Paris School of Economics and
École des Ponts ParisTech

Shifts and profits within multinationals and across jurisdictions

Francisco Gabriel

Supervisor: Prof. Angelo Secchi

Submitted in partial fulfilment of the requirements for the degree of
Master of Science in Economic Decision and Cost-Benefit Analysis of the
École de Ponts ParisTech and Paris School of Economics, November 2022

Abstract

This text analyzes multinational's country-by-country data in which revenues are partitioned into related - originating from within the firm - and unrelated revenues. A systematic relationship is uncovered: we find that those subgroups in which most of the revenue originates from within the multinational (but in other jurisdictions) tend to book higher profits than their counterparts, *ceteris paribus*.

The aforementioned effect does not appear to be homogeneous: stronger effects are found for the most profitable subgroups and in fact, the effect is the opposite for the less profitable subgroups. This suggests that there are two kinds of subgroups within the ones with mostly related revenues: those where there are no incentives to artificially inflate profit and those where there are.

The analyses herein were conducted on a novel dataset published by the EU Tax Observatory earlier this month[1].

[1]Data available at: `https://taxobservatory.shinyapps.io/company_cbcr_data/`

Acknowledgements

I am grateful to more people than I can address here.

I am very grateful to Prof. Angelo Secchi - for his insightful classes, his availability to discuss this thesis and his ability to uplift his students.

This work would not have been possible without the creation of the database of voluntarily published country-by-country reports. For that, I am grateful to the past and current members of the EU Tax Observatory involved in that project - in particular, Giulia Aliprandi and Gerrit von Zedlitz. The flexibility Panayiotis Nicolaides gave me to work on this text during office hours was fundamental and has been very much appreciated.

This work rests on a lot of open-source code: from my text-editor, to the R programming language and its great packages, to 'git' and to LaTeX.

Finally, I'm very very grateful to those closest to me, who have endured my difficult moods. Obrigado, Laura!

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This text is about profit shifting and tax avoidance. The Tax Justice Network - an important advocacy group on tax matters - elucidatively defines profit shifting as "a technique used by multinational corporations to pay less tax than they should that involves a multinational corporation moving the profit it makes in the country where it manufactures products or sells good and services into to a tax haven"[1]. This shift of profits causes multinationals to pay less tax in the countries where real economic activity happens. Profits are then taxed at much lower rates - if taxed at all, given that some countries have no taxes on profit (Bermuda is an example[2]) - in the jurisdictions in which they re-appear.

Historically speaking, shifting profits - and thus avoiding tax - has not been a difficult activity. And there have been multiple legal ways of achieving incredible reductions in the amount of tax due.

Nike's aggressive tax planing has been exposed by the Paradise papers[3]. From 2005 to 2014, Nike International Ltd - an entity based in Bermuda and with no employees - held Nike's intellectual property (including designs, branding, logo), enabling it to charge royalty fees to other Nike subsidiaries. As such, worldwide, Nike subsidiaries would have artificially lower profits, causing them to pay less tax to the local tax authorities. On the other hand, Nike

[1] https://taxjustice.net/faq/what-is-profit-shifting/
[2] https://taxsummaries.pwc.com/bermuda/corporate/taxes-on-corporate-income
[3] https://www.icij.org/investigations/paradise-papers/

International Ltd had tremendous profits year after year, taxed at rate of zero percent[4].

Tackling tax avoidance by multinationals involves international cooperation that has been proven difficult to obtain. Bermuda - despite being a British Overseas Territory - has fiscal autonomy. With a population of about sixty thousand, a zero tax rate on corporate profits can turn out to be extremely helpful for the local authorities, provided they charge multinationals some fee. The incentives for countries to undercut one another are there, irrespective of the country size or of the country's wealth. Poor countries undercut each other to attract investment; rich countries do it too - although some more than others.

Profit shifting and tax avoidance is not perpetrated by a few multinational enterprises - it is a systematic issue. Recent developments in research estimate that, globally, multinationals shift 36% of their foreign profits - about $600 billion - to tax havens (Tørsløv et al. (2022)). Several other studies have arrived at other figures for the amount of profit shifted globally to tax havens: Álvarez-Martínez et al. (2022) find that tax revenue losses for the EU, Japan, the UK, and the US amount to EUR 160.8 billion or 9.85% of corporate income tax revenues; Clausing (2016) estimates that, in 2012, over a thousand billion USD of profits were shifted to low-tax jurisdictions.

Multiple countries and international organizations have been attempting to stop the aforementioned rate to the bottom. Given the incentives at play and the magnitude of the issue, more scrutiny of multinationals' global activities is warranted in order to, on the one hand, identify and plug legal loopholes and, on the other, prevent foul-play by multinationals.

Action 13 of the OECD/G20 Base Erosion and Profit Shifting (BEPS) Project requires large multinationals to report to tax authorities country-by-country data on the global allocation of income, profit, taxes paid and economic activity among the countries in which they operate. These country-by-country reports are currently required to be filed by multinationals with consolidated group revenue upward of 750 million euros with the tax authorities of the jurisdiction in which they are headquartered. The tax authorities have been collecting the reports since 2016 and do not make the reports public.

[4] `https://www.theguardian.com/news/2017/nov/06/nike-tax-paradise-papers`

A key feature of these reports is that, for each country in which a multinational enterprise operates, revenues are partitioned into related and unrelated revenues.

Unrelated revenues are sales to third-parties. When a consumer goes into a Shell gasoline station in France and spends 100€, Shell-France's unrelated revenues grow by 100€. Summing unrelated revenue across all jurisdictions results in the Shell's consolidated revenue.

As France consumes over ten times the oil it produces [5], let us assume the gasoline the consumer put in her tank comes from Oman. Shell-France would then pay Shell-Oman 50€, which would appear under related-revenues of Shell-Oman.

This partitioning of revenues in the country-by-country reports filed with tax authorities allows the inspection of money flows within a corporation across jurisdictions - these flows may or may not be due to profit shifting.

There are international tax conventions that determine how much and where multinationals pay their taxes. Among other things, these conventions protect firms from double-taxation on their profits - *i.e.* profits of subsidiaries that have been taxed locally will not be taxed again at the headquarter country of the multinational. These conventions generally mandate enterprises to adhere to the arm's length principle, which states that the parties in a transaction should act in their own best interest. In other words, international transactions between entities of the same multinational company should mirror market prices.

According to the arm's length principle, Shell-France should pay the fair market price to Shell-Oman for the oil it buys - no more, no less. As the prices of oil and its derivatives are readily available, mispricing its transfers would be to blatantly violate this principle, with the ensuing legal consequences.

If I was the CEO of a multinational in the petroleum industry, with extracting operations in Oman and retail sales in France - I shall call my hypothetical multinational 'Concha' - and I was attempting to dogde France's relatively high corporate income taxes, I would set up a branch in the Bahamas (Concha-Bahamas), move my company's intellectual property (IP) there and

[5] `https://www.worldometers.info/oil/france-oil/`

then charge Concha-France approximately the amount of profits it would otherwise book in France for the use of the valuable 'Concha' logo. This way, my profit would be shifted to the Bahamas, where the corporate income tax rate is - like in Bermuda - zero percent. It would then be much harder for tax authorities to prove any foul-play.

Table 1.1: Shell's country-by-country report partitions revenues into related and unrelated.

Country	Unrelated Revenues	Related Revenues	Employees	Pre-tax Profit	Tax Paid
Bahamas	4 880M$	10 978M$	35	653M$	0M$
France	1 140M$	269M$	349	-60M$	3M$
Oman	985M$	5 244M$	475	1 324$	1 083$

Source: Shell Tax Contribution Report 2020
Available at: `https://reports.shell.com/tax-contribution-report/2020/`.

Table 1.1 shows an abridged version of Shell plc - not Concha - country-by-country data. While the reports filed with the local tax authorities are not made public, Shell plc has voluntarily published what it claims to be its country-by-country data. One would expect Shell-Oman to have, as it does, high related-revenues: Oman is a oil-rich country and Shell has extractive operations there. As for Shell-Bahamas, the case is not as straightforward: the 'Shell Bahamas Power Company Inc.' - previously named 'Shell E&P Ireland Offshore Inc.' - books about sixteen billion US dollars in total revenues, despite having only 35 employees. To put things into perspective, that is about half of the worldwide consolidated revenue of L'Oréal in the same year[6] and the profitability per employee is of about nineteen million US dollars that year. For the interested reader, Shell's country-by-country data for the years of 2018 and 2019 are all the more interesting [7][8][9]. Shell-Bermuda appears to be another gem. In 2018, this subgroup had more profit than it had revenues - dividends are not included in revenues -, with each of the 3 employees bringing more than 200M$ of profit, on average.

From the inspection of the abridged version of its 2020 Tax Contribution report, we can divide the Bahamas, France and Oman into two groups - one, comprised of the Bahamas and Oman, where share of related revenues over the total revenues is quite high; another, comprised of France, where it isn't. Note how this share is actually higher in Oman relative to the Bahamas

[6] `https://www.loreal-finance.com/system/files/2022-03/LOREAL_2021_Annual_Report.pdf`
[7] `https://www.reuters.com/article/uk-global-oil-tax-havens-specialreport-idUKKBN28J1IC`
[8] `https://reports.shell.com/tax-contribution-report/2018/`
[9] `https://reports.shell.com/tax-contribution-report/2019/`

(84% *vs* 69%, respectively).

Let us transpose these country-by-country results to the hypothetical Concha company. The figures in Oman should not arise any particular suspicion given that high related revenues may be perfectly justifiable. Rather, it is another figure in the table that would make Concha's operation in the Bahamas questionable: there is almost five billion US dollars in unrelated revenue. Given that the Bahamas has a population of under 360 000 inhabitants [10], this would mean that if were Concha to be selling gasoline to the locals, the average Bahamian would be spending an yearly amount upward of 13 000$.

Considering how profit-shifting involves transferring money within company and across jurisdictions, reliable country-by-country data in which revenues are partitioned into related and unrelated revenues data must reflect profit shifting, when existing. In this case, the difficulty lies in discerning it from 'normal' economic activity in which the arm's length principle is upheld.

As mentioned, the country-by-country reports files with the the tax authorities are not public. However, due to mounting pressure by some investors and civil society, there are multinationals publishing their country-by-country reports, making them freely available online. These reports are most often shared in PDF format, which impedes the analysis of the data therein. The EU Tax Observatory[11] has aggregated and standardized such reports into a public database[12].

The remainder of this text studies the relationship between related revenues (the 'shift') and profit (before tax) for large multinationals.

The text is organized as follows:

- Chapter 2 reviews the literature on profit shifting, focusing particularly on the most relevant empirical approaches, on the magnitude of the issue and on the main channels for tax evasion;

[10] `https://www.cia.gov/the-world-factbook/countries/bahamas-the/`
[11] `https://www.taxobservatory.eu/`
[12] The data is available at `https://taxobservatory.shinyapps.io/company_cbcr_data/`

- Chapter 3 presents the dataset - country-by-country data for Shell in 2020 is used as a motivating example - and regresses profits on the relative weight of related revenues;
- Chapter 4 discusses the main results and the limitations of the models;
- Chapter 5 concludes.

Appendix A details the construction of the database of voluntarily published country-by-country reports.

Chapter 2

Review of the literature

The literature relative to profit shifting and corporate tax avoidance is considerable and, as shown below, is evolving at a high pace.

2.1 Empirical approaches

An often used approach to empirically estimate base erosion and profit shifting follows the seminal work by Hines Jr and Rice (1994) and Grubert and Mutti (1991). It stems from the observation that the profit before tax as reported by an affiliate can be decomposed into the 'real' - the fruit of the economic activity, using labor and capital as inputs - and into the 'shifted' profit. Measures of this real economic activity are included in the regression specification as controls; 'shifted' profit, on the other hand, is determined by the incentive the affiliate has to artificially inflate or deflate profits. In some regression models, differentials between the tax rates to which the affiliates are exposed and the tax rates that the parent entity faces serve as a proxy for the 'incentive' part of the model. More intricate mutations of this framework take into account the tax rates faced by all affiliates - see Heckemeyer and Overesch (2017).

This approach can be represented by a model in which the logarithm of the profit (Π) is regressed on the tax-rate differential (τ) - the variable of interest -, on the logarithm of the

capital inputs (K), on the logarithm of the labor input (L) and potentially on other controls (X), as shown in Equation 2.1. It is customary to omit loss-making affiliates, given that the incentives for such affiliates are attenuated by the asymmetries of tax law. Such models are very similar to the one use later in this text, in Chapter 3.

$$\log \Pi_i = \beta_0 + \beta_1 \tau_i + \beta_2 \log K_i + \beta_3 \log L_i + X_i + \epsilon_i \tag{2.1}$$

Accounting literature has come up with an alternative. To estimate profit shifting, Dyreng and Markle (2016) rely on the assumption that the allocation of a multinational's sales is difficult to manipulate. Under this - quite stringent - assumption, it is possible to study profit shifting by analyzing differences between the location of multinational's sales and the location of their reported earnings. When panel data is available, other empirical strategies are possible, such as the one described in Dharmapala and Riedel (2013).

2.2 Overview of findings of the empirical literature

It can be demonstrated that, due to its specification, the aforementioned model pioneered by Hines Jr and Rice (1994) estimates the semi-elasticity of profit before tax with respect to the differential of the tax rates. This semi-elasticity represents the percentage change in pre-tax profits associated with a rise of one percentage point in the tax rate differential. For example, suppose that this semi-elasticity is estimated to be equal to 1, that Concha-Portugal operates under the same tax rate as Concha-Spain (*ergo* the differential is zero), and that Concha-France books one billion euros in pre-tax profit. This means that a drop of 10 percentage points in the tax rate in France (and hence a increase of 10 percentage points in the differential with respect to Spain) would be associated with a 10 percent increase in Concha-France's pre-tax profit, elevating it to 1.1B€.

