RUHR ECONOMIC PAPERS

Philipp Großkurth

Dynamic Structure – Dynamic Results? Re-estimating Profit Shifting with Historical Ownership Data

rwi

#811

Imprint

Ruhr Economic Papers

Published by

RWI – Leibniz-Institut für Wirtschaftsforschung
Hohenzollernstr. 1-3, 45128 Essen, Germany

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

Editors

Editorial Office

Ruhr Economic Papers #811

Responsible Editor: Roland Döhrn

ISSN 1864-4872 (online) – ISBN 978-3-86788-940-7

Ruhr Economic Papers #811

Philipp Großkurth

Dynamic Structure - Dynamic Results? Re-estimating Profit Shifting with Historical Ownership Data

Bibliografische Informationen der Deutschen Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at **http://dnb.dnb.de**

RWI is funded by the Federal Government and the federal state of North Rhine-Westphalia.

http://dx.doi.org/10.4419/86788940
ISSN 1864-4872 (online)
ISBN 978-3-86788-940-7

Philipp Großkurth[1]

Dynamic Structure - Dynamic Results? Re-estimating Profit Shifting with Historical Ownership Data

Abstract

Ownership structures of multinational enterprises are commonly assumed to remain constant over time, both due to a lack of easily accessible panel data and to facilitate empirical analyses. This paper discusses the validity of this assumption and assesses its relevance in the context of profit shifting. A new method of reconstructing historical ownership information in Bureau van Dijk's ORBIS database reveals a highly dynamic environment. The validity of the assumption collapses with increasing panel length; ownership structures are rarely constant over time. Moreover, about 9 percent of firms with observed ownership data change owners in each year. The relevance of the assumption is tested by re-estimating indirect measures of profit shifting for selected benchmark samples. Assuming ownership structures as constant has a strong impact on sample composition, adding almost 29 percent of additional observations compared to historical ownership data. In the context of profit shifting, estimates based on constant ownership data are found to be larger in absolute magnitude compared to estimates based on historical ownership data.

JEL Classification: F23, H25, H26

Keywords: ORBIS; historical ownership data; MNE; profit shifting

June 2019

1 Philipp Großkurth, RWI. – All correspondence to: Philipp Großkurth, RWI, Büro Berlin, Invalidenstr. 112, 10115 Berlin, Germany, e-mail: philipp.grosskurth@rwi-essen.de

1. Introduction

The accuracy and reliability of empirical research is limited by the availability and quality of data. Research on multinational enterprises (MNE) is particularly demanding, because many questions of interest can only be answered by combining data from several countries. Compared to research on independent national firms, research on MNE requires a global scope as well as reliable information on the structural connections between the firms. Due to its global coverage and inclusion of information about corporate ownership structures, Bureau van Dijk's (BvD) ORBIS database, an extension of its predecessor AMADEUS, has established itself as a cornerstone of empirical research on MNE. Among comparable data sources it currently comes closest to meeting these requirements.

Consequently, a large body of literature both on firms in general and on MNE in particular rely on BvD data. In the context of profit shifting, seminal papers such as Huizinga and Laeven (2008) and Dharmapala and Riedel (2013) have made use of the data as well as a substantial amount of literature that built upon them. In a meta-analysis of articles on profit shifting behavior by Heckemeyer and Overesch (2017) provide an extensive review of the literature and identify 25 studies of relevance. Out of the 11 articles which do not exclusively focus on US MNEs 10 use BvD data in some form.[1]

Until recently, however, this data came with a major restriction. Ownership information was not available in each year, but only on a most-recent basis. Similar to other static variables (e.g. a firm's industry classification) this data then had to be copied for each year to allow for the use of financial panel data in all periods. Since it was not possible to examine changes in group structures over time, two assumptions became inevitable. First, group structures had to be constructed as unchanging (by copying the ownership information of the most recent year to all previous years) and thus assumed to have remained constant. This implies that corporate networks do not expand, which is a precondition for the existence of MNEs in the first place. Second, the observed group structures had to be assumed as exogenous. Among other things, this implies that MNE did not cre-

[1]The other articles use data from the IRS, the BEA or Compustat. A notable exception is Weichenrieder (2009), who examines the profit shifting behavior of German multinationals by using MiDi.

ate new firms in countries due to low tax rates or shift headquarters and thus refrained from profit shifting at the extensive margin. Neither assumption is particularly convincing.

Constructing constant structures thus misclassifies firms by definition, but it can be a viable empirical strategy. After all, wrongfully classified independent firms cannot take advantage of the channels MNE affiliates have access to. Budd et al. (2005) argue accordingly that misclassified firms would only bias estimates of MNE-level effects towards zero and that the extent of misclassification would be small.[2] However, this paper finds that the extent of induced misclassification is considerable, bidirectional, and potentially consequential.

Furthermore, many papers work with constant ownership data without explicitly discussing its implications (e.g. Navaretti et al. (2003), Huizinga and Laeven (2008), Arulampalam et al. (2012), De Simone (2016), Loretz and Mokkas (2015) and Markle (2016)).[3] This paper complements the literature with a descriptive analysis of the ORBIS ownership data between 2002 and 2012, sheds light on the construction of the database and discusses its implications. Finally, several measures of profit shifting are re-estimated and discussed in terms of their sensitivity to this issue.

The structure of the paper is as follows. Section 2 introduces the data, the identification algorithm to track global ultimate owners (GUO) as well as the identification strategy. Section 3 describes the data, evaluates the assumption of constant ownership structures and discusses its theoretical implications. Section 4 reports the results for different measures of profit shifting. Section 5 concludes. An extended appendix elaborates on a range of technical aspects.

[2] This claim has since been referenced and repeated in several other studies (e.g. Dischinger (2008), Dischinger and Riedel (2011), Becker and Riedel (2012), Dharmapala and Riedel (2013), Dischinger et al. (2014), Brandstetter (2014)).

[3] There are exceptions when it comes to the preparation of ownership data. Maffini and Mokkas (2011) construct a panel by adding information on Mergers and Acquisitions from BvD's ZEPHYR database. Alexander et al. (2017) combine historical updates of ORBIS. Neither contrasts the effects of using yearly instead of constant structures.

2. Data and Methodology

Bureau van Dijk's ORBIS database constitutes the most advanced representation of corporate ownership structures. Both the size and the complexity of the database require extended data preprocessing. Appendix A.1 details the data extraction and integrity verification strategy. Appendix A.2 outlines the reconstruction of the business groups, a discussion of the implications of different boundary definitions and the tracing algorithm. Appendix A.3 tests the accuracy of the algorithm by drawing a comparison to the results of Jaraite et al. (2013).

The data presented in this paper was extracted as part of a larger effort to map MNEs within the European Union's Emissions Trading System, detailed in aus dem Moore et al. (2019). The ownership data was extracted manually using ORBIS' online interface, which allows for a batch-wise extraction of top shareholder information in specific years. The yearly ownership data was then reconstructed with a custom algorithm to replicate ORBIS' GUO identification.

In short, ORBIS ownership structures are reconstructed for each year and then merged with financial data on the firm level. Hierarchical structures are constructed by linking firms with ownership connections to each other, thus forming business groups. Each business group can only have one global ultimate owner (GUO) at the top of its hierarchy of firms. In this context, multinational enterprises (MNE) are business groups which include at least two firms from different countries. Affiliates of multinational enterprises are all firms which are part of such a group. Throughout this paper, the unit of observation always remains the individual firm.

Table 1: GUO identification accuracy

GUO evaluation results	No.	%	%
matched hit	2,946,188.0	88.3	88.3
matched miss	12.0	0.0	88.3
mismatched miss	360,031.0	10.8	99.1
other GUO found	29,892.0	0.9	100.0
Total	3,336,123.0	100.0	

The core benchmark for the reconstructed ownership data remains ORBIS itself. Consequently, the performance of the algorithm is evaluated with a benchmark data set which was extracted on a most-recent basis.[4] Reconstructing the GUOs then allowed for a direct comparison between original ORBIS GUO information and reconstructed GUO information, eliminating a wide range of potentially confounding factors. The algorithm correctly identifies 88.3 percent of all GUOs in the evaluation dataset, which corresponds to 99 percent accuracy for cases in which a GUO is found (see Table 1). The remaining differences in total coverage are explained by the selective scope of the evaluation export. In contrast to ORBIS as a whole, the evaluation dataset largely included firms of at least medium size. This means that firms with linked subsidiaries below this threshold cannot be identified as GUOs, but it also means that all mismatched misses correspond to GUOs without a group structure that can be assessed empirically.[5]

Correspondingly, 97.7 percent of the firms that are not identified as GUOs are listed in ORBIS as GUOs themselves, but none of their subsidiaries reports financial data. The cases in which a different GUO is found (1 percent) can be explained by differences between the availability of top shareholder information (on which the GUO construction is based) and GUO status in the original data. Overall, the algorithm is able to replicate ORBIS' ownership identification technology.[6] The algorithm is then applied to the ownership data extracted in each year. The yearly waves from 2002-2012 are then combined to a full panel. Constant ownership structures are created by simply copying the ownership information in one year to all other years. Effectively, this process does fill in a large a mount of missing information and creates substantially larger datasets.

[4] Selected were all firms of at least medium size which were either a GUO or a subsidiary, as identified by ORBIS. This selection was then complemented by all firms with at least one subsidiary, regardless of size. The data included current global ultimate ownership as well as top shareholder information at the time of extraction. The evaluation sample of 3.3m firms was extracted in July 2015.

[5] A firm is also classified as a GUO itself if it can be proven that another firm within the evaluation dataset is its subsidiary. This cross-sample correction at the first hierarchy level is responsible for 6 percent of the correctly identified GUOs.

[6] The amount of identified GUOs is, however, limited by the amount of data used for the identification of economically relevant business groups. Given that only a subset of ORBIS is extracted to gauge the quality of the ownership data (14.379m firms, see appendix A.1), not all GUOs can be found. These cases require missing financial data on the part of the subsidiary and a connection of two links to the firm which is a GUO itself.

