influence of extreme outliers on both ends of the distribution, we winsorized the equity ratio at the 1st and 99th percentiles.

## 5.5 Research Design

### 5.5.1 Research Question

Is there a significant difference in firms' total assets and equity ratios between firms in the 13353 postcode and those in other Berlin regions?

### 5.5.2 Hypotheses

**Null Hypothesis 1 ( $H_0$ ):** There is no significant difference in firms' total assets between those located in the 13353 postcode and those located in other Berlin regions.

**Alternative Hypothesis 1 ( $H_1$ ):** There is a significant difference in firms' total assets between those located in the 13353 postcode and those located in other Berlin regions.

**Null Hypothesis 2 ( $H_0$ ):** There is no significant difference in firms' equity ratios between those located in the 13353 postcode and those located in other Berlin regions.

**Alternative Hypothesis 2 ( $H_1$ ):** There is a significant difference in firms' equity ratios between those located in the 13353 postcode and those located in other Berlin regions.

### 5.5.3 Data Source

This study uses the Orbis database, covering annual financial data of firms located in Berlin.

### 5.5.4 Variable Construction

#### Independent Variable

- **postcode\_13353:** A binary indicator equal to 1 if the firm is located in the 13353 postal code area, and 0 otherwise.

#### Dependent Variables

- **EquityRatio\_wins:** The equity ratio (shareholders' equity divided by total assets), winsorized at the 1st and 99th percentiles.
- **log\_toas:** The natural logarithm of the firm's total assets.

### 5.5.5 Research Model

$$\log\_toas_{it} = \alpha_s + \beta \cdot \text{postcode\_13353}_i + \gamma_t + \varepsilon_{it}$$

$$\text{EquityRatio\_wins}_{it} = \alpha_o + \beta \cdot \text{postcode\_13353}_i + \gamma_t + \varepsilon_{it}$$

where

- $\alpha$ : the intercept term
- $\gamma_t$ : captures year fixed effects.
- $\varepsilon_{it}$ : the error term for firm  $i$  in year  $t$ .
- $\text{postcode\_13353}_i$ : a binary indicator equal to 1 if firm  $i$  is located in postcode 13353; otherwise, 0. This is a time-invariant firm characteristic.
- $\log\_toas_{it}$ : the natural logarithm of total assets for firm  $i$  in year  $t$ .

### 5.5.6 Fixed Effects and Robustness

- **Year Fixed Effects:** To control for time-variant macroeconomic factors.
- **Clustered Standard Errors:** Standard errors are clustered at the postcode level to account for intra-cluster correlation.

## References

- [1] **Breuer, M.** and **DeHaan, E.** (2024). *Using and interpreting fixed effects models*. *Journal of Accounting Research*, 62(4), 1183–1226. <https://doi.org/10.1111/1475-679X.12559>
- [2] Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. M. (2023). When should you adjust standard errors for clustering? *The Quarterly Journal of Economics*, 138(1), 1–35. <https://doi.org/10.1093/qje/qjac038>