Heckemeyer and Overesch (2017) perform a meta-regression in which 238 estimated semi- elasticities from 25 separate academic studies of profit shifting are collected. Their 'consensus'

regression estimates the semi-elasticity to be of approximately 0.8, after controls. Huizinga and Laeven (2008) and Hines Jr and Rice (1994) compute semi-elasticity estimates of 1.31 and and 2.25, respectively.

2.2.1 The scale of tax avoidance

The literature on the scale of tax avoidance is extensive and growing rapidly. Tørsløv et al. (2022) estimate that, globally, multinationals shift 36% of their foreign profits - about $600 billion - to tax havens. Cobham and Janskỳ (2018) investigates how profit shifting affects different areas of the globe. The authors find that poor regions - in particular South Asia, Sub-Saharan Africa, Latin America and the Caribbean - are hit the hardest in terms of tax revenue losses as a fraction of the size of their economy. Tax revenue losses are estimated - through the use of regressions - for both OECD and non-OECD countries, as is the case in Crivelli et al. (2016). Cobham and Janskỳ (2018) finds tax revenue losses of amount to .7% in OECD countries and 1.3% in non-OECD; Crivelli et al. (2016) estimates 1% and 1.3%, respectively. Janskỳ and Palanskỳ (2019) corroborates the finding that poor countries lose more tax revenue as a fraction of their economy. It further estimates that, in 2016, 37% of multinationals profit was shifted.

Using aggregated data from annual surveys of multinationals headquartered in the US, Clausing (2016) finds that seven tax havens with effective tax rates under 5% are responsible for half the foreign income of US-based multinational, while accounting for 5% of their foreign employment. Further, this study estimates that profit shifting cost the US almost half of the federal corporate tax revenue in 2012 - with the figures ranging from $77 billion to $111. Dyreng et al. (2022) leverage publicly available consolidated financial statements to ascertain that a small number of firms account for most of the profits shifted. Their estimates for profit shifted out of the US for the year of 2012 is significantly lower than Clausing (2016): $107 billion, corresponding to $37 billion in tax revenue losses.

Relying on country-by-country data filed by German multinationals to the German tax authority for the years of 2016 and 2017, Fuest et al. (2022) find that these enterprises shift about €5.4

billion of pre-tax profits to tax havens - a loss of €1.6 billion to the German tax authorities. This study further states that the majority (87%) of the total profits recorded in tax havens are booked in European tax havens. Álvarez-Martínez et al. (2022) finds that the tax revenue losses by the EU due to profit shifting amount to 36 billion€a year, corresponding to over 7% of corporate tax revenues. The figures would be higher were not for the fact that some EU countries attract artificial profits from other rich countries. These results are obtained from a computable general equilibrium model.

Bratta et al. (2021) estimates the semi-elasticity of profits relative to income taxes. This study finds that about 80% of global profit shifting involves just seven 'source' and eight 'sink' countries. It estimates that global tax revenues were cut by €245 billion in 2017 (from €887 billion of profits shifted). In the same vein, Garcia-Bernardo and Janskỳ (2022) calculates that a small group of countries with effective tax rates under 1% attracts 40% of the shifted profits and that the global scale of profit shifting amounts to close to one trillion US dollars - this estimate is consistent with tax revenue losses ranging from $200 million to $300 million. These estimates rely on country-by-country data published by the OECD for the year of 2016 and by the US' Internal Revenue Service for 2017.

An important article that sheds light on tax avoidance is Bilicka (2019). By matching UK subsidiaries of multinationals with domestic firms, the author finds that the former are 30 percentage points more likely to report zero taxable profits than the latter. The study further finds that the mean ratio of taxable profits to total assets for subsidiaries of multinationals is half that of matched domestic firms.

Finally, two relevant texts on the topic of the magnitude of profit shifting are Wier and Reynolds (2018) - which analyzes data from the South African Revenue Service to show that the propensity to shift profits is for the top 5% largest subsidiaries - and Fatica and Gregori (2020), in which country-by-country data relative to banking sector is used to estimate that 21% of the profits in the sector are shifted for tax reasons.

2.2.2 Channels of tax avoidance

The literature on tax avoidance focuses on the main channels for profit shifting: transfer mis-pricing, strategic location of intellectual property (IP) and international debt shifting (Beer et al. (2020) and Dharmapala (2014)).

Regarding debt-shifting, Bilicka (2019) finds that UK subsidiaries of multinational firms take on significantly more debt than comparable domestic firms. From her computations, debt-shifting accounts for approximately 40% of the profit ratio gap between the two types of firms. Johansson et al. (2017), on the other hand, estimate debt-shifting to be the cause of at least 20% of profit shifting by multinationals.

Regarding transfer pricing, Cristea and Nguyen (2016) use a triple difference estimation method to arrive at the conclusion that once a Danish multinational owns an affiliate in a country with lower corporate tax rate, they reduce the unit value of their exports there by a figure ranging between 6% and 9%. They further find that this effect is stronger for differentiated goods.

On the topic of IP - both of patents and intangible assets in general -, Baumann et al. (2020) reports that 79% of patents in held tax havens were invented in another foreign country, while this figure is 6% for patents held elsewhere. Johansson et al. (2017) finds that tax planning reduces effective tax rate of large multinationals by four to eight percentage points and that the effect is more preeminent for firms with intensive use of intangible assets.

Chapter 3

Data and Analysis

This chapter starts by introducing the dataset used: first, by using country-by-country data relative to Shell in 2020 as a motivating example; then, by presenting summary statistics for a rich subset of the data published by the EU Tax Observatory in November of 2022[1].

Once that is achieved, the dataset is analysed in Section 3.2.

3.1 The dataset

This text relies on the data from voluntarily published country-by-country reports. These often come in PDF and are difficult to study. A team of past and current members of the EU tax observatory - of which I am part - compiled these documents into a public database. Appendix A goes into detail about the creation of the database.

The dataset used is comprised of over 200 reports - collected between 2016 and 2021 - from over 100 multinationals. Most of them have consolidated (unrelated) revenues upward of €750M.

Data is collected on unrelated and related revenues; profits before income tax; tax paid and accrued; number of employees and tangible assets other than cash and equivalents; and on

[1] data available at here.

stated capital and accumulated earnings. Not all observations have data on all these variables - far from it.

In order to analyse the data in regressions with sufficient controls, the observations kept for analysis purposes were required to contain information on all of the following variables: unrelated and related revenues, profits before tax, number of employees and tangible assets.

Access to panel data would be ideal. However, it is rarely the case that there is rich data for multinationals spanning over several years. As such, for the remainder of the chapter, only data for the year 2020 is kept - 2020 is by far the year for which more data is available.

3.1.1 Shell as a motivating example

Table 3.1 shows illustrative country-by-country data from Shell relative to the Bahamas (BHS), Bermuda (BMU), France (FRA), Kazakhstan (KAZ), Oman (OMN) and Spain (ESP).

In France and Spain, most of the sales occur to other parties (the figures for unrelated revenues are significantly higher than related revenues) and the corporate income tax rates were, respectively, 29% and 25% - including social contribution payments - in 2020.

Kazakhstan and Oman are resource-rich countries where extraction takes place. The corporate income tax rates are set at 20% and 15%, respectively, but there are production-sharing agreements that increase tax rates for upstream projects. In fact, in Oman, the effective tax rate is of 84%. In both these countries, most revenue comes from within Shell, across jurisdictions, but there is little incentive to shift profits into these countries.

Finally, neither the Bahamas nor Bermuda extract petroleum and neither had corporate income tax in 2020. Most revenues come from within Shell, from other countries.

Figure 3.1 shows how the different subgroups at Shell partition the revenue they book into related and unrelated revenues. Note that both scales are logarithmically transformed and are in millions of euros. The black line splits the subgroups into two: those with mostly related revenues - a minority - are above it, those with mostly unrelated are below. Jurisdiction below

Table 3.1: Abridged country-by-country data relative to Shell 2020 (monetary values in euros).

jurisdiction	unrelated_revenues	related_revenues	profit_before_tax	employees	tangible_assets
BHS	4 280 591 227	9 629 514 302	572 475 690	35	626 354 146
BMU	3 509 315	61 965 331	31 653 673	2	80 067
FRA	1 000 434 145	236 619 057	-52 361 708	349	443 987 741
KAZ	29 079 917	1 316 518 163	81 750 435	393	9 426 328 934
OMN	864 389 822	4 599 895 609	1 161 714 560	475	231 316 588
ESP	448 732 757	144 576 952	14 302 976	132	38 067 971

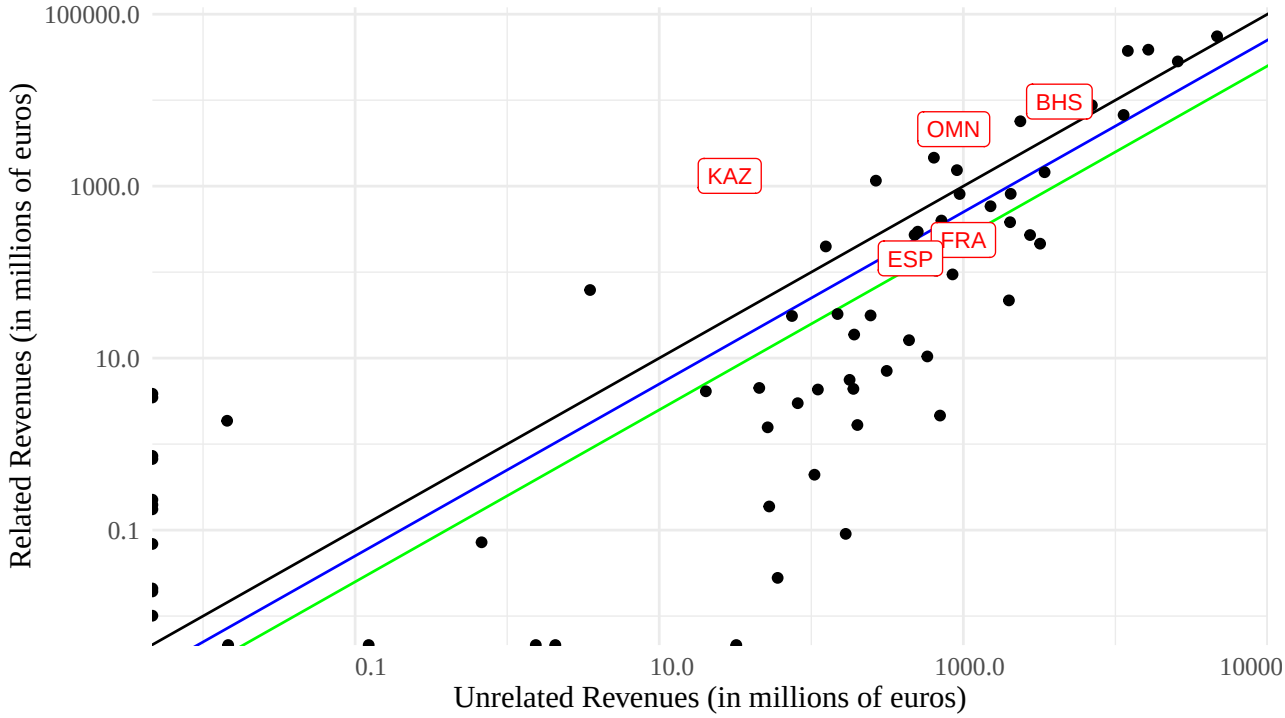


Figure 3.1: Related revenues to unrelated revenues for Shell's subgroups.

the blue line have at least twice as much unrelated revenues relative to related revenues; the figures are four times as much for the green line. Both Spain and France are close to the green line.

Next, it is interesting to see if there are any patterns with respect to profitability. Figure 3.2 is unequivocal regarding which of the representative subgroups book more profit per employee - note that the French subgroup is not present as it recorded losses in 2020 and that, as before, the scales have been transformed.

Figure 3.3 presents profitability per capital employed (tangible assets). What is surprising in this case is how little tangible assets Shell holds in Oman.

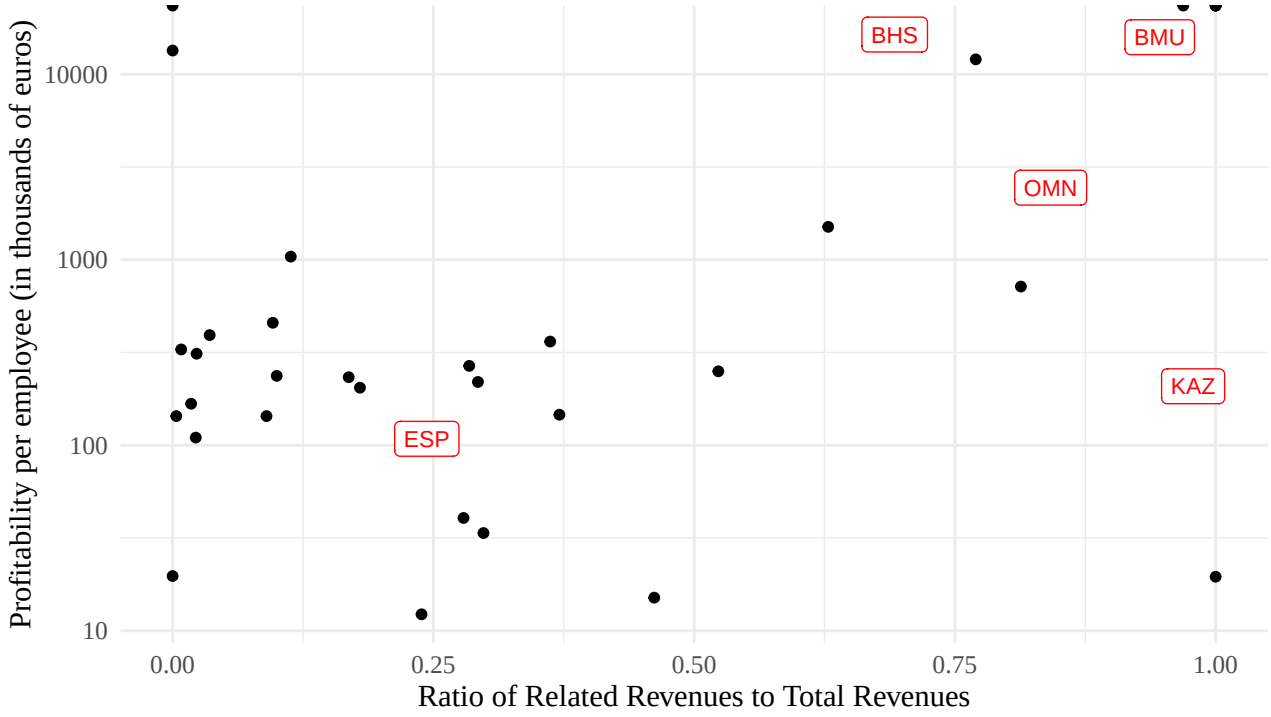


Figure 3.2: Shell's profitability per employee to ratio of related revenues.