Data for Earnings before Interest and Taxes (EBIT), Profit and Loss before Taxes (PLBT), Fixed Assets, and Cost of Employees were extracted from unconsolidated local registry filings in ORBIS, adjusted for inflation with IMF IFS PPI data, and taken in logs. All financial data was extracted from a single bulk export provided by Bureau van Dijk in November 2015. Corporate tax rates originate from the Oxford University Centre for Business Taxation's CBT Tax Database. Data for GDP, GDP per capita, GDP growth, and unemployment was extracted from the World Bank's World Development Indicators database, but not transformed in any way. Corruption was measured by Transparency International's Corruption Perception Index, which was rescaled for 2012 to account for a change in methodology. Appendix A.6 provides summary statistics and an overview of the country distributions for the core samples.[7]

2.1. Empirical model

In line with Huizinga and Laeven (2008) and Lohse and Riedel (2013), the estimated outcome takes the following form:

$$
\begin{aligned}
ln(EBIT)_{it} = \beta_0 + \beta_1 TAX_{kt} + \beta_2 ln(FIAS)_{it} + \beta_3 ln(STAF)_{it} \\
+\beta_4 X_{kt} + \gamma_{jt} + \delta_i + \epsilon_{it}.
\end{aligned} \tag{1}
$$

$ln(EBIT)_{it}$ denotes the logarithm of earnings before interest and taxes (EBIT) of a given firm i at time t.[8] As suggested by Huizinga and Laeven (2008), the sample is limited to affiliates with positive operating pre-tax profits. β_0 is a constant, TAX_{it} is the tax measure of a firm i in country k[9], $ln(FIAS)_{it}$ is the logarithm of a firm's fixed assets and $ln(STAF)_{it}$ is the logarithm of a firm's costs of employees. Furthermore, a set of country-level control variables was included (X_{kt}), which consists of GDP, GDP per capita, GDP

[7] In line with the literature, the intersection of ownership data, financial information, and country-level variables leads to final samples that are highly eurocentric for the time period of this study. Since the country selection in Lohse and Riedel (2013) (26 European countries) covers over 96 percent of observations from 2002-2012, observations outside of Europe are discarded to increase the comparability of results.

[8] All estimations were repeated for an alternative dependent variable, profit and loss before taxes (PLBT). The results were either similar or even more pronounced.

[9] Since the corporate tax rate is merged with the firm level data on the country level, all firms i in country k carry the same value for the corporate tax rate, but may differ in the other measures. This multilevel aspect of the data would suggest clustering the standard errors at the country level, but the small number of clusters renders this infeasible. The subsequent clustering on the firm level should thus be taken with a grain of salt.

growth, corruption, and unemployment.[10] Fixed effects are included at the industry-year (γ_{jt}, NACE Rev. 2 two-digit level) and the firm level (δ_i), ϵ_{it} denotes the error term.

Different TAX_{it} have been proposed in the literature on profit shifting. This paper examines the sensitivity of four prominent measures of indirect profit shifting to different ownership structures:

This paper examines the sensitivity of indirect profit shifting estimates to the chosen method of identifying multinational enterprises within firm-level data. Four different measures of indirect profit shifting are used as benchmarks: the corporate tax rate, the tax differential (I) of a firm to its GUO, the tax differential (II) of a firm to the average of the MNE it belongs to and Huizinga & Laeven's C, a profit-weighted aggregate differential.

The calculation of the corporate tax rate τ_i for firm i does not depend on the structure of the firm's business group because it is determined on the country level. Consequently, if estimated for a sample of MNE affiliates its coefficient can be interpreted as the affiliates' semi-elasticity of EBIT to the tax rate in a country, but it is not sufficient to measure profit shifting without contrasting MNE affiliates to independent firms. The tax differential is constructed as (I) the difference between the corporate tax rate, τ of a firm i and its respective GUO g

$$TaxDiff(I)_i = (\tau_i - \tau_g) \tag{2}$$

as well as (II) the difference between a firm's corporate tax rate, τ_i, and the average corporate tax rate $\frac{1}{n}\sum_{i=1}^{n}\tau_i$ of its business group:

$$TaxDiff(II)_i = \left(\tau_i - \frac{1}{n}\sum_{i=1}^{n}\tau_i\right). \tag{3}$$

Both measures depend on the structure of the group that firm i is a part of and thus improve upon the identification problem outlined above. Yet for the same reason they are

[10]The selection of country control variables follows standard practice in the literature. Since it has shortcomings, no interpretation of the coefficients is undertaken. The selection remains unchanged in all estimations.

also sensitive to the construction of different business group structures. Both variables' construction based on constant ownership data thus introduces a measurement error. And while they take a firm's relative position within a business group into account, they ignore a firm's relative profit shifting capacity. To remedy this, Huizinga & Laeven's C is constructed according to the following definition:

$$C_i = \frac{1}{(1-\tau_i)} \frac{\sum_{i \neq k}^{n} \left(\frac{B_k}{1-\tau_k}\right)(\tau_i - \tau_k)}{\sum_{k=1}^{n} \left(\frac{B_k}{1-\tau_k}\right)} \tag{4}$$

Here, τ_i denotes the tax rate of firm i, τ_k is the tax rate of another firm k in the group with n firms in total and B_k are the profits of firm k.

3. Descriptive Analysis

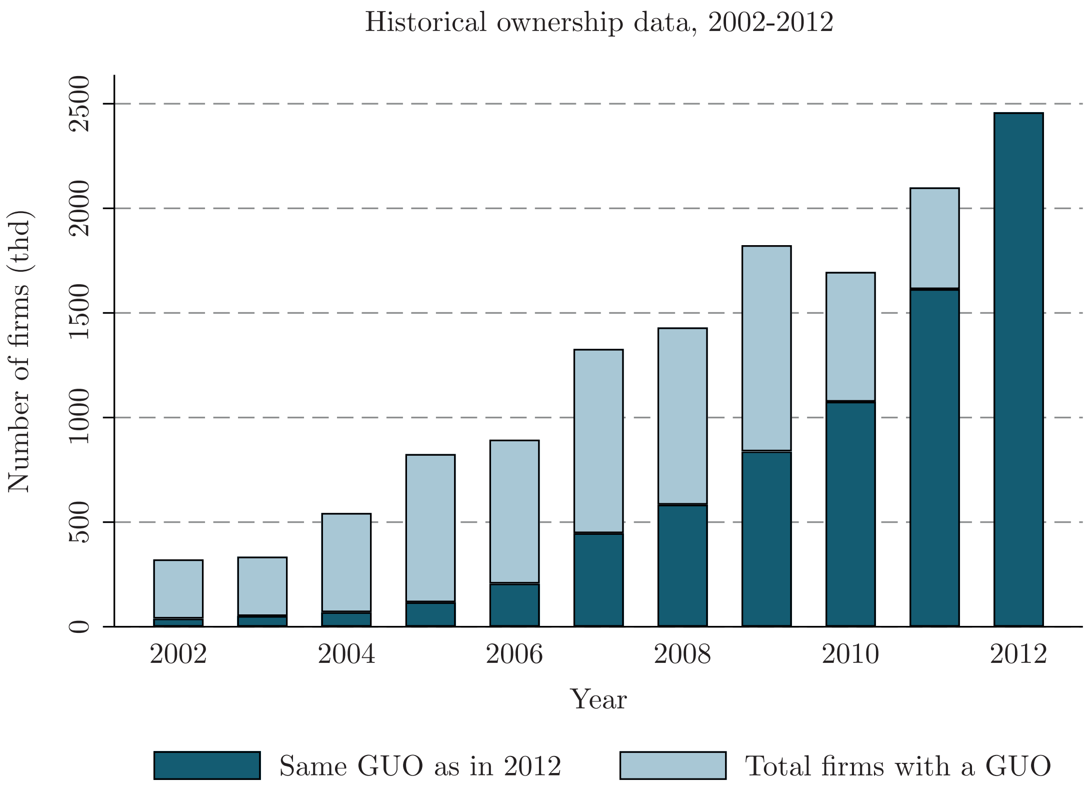

Figure 1: Historical ownership data, 2002-2012

Figure 1 illustrates the GUO data in each year. The number of firms with nonmissing ownership information grew at an average rate of about 24 percent per year. A similar trend is reported within the ORBIS ownership guide. The lower figures reported here, however, are a result of the selective data extraction process. Unfortunately, in the absence of an initial inventory of ownership connections it remains challenging to disentangle the growth and expansion patterns of MNE from mechanical data quality improvements. The latter is constantly implemented by adding newly found ownership connections and updating the existing data in the process. This structure supports the hypothesis that the coverage of the database is converging to an accurate representation of the real present ownership structure. The trend does not suggest that this representation had been achieved by 2012 and it challenges the assumption that current firm structures can be assumed as constant over all years. Figure 1 also illustrates the share of verifiably constant ownership information compared to the most recent year (2012). Looking three years into the past only about a third of the ownership data is identical to the first year. Virtually none of the firms reported the same GUO in 2002 compared to 2012.

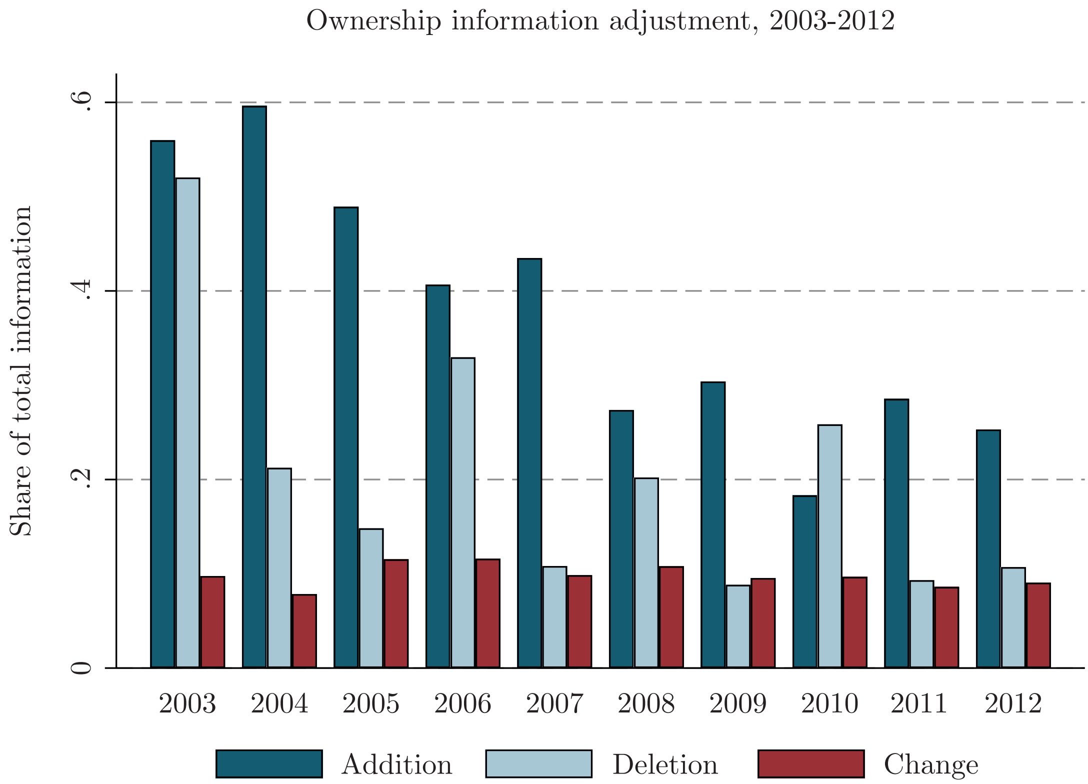

Figure 2: Ownership information adjustments in ORBIS, relative

One could argue that the additions merely reflect the process of building a dataset. But even if both the quality of the data were increasing over time and the database converged to an accurate representation of the firms' current ownership structures, assuming firm ownership as constant would still misclassify a large number of firms. Figure 2 illustrates the yearly changes in relation to the total amount of ownership data in each year. As the size of the ownership database grows, the share of newly added information declines as expected. Similarly, the amount of deletions (changes to missing values) declines. The share of *changing* information, however, remains constant. In these cases, ownership data is *nonmissing but different* in both years. On average, about 9 percent of firms with observed ownership data change owners in each year, regardless of the size of the ownership database. Consequently, identical selection criteria can return differing samples of multinational corporation's affiliates in different years.