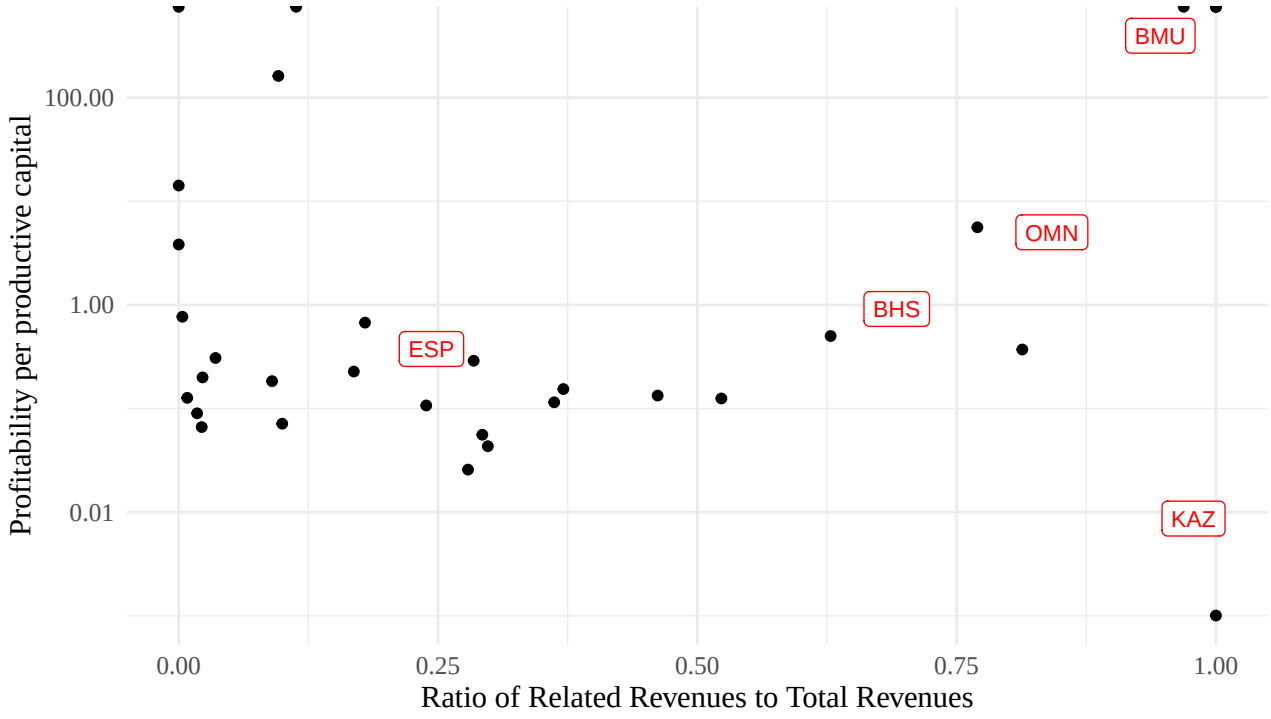


Figure 3.3: Shell's profitability per productive capital to ratio of related revenues.

Table 3.2: Economic Activities most frequently associated with the multinationals in the dataset.

Economic Activity	# of Multinationals
C - Manufacturing	16
B - Mining and quarrying	9
Other	9
J - Information and communication	3
H - Transportation and storage	3
D - Electricity, gas, steam and air conditioning supply	3

Table 3.3: HQ countries most frequent regarding the multinationals in the dataset.

HQ jurisdiction	# of Multinationals
ITA	16
Other	7
GBR	6
AUS	5
ESP	5
NLD	4

Throughout the previous Figures, it is apparent that there are - broadly speaking - three sets of countries of which these six countries are representatives. The group of countries in which Spain and France are included has considerably more elements than the rest.

3.1.2 The dataset used in the analysis

This is the point in which Shell is left behind and we shift our attention to the whole dataset that will be used for analysis. Table 3.2 presents the economic activities most frequent among the multinationals that make up the dataset that will be analysed. Specifically, it states the NACE[23] (Nomenclature of Economic Activities) codes for the parent entity of the multinational. Manufacturing is the most frequent activity, followed by mining and and quarrying. Table 3.3 indicates in which jurisdictions most of the multinationals present in the dataset are headquartered.

Figure 3.4 shows the histogram of the share of related revenues over the total revenues for the

[2] https://nacev2.com/en

[3] https://ec.europa.eu/eurostat/documents/3859598/5902521/KS-RA-07-015-EN.PDF.pdf/dd5443f5-b886-40e4-920d-9df03590ff91?t=1414781457000

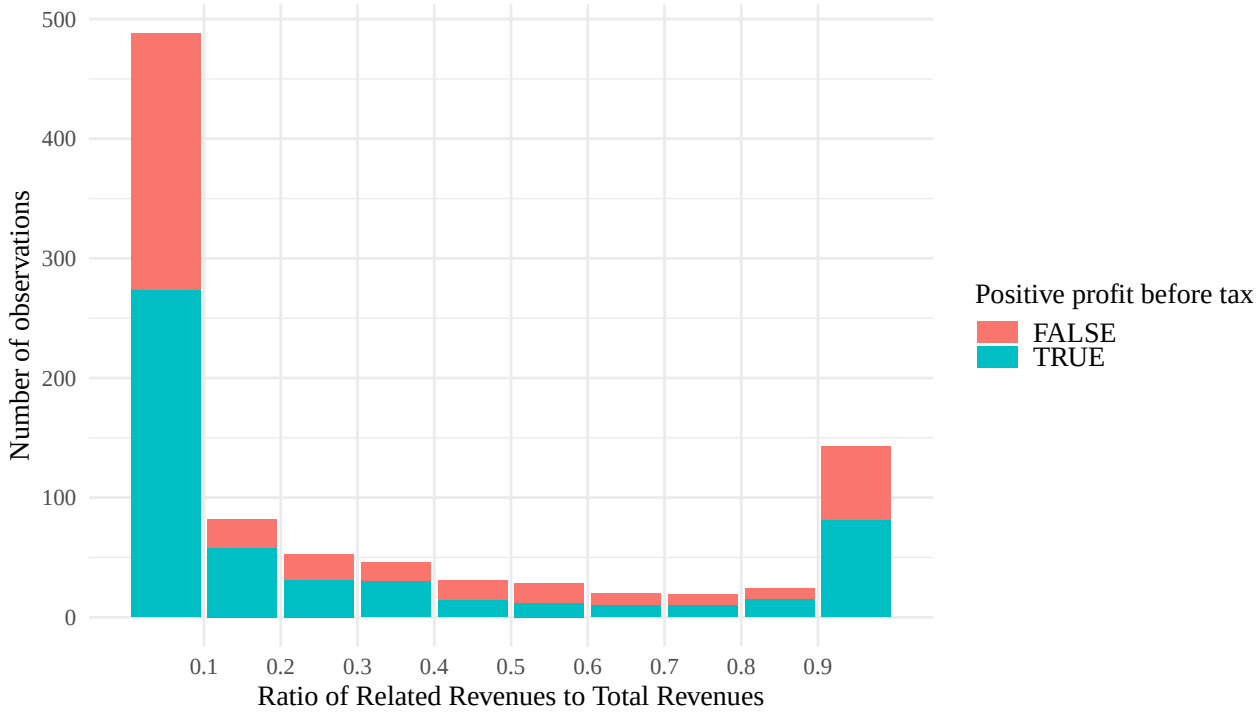


Figure 3.4: Histogram of the ratio of related revenues across the dataset.

dataset under consideration. Considering that only the profitable subgroups will be included in part of the analysis, different colors are used for the profitable and unprofitable subgroups. In any case, the main point is clear: there are many subgroups with less than 10% of related revenues. There is then a sharp drop to the number of subgroups with 10% to 20% of related revenues and the numbers keep trending downward. There is a significant number of subgroups with more than 90% of related revenues.

Table 3.4 presents summary statistics relative to the subgroups. The first thing that is apparent is how far the median subgroup is from the average subgroup: the former books about €80M in revenue, has about one hundred employees and €9M in tangible assets; for the latter, the figures are €1,75B, 1 400 and €1B.

Figure 3.5, 3.7 and 3.6 are the counterparts of figures 3.1, 3.3 and 3.2 (respectively), with the difference that all subgroups that will undergo analysis have been included. Subgroups corresponding to headquarter countries are marked with a '+' sign.

Table 3.4: Summary statistics for the variables of interest in the dataset under study.

Statistic	Mean	St. Dev.	Median
total_revenues	1,745,741,981.0	7,233,009,183.0	78,295,597.0
related_revenues	628,722,929.0	3,593,608,177.0	3,317,454.0
unrelated_revenues	1,117,019,052.0	4,216,948,789.0	47,121,231.0
profit_before_tax	−682,588,879.0	42,679,129,808.0	424,615.7
employees	1,424.3	5,298.7	99
tangible_assets	1,047,008,964.0	3,987,863,598.0	8,911,000.0

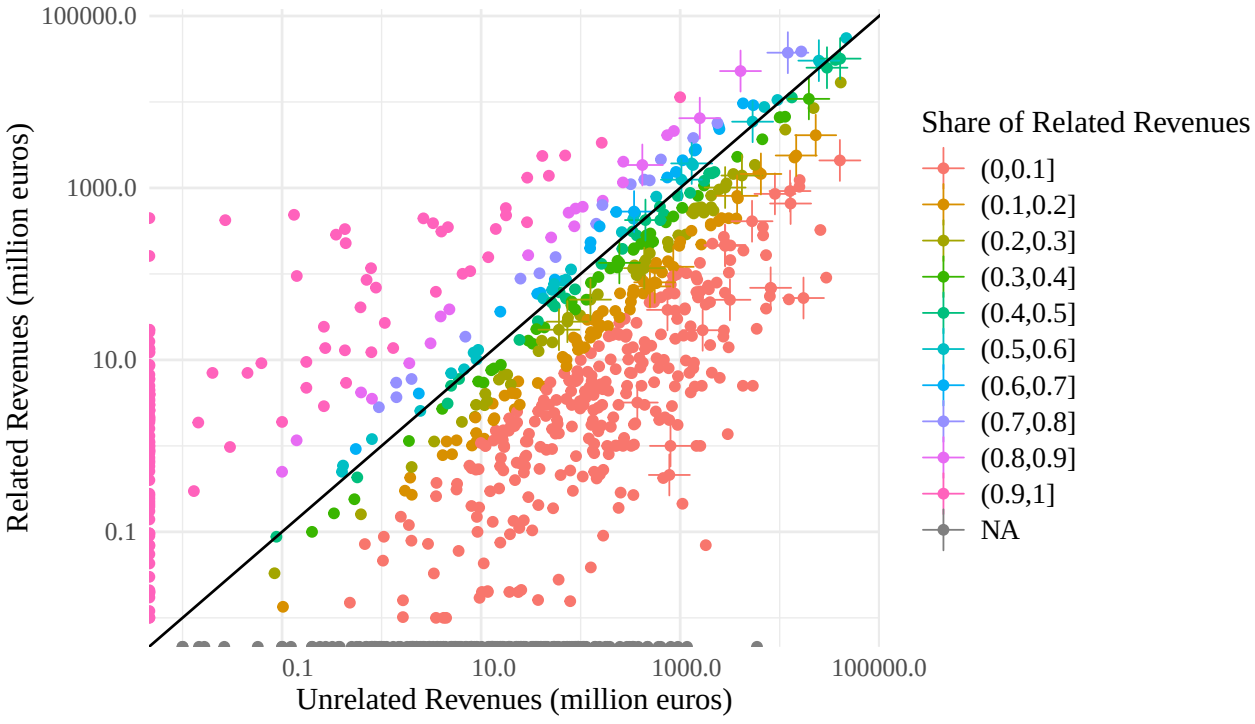


Figure 3.5: Related to Unrelated across the dataset.

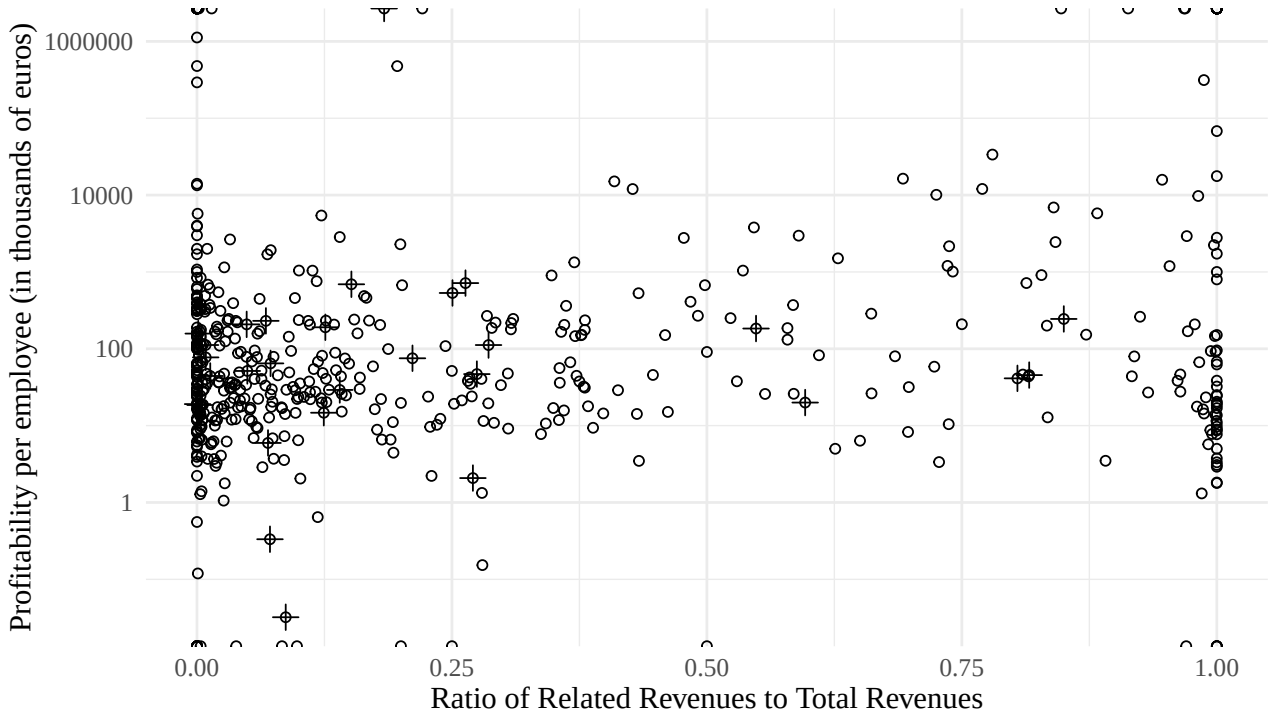


Figure 3.6: Profitability per emploee to the ratio of related revenues across the dataset.

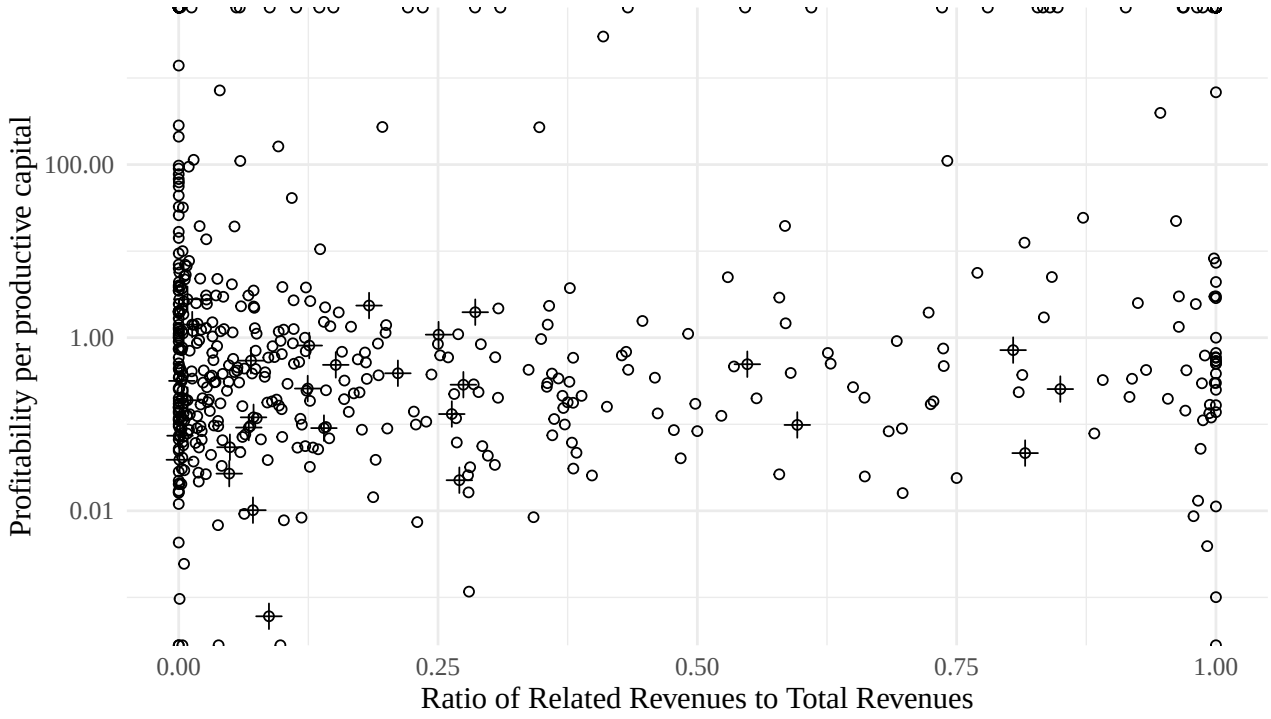


Figure 3.7: Profitability per value of tangible assets to the ratio of related revenues across the dataset.

3.2 Analysis

In order to study the relationship between the share of related revenues and pre-tax profit, a model closely related to the one presented in Equation 2.1 was chosen.

This model, as specified in Equation 3.1, regresses pre-tax profit (Π) on a binary variable that flags whether or not most of the revenue of a subgroup originates from within the multinational - the regressor of interest -, on the amount of tangible assets other that cash and cash-equivalents (K), and on number of employees (L). MR, which stands for 'mostly related', denotes the regressor of interest and ϵ_i the residual.

The rationale for choosing a binary variable as the regressor of interest instead of the ratio of related revenues goes as follows: while the share of related revenues may contain information relative to profit shifting, should it exist, I don't expect this relationship to be linear. As seen in the introductory chapter, this ratio can be higher in resouce-rich countries than in known tax-havens. Further, Shell does have a share of close to 20% related revenues in France - it operates a lubricants blending plant in Nanterre. As it is unlikely that any multinational shifts profits into France, these 20% are not likely to be indicative of any form of profit shifting.

$$\log \Pi_i = \beta_0 + \beta_1 \mathrm{MR}_i + \beta_2 \log K_i + \beta_3 \log L_i + \epsilon_i \tag{3.1}$$

A regression based on Equation 3.1 is liable to omit variable biases. Important potential sources of bias would be associated with the multinational companies themselves: the sector in which they operate, their corporate structure or their overall competence. As such, multinational fixed-effects were introduced as dummy variables in order to address factors that vary from one enterprise to the next.

Another potential ommited variable bias (OVB) would be the total revenue each subgroup books. Suppose Concha-Oman extracts petroleum and sells gasoline to Concha-Spain, which then sells it to the public. Suppose both subgroups employ the same amount of capital and labor. Concha-Spain will sell the gasoline to the final consumer at a higher price than the one

it bought it from Concha-Oman - after all, Concha-Spain added value. As revenues are the highest at the end of the production chain, the regressor of interest is expected to be negatively correlated with the total amount of revenues, conditional on the amount of labor and capital employed. As it is likely that, for a given amount of employees and capital, the total amount of revenues is correlated with profit - particularly in a setting in which only profitable subgroups will be used in the regression - omitting the total revenues would most probably lead to an OVB, causing the effect of the regressor of interest to be underestimated.

Finally, given that firms may double count profits by including dividends, a binary variable stating whether the subgroup is the headquarter of the multinational is included. Godar et al. (2022) finds no evidence of this issue when analysing voluntarily published country-by-country data. However, as the samples differ - with the sample in Godar et al. (2022) being significantly smaller - a dummy variable on whether the subgroup is the headquarter seems reasonable, in order to stay on the safe side.

Taking the these potential OVBs into consideration, the regression specification can be written as in Equation 3.2, in which 'T_i' denotes the total amount of revenues, 'isHQ' the binary variable that flags headquarter subgroups and '$\overline{X_i} \cdot \overline{\gamma}$' the company fixed-effects.

$$\log \Pi_i = \beta \mathrm{MR}_i + \beta_1 \log K_i + \beta_2 \log L_i + \beta_3 \log T_i + \beta_4 \mathrm{isHQ}_i + \overline{X_i} \cdot \overline{\gamma} + \epsilon_i \tag{3.2}$$

3.2.1 Ordinary Least Squares

Tables 3.5 and 3.6 present, respectively, summary statistics and the correlation matrix for the data used in the OLS regression. As the OLS is very sensity to large outliers and the original data varied by several orders of magnitude (see 3.4), taking the logarithm of the variables reduces the variance - and other moments of the distribution - into values better suited for the OLS. Furthermore, following the literature (see for example Hines Jr and Rice (1994)), only profitable subgroups were considered.

The ordinary least squares (OLS) estimates for the coefficients can be found in Table 3.7. Tables

Table 3.5: Summary statistics for the data used in the OLS regression.

Statistic	Mean	St. Dev.	Median
total_revenues	18.2	2.8	18.5
related_revenues	13.7	6.6	15.4
profit_before_tax	16.0	2.8	15.9
employees	4.8	2.6	4.9
tangible_assets	14.7	6.8	16.3

Table 3.6: Correlation matrix for the data used in the OLS regression.

	profit_before_tax	employees	tangible_assets	related_revenues	total_revenues
profit_before_tax	1	0.609	0.536	0.391	0.792
employees	0.609	1	0.693	0.473	0.784
tangible_assets	0.536	0.693	1	0.329	0.683
related_revenues	0.391	0.473	0.329	1	0.506
total_revenues	0.792	0.784	0.683	0.506	1
mr	-0.117	-0.224	-0.302	0.275	-0.237

3.8 and 3.9 present the same results - fixed effects at the multinational-level are always included - but cluster the standard errors at the multinational and jurisdiction levels, respectively.

	Estimate	Robust s.e	t value	Pr(>|t|)
mr	0.3609	0.1660	2.17	0.0301
total_revenues	0.7552	0.0679	11.12	0.0000
employees	0.1301	0.0742	1.75	0.0803
tangible_assets	-0.0260	0.0176	-1.47	0.1410
is_hqTRUE	0.2979	0.3225	0.92	0.3561

Table 3.7: Regression coefficients, together with the heterosketasticity robust standard errors.

All of the aforementioned computations result in standard error estimates of less than half of the coefficient associated with the regressor of interest - the 'mostly related' binary variable. As such, the hypothesis of β being zero is rejected at the significante level of 5%.

As shown in 3.3 below, where 'MR' and 'MU' denote 'mostly related' and 'mostly unrelated', respectively, the value obtained for the parameter β is substantial: *ceteris paribus*, the profits of a profitable subgroup are 43% higher when most of the revenue comes from within the multinational, from other jurisdictions.

Table 3.8: Regression coefficients - standard errors clustered at the multinational-level.

	Dependent variable:
	profit_before_tax
mr	0.361**
	(0.174)
total_revenues	0.755***
	(0.075)
employees	0.130
	(0.088)
tangible_assets	−0.026
	(0.021)
is_hq	0.298
	(0.355)
Observations	534
R^2	0.781
Adjusted R^2	0.759
Residual Std. Error	1.372 (df = 486)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 3.9: Regression coefficients - standard errors clustered at the jurisdiction-level.

	Dependent variable:
	profit_before_tax
mr	0.361**
	(0.152)
total_revenues	0.755***
	(0.074)
employees	0.130*
	(0.078)
tangible_assets	−0.026
	(0.020)
is_hq	0.298
	(0.269)
Observations	534
R^2	0.781
Adjusted R^2	0.759
Residual Std. Error	1.372 (df = 486)
Note:	*p<0.1; **p<0.05; ***p<0.01

$$\begin{aligned} \log \Pi_{\text{MR}} - \log \Pi_{\text{MU}} = \beta \Leftrightarrow \\ \Leftrightarrow \Pi_{\text{MR}} = e^{\beta} * \Pi_{\text{MU}} \Leftrightarrow \\ \Leftrightarrow \Pi_{\text{MR}} \sim 1.43 * \Pi_{\text{MU}} \end{aligned} \tag{3.3}$$

3.2.2 Quantile regression

The identification strategy discussed above does not differentiate subgroups such as Shell-Kazakhstan - which report high shares of related revenues and are subject to high tax rates - from subgroups such as Shell-Bahamas, unburdened by corporate income taxes.

In this context, a quantile regression (see Koenker and Hallock (2001)) may offer valuable insight. Instead of computing the coefficients that provide the minimum mean squared error - as is the case for the OLS -, it allows for the estimation of the conditional quantiles of the response variable.

In a setting in which related revenues are sometimes but not always associated with profit shifting, we would expect the regressor of interest not to be statistically significant for the lower quantiles of the profit distribution but to be so in the higher quantiles. This can indeed be observed in Figure 3.8. The lowest value for the coefficient of the regressor of interest in the right-hand side of Figure 3.8 corresponds to 0.493 (for the eightieth percentile). A re-computation of 3.3 for this figure yields a 64% increase in profits before tax when revenues are mostly related relative to when they are mostly unrelated, everything else being constant.

It should be noted that, unlike for the OLS regression, the estimates for the coefficients in the quantile regression were obtained from the complete dataset - there was no filtering out of unprofitable subgroups[4].