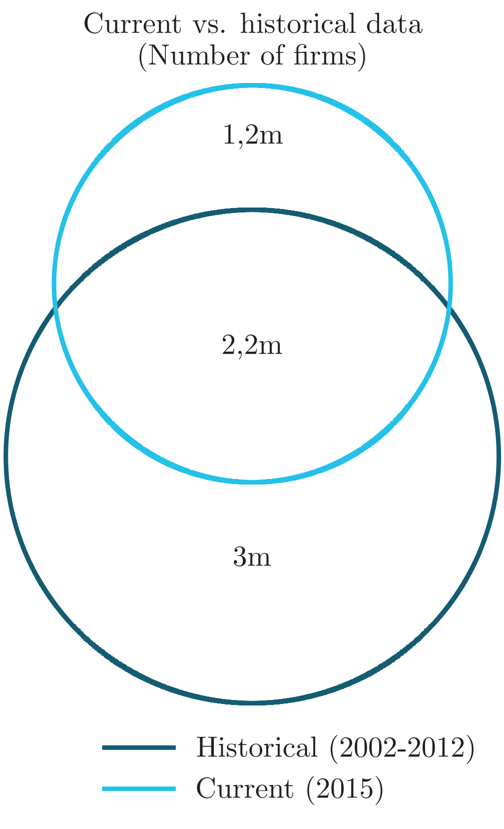

Figure 3: The Tip of the Iceberg

If few firms are responsible for a large share of the changes the number of affected firms could still turn out to be negligible. However, Figure 3 illustrates that this is not the case. The large turnover within the database creates an "iceberg-effect", whereby the most recent state of the data only paints an incomplete picture. Comparing the historical sample of all firms with ownership information in any year from 2002-2012 to another sample of firms with current ownership information (end of 2015) reveals substantial differences. A sizeable share of firms (1,2m) is only present in the most-recent export, suggesting that these firms started to report ownership connections only after 2012. A much larger share (3m) is no longer included.[11] Those firms had been part of business groups in the past, but no longer reported ownership connections at the date of the most-recent export. While the conclusion that a snapshot of the data yields a selective sample is somewhat trivial, the magnitude of the difference is striking.

Unfortunately, the “true” historical ownership structures remain unknown. Without an initial inventory the database cannot accurately reflect the historical reality. It thus remains unclear how many firms were MNE affiliates in the past, but not yet included in the database. Consequently, even datasets built upon the combination of yearly ownership information (i.e. merging several years of historical versions of ORBIS) remain incomplete because they incorrectly exclude group structures that existed before they were added to the database. Correspondingly, datasets with ownership information constructed as constant over time correctly include group structures that existed, but were not yet added to the database. Unfortunately it is not possible to identify the affected firms either. The yearly approach, however, becomes more accurate as the database matures while the constant approach does not. In sum, the construction of constant ownership structures a) classifies some firms as multinationals who were not, b) does not classify some firms as multinationals who were and c) ignores changes over time.

The first point, as argued by Budd et al. (2005), could be unproblematic. If the variables of interest capture effects exclusive to MNE, this would indeed introduce a bias towards zero. After all, an effect exclusive to MNE affiliates would not be observable for independent firms by definition. However, with increasing frequency the misclassification could also lead to insignificant results. The second point would be unproblematic if firms which no longer report ownership connections at the date of examination were not systematically different from the ones that do. The third point, however, remains a key issue.

If group structures are constructed as constant over time, they have to be assumed as exogenous to the variables of interest. This means that the extensive margin of profit

[11] Extending the historical sample to 2014 would close the gap to the current export, but can not reduce the number of missing firms because the historical dataset would only grow larger.

shifting, which is available to all MNE, is assumed to be irrelevant. The construction of constant ownership structures based on current information creates a sample of firms which could have been shaped, among other things, by the development of national corporate tax rates over time. If MNE do indeed adjust to national corporate tax rates, any sample created in this way then suffers from endogenous survivorship bias. Consequently, a sample constructed from current ownership information includes a selection of firms that might have been determined at least to a certain extent by the effects it is supposed to estimate. Furthermore, it would be impossible to disentangle the effects of sample composition from the effects of variable measurement.

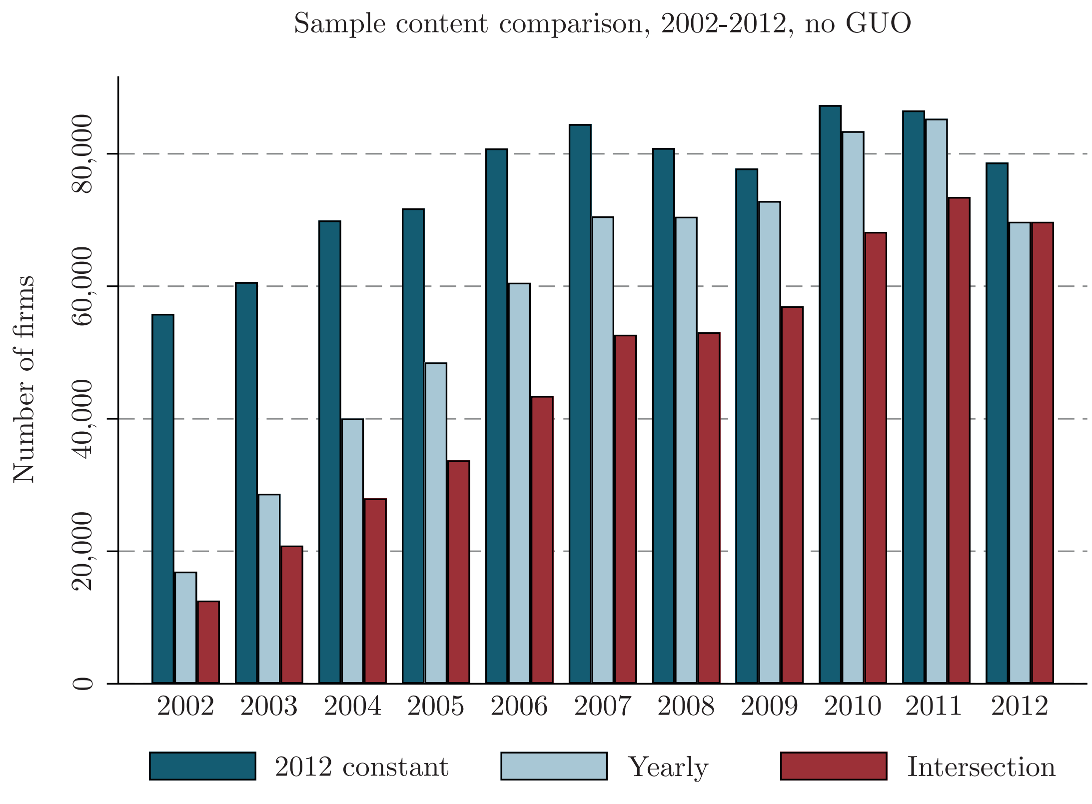

Figure 4: Sample composition according to ownership structure, 2002-2012

To better understand the magnitude of these effects, three benchmark samples are constructed. Since business group structures are merged with the financial data, both the quality of ownership data as well as the quality of financial data determine the sample composition.[12] Figure 4 illustrates their composition over time after estimation. The first

[12]The quality of the ownership data is improved by closing gaps between existing ownership links. A

sample (2012 constant) copies ownership information from 2012 to all other periods, which is the most common approach in the literature.[13] It adds a large amount of observations in past periods, limited only by the quality of the financial data. The drop in observations around the financial crisis in the years 2008 and 2009 is a result of the aforementioned log transformation of EBIT. The second sample (Yearly) uses the full panel of ownership information, which can change in each year. It includes firms which no longer reported ownership data in 2012 and does not extrapolate from the original data. The number of firms in the constant sample is higher in 2012 because the dynamic sample includes a larger amount of singletons, which drop out during estimation. The third sample (Intersection) only includes observations which are present in both previous samples. This sample enforces the same sample composition, thus allowing for a clean identification of the sensitivity of variable measurement to changes in MNE structures. The only difference between these two samples is the way in which the ownership structures are constructed.

4. Results

In the context of MNE research, the method of identifying ownership structures influences estimates in two important ways. First, the chosen criterion of what constitutes an MNE affiliate decides which firms enter a sample. Second, the method of identifying ownership structures determines how firms in the sample are connected to each other. Both of these effects have to be disentangled before their relevance can be assessed. Tables 2-5 report the results for the four alternative tax variables. Within each table, column (1) refers to the 2012 constant ownership sample. Column (2) refers to the ownership sample with yearly changing information. Columns (3)-(5) report the results for the intersection of both datasets.

A comparison between the coefficients in column (3) and (1) highlights the effect of including additional firm years that were constructed with constant ownership data, but

discussion of the gap-closing algorithm is provided in Appendix A.4. The financial data is not subjected to a similar interpolation process. An overview of the ownership datasets before merging with the financial data is provided in Appendix A.5.