[4]the pre-tax profit underwent the following transformation:

$$x \to \text{sign}(x) * \log(\text{sign}(x) * x + 1)$$

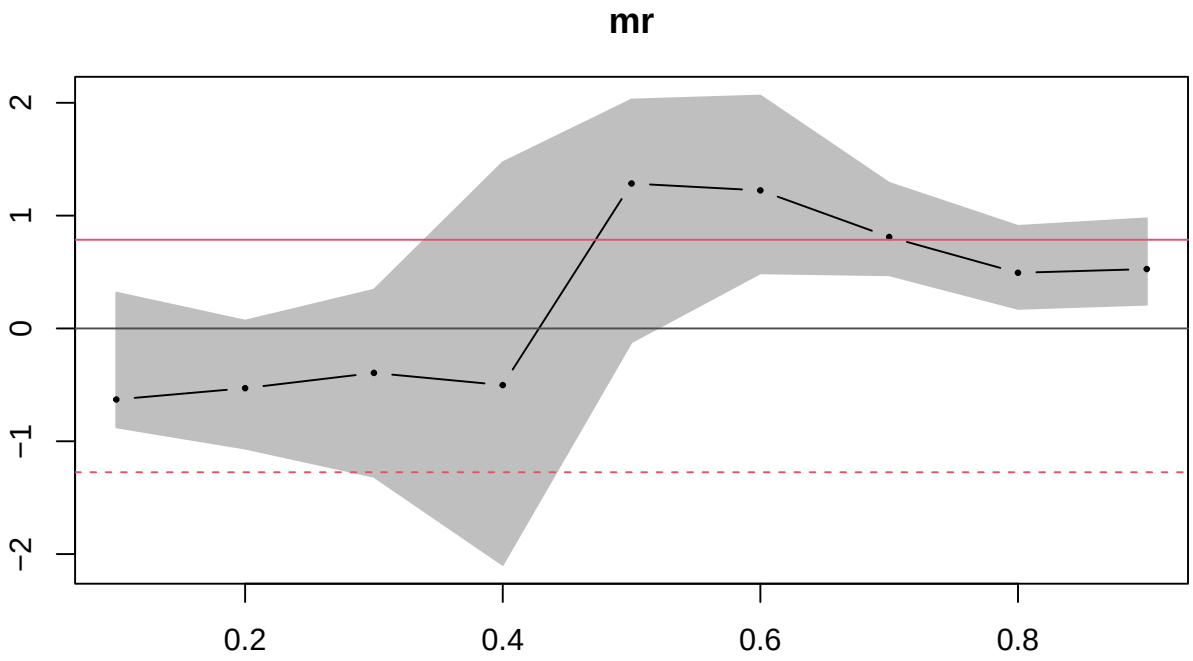


Figure 3.8: Estimates of the coefficient associated with the 'mostly related' variable across profit deciles.

Chapter 4

Discussion

Chapter 3 showed that a variable indicating whether or not a subgroup has mostly related revenues is a statistically significant predictor of the profits booked by the subgroup in question. Further, the effect found was substantial - it increased profits by 43%, everything else equal.

Questions remain as to the reliability of these results and of how they should be interpreted. Can this extra profit be attributed, as is, to profit shifting? Furthermore, this text has brought to attention the heterogeneity within subgroups with mostly related revenues - a significant amount of them is expected to be well represented by the Bahamian subgroup of Shell, while others should resemble Shell Kazakhstan. Figure 3.8 corroborated this intuition - the effect on profits of having mostly related revenues is far from constant across the sample with mostly related revenues. Is it possible to provide profit shifting estimates only for the subgroups that engage in profit shifting?

4.1 Is the estimate for the coefficient biased?

This section leaves aside the heterogeneity of subgroups with mostly related revenues and whether or not these excess profits are due to profit shifting and focuses on the estimate for the coefficient associated with the 'mostly related' binary variable.

The estimate of interest would be biased if there is an omitted variable that is correlated, within the same multinational, with both the 'mostly related' binary variable and with the profits before tax.

An important potential omitted variable would be one that took into account which activities take place in each subgroup. Taking fossil fuels out of the ground generally costs much less than their market value - and hence the additional taxes in production-sharing agreements oil-rich countries tend to engage in. As extraction is likely to be correlated with having mostly related revenues and with pre-tax profits, the estimate may be biased upward. Additional controls for the activities carried out in each subgroup would address this issue satisfactorily - unfortunately, such data is not available at the moment. Whle some multinationals in the dataset are in the extractive industry, this is far from being the case for all of them - the dataset is comprised of firms that operate in multiple sectors.

There are other variables not present in the regression that would drive down the value of the estimate of interest, should they be included. The statutory tax rates [1] to which the subgroups are subject would be one such example. As seen in Chapter 2, it is costumary in the literature to use this variable as a proxy for profit shifting. The purpose of this text is to introduce an alternative identification mechanism - with its strengths and weaknesses relative to other identification mechanisms that have been already studied. It would be of limited interest to add regressors used elsewhere to study profit shifting.

Finally, it is possible to create narratives in which a variable not present in the regression equation is associated both with the share of related revenues and the subgroup's profit - but it is another thing to make such narratives convincing. One such narrative could go as follows: when opening up a branch in a new country, multinationals send over their most competent employees to lay the groundwork and infant subgroups will mostly work for other subgroups. As such, the age of the subgroup would be correlated with the share of related revenues and, via the extraordinary competence of the subgroup's employees, to the profit. However, I find this narrative to be unconvincing, just as I found others I came up with in order to try to

[1] https://www.icij.org/investigations/luxembourg-leaks/leaked-documents-expose-global-companies-secre

challenge the regression specification.

4.2 Are higher profits indicative of profit shifting?

Taking for granted that having mostly related revenues is associated with higher profit, can one attribute that effect to profit shifting?

Considering that profit is defined as the revenues minus the costs, higher profits can be caused by either more revenue, keeping costs constant; or with less cost, for a given level of revenue. As the revenue is 'quantity times price', we are left with three options: either these subgroups find a way of producing more - while keeping capital and labor inputs constant -, they find ways of cutting costs or they overcharge other subgroups. The latter would be a violation of the arm's length principle multinationals are required to adhere to and would, in essence, be profit shifting.

The analysis in Chapter 3 does not enable ruling out the other two hypothesis. However, one thing is certain: if the higher profits were due to them, the subgroups from which the money comes (the origin of the transfer) would not be hurt in any way. Ideally, the country-by-country data would not only show the amount of related revenues, it would also show the amount paid to related subgroups.

Under the arm's length principle, it should be irrelevant whether the inputs of a subgroup come from related or unrelated parties. The existing literature is elucidating on this topic: Bilicka (2019), leveraging confidential tax returns data from the UK tax authority - a gold standard with respect to data quality and richness -, finds that UK subgroups of multinational firms are less profitable than their matched domestic counterparts (and are 30 percentage points more likely to report zero taxable profits). Given that multinational subgroups are the ones with the opportunity to buy from related parts, this result leads us to think that it is not so much the case that subgroups with mostly related revenues are extremely productive but rather that they are over-charging other subgroups in high-tax countries.

4.3 Distinguishing Kazakhstan from the Bahamas

The OLS regression specified is unable to distinguish between subgroups with mostly unrelated revenues and estimates the coefficients that minimize the mean squared error. While this is not ideal, separating these subgroups based on the associated jurisdictions - *i.e.* adding variables to control for tax havens - would incur in grave endogeneity problems.

As such it was deemed preferable to use a quantile regression. After controlling for labor and capital inputs and for total revenues, multinational-level fixed effects and headquarter effects, the coefficient associated with having mostly related revenues shows a jump between two values, from one close to zero (a statistically insignificant negative) to one unambiguously positive and corresponding to increases in profit of at least 60%. This would be consistent with the narrative of the two types of subgroups, one in which there is no effect and one in which the effect is very substantial. Note that these estimates are after controls for the total amount of revenue - how Shell-Bahamas books such revenue figures in the Bahamas is another issue not in the scope of this text.

Chapter 5

Conclusion

I find country-by-country data on multinational's activities to be extremely interesting. The intricacies of the data relative to several subgroups is so intriguing it was difficult to keep in mind the bigger picture.

While I may have been tempted to write this thesis only about Shell - there certainly is enough material for that -, profit shifting is much bigger than Shell. Recall the estimates of about $600B of profit shifted to tax havens - that figure doubles the revenue of Shell (or the GDP of Portugal).

Analyzing this phenomenon from a mathematical perspective allows us to get a better grasp at the systematic issue - and I hope this text contributes in that regard. Mathematical abstraction does not come without a cost: revenues are not all alike, nor are subgroups with mostly related revenues.

The results in Chapter 3 are not without their caveats, and the data on which they are based is not without its faults. Nonetheless, I believe there is value in these results; and that the analysis of these reports can lead to better data collection. In particular, this exercise has lead me to conclude that country-by-country data should reflect both sides of transactions within multinationals across jurisdictions. The addition of this data would better equip tax authorities and researches to identify channels for tax avoidance.

Bibliography

Adebayo, E., A. A. Lashitew, and E. Werker (2021). Is conventional wisdom about resource taxation correct? mining evidence from transparency reporting. *World Development 146*, 105597.

Álvarez-Martínez, M. T., S. Barrios, D. d'Andria, M. Gesualdo, G. Nicodème, and J. Pycroft (2022). How large is the corporate tax base erosion and profit shifting? a general equilibrium approach. *Economic Systems Research 34*(2), 167–198.

Baumann, M., T. Böhm, B. Knoll, and N. Riedel (2020). Corporate taxes, patent shifting, and anti-avoidance rules: empirical evidence. *Public Finance Review 48*(4), 467–504.

Beer, S., R. De Mooij, and L. Liu (2020). International corporate tax avoidance: A review of the channels, magnitudes, and blind spots. *Journal of economic surveys 34*(3), 660–688.

Bilicka, K. A. (2019). Comparing uk tax returns of foreign multinationals to matched domestic firms. *American Economic Review 109*(8), 2921–53.

Bratta, B., V. Santomartino, P. Acciari, et al. (2021). Assessing profit shifting using country-by-country reports: a non-linear response to tax rate differentials. *Dipartimento delle Finanze WP* (11).

Clausing, K. A. (2016). The effect of profit shifting on the corporate tax base in the united states and beyond. *National Tax Journal 69*(4), 905–934.

Clausing, K. A. (2020). Profit shifting before and after the tax cuts and jobs act. *National Tax Journal 73*(4), 1233–1266.

Cobham, A. and P. Janskỳ (2018). Global distribution of revenue loss from corporate tax avoidance: re-estimation and country results. *Journal of International Development 30*(2), 206–232.

Cristea, A. D. and D. X. Nguyen (2016). Transfer pricing by multinational firms: New evidence from foreign firm ownerships. *American Economic Journal: Economic Policy 8*(3), 170–202.

Crivelli, E., R. De Mooij, and M. Keen (2016). Base erosion, profit shifting and developing countries. *FinanzArchiv/Public Finance Analysis*, 268–301.

Dharmapala, D. (2014). What do we know about base erosion and profit shifting? a review of the empirical literature. *Fiscal Studies 35*(4), 421–448.

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

Dowd, T., P. Landefeld, and A. Moore (2017). Profit shifting of u.s. multinationals. *Journal of Public Economics 148*, 1–13.

Dyreng, S., R. Hills, and K. Markle (2022). Tax deficits and the income shifting of us multinationals. *Available at SSRN*.

Dyreng, S. D. and K. S. Markle (2016). The effect of financial constraints on income shifting by us multinationals. *The Accounting Review 91*(6), 1601–1627.

Fatica, S. and W. D. Gregori (2020). How much profit shifting do european banks do? *Economic Modelling 90*, 536–551.

Fuest, C., F. Hugger, and F. Neumeier (2022). Corporate profit shifting and the role of tax havens: Evidence from german country-by-country reporting data. *Journal of Economic Behavior & Organization 194*, 454–477.

Garcia-Bernardo, J. and P. Janskỳ (2022). Profit shifting of multinational corporations worldwide. *arXiv preprint arXiv:2201.08444*.

Garcia-Bernardo, J., P. Janskỳ, and T. Tørsløv (2021). Multinational corporations and tax havens: evidence from country-by-country reporting. *International Tax and Public Finance 28*(6), 1519–1561.

Godar, S., G. Aliprandi, T. Faccio, and P. Janský (2022). The long way to tax transparency: lessons from the early publishers of country-by-country reports.

GRI. Comparison of gri 207: Tax 2019 and oecd action 13 beps country-by-country report.

GRI (2020). *GRI 207: Tax 2019*.

Grubert, H. and J. Mutti (1991). Taxes, tariffs and transfer pricing in multinational corporate decision making. *The Review of economics and Statistics*, 285–293.

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

Hines Jr, J. R. and E. M. Rice (1994). Fiscal paradise: Foreign tax havens and american business. *The Quarterly Journal of Economics 109*(1), 149–182.

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

Janskỳ, P. and M. Palanskỳ (2019). Estimating the scale of profit shifting and tax revenue losses related to foreign direct investment. *International Tax and Public Finance 26*(5), 1048–1103.

Johansson, Å., Ø. B. Skeie, S. Sorbe, and C. Menon (2017). Tax planning by multinational firms: Firm-level evidence from a cross-country database.

Koenker, R. and K. F. Hallock (2001). Quantile regression. *Journal of economic perspectives 15*(4), 143–156.

OECD (2015). *Transfer Pricing Documentation and Country-by-Country Reporting, Action 13 - 2015 Final Report*.

OECD (2022). Guidance on the implementation of country-by-country reporting: Beps action 13.

Tørsløv, T., L. Wier, and G. Zucman (2022, 07). The Missing Profits of Nations. *The Review of Economic Studies*. rdac049.

Wier, L. and H. Reynolds (2018). Big and" unprofitable": How 10% of multinational firms do 98% of profit shifting. Technical report, WIDER Working Paper.

Appendix A

Note describing the database

Introduction

Profit shifting by multinationals has been receiving increasing attention from governments, researchers and the civil society in general. Recent developments in the field indicate that more scrutiny of multinationals' global activities is warranted: Tørsløv et al. (2022) estimate that, globally, multinationals shift 36% of their foreign profits - about $600 billion - to tax havens.

An in-depth analysis of where these companies create real economic activity and book their profits would provide sound foundations for an open and informed debate on the taxation of MNEs. To date, such undertakings have been hindered by the lack of comprehensive data.

Country-by-Country reporting is an important instrument that can be leveraged for this purpose. Notwithstanding the fact that such reporting has been required for enterprises in the extractive and banking sector for some time, the OECD/G20 Base Erosion and Profit Shifting (BEPS) Project is a breakthrough with regard to data collection, as it widens the scope of country-by-country data disclosure to all multinationals with consolidated group revenue upward of 750 million euros.

The BEPS Project prompts its signatories to collect - and provides a template for enterprises to report - information on the global allocation of multinationals' profit, taxes and proxy variables

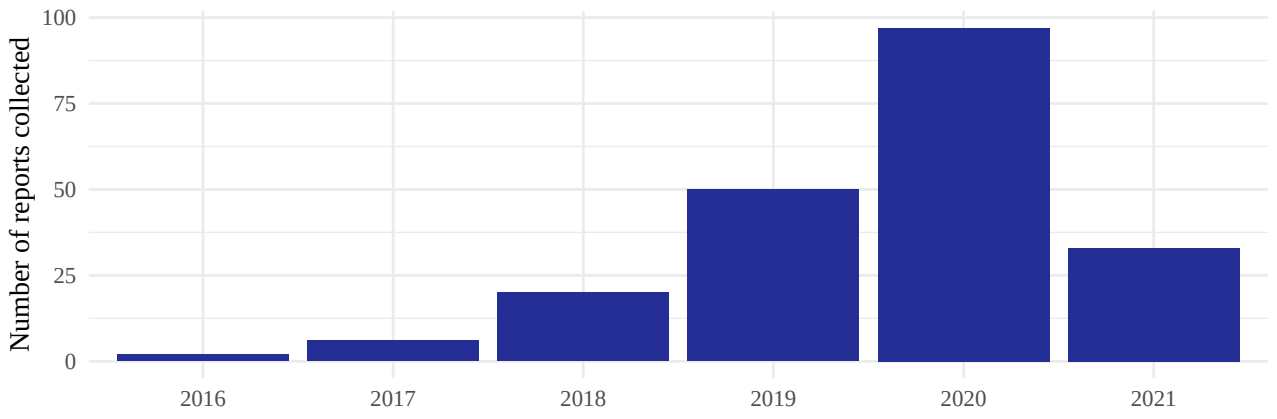


Figure A.1: Country-by-Country Reports collected for the fiscal years between 2016 and 2021

for economic production. These Country-by-Country reports are currently required to be filed with the tax authorities of the jurisdiction in which multinationals are headquartered. The tax authorities have been collecting the reports since 2016 and do not make the reports public.

In recent years, pressure - applied by both the civil society and investors - for tax transparency has been mounting and this is particularly the case for the disclosure of CbC reports. This has resulted in the voluntary diclosure, by some multinationals of their CbC data. The number of companies voluntarily publishing these data has been steadily increasing, as can be seen in Figure A.1. Over 80 reports have been collected relative to the year of 2020, compared with less than 20 for 2018. These reports are not published immediately after the end of the fiscal year: significantly more reports are expected to be published for 2021.

The recent shift towards tax transparency enables the creation of a rich database comprised of voluntarily disclosed reports, thus providing an novel way of studying corporate tax avoidance. The database here described aggregates and standardizes these reports into CSV format to facilitate statistical analysis.

We expect the voluntary release of country-by-country reports to continue rising in the coming years. In 2022, some of Amazon's shareholders presented a proposal in the company's annual meeting for the board to release a tax transparency report[1]. Shortly afterwards, Cisco and Microsoft received similar proposals by their investors[2].

[1]https://s2.q4cdn.com/299287126/files/doc_financials/2022/ar/Amazon-2022-Proxy-Statement.pdf
[2]https://www.forbes.com/sites/taxnotes/2022/06/28/microsoft-and-cisco-face-shareholder-pressure-over-

In addition to being important for the tax authorities, data disclosed in CbC reports serves as input for datasets used by researchers. The OECD - as well as some tax authorities - publish aggregate data obtained the reports filed with the tax authorities. Clausing (2020) and Garcia-Bernardo et al. (2021), for example, use aggregate CbC data published by the US' Internal Revenue Service; Fuest et al. (2022) examine data on the multinationals headquartered in Germany and Bratta et al. (2021) benefits from data obtained from the Italian tax authority. Garcia-Bernardo and Janskỳ (2022) analyse CbC data published by the OECD, which is comprised of multiple sets of data aggregated at the country level, each one gathered by a different tax authority. The main limitation of these CbC datasets its aggregation at the country level.

Up until now, researchers generally have had two alternatives: either use the data published by the tax authorities, aggregated at the country level; or resort to proprietary databases that compile financial figures at the enterprise-level. Exceptions include the study of data from confidential tax returns, which has enourmous value - Bilicka (2019) uses UK data and Dowd et al. (2017) US data - but is very scarce; and research focused on particular industries - namely the banking (Fatica and Gregori (2020)) and the extractive sectors (Adebayo et al. (2021))- for which public, company-level CbC data exists. Given this paradigm, we consider this novel database to add value to the field by allowing the study of disaggregated CbC data from a sample of companies from multiple sectors.

There have been developments at the institutional level that will enhance future data coverage and quality. The EU directive 2021/2101[3] requires country-by-country reports of multinationals with revenues upward of €750 million to be publicly available from the fiscal year of 2025 onward and the Australian government has shown its intention to implement a similar measure to enter into effect 18 months earlier[4].

The rest of the paper proceeds as follows: Section A.1 characterizes the content and the scope of the database; Section A.2 provides descriptive statistics; Finally, Section A.3 describes the workflow that results in the uniform database.

public-disclosures/

[3]https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32021L2101&from=EN

[4]https://www.law360.com/tax-authority/articles/1542790/australia-to-have-first-public-country-by-country-tax-rules

A.1 Content and scope of the database

This section presents the variables collected and the some key characteristics of the set of multinationals for which we have extracted information: from their sectoral composition to their geographical distribution to the frequency and richness of their country-by-country reports.

A.1.1 Data collected

OECD's Action 13 final report provides a template (see Figure A.9, in Appendix A.4.1) multinationals should use to disclose their country-by-country reports to the appropriate tax authorities. To comply with this standard, for each tax jurisdiction in which they operate, multinationals have to disclose data on: related party, unrelated party and total revenues; profits before income taxes; income tax paid and accrued; stated capital; accumulated earnings' number of employees; and tangible assets other than cash and cash equivalents.

In the current context of increased demand for tax transparency, another "standard" format to disclose their fiscal and economic activities has emerged: the Global Reporting Initiative (GRI) standard 207-4. The OECD template and the GRI standard are closely related. Table A.1 provides a detailed comparison of the numerical variables required by each. The set of the numerical variables present in the to OECD's standard encompasses the numerical variables required by the GRI standard. In order to comply with GRI-207-4, it is not necessary for multinationals to publish data on Accumulated Earnings nor Stated Capital.

This database is comprised of publicly accessible data on the variables introduced by the OECD template. The reasons for the choice of mirroring OECD's template are threefold: it includes the mandatory numerical information reports need to disclose in order to conform with the GRI-207 standard; it disregards variables that appear only sporadically in country-by-country reports; and enables comparisons with third sources that publish data originating from country-by-country reports (such as the aggregate data published by the USA's IRS or the OECD). Information found in incomplete reports (with regard to either of the aforementioned standards) will be added to the database as long as it is aligned with the definition of the variables.

Variable	OECD	GRI 207-4
Unrelated Party Revenues	✓	✓
Related Party Revenues	✓	✓
Total Revenues	✓	-
Profit Before Tax	✓	✓
Income Tax Paid (on Cash Basis)	✓	✓
Income Tax Accrued - Current Year	✓	✓
Stated Capital	✓	-
Accumulated Earnings	✓	-
Number of Employees	✓	✓
Tangible Assets Other Than Cash and Equivalents	✓	✓

Table A.1: The OECD template includes all numerical variables present in the GRI-207-4 standard.

Table A.2 provides notes that allow for a rigorous understanding of what each variable captures. There are two other clarifications we address now: the time frame and the region for which data is reported.

Regarding the former, it should be borne in mind that these reports disclose data for a given fiscal year, which more often than not corresponds to a calendar year. Exceptional reports are usually published by multinationals headquartered in either Australia, Japan or the UK. Version 1 of the database associates each observation with a calendar year. Reports in which the fiscal year does not match the calendar year will be associated with calendar years as follows: if the fiscal year ends in the last of June or later of some year, the report will be associated with that same calendar year; otherwise, with the previous year.

Regarding the regions for which data is published, multinationals report data not necessarily on countries but rather on tax jurisdictions. These are defined as either a State or as a non-State jurisdiction which has fiscal autonomy. As such, there is data relative to Hong Kong and Macau, for example. Furthermore, not all multinationals exhaustively list the jurisdictions in which they operate: some of them aggregate figures for continents or sub-continents or bundle together jurisdictions with little activity. Another caveat is that we have not collected BT's data relative to the UK, where they book about 85% of the revenue.

Variable	Notes
(Unrelated, Related and Total) Revenues	OECD: Should include revenues from sales of inventory and properties, services, royalties, interest, premiums and any other amounts. Should exclude payments received from other entities that are treated as dividends in the payor's tax jurisdiction. GRI-207-4: Can report other sources of revenue, for example, dividends, interest, and royalties, where this is standard practice in the sector
Profit Before Tax	OECD: should include all extraordinary income and expense items.
Income Tax Paid (on Cash Basis)	Both: should include cash taxes paid to all tax jurisdictions. If entity A, tax-resident in Y, earns interest in Z, tax withheld in Z should be reported in Y. GRI-207: Taxes paid to other jurisdictions may be reported separately.
Income Tax Accrued - Current Year	Both: should exclude deferred taxes and provisions for uncertain tax liabilities.
Stated Capital	-
Accumulated Earnings	-
Number of Employees	Both: Independent contractors may count. Approximation permissible. OECD: Number of employees reported on a full-time equivalent basis.
Tangible Assets Other Than Cash and Equivalents	-

Table A.2: Notes on the variables collected for the database gathered from OECD (2015) and GRI (2020). Information also summarized in GRI (GRI).

A.1.2 Coverage

Version 1 of the database is comprised of 208 country-by-country reports, from 111 multinational companies, amounting to 4756 jurisdiction-level observations. As seen in Figure A.1, the trend is for the number of reports published to increase. The year of 2020 is the one for which more reports are available and we anticipate the number of reports published regarding 2021 to grow significantly over the coming months.

Geographical distribution and sectoral composition

Country-by-country data has been collected for multinationals headquartered in 24 different jurisdictions. As can be observed in Table A.3, Spain and Italy are the countries with the

Table A.3: Distribution of HQ location across the multinationals in the dataset.

HQ country	Number of multinationals
ITA	23
ESP	18
GBR	15
NLD	9
NOR	6
AUS	5
DEU	5
JPN	5
SWE	3
CAN	3
DNK	3
Other	16

Table A.4: Composition of the sample of multinationals across sectors.

Sector	Number of multinationals
Mining & Extraction	17
Chemicals, Petroleum, Rubber & Plastic	13
Banking, Insurance & Financial Services	12
Business Services	10
Communications	9
Utilities	9
Industrial, Electric & Electronic Machinery	8
Transport, Freight & Storage	5
Metals & Metal Products	5
Other	23

higher number of headquarters of the companies in our dataset.

Regarding the spread of these multinationals across the globe, the majority of firms operate on a relatively small number of countries, the median number of countries for which data is published being 14. Most of the multinationals in the database report their activity in a number of jurisdictions ranging from 7 to 31.5. Shell is the multinational disclosing data in the highest number of countries: 101, all years combined.

Regarding the sectoral composition of multinationals in the database, we note that mining and extraction is the sector with most firms from our dataset. Table A.4 presents other relevant sectors.