[13] Note that this approach prevents the comparison of results from research projects undertaken at different points in time. Ownership information extracted from ORBIS on a most-recent basis is only identical for the same week and can change considerably between weeks if new batches are added.

not part of the historical data. This can be interpreted as the sample composition effect in the case of constant ownership structures. Correspondingly, a comparison between the coefficients in column (4) and (2) highlights the effect of including firm years that were part of the historical data, but not constant in their ownership structures. This can be interpreted as the sample composition effect in the case of using yearly changing ownership structures. A comparison between column (3) and column (4) reveals the net impact of different ownership structures on the estimates if the sample compositions are identical. Column (5) tests if this difference is significant. Each estimation uses the same clustering on the firm level, the same country control variables (GDP, GDP per capita, growth, corruption, and unemployment) and the same sector-year and firm fixed effects. Coefficients for log Fixed Assets and log Cost of Employees are always significant and within plausible ranges.[14]

Taking a closer look at the number of firms and observations per sample, several preliminary conclusions can be drawn. While the sample with constant ownership information consists of the largest number of observations, it includes fewer firms than the dynamic sample. This is a result of the 'iceberg-effect' described in figure 3 and illustrates the extent to which reconstructing constant ownership structures reshapes the data. The large difference in observations between the sample with constant ownership information and the intersection suggests that the extent of misclassification introduced by constructing ownership structures as constant is considerable. In the benchmark case based on the year 2012, reconstructing constant ownership structures retroactively adds around 29 percent of additional observations compared to using historical ownership data. Furthermore, the difference in firms between the sample with yearly changing data and the intersection of both suggests that the amount of excluded firms is also considerable. Overall, sample composition is strongly affected by the chosen method to construct ownership structures.

[14]All results shown in this section have been estimated after gap-closing, but the coefficients remain qualitatively and quantitatively similar if the original data were used. Gap-closing does not effect the sample with constant ownership information.

Table 2: Measurement error vs. sample composition, tax rate

Sample:	Cons. '12	Yearly	Intersec.	Intersec.	Intersec.
Dependent Variable:	Log EBIT	Log EBIT	Log EBIT	Log EBIT	Log EBIT
A) Tax Rate, con.	-0.372^{***}		-0.454^{***}		
B) Tax Rate, dyn.		-0.314^{***}		-0.454^{***}	-0.454^{***}
A) - B)					0.000
Log Fixed Assets	0.088^{***}	0.073^{***}	0.073^{***}	0.073^{***}	0.073^{***}
Log Cost of Employees	0.407^{***}	0.379^{***}	0.384^{***}	0.384^{***}	0.384^{***}
Firms	138015	142057	105327	105327	105327
Obs.	834590	647053	512718	512718	512718
Within R Squared	0.076	0.053	0.055	0.055	0.055

Significance levels: $*$ p <0.10, $**$ p <0.05, $***$ p <0.01. Standard errors are clustered on the level of the firm.

Sector-Year fixed effects, firm-level fixed effects, and country control variables included, but not shown.

Table 2 examines the influence of the corporate tax rate on Log EBIT, which is determined at the country level. There is no difference between the coefficients in columns (3)-(4) because an MNE's structure is irrelevant for the measurement of the corporate tax rate. The coefficient of -0.454 means that an increase in the corporate tax rate by 10 percent is associated with a reduction of EBIT by 4.54 percent.

The reduction in absolute magnitude from column (3) to (1) could be interpreted as evidence for the hypothesis in Budd et al. (2005); the addition of irrelevant firms through the construction of constant ownership structures seems to induce a measurement error that biases the results towards zero. However, the reduction from column (4) to (2) indicates that the inclusion of previous MNE affiliates which no longer reported ownership information in 2012 reduces the coefficient even further. This suggests that the link between tax rates and EBIT was weaker for actual MNE affiliates in the past compared to retroactively constructed hypothetical MNE affiliates. The differences are fully explained by different sample compositions.

Table 3 examines the effect of the tax differential (I) between a firm and its global ultimate owner on Log EBIT. The measurement of this variable depends on the chosen corporate structures, which in turn allows for the identification of the impact of different reconstruction methods on the estimated effect in columns (3)-(5). For the same observa-

Table 3: Measurement error vs. sample composition, tax differential (I)

Sample: Dependent Variable:	Cons. '12 Log EBIT	Yearly Log EBIT	Intersec. Log EBIT	Intersec. Log EBIT	Intersec. Log EBIT
A) Tax Diff. (I), con.	-0.237^{***}		-0.078		
B) Tax Diff. (I), dyn.		-0.010		-0.050	-0.081
A) - B)					-0.052
Log Fixed Assets	0.088^{***}	0.073^{***}	0.073^{***}	0.073^{***}	0.073^{***}
Log Cost of Employees	0.407^{***}	0.379^{***}	0.384^{***}	0.384^{***}	0.384^{***}
Firms	138015	142057	105327	105327	105327
Obs.	834590	647053	512718	512718	512718
Within R Squared	0.076	0.053	0.055	0.055	0.055

Significance levels: $*$ p <0.10, $**$ p <0.05, $***$ p <0.01. Standard errors are clustered on the level of the firm.

Sector-Year fixed effects, firm-level fixed effects, and country control variables included, but not shown.

tions, calculating the tax differential based on yearly ownership data reduces the coefficient slightly. However, the coefficients as well as the difference between the two methods, reported in column (5), are insignificant.

Investigating the effects of sample composition now reveals a different picture. The coefficient for the tax differential (I) is negligible and insignificant in column (2), but large and highly significant in column (1). This suggests that the different sample compositions determine the effect of the tax differential to a firm's GUO on log EBIT. While the effect remains small and insignificant for the actual historical ownership data, the construction of constant structures leads to the estimation of a negative effect. The significant coefficient estimated for the constant sample means that a 10 percent increase in the tax difference between firm and GUO is associated with a reduction of a firm's EBIT by 2.37 percent, which is slightly lower than other findings in the literature.

While the tax differential (I) between firm and GUO only requires information on two firms (and their link), the tax differential (II) between the firm and its business group requires information on the entire MNE. Table 4 reports the results for the tax differential (II). Comparing column (3) and (4) reveals a substantial difference between the two methods. While the estimation based on yearly ownership structures (column 4) returns a small and insignificant coefficient, the estimated effect based on ownership structures

Table 4: Measurement error vs. sample composition, tax differential (II)

Sample:	Cons. '12	Yearly	Intersec.	Intersec.	Intersec.
Dependent Variable:	Log EBIT	Log EBIT	Log EBIT	Log EBIT	Log EBIT
A) Tax Diff. (II), con.	-0.440^{***}		-0.482^{***}		
B) Tax Diff. (II), dyn.		-0.066		-0.098	-0.434^{***}
A) - B)					-0.639^{***}
Log Fixed Assets	0.088^{***}	0.073^{***}	0.073^{***}	0.073^{***}	0.073^{***}
Log Cost of Employees	0.407^{***}	0.379^{***}	0.384^{***}	0.384^{***}	0.384^{***}
Firms	138015	142057	105327	105327	105327
Obs.	834590	647053	512718	512718	512718
Within R Squared	0.076	0.053	0.055	0.055	0.055

Significance levels: * p <0.10, ** p <0.05, * * * p <0.01. Standard errors are clustered on the level of the firm.

Sector-Year fixed effects, firm-level fixed effects, and country control variables included, but not shown.

constructed as constant (column 3) is large and significant. The difference between the two methods, reported in column (5), is significant as well. This difference persists when the samples are extended (columns 1 and 2), suggesting that the bias seems to originate mostly from the measurement of the variable and not from the composition of the samples.

Table 5 shows a similar pattern for Huizinga & Laeven's C. The coefficient estimated based on constant ownership structures (column 3) is more than three times as large as the one estimated based on yearly ownership structures (column 4). The difference between the two methods is once again significant (column 5). The inclusion of additional firm years increases the coefficient for the sample constructed as constant, but reduces it in the case of the historical ownership structures. Once again sample composition seems to have a smaller impact on the estimates than the measurement of the variable. In all cases, the combined effects of sample composition and differences in measurement lead to smaller coefficients in the yearly sample (2) when compared to the constant sample (1). More importantly, the inclusion of irrelevant firms as a result of constructing constant ownership structures does not always seem to introduce a bias towards zero. Instead the retroactive reconstruction of MNEs based on their most recent image seems to inflate estimates for indirect profit shifting. In particular, estimates based on constant ownership structures cannot be interpreted as conservative lower bounds if the measure of interest depends on an MNE's group structure.

Table 5: Measurement error vs. sample composition, Huizinga & Laeven's C

Sample:	Cons. '12	Yearly	Intersec.	Intersec.	Intersec.
Dependent Variable:	Log EBIT	Log EBIT	Log EBIT	Log EBIT	Log EBIT
A) H&L's C, con.	-0.278^{***}		-0.242^{***}		
B) H&L's C, dyn.		-0.039		-0.069	-0.220^{***}
A) - B)					-0.308^{***}
Log Fixed Assets	0.088^{***}	0.073^{***}	0.073^{***}	0.073^{***}	0.073^{***}
Log Cost of Employees	0.407^{***}	0.379^{***}	0.384^{***}	0.384^{***}	0.384^{***}
Firms	138015	142057	105327	105327	105327
Obs.	834590	647053	512718	512718	512718
Within R Squared	0.076	0.053	0.055	0.055	0.055

Significance levels: ∗ p <0.10, ∗∗ p <0.05, ∗ ∗ ∗ p <0.01. Standard errors are clustered on the level of the firm.

Sector-Year fixed effects, firm-level fixed effects, and country control variables included, but not shown.

Since it is plausible to assume that MNEs adjust their structures over time in response to the tax rate, the data would likely reflect this endogenous response. Ownership structures observed on a most-recent basis then describe the final outcome of all previous structural changes undertaken in response to the tax rate. However, constant ownership structures can also be constructed based on the initial state of the business groups in 2002. The methodology is identical (ownership information from 2002 is copied to all other periods), but the result is a dataset which has not been shaped in response to changes of the tax rate. In Table 6, all four measures were re-estimated with this sample. The number of

Table 6: 2002 constant ownership data

Sample:	Cons. 2002	Cons. 2002	Cons. 2002	Cons. 2002
Dependent Variable:	Log EBIT	Log EBIT	Log EBIT	Log EBIT
Corporate Tax Rate	-0.357^{**}			
Tax Rate Differential (I)		-0.002		
Tax Rate Differential (II)			-0.471^{***}	
Huizinga & Laeven's C				-0.200^{*}
Log Fixed Assets	0.081^{***}	0.082^{***}	0.082^{***}	0.082^{***}
Log Cost of Employees	0.380^{***}	0.381^{***}	0.380^{***}	0.380^{***}
Firms	33743	33743	33743	33743
Obs.	214504	214504	214504	214504
Within R Squared	0.061	0.061	0.061	0.061

Significance levels: ∗ p <0.10, ∗∗ p <0.05, ∗ ∗ ∗ p <0.01. Standard errors are clustered on the level of the firm.