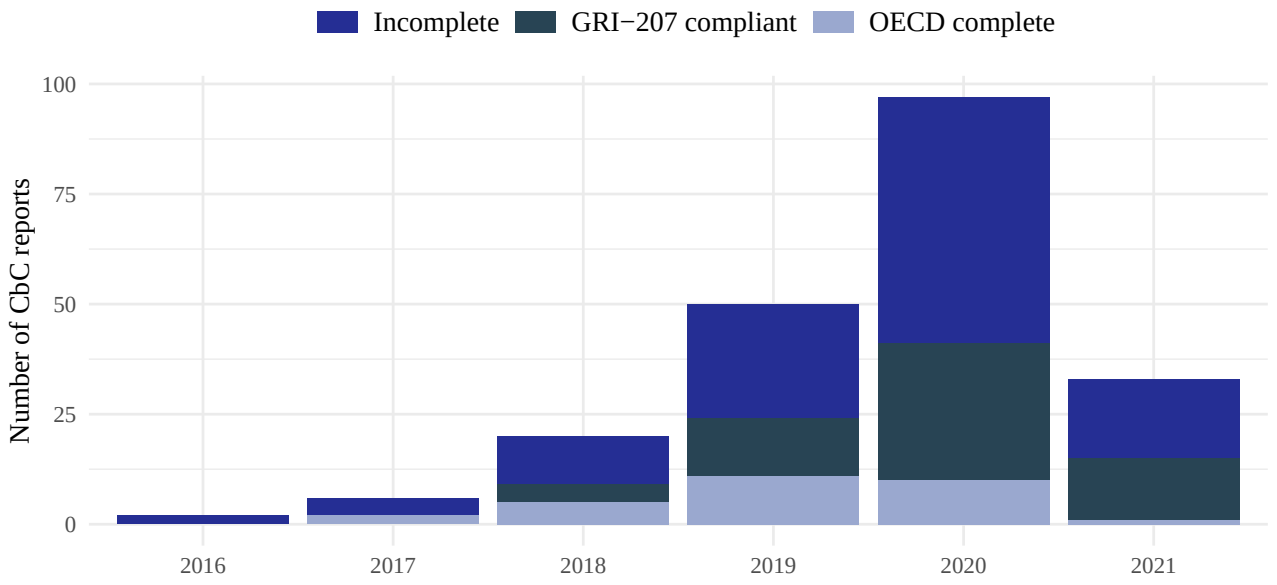


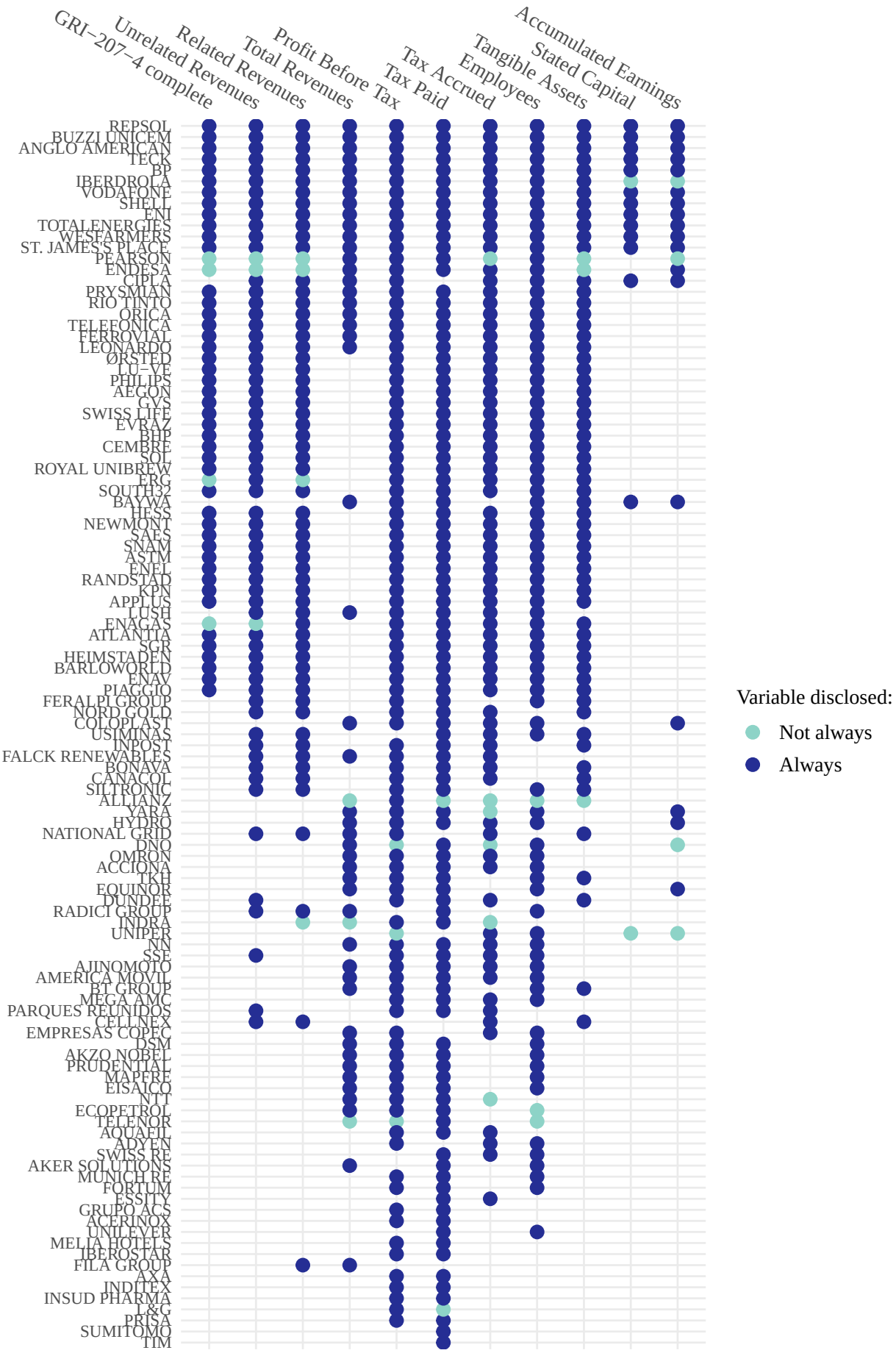
Figure A.2: The publication of reports that comply to the GRI-207-4 standard is surging.

Frequency and richness of the reports

As for the frequency with which the multinationals in our database have published data, 64 have done so at least twice, including 21 for which we have collected 3 reports and 5 for which we have at least 4 reports - see Figure A.10, in appendix A.4.2.

It is rare for multinationals to publish reports with data on all OECD variables, but it is often the case that they are GRI-207-4 compliant. Figure A.2 highlights the fact that, in addition to the observed surge in the total number of reports, the disclosure reports rich in data is on the rise as well. The trend for number of the GRI-207 complete reports published closely follows that of the overall figure - it, too, more than quadrupled from 2018 to 2020. This is not the case for OECD-complete reports; in fact, some recent reports explicitly mention their decision to change the from the OECD to the GRI-207 reporting standard.

Considering that a significant number of the publicly available reports provides information on just a subset of the variables the database aims to collect, Figure A.3 presents, for each variable of interest, which multinationals disclose it - in a manner aligned with the OECD standard. The majority of the reports collected disclose data on the profit before tax; tax paid and accrued; unrelated, related and total revenues; and on number of employees and the magnitude



Notes: Usiminas and FILA report profits in a way not aligned with the DB.
Allianz and others report tax accrued including deferred taxes.
The GRI−207−4 complete column indicates whether the report complies with the standard.

Figure A.3: Variables disclosed by multinational.

of tangible assets (excluding cash and cash equivalents). Data on accumulated earnings and stated capital is less frequently reported.

Considering that the number of multinationals for which we have data is currently limited, we are able to comprehensively list them in appendix A.4.2, together with information on how rich each report is. Figure A.3 illustrates which variables are always disclosed, never disclosed or neither, at the multinational level. Figures A.11 and A.12 indicate whether or not each variable of interest is disclosed in each report.

A.2 Descriptive statistics

This section presents descriptive statistics of the dataset. Table A.5 shows summary statistics for all numerical variables in the database. The column on the right ('N') indicates the number of firms for which data has been collected. In what follows, observations relative to the same multinational and jurisdiction were averaged across the years for which data is available.

From Table A.5, it is possible to form a broad idea about the average multinational: it operates in 23 jurisdictions, employs about 31 thousand people and as annual turnover in the order of 23 billion euros. These average multinational is also significantly larger than the median multinational in the sample, indicating the presence of some extremely large enterprises - some of these companies are, in fact, among the largest in the world. Another important take away from Table A.5 is the extreme variability in size of the multinationals in the sample: the turnover of the largest is over a thousand times larger than that of the smaller.

A.2.1 Additional summary statistics

Multinational size

The figures disclosed by a few multinationals hold sway over the aggregate values of the variables of interest. Figure A.4 shows how the ten biggest multinationals - in terms of unrelated revenue

Table A.5: Summary statistics for the variables collected.

Statistic	N	Mean	Median	Min	Max
Number of subgroups	111	23	14	2	98
Tax Paid	104	581	110	−18	5,798
Profit Before Tax	101	−2,133	509	−316,594	18,365
Employees	86	30,665	15,668	478	189,158
Tax Accrued	81	666	140	−22	5,460
Unrelated Revenues	65	21,661	3,845	137	272,537
Related Revenues	64	11,702	413	0	283,285
Tangible Assets	64	18,604	3,571	69	202,809
Total Revenues	51	48,265	17,047	518	555,823
Accumulated Earnings	22	8,484	4,822	−330,475	182,877
Stated Capital	15	275,841	36,881	556	2,486,316

Notes: All monetary values are expressed in millions of euros.
The monetary values herein were obtained from averaging observations for any given multinational-jurisdiction pair across the years for which data is available. The number of subgroups states, for a given multinational and across the years for which reports have been published, the maximum amount of jurisdictions for which data has been disclosed.

- account for about 75% of aggregate sum. Even within the top-10, three companies (Shell, BP and Total) clearly stand out.

The central part of Figure A.4 presents the shares of pre-tax profits. Despite being disclosed by almost every MNE in the dataset, the profits of the top-10 account for about half of the total. Some of the companies that book the most profit (Allianz, America Movil and NTT) do not report values for unrelated revenues and have thus been excluded from the left side of Figure A.4. The concentration of tax paid per MNE (Figure A.4, on the right) is somewhat similar to that of the profits booked.

It would be a mistake to assume that, apart from the top-10 largest multinationals, the database consists of rather homogeneous enterprises. There is a cluster of firms that book external revenue in the lower part of the order of magnitude of the tens of billions euros and another, more sizable cluster booking revenues in the neighborhood of one billion euros. Figure A.5 presents the distribution of the unrelated revenues across the dataset.

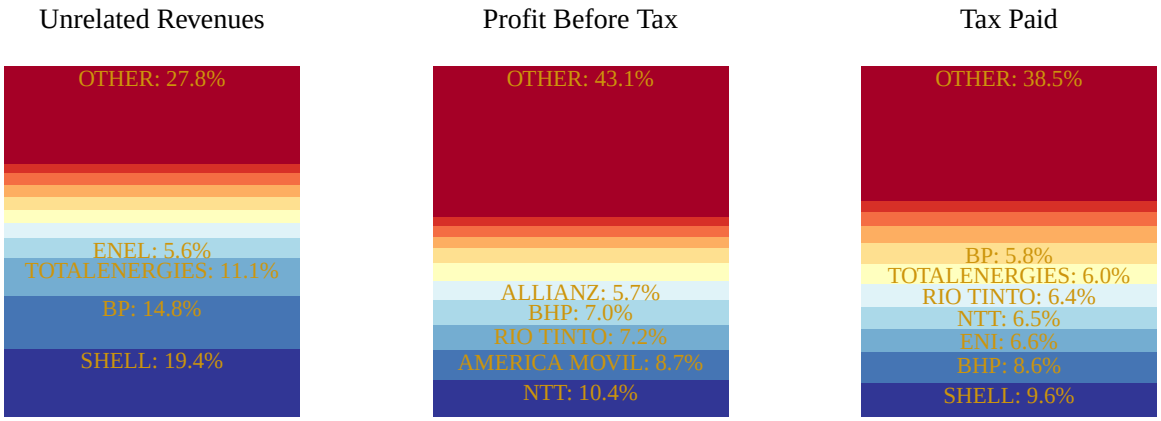


Figure A.4: Top-10's share of external revenues (left), profit before tax (center) and tax paid (right) relative to the whole.

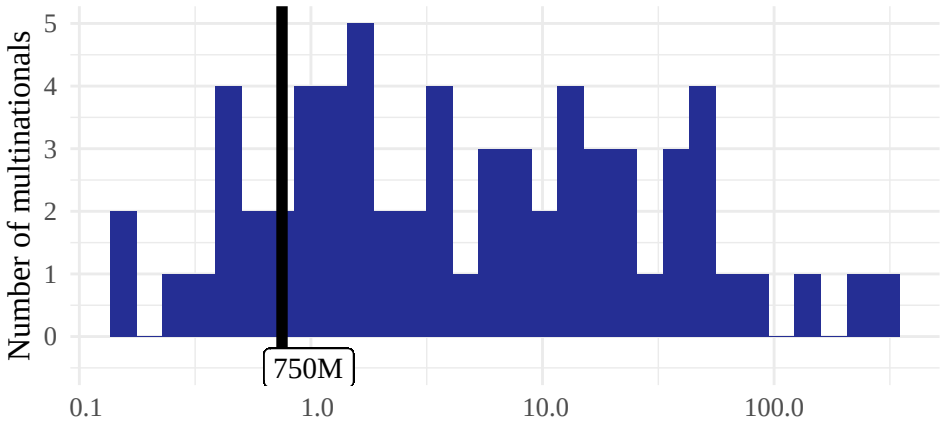


Figure A.5: Histogram of unrelated revenues in billions of euros (after logarithmic transformation).

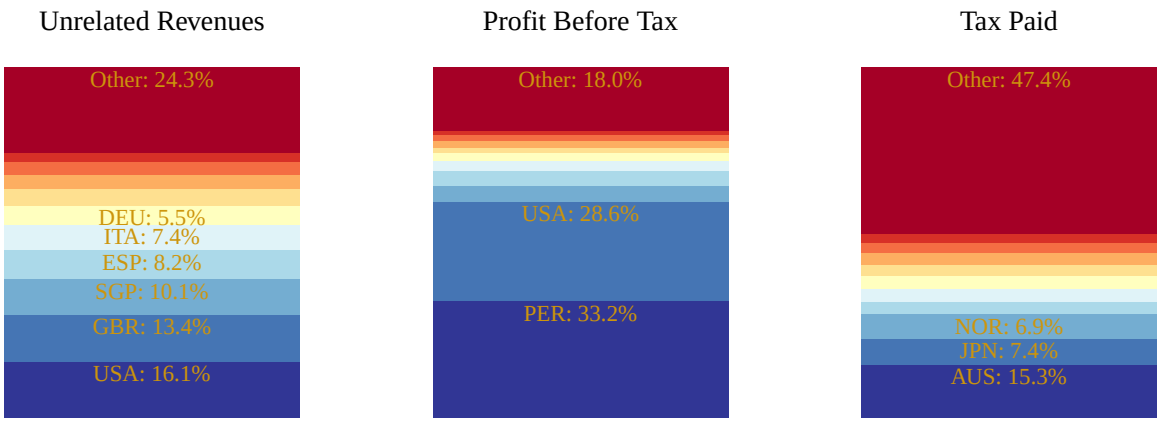


Figure A.6: Shares of external revenues (left), profit before tax (center) and income tax paid (right) across jurisdictions.