Sector-Year fixed effects, firm-level fixed effects, and country control variables included, but not shown.

firms and observations is considerably smaller and the coefficient for the corporate tax rate is significant and slightly smaller than the 2012 constant sample. The change is once again fully explained by the altered sample composition. Firms which have been part of MNEs in the past seem to exhibit a similar sensitivity to changes in the tax rate. The tax differential (I) to the GUO, however, is now small and insignificant. This could indicate that firms adjusted their headquarters in response to changes in the tax rate. The coefficient of the tax differential (II) to the group grows even larger and stays significant. Huizinga & Laeven's C remains significant at a lower level with a slightly smaller coefficient.

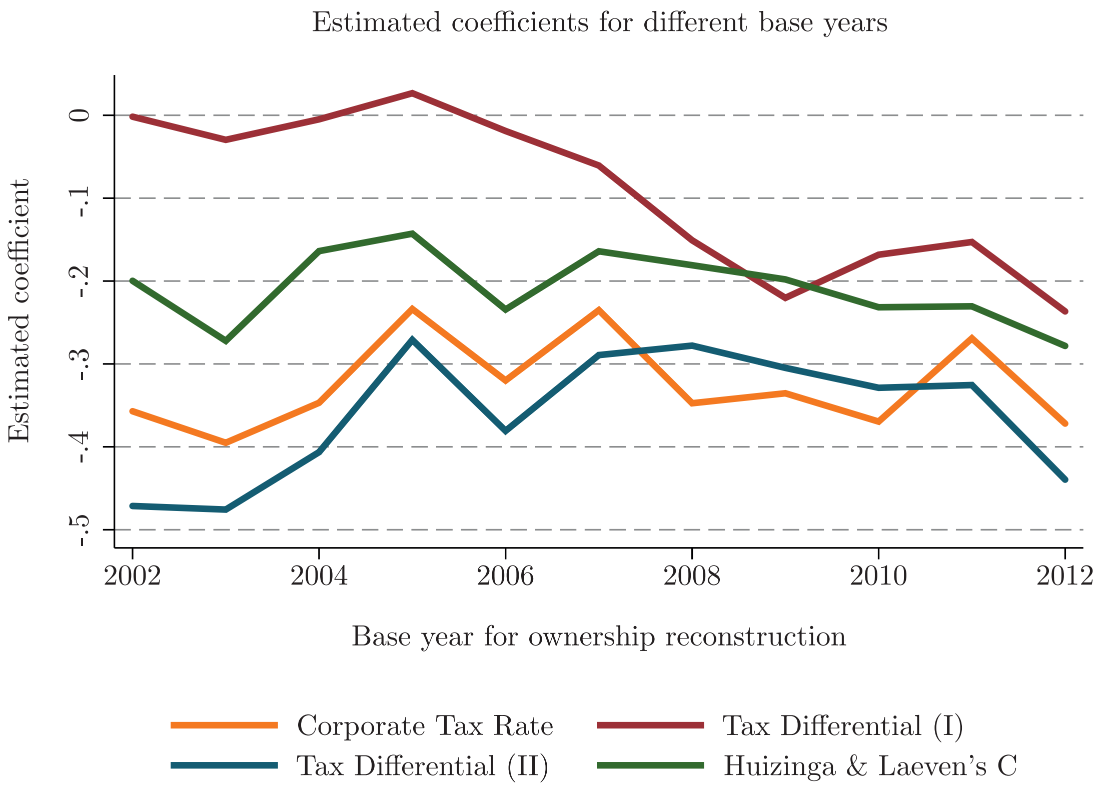

Figure 5: Estimated coefficients for different base years

The different estimates point towards the potential relevance of endogenous structural adjustments, but this also implies that estimates based on constant ownership structures are not directly comparable across years. Figure 5 examines the behaviour of the estimated coefficients over all possible base years for ownership reconstruction. The X-axis here refers to the year in which the ownership information was assumed as constant before copying this information to all other years.[15] The coefficients in base year 2002 correspond to those reported for each tax measure in table 6 while the coefficients for base year 2012 correspond to the coefficient reported for each tax measure in each column (1) in tables 2-5.

For datasets constructed based on ownership structures before 2007, all four measures follow somewhat similar trends. Most notably, while the other measures are always significant at least at the 5 percent level, the tax differential (I) to the GUO remains insignificant until 2008. The financial crisis then seems to shake up the corporate structures. This follows from the drop in the coefficient for the tax rate while both the tax differential (II) to the group and Huizinga & Laeven's C remain stable. The latter measures only differ in levels, but follow the same trend over all years. Figure 5 emphasizes that great care should be taken during the data preparation process whenever the measurement of a variable of relevance for the estimation of an effect depends on an MNE's corporate structure. It also reiterates that estimates for indirect measures of profit shifting in particular can only be compared if the underlying data is of the same vintage. The results for the benchmark samples also highlight the sensitivity of some estimates to changes in sample composition.

The results have implications beyond the indirect assessment of profit shifting. Data aggregation procedures in particular are sensitive to the chosen ownership reconstruction method and should be evaluated against different benchmark structures. The large impact of small choices at the beginning of a research process involving ownership data also calls for a stronger emphasis on illustrating data preparation and cleaning choices. Future research could aim for a more detailed empirical assessment of MNE's structural changes over time. Given the outlined challenges at the level of the firm, new data structures could allow for the use of different tools to manage the complexity and help to avoid some of the pitfalls on the way ahead.

[15] The respective sample sizes thus roughly correspond to the pattern described by figure 1, ranging from 214504 observations for base year 2002 to 834590 observations for base year 2012. All corresponding calculations are available upon request.

5. Conclusion

The assumption of constant firm ownership structures is a cornerstone in MNE research, yet neither plausible, nor without consequences. Although arcane at first glance, the method chosen for the reconstruction of ownership structures defines the sample of interest and has a meaningful impact on the estimated effects. The retroactive reconstruction of constant ownership structures a) classifies some firms as multinationals who were not, b) does not classify some firms as multinationals who were and c) ignores changes over time. Using ownership data from 2012 adds around 29 percent additional observations compared to using yearly ownership data. Since about 9 percent of firms change owners in each year, results based on constant ownership structures are only comparable across datasets of the same vintage. Different indirect measures of profit shifting are re-estimated, revealing that both sample composition and variable measurement influence the estimated effects. The more a variable's measurement depends on a business group's structure, the more sensitive it becomes to the chosen ownership reconstruction method. Future work on MNE could benefit from more detailed discussions of the results' sensitivity to different ownership structures.

References

Alexander, A., Vito, D., Antonio, and Jacob, M., 2017. Corporate Tax Reforms and Tax-Motivated Profit Shifting: Evidence from the EU. SSRN Scholarly Paper ID 2841246, Social Science Research Network, Rochester, NY.

Aminadav, G., Bachrach, Y., Kosenko, K., Rosenschein, J. S., and Wilf, Y., 2011. Rebuilding the Great Pyramids: A Method for Identifying Control Relations in Complex Ownership Structures. *SSRN Electronic Journal*.

Arulampalam, W., Devereux, M. P., and Maffini, G., 2012. The direct incidence of corporate income tax on wages. *European Economic Review*, 56(6):1038–1054.

aus dem Moore, N., Großkurth, P., and Themann, M., 2019. Multinational corporations and the EU Emissions Trading System: The specter of asset erosion and creeping deindustrialization. *Journal of Environmental Economics and Management*, 94:1–26.

Becker, J. and Riedel, N., 2012. Cross-border tax effects on affiliate investment—Evidence from European multinationals. *European Economic Review*, 56(3):436–450.

Brandstetter, L., 2014. Do Corporate Tax Cuts Reduce International Profit Shifting? *SSRN Electronic Journal*.

Budd, J. W., Konings, J., and Slaughter, M. J., 2005. Wages and International Rent Sharing in Multinational Firms. *Review of Economics and Statistics*, 87(1):73–84.

De Simone, L., 2016. Does a common set of accounting standards affect tax-motivated income shifting for multinational firms? *Journal of Accounting and Economics*, 61(1):145–165.

Dharmapala, D. and Riedel, N., 2013. Earnings shocks and tax-motivated income-shifting: Evidence from European multinationals. *Journal of Public Economics*, 97:95–107.

Dischinger, M., 2008. Profit Shifting by Multinationals and the Ownership Share: Evidence from European Micro Data. *Münchener Wirtschaftswissenschaftliche Beiträge (VWL)*, 17.

Dischinger, M., Knoll, B., and Riedel, N., 2014. The role of headquarters in multinational profit shifting strategies. *International Tax and Public Finance*, 21(2):248–271.

Dischinger, M. and Riedel, N., 2011. Corporate taxes and the location of intangible assets within multinational firms. *Journal of Public Economics*, 95(7-8):691–707.

Heckemeyer, J. H. and Overesch, M., 2017. Multinationals' profit response to tax differentials: Effect size and shifting channels. *Canadian Journal of Economics/Revue canadienne d'économique*, 50(4):965–994.

Huizinga, H. and Laeven, L., 2008. International profit shifting within multinationals: A multi-country perspective. *Journal of Public Economics*, 92(5-6):1164–1182.

Jaraite, J., Jong, T., Kažukauskas, A., Zaklan, A., and Zeitlberger, A. C. M., 2013. Matching EU ETS Accounts to Historical Parent Companies. A Technical Note. SSRN Scholarly Paper ID 2384537, Social Science Research Network, Rochester, NY.

Kalemli-Ozcan, S., Sørensen, B. E., Villegas-Sanchez, C., Volosovych, V., and Yesiltas, S., 2015. How to Construct Nationally Representative Firm Level Data from the ORBIS Global Database. SSRN Scholarly Paper ID 2663407, Social Science Research Network, Rochester, NY.

Lohse, T. and Riedel, N., 2013. Do Transfer Pricing Laws Limit International Income Shifting? Evidence from European Multinationals. SSRN Scholarly Paper ID 2334651, Social Science Research Network, Rochester, NY.

Loretz, S. and Mokkas, S., 2015. Evidence for Profit Shifting with Tax-sensitive Capital Stocks. *FinanzArchiv: Public Finance Analysis*, 71(1):1–36.

Maffini, G. and Mokkas, S., 2011. Profit Shifting and Measured Productivity of Multinational Firms. *Oxford Bulletin of Economics and Statistics*, 73(1):1–20.

Markle, K., 2016. A Comparison of the Tax-Motivated Income Shifting of Multinationals in Territorial and Worldwide Countries. *Contemporary Accounting Research*, 33(1):7–43.

Navaretti, G. B., Checchi, D., and Turrini, A., 2003. Adjusting Labor Demand: Multinational Versus National Firms: A Cross-European Analysis. *Journal of the European Economic Association*, 1(2-3):708–719.

Vitali, S., Glattfelder, J. B., and Battiston, S., 2011. The Network of Global Corporate Control. *PLoS ONE*, 6(10):e25995.

Weichenrieder, A. J., 2009. Profit shifting in the EU: Evidence from Germany. *International Tax and Public Finance*, 16(3):281–297.

A. Appendix

The ORBIS database provided by Bureau van Dijk is the world's largest firm-level database and constitutes the leading resource for empirical research involving global business group structures. Over the last two decades it has emerged from the eurocentric AMADEUS database and is constantly being expanded in terms of its scope and functionality. As of August 2017, ORBIS included information on more than 200m firms worldwide. The data is being constantly improved and ORBIS moves at a steady pace towards becoming the first global firm-level database with comprehensive coverage. However, this herculean task was not taken on with the intention to provide researchers with more data. As pointed out by Kalemli-Ozcan et al. (2015), researchers who intend to use the database for academic purposes need to make extensive preparations.[16]

A.1. Extraction

To allow for consistent data extraction from a constantly updating database, the scope of exports had to be restricted and was carefully monitored. As a first step, the sample was restricted to firms of at least medium size. To belong to this category in ORBIS, a firm either has to have an operating revenue of at least 1m EUR, total assets of at least 2m EUR, or at least 15 employees in any year.[17] This category is comparable to the European Commission's definition of a small firm, which sets the respective values at turnover of more than 2m EUR, total assets of more than 2m EUR and more than 10 employees.[18] This initial selection of firms reduced the number of available firms to about 12.3m.[19]

To construct reliable ownership links, potential gaps have to be closed. Ownership data of firms that do not satisfy the former criteria, but might still be connected to our

[16] Kalemli-Ozcan et al. (2015) provide a comprehensive guide for the reconstruction of ownership data based on historical ORBIS updates. The data in this paper was extracted manually via the database's online interface, which is an alternative for empirical researchers without access to historical updates.

[17] ORBIS also requires two additional criteria to be satisfied to include a firm in these categories. Companies with ratios of Operating Revenue per Employee or Total Assets per Employee below 100 Euro are excluded. Companies with missing values for all of the variables above, but with data for "level of Capital" beyond certain thresholds are included. This latter criterion in particular leads to the inclusion of almost three million additional firms. We do not include these two critera in our search strategy and instead opt for the variable thresholds to be satisfied.

[18] A more detailed description of the European Commission's SME criteria can be found here: http://ec.europa.eu/growth/smes/business-friendly-environment/sme-definition_de

[19] The initial backbone search was conducted on 20 February 2015. All exports were undertaken between February and August 2015.

business groups, also have to be considered. Since a sizeable share of the database consists of missing information, this matters. Consequently, we added the firms in our ownership chains one level above and below the starting level which were missing in our original selection. We discarded double-missing links as unreliable information. The final sample used for the identification of ownership links consisted of 14,379m firms.

Due to technical restrictions, exports had to be undertaken in yearly batches of 25000 firms.[20] Each download was prepared with a search file that can be downloaded and uploaded to another account, thus making sure that the potential for human error was minimized. All files were extracted with filenames that allowed for seamless appending and automated verification of integrity. Download and verification of the data was done by different researchers to avoid individual blind spots.[21]

The integrity of the exports was verified in an automated process. Initially, for each batch we stored the list of bvdids of all firms that were to be extracted. This backbone was itself checked manually before using it to verify the exports. By merging each individual export with this one-column backbone we ensured that no firms were lost or exported twice (horizontal verification). Furthermore, from the first batch in each year we extracted the list of variables that each batch was supposed to include and checked it manually. Cross-checking the variables of each following batch ensured exact matches and verified that no mistakes were made concerning the selection of variables (vertical verification). Since some of the variables also included the chosen export year in their string, this process also ensured that no batches were mixed up and no mistakes were made in the process of extracting the period. Errors found in the process were remedied by additional manual exports until the integrity of each wave was ensured. Furthermore, all exports were conducted after sorting the companies in our sample according to their total assets.

[20] ORBIS applies an export restriction of the type $E = \frac{1000000}{2+n}$ where E is the maximum number of firms that can be exported and n represents the number of selected variables. Ownership variables, however, carry 10-fold weight in this equation. Furthermore, a row counter as well as the company name are always exported. For our ownership sample we exported three ownership variables, seven financial variables and the bvdid of the firm, resulting in a maximum batch size of 25000 in yearly waves.

[21] Not all of the extracted information was used to construct the ownership links. Some batches had to be exported more than once to remedy errors. Aside from that, differences in the selected firms between export vintages reflect both changes in bvdids and changes in the scope of ORBIS itself, which had to be corrected for. Furthermore, the database is constantly being extended in an effort to expand its global coverage, which makes later exports larger than earlier ones. In an effort to eliminate a wide range of problems and to ensure consistency between exports, we reduced the final backbone list of bvdids to those which were present in all batches.

This sorting ensured that any human error in the extraction process would not lead to systematic differences in data between countries.

Since the exports were undertaken over the course of several months, additional features of the database had to be taken into account. The database is essentially dynamic, with weekly updates adding constantly to the body of information. The database could be best described as a repository of firms in their current state, focused on providing an accurate image of the present. Historical information varies in quality, but gets updated and improved over time. Consequently, an advantage of the approach we followed compared to merging old ORBIS updates is that it is based on current historical information, not historical historical information. However, the data used here could be easily complemented by running the same algorithm on historical waves of the database (which were not available to us).[22] The dynamic nature of the database also affects the bvdids. While these identification numbers are unique for each firm at any given point in time, they can vary between database updates. There are many possible reasons for this, some of them more impactful than others. For example, an identification number can change in case the reporting data provider adjusts its reporting system. This is largely unproblematic, because the economic information connected to the number is not affected and changes can simply corrected for by updating the number. However, there can also be cases in which the change of an identification number is highly impactful. For example, in cases where two separate entities merge into a new one two old numbers are replaced by one new number. The corresponding time series, however, is in some cases altered substantially and could not be used by simply updating the numbers. To limit the potential impact on our analysis we applied a series of corrections. In cases where one bvdid changed to another, the new bvdid was used. In cases where a bvdid changed into an already existing number it was assumed that an acquisition took place and the already existing bvdid was used. In cases where two bvdids merged into a new number all numbers were discarded.[23]

[22] While this may sound arcane it has some far-reaching consequences. Since historical information is continuously updated with data of better quality, our approach is superior to using a combination of vintage ORBIS exports. Merging vintage waves of ownership data could also inflate the reported connections, because information on terminated links would not be considered in waves exported in the past.

[23] An alternative approach could be to continue the existing time series with the data from the new number, essentially cloning the series. Mergers and acquisitions induce fundamental structural changes in a firm-level dataset. We assume that one methodology was applied consistently and take the current state of the database as given. A more detailed assessment of these phenomena is beyond the scope of this paper.

A.2. Construction

After the raw data export, the business groups had to be reconstructed for each year. For this approach, using the GUO reported in ORBIS was not an option because it is not available by year. Consequently, the ownership identification algorithm of Bureau van Dijk had to be replicated, tested, and then applied to the full panel.

A.2.1. Business Group Identification

Throughout this paper, a hierarchy of control is assumed to be the guiding framework to describe a business group. In this hierarchy, the GUO stands at the top. Chains of ownership links then group the controlled firms on different levels. In this framework, the identification of a business group in the data can be done either from the top or from the bottom, but only the latter approach leads to uniquely identified GUOs.

In the first case, each firm is assumed to be the GUO of a business group consisting of all the firms it controls. The advantage of this approach is a relatively low data requirement, because it requires data only of the firms of interest as well as their subsidiaries. Mapping all connections requires an iterative process, but it is still fast to construct business groups in this way. The initial selection of firms, however, is only sufficient if the information at the level of the controlling firms is correct, complete, and if the controlling firms are independent. Still, even if all of these criteria are satisfied the resulting business groups will be incomplete and overlapping.

Drawing a more accurate picture requires the application of a more time-consuming bottom-up approach. Here each firm is initially treated as a potential subsidiary and all shareholder information is extracted. Depending on a criterion chosen by the researcher, the controlling firms are then identified. The groups are ultimately constructed by repeatedly merging this data with itself.

A.2.2. Boundary Definition

The selection criterion for the controlling firm critically determines the shape of the business groups. Without any restrictions, business groups are interlinked network structures and it becomes much more difficult to disentangle them. The more restrictive the selection for the controlling firms gets the more it simplifies the group structure. At its extreme, the identification of only one controlling firm enforces a tree-shaped hierarchy with only one GUO. Although it is possible in ORBIS to extract an ownership profile at the firm level, the definition of GUOs follows the same logic.

The rule which decides what's inside and what's outside of a group is at the heart of any tracing algorithm. For this very reason the definition of what constitutes a multinational business group is entirely dependent on the choice of the observer. This has far-reaching consequences.

First of all, aside from full ownership the boundaries of a group are always fuzzy. Every firm that's part of one group is potentially part of another. A seminal empirical analysis of this concept has been undertaken by Vitali et al. (2011), who find that a core of very few firms in the financial sector control a substantial share of all other firms through their ownership relations.

> *"In detail, nearly 40 percent of the control over the economic value of TNCs in the world is held, via a complicated web of ownership relations, by a group of 147 TNCs in the core, which has almost full control over itself."*
> (Vitali et al., 2011)

On the one hand, without a restriction the group structures will overlap, which allows for a complete identification of each firms' connections, but not for a distinction between business groups. On the other hand, with a restriction to a single controlling firm the business groups will be uniquely identified, but each firm's connections are reduced to a minimum. Inevitably, a decision has to be made by the researcher depending on the context of the analysis.

To identify distinguishable business groups we assume that connection implies coordination and that there is a top-down flow of control along the shareholding structures of a business group. Consequently, we require a firm to control at least 50.01 percent of another firm to be considered as being the controlling entity. We also assume that this implies full control and exclude the possibility of veto rights and other forms that are independent of majority shares.