Revenues, profits and taxes across jurisdictions

Not only is the third-party revenue highly concentrated in a few multinationals, it is also highly concentrated with respect to geographical distribution. While revenue data has been collected for over 150 jurisdictions, over half the unrelated revenue is booked on just five jurisdictions - see Figure A.6, on the left.

Neither the profits nor taxes paid mirror the distribution of revenues across the globe, as can be seen in Figure A.6. It should be kept in mind that three parts of Figure A.6 aggregate values over the data available: because different multinationals publish data on different sets of variables, each part of the image concerns a different set of firms.

A.3 Information extraction

This Section details the information extraction pipeline that, starting from repository of country-by-country reports - mostly in PDF -, leads to a uniform database. The process starts by rearranging, if needed, tables into a standard format such that a single table contains all the data relative to one given year. Then, variable and jurisdiction names are made uniform and currency, units and cell-level data harmonized.

A.3.1 Table formatting

Not all reports show one year of country-by-country data per table. It is sometimes the case that a table has information on multiple years. We have also extracted data from reports that spread data across multiple tables in the document - data for profit before tax may be on a given page and data for tax paid somewhere else. Inditex is a prime example of this non-standard reporting: different tables contain information on different variables (one for income tax, another for profits) and each table encompasses reporting for the years of 2018, 2019 and 2020 - see Figure A.7.

A significant minority of the reports we have collected have jurisdictions as columns and the variables as rows (see, the example, Figure A.8). Such reports are transposed automatically by the software we developed.

A.3.2 Standardization of variable names and values

Across the reports collected, there is wide variability of the names used in the PDF to refer to the standard variables of interest to this database. 'Third-party revenue', 'Revenues from unrelated parties', 'Revenues from third party sales', 'Income from sales to third parties', or 'Income from intra group operations with other tax jurisdictions' all have been used to refer to the same variable.

Whenever a new variable name is encountered, the software developed prompts the person updating the database, instructing her to either make the correspondence between that name and a standard one or to state that that column should be disregarded. This intervention is then automatically recorded and the rule is given a scope of application - 'just once', 'for this multinational' or 'global'. For the sake of transparency and reproducibility, a text file has been kept that records all rules that have been applied.

The case is analogous for the jurisdiction names. For example, different reports refer to the United States of America either as 'The US', 'USA' or 'United States', 'United Stated of

MARKETS	2020	2019	2018
AMERICA	**-129**	**359**	**294**
BRAZIL	-30	63	74
CANADA	-9	27	17
UNITED STATES	-48	84	50
MEXICO	-43	146	116
OTHER	1	39	37
ASIA AND ROW	**60**	**657**	**561**
AUSTRALIA	-2	13	11
CHINA	1	375	364
SOUTH KOREA	6	57	24
JAPAN	17	83	56
KAZAKHSTAN	14	22	15
OTHER	24	107	91
SPAIN	**640**	**1,805**	**1,650**
SPAIN	640	1,805	1,650
EUROPE	**388**	**1,720**	**1,417**
GERMANY	1	51	14
BELGIUM	-2	26	83
FRANCE	28	101	139
GREECE	15	41	31
NETHERLANDS	9	328	274
HUNGARY	-2	11	12
ITALY	-48	93	83
POLAND	3	44	38
PORTUGAL	-1	77	63
ROMANIA	31	67	67
RUSSIA	86	229	154
SWITZERLAND	145	307	257
UKRAINE	39	54	28
OTHER	84	291	174
PROFIT BEFORE TAX	**959**	**4,541**	**3,922**
CONSOLIDATION	442	140	506
PROFIT BEFORE TAXES CONSOLIDATION	**1,401**	**4,681**	**4,428**

MARKETS	2020	2019	2018
AMERICAS	**38**	**120**	**93**
BRAZIL	-	14	18
CANADA	6	6	5
UNITED STATES	6	41	20
MEXICO	15	42	38
OTHER	11	17	12
ASIA AND ROW	**62**	**123**	**119**
AUSTRALIA	4	4	3
CHINA	20	56	71
SOUTH KOREA	7	14	6
JAPAN	19	28	21
KAZAKHSTAN	5	5	4
OTHER	7	16	14
SPAIN	**103**	**372**	**360**
SPAIN	103	372	360
EUROPE	**222**	**392**	**383**
GERMANY	-6	12	5
BELGIUM	-1	6	24
FRANCE	15	30	52
GREECE	1	10	12
NETHERLANDS	84	127	101
HUNGARY	1	1	1
ITALY	6	21	25
POLAND	18	14	12
PORTUGAL	5	16	14
ROMANIA	1	9	10
RUSSIA	18	41	34
SWITZERLAND	48	58	60
UKRAINE	7	10	5
OTHER	25	37	28
	425	**1,007**	**955**
CONSOLIDATION	42	116	110
INCOME TAX	**467**	**1,123**	**1,065**

Figure A.7: Inditex spreads data across different tables - each including multiple years.

Economic value generated

2021	Nether-lands	Switzer-land	Rest of Europe	North America	Latin America	China	Rest of Asia	Rest of the world	Total
Net sales by origin									
In € million	2,017	2,370	1,336	1,351	666	1,208	423	98	9,468
In %	21	25	14	14	7	13	4	1	100
Adjusted operating profit									
In € million	270	347	131	130	96	116	71	6	1,167
In %	23	30	11	11	8	10	6	1	100

Figure A.8: The CbC report published by DSM in 2021 has countries and other aggregate geographies as columns and variables as rows.

America'.

On the topic of standardizing jurisdiction names, we note that some reports disclose data on *de facto* states not internationally recognized, such as Somaliland or Kurdistan; others, instead of publishing data on a strict country-by-country basis, aggregate multiple jurisdictions together (*e.g.* 'Ireland and the Netherlands' or 'Hong Kong and Macao' , see, respectively, Cellnex 2020[5] and Lush 2019[6]). Many CbCRs aggregate countries together either as "rest of the world" or "rest of" some continent.

The approach taken was to we convert jurisdictions names to its standardized 3-letter code (ISO 3166-1 alpha-3). When that is not possible, we set the jurisdiction name to 'other' or 'other' followed by additional information - *e.g.* 'other apac' or 'other europe'. We have used dictionaries that make the correspondence between country names and their ISO 3166-1 alpha-3 code and, as is the case for variable names, the software prompted us to decide on a juridisction name for the more difficult cases. Once more, all rules are recorded automatically.

We have collected country-by-country reports in multiple languages (*e.g.* Spanish, Italian), which implies that variable and country names had to be translated. We resorted to dictionaries of country names in the corresponding languages and to the creation of the aforementioned rules to translate the variables. As stated before, all rules are recorded and can be inspected and/or altered.

A.3.3 Units, currency and signs

More than ten different currencies are used by the multinationals in our database to report country-by-country figures. Furthermore, it is often the case that these companies report their financial values not in unitary amounts, but rather in thousands, millions or even billions - for currencies orders of magnitude less valuable than the euro - of whichever currency the report uses.

[5]https://informeanual2020.cellnextelecom.com/2020/assets/documentos/informe-anual-integrado-2020-web-con-informe-eng-def.pdf

[6]https://find-and-update.company-information.service.gov.uk/company/04162033/filing-history/MzI3MjQ3MTg0N2FkaXF6a2N4/document

In order to have a consistent database, the units and the currencies are manually collected for each report. This information is then used by the code we developed to automatically transform all figures into (unitary) euros. The conversion rates used have been computed by averaging the daily rates found in XE.com throughout the year in question.

Another source of inconsistency across the PDFs collected is the sign attributed to some variables. Frequently, multinationals pay (a positive) tax on the profits they have made but record the symmetric value in their report. This behavior is also often observed for the tax accrued in the current year, and less frequently associated with other variables - such as the pre-tax profits. Similarly to the case of units and currencies, we manually collect information regarding which columns of the report should be multiplied by -1 and rely on the code to make sure tax paid to the authorities is always shown as a positive value - and, in general, that the values found in each column align with OECD's guidance on implementation of country-by-country reporting OECD (2022).

The aforementioned problem with signs is due to the fact that multinationals not strictly adhere to neither the template provided by the OECD nor to the GRI-207 standard and sometimes use variations of the variable names, as discussed above. For example, South32 names the 'profit before income tax' column as "(Profit)/loss before tax" - hinting to the fact positive profits are shown as negative values.

A.3.4 Cell-level issues

Finally, there were small within-cell issues that have had to be addressed: ranging from the notation used to denote a negative value ('-5' *vs* '(5)') , to the use of characters to ease readability (3000 is often presented as '3,000', '3.000' or '3 000'), to the removal of sub- and superscripts and of other characters present in the cells. All of these issues were handled by the code without the need of any human intervention.

A.4 Appendices to the note

A.4.1 OECD's template

A. Model template for the Country-by-Country Report

Table 1. **Overview of allocation of income, taxes and business activities by tax jurisdiction**

Name of the MNE group: Fiscal year concerned: Currency used:										
Tax Jurisdiction	Revenues			Profit (Loss) before Income Tax	Income Tax Paid (on Cash Basis)	Income Tax Accrued – Current Year	Stated Capital	Accumulated Earnings	Number of Employees	Tangible Assets other than Cash and Cash Equivalents
	Unrelated Party	Related Party	Total							

Figure A.9: The OECD template for the Country-by-Country Report - directly extracted from OECD (2015).

A.4.2 Variables disclosure per multinational and report

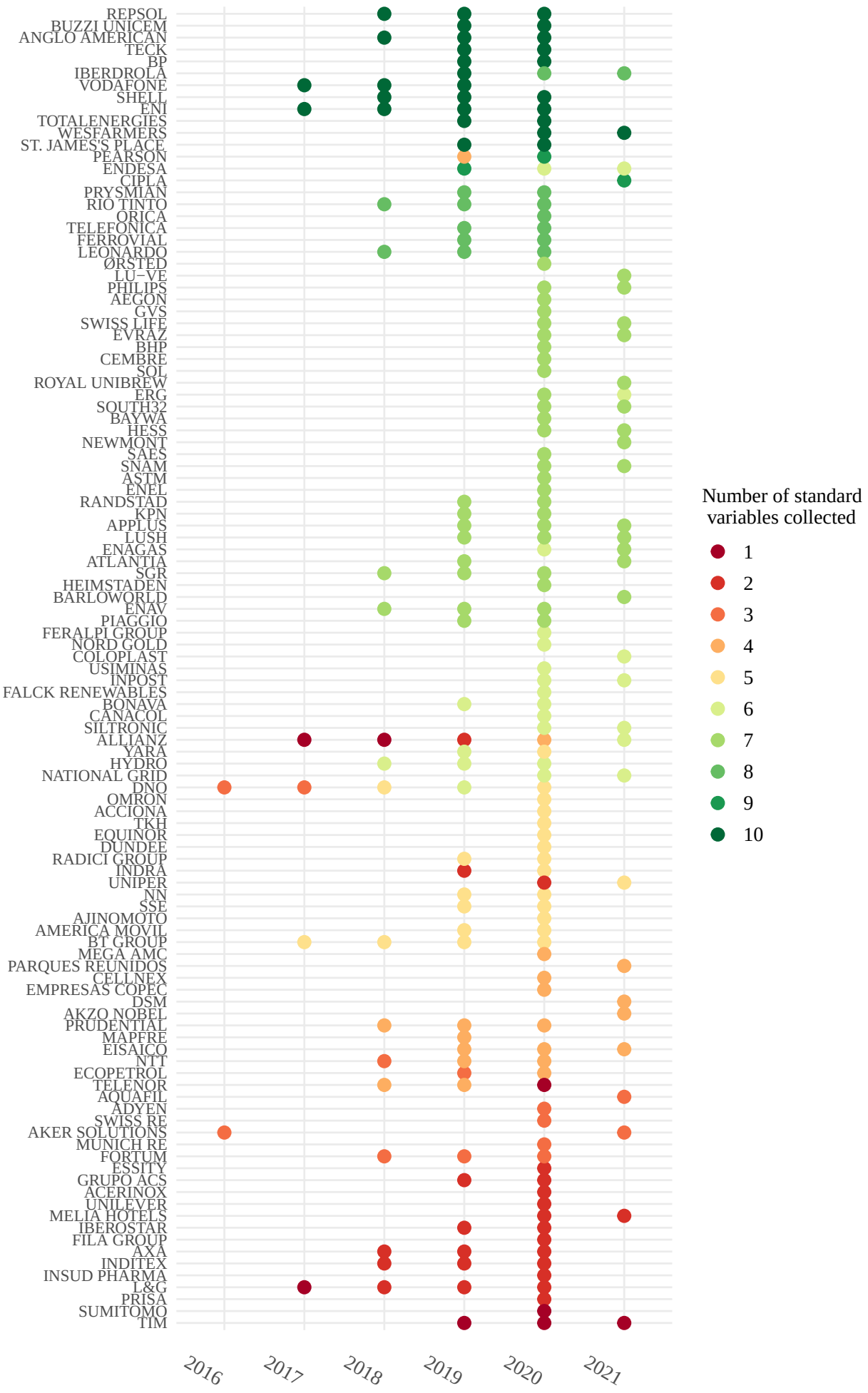