A major advantage is that the criterion allows for the identification of one-directional chains of ownership, because each firm can only have one firm controlling it. It has to be noted that this is a substantial simplification of the complex network structures of contemporary multinational business groups.[24] The resulting group boundaries are sharp because no group can be connected to another. However, the resulting tree-shaped firm hierarchies can still be of great complexity.

The criterion also has a series of disadvantages that have to be taken into account. Most importantly, it excludes joint ventures with 50/50 share distributions. While these are an important form of entrepreneurial cooperation, they also violate the assumption of one-directional top-down control. The criterion also excludes all forms of portfolio investment under the assumption that these are not undertaken to gain control over another company.

A.2.3. Tracing Algorithm

Building upon Jaraite et al. (2013), *chains of control* are constructed in a bottom-up approach. For each year the ownership data is repeatedly merged with itself until the number of successfully merged firms approaches zero.[25] This process creates a large repository of links between individual companies and their respective ultimate owners. Aggregating these links then returns GUOs for each firm. In contrast to extracting GUO data only, this approach allows for the identification and complete reconstruction of the business group structures. The data can then be used to investigate different hierarchy levels and obtain meta-information of the business groups.

Following Jaraite et al. (2013), certain shareholders are excluded before constructing the chains of control.[26] In those cases the penultimate firm was identified as the group's GUO. This is done to ensure that the GUO is identified as the last corporate entity in the chain of control. Furthermore, the methodology applied in Jaraite et al. (2013) was extended by implementing a cross-sample correction to identify GUOs without any top

[24] In the absence of this criterion, identification of the ultimate owner becomes considerably more difficult. Aminadav et al. (2011) introduce a method to do this in their aptly named paper "Rebuilding the Great Pyramids: A Method for Identifying Control Relations in Complex Ownership Structures" based on weighted voting games literature.

[25] Most links are terminated quickly and very few extend beyond 10 connections. However, a small number of recurring loops were discovered in the process. I check for loops up to a length of 8 firms within a moving window and retain the first firm that's part of a loop structure.

[26] These are "One or more named individuals or families", "Employees/Managers/Directors", and "Public authority, State, Government". For the exact replication of the GUO information provided by ORBIS for the most recent year, three additional types were excluded. These are "Other unnamed shareholders, aggregated", "Public (publicly listed companies)", and "Unnamed private shareholders, aggregated".

shareholder information.[27]

The speed and simplicity of this algorithm is made possible by applying a very restrictive boundary definition. If this criterion is relaxed, things become more complex and computationally intensive. Aminadav et al. (2011) discuss both the concept of control and its implications under relaxed control criteria in more detail.

A.3. Verification

The results of the algorithm are evaluated both by comparing it to the data from Jaraite et al. (2013) and by comparing it to the current GUO information provided by ORBIS (see section 2).

Table 7 compares the results of this paper's ownership identification when applied to the EU ETS data used in aus dem Moore et al. (2019) with the ownership identification results of Jaraite et al. (2013) for the same selection of firms. Five cases are distinguished to compare the data. Cases in which this algorithm identifies the same GUO are classified as a "matched hit". If the information is determined similarly as missing the firm is flagged as a "matched miss". If no GUO is found even though there should have been one, the case is labelled as a "mismatched miss". Correspondingly, if a GUO is found where there should be none, the case is flagged as a "mismatched hit". If a different GUO is found, the case is classified as "other GUO found".

In 2007, the GUO data found in this paper matches in 80.8 percent of all cases. The changes in the categories between years can be explained by data gaps which have to be present in both studies (Germany and the UK), because applying a gap closing correction does not affect mismatched misses or other GUOs found (see table 8). However, applying a gap closing correction also leads to a large amount of additionally identified GUOs.

[27] After the first round of GUO identification the results are merged with the original data. Firms for which subsidiaries had been identified before are then directly classified as GUO themselves.

Table 7: Comparison to Jaraite et al. (2013)

	Selected Years											
GUO similarity	**2005**			**2006**			**2007**			**Total**		
	No.	%	%	No.	%	%	No.	%	%	No.	%	%
matched hit	2,388	32.8	32.8	2,433	33.4	33.4	2,814	38.7	38.7	7,635	35.0	35.0
matched miss	3,601	49.5	82.3	3,716	51.1	84.5	3,066	42.1	80.8	10,383	47.5	82.5
mismatched miss	333	4.6	86.9	158	2.2	86.6	356	4.9	85.7	847	3.9	86.4
mismatched hit	603	8.3	95.1	662	9.1	95.7	588	8.1	93.7	1,853	8.5	94.9
other GUO found	354	4.9	100.0	310	4.3	100.0	455	6.3	100.0	1,119	5.1	100.0
Total	7,279	100.0		7,279	100.0		7,279	100.0		21,837	100.0	

Table 8: Comparison to Jaraite et al. (2013), after gap closing

	Selected Years											
GUO similarity,	**2005**			**2006**			**2007**			**Total**		
adjusted	No.	%	%	No.	%	%	No.	%	%	No.	%	%
matched hit	2,394	32.9	32.9	2,440	33.5	33.5	2,818	38.7	38.7	7,652	35.0	35.0
matched miss	2,988	41.0	73.9	2,874	39.5	73.0	2,415	33.2	71.9	8,277	37.9	72.9
mismatched miss	315	4.3	78.3	143	2.0	75.0	336	4.6	76.5	794	3.6	76.6
mismatched hit	1,216	16.7	95.0	1,504	20.7	95.6	1,239	17.0	93.5	3,959	18.1	94.7
other GUO found	366	5.0	100.0	318	4.4	100.0	471	6.5	100.0	1,155	5.3	100.0
Total	7,279	100.0		7,279	100.0		7,279	100.0		21,837	100.0	

Table 9 takes a closer look at the remaining mismatches and examines the different mechanics of the algorithm. Because of the lowest overall accuracy, the focus here is again on 2007 but results are similar for the other years. Applying the shareholder type corrections suggested by Jaraite et al. (2013) explains almost all of the mismatched misses. The adjustment at the very first level was not remedied by the cross-sample correction, indicating that the lost GUOs have no subsidiaries of meaningful size themselves. In addition to that, the remaining mismatched misses can be explained both by the different scope of the initial data export and the fact that no manual corrections were applied to the data. Mismatched hits, however, are largely explained by the cross-sample correction. This suggests that this method is an improvement in cases where a firm with subsidiaries is not identified as a GUO itself. In cases where another GUO was found neither the cross-sample correction nor the GUO type correction seem to drive the results. However, in a large share of these cases the end of the ownership chain could not be verified. This means that the firm indicated as the top shareholder was not of at least medium size and thus not part of our original export. In cases where the end of the chain could be verified and the results are still different it is possible that the ownership information was updated.

Table 10 details all firms in 2007 where the results were different, but unchanged by neither the cross-sample correction nor the GUO type adjustment. Mismatched misses are largely explained by missing top shareholder information. The results in the last two categories can be explained by firms not being included in our original sample. In this case, finding different GUOs is to be expected.

Table 9: Assessment of adjustments, 2007

	GUO similarity					
	matched hit	**matched miss**	**mismatched miss**	**mismatched hit**	**other GUO found**	**Total**
	No.	No.	No.	No.	No.	No.
Cross-sample correction						
Corrected	120	0	0	481	7	608
Unchanged	2,694	3,066	356	107	448	6,671
Total	2,814	3,066	356	588	455	7,279
GUO type correction						
Replicating Jaraite et al. (2013)	387	10	322	8	20	747
Replicating ORBIS	159	0	5	9	4	177
Unchanged	2,268	3,056	29	571	431	6,355
Total	2,814	3,066	356	588	455	7,279
Ownership chain tracing						
End not verified	632	10	325	34	267	1,268
End verified	2,182	0	4	78	184	2,448
no info	0	3,056	27	476	4	3,563
Total	2,814	3,066	356	588	455	7,279
N	2,814	3,066	356	588	455	7,279

Table 10: Assessment of unadjusted mismatches, 2007

	GUO similarity			
Ownership chain tracing	**mismatched miss**	**mismatched hit**	**other GUO found**	**Total**
	No.	No.	No.	No.
End not verified	2	16	244	262
End verified	0	78	183	261
no info	27	0	0	27
Total	29	94	427	550

A.4. Gap closing

Figure 6 illustrates the amount of nonmissing ownership information in each year. Since ownership links in ORBIS are terminated only if no new information is posted for an extended period of time, some of the gaps in the data could be cases of constant ownership. Figure 6 also illustrates the magnitude of the correction, which is applied to fill gaps between two years with identical ownership information. Technically, data is carried forward to fill gaps only if there is an existing data point in the future. Deletions of information remain unchanged.

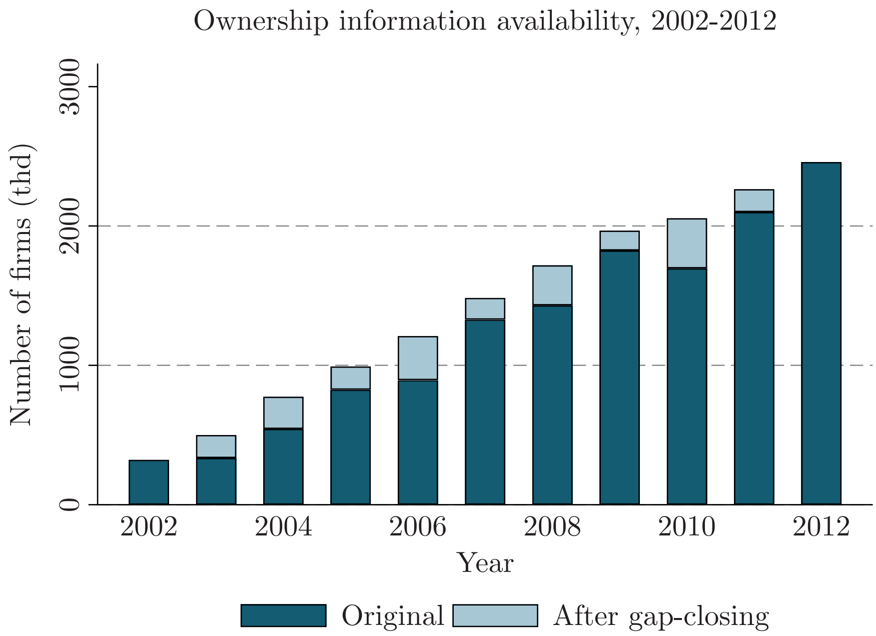

Figure 6: Historical GUO Info in ORBIS

Even if there had not been any gaps in the data, the inherent growth of the dataset would not allow for backward extrapolation. Although the gap-closing procedure also increases the share of verifiably constant information in the past, the overall result (that there is barely any) is not affected. Since some countries can miss certain years entirely, gap-closing should be applied at an early stage of ownership reconstruction.

A.5. Ownership data construction

Figure 7 illustrates the structure of the ownership data samples before they were merged with the financial data. Included are only affiliates of MNEs, which explains the difference to the total amount of ownership data in 2012 as illustrated in Figure 1. The constant 2012 ownership sample includes the same firms in each year. The intersection and the sample with yearly changing information converge to the constant sample in 2012. The difference between the yearly sample and the intersection sample are firms which reported ownership data in the past, but no longer do so. This difference explains the larger total number of firms in the yearly sample after estimating the effects. Merging the ownership data with the financial data greatly reduces the samples.

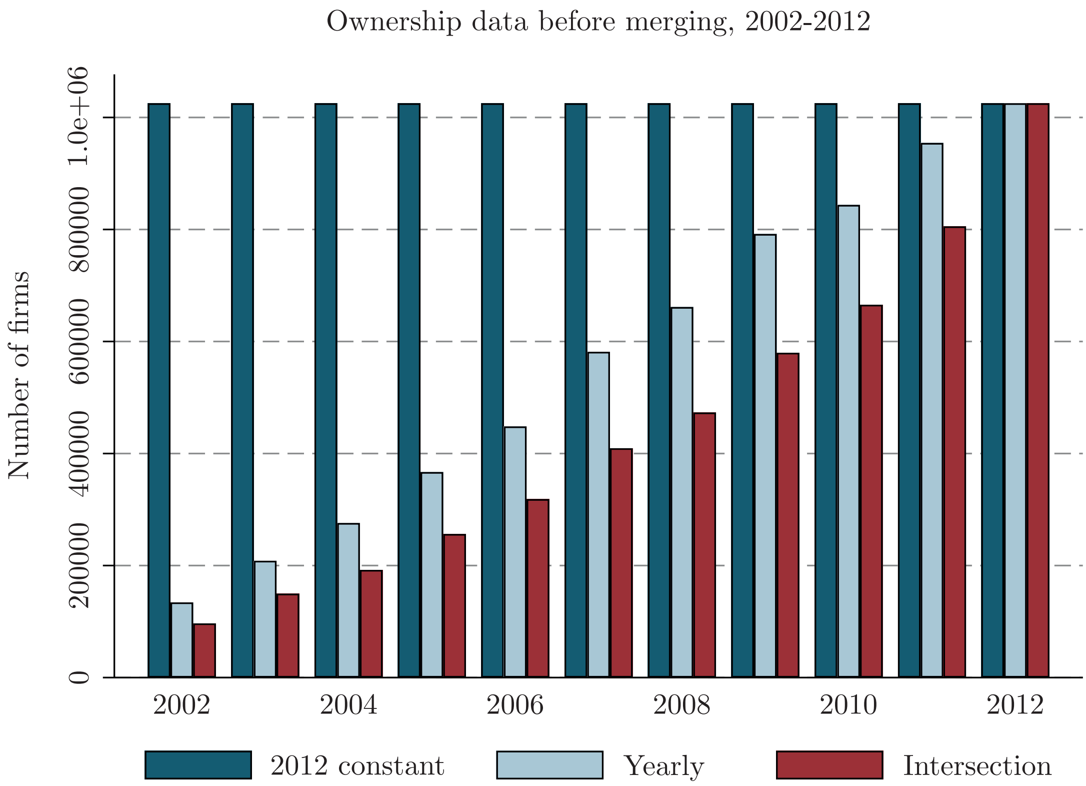

Figure 7: Ownership data construction, 2002-2012

A.6. Summary statistics

A.6.1. 2012 constant ownership

Table 11: Summary statistics, 2012 constant ownership

Variable	**Mean**	**Std. Dev.**	**Min**	**Max**	**N**
Corporate Tax Rate	0.31	0.057	0.1	0.409	881199
Tax Rate Differential (I)	-0.012	0.065	-0.308	0.304	881199
Tax Rate Differential (II)	0.002	0.038	-0.269	0.237	881199
Huizinga & Laeven's C	0.001	0.051	-0.307	0.412	881199
PLBT, in thd	9858.246	130804.565	1	70421232.984	881199
Fixed Assets, in thd	67891.476	1239678.177	1	352380280.346	881199
Cost of Employees, in thd	12453.192	123982.015	1	78895059.173	881199
GDP, in bn	1637.191	1086.356	15.927	3559.587	881199
GDP per capita, in thd	40.086	15.666	2.654	91.594	881199
GDP growth	1.379	2.635	-14.8	12.1	881199
Unemployment	8.013	3.128	2.493	24.787	881199
Corruption	6.971	1.654	2.2	9.700	881199

Table 12: Country distribution, constant 2012 ownership

Country	No.
Austria	2,725
Belgium	8,095
Bulgaria	400
Croatia	1,317
Czech Republic	2,414
Denmark	3,799
Estonia	941
Finland	2,947
France	22,737
Germany	13,155
Hungary	1,231
Ireland	1,111
Italy	18,555
Netherlands	2,860
Norway	7,434
Poland	4,760
Portugal	2,223
Romania	2,517
Slovak Republic	919
Spain	9,177
Sweden	8,005
Switzerland	32
Ukraine	1,353
United Kingdom	19,308
Total	138,015

A.6.2. Yearly ownership

Table 13: Summary statistics, yearly ownership

Variable	Mean	Std. Dev.	Min	Max	N
Corporate Tax Rate	0.305	0.06	0.1	0.409	686344
Tax Rate Differential (I)	-0.016	0.07	-0.308	0.304	686344
Tax Rate Differential (II)	0.002	0.041	-0.3	0.244	686344
Huizinga & Laeven's C	0.001	0.054	-0.306	0.449	686344
PLBT, in thd	11824.59	145408.616	1	70421232.984	686344
Fixed Assets, in thd	81366.068	1372636.552	1	352380280.346	686344
Cost of Employees, in thd	14127.549	133059.399	1	78895059.173	686344
GDP, in bn	1654.19	1104.401	15.927	3559.587	686344
GDP per capita, in thd	37.982	14.336	2.654	91.594	686344
GDP growth	1.296	2.792	-14.8	12.1	686344
Unemployment	8.372	3.445	2.493	24.787	686344
Corruption	6.893	1.609	2.2	9.700	686344

Table 14: Country distribution, yearly ownership

Country	No.
Austria	2,763
Belgium	7,787
Bulgaria	911
Croatia	1,003
Czech Republic	4,222
Denmark	3,514
Estonia	1,049
Finland	3,054
France	24,790
Germany	13,924
Hungary	1,431
Ireland	983
Italy	17,153
Netherlands	2,596
Norway	4,133
Poland	5,537
Portugal	3,264
Romania	3,680
Slovak Republic	1,008
Spain	14,193
Sweden	5,998
Switzerland	30
Ukraine	709
United Kingdom	18,325
Total	142,057

A.6.3. Intersection

Table 15: Summary statistics, intersection between 2012 constant and yearly data

Variable	Mean	Std. Dev.	Min	Max	N
D: Corporate Tax Rate	0.303	0.056	0.1	0.409	544032
C: Corporate Tax Rate	0.303	0.056	0.1	0.409	544032
D: Tax Rate Differential (I)	-0.016	0.068	-0.308	0.304	544032
C: Tax Rate Differential (I)	-0.016	0.069	-0.308	0.304	544032
D: Tax Rate Differential (II)	0.002	0.041	-0.266	0.244	544032
C: Tax Rate Differential (II)	0.002	0.042	-0.269	0.237	544032
D: Huizinga & Laeven's C	0.001	0.054	-0.306	0.449	544032
C: Huizinga & Laeven's C	0.001	0.056	-0.307	0.368	544032
PLBT, in thd	13234.634	158834.188	1	70421232.984	544032
Fixed Assets, in thd	92239.783	1516145.715	1	352380280.346	544032
Cost of Employees, in thd	15820.111	148420.815	1	78895059.173	544032
GDP, in bn	1705.412	1113.087	15.927	3559.587	544032
GDP per capita, in thd	39.158	14.3	2.654	91.594	544032
GDP growth	1.184	2.737	-14.8	12.1	544032
Unemployment	8.193	3.311	2.493	24.787	544032
Corruption	6.995	1.594	2.2	9.700	544032

Table 16: Country distribution, intersection

Country	No.
Austria	2,437
Belgium	6,057
Bulgaria	310
Croatia	840
Czech Republic	1,924
Denmark	2,964
Estonia	778
Finland	2,287
France	18,189
Germany	11,239
Hungary	1,010
Ireland	878
Italy	13,449
Netherlands	2,225
Norway	3,630
Poland	3,944
Portugal	1,967
Romania	2,090
Slovak Republic	750
Spain	7,205
Sweden	4,839
Switzerland	25
Ukraine	559
United Kingdom	15,731
Total	105,327

A.6.4. 2002 constant ownership

Table 17: Summary statistics, 2002 constant ownership

Variable	Mean	Std. Dev.	Min	Max	N
Corporate Tax Rate	0.312	0.058	0.1	0.409	214504
Tax Rate Differential (I)	-0.018	0.07	-0.308	0.304	214504
Tax Rate Differential (II)	0.001	0.039	-0.26	0.243	214504
Huizinga & Laeven's C	0.001	0.053	-0.306	0.376	214504
EBIT, in thd	16173.881	203529.026	1	69500417.869	214504
Fixed Assets, in thd	108981.215	1403428.649	1	204020645.802	214504
Cost of Employees, in thd	21894.07	124448.291	1	18689863.642	214504
GDP, in bn	1749.046	1091.288	15.927	3559.587	214504
GDP per capita, in thd	37.989	12.986	2.654	91.594	214504
GDP growth	1.567	2.478	-14.8	12.1	214504
Unemployment	8.01	3.205	2.493	24.787	214504
Corruption	7.098	1.501	2.2	9.700	214504

Table 18: Country distribution, 2002 constant ownership

Country	No.
Austria	931
Belgium	2,391
Bulgaria	134
Croatia	169
Czech Republic	760
Denmark	438
Estonia	86
Finland	550
France	4,852
Germany	4,590
Hungary	332
Ireland	155
Italy	2,523
Netherlands	804
Norway	817
Poland	846
Portugal	817
Romania	951
Slovak Republic	160
Spain	2,717
Sweden	900
Switzerland	6
Ukraine	27
United Kingdom	7,787
Total	33,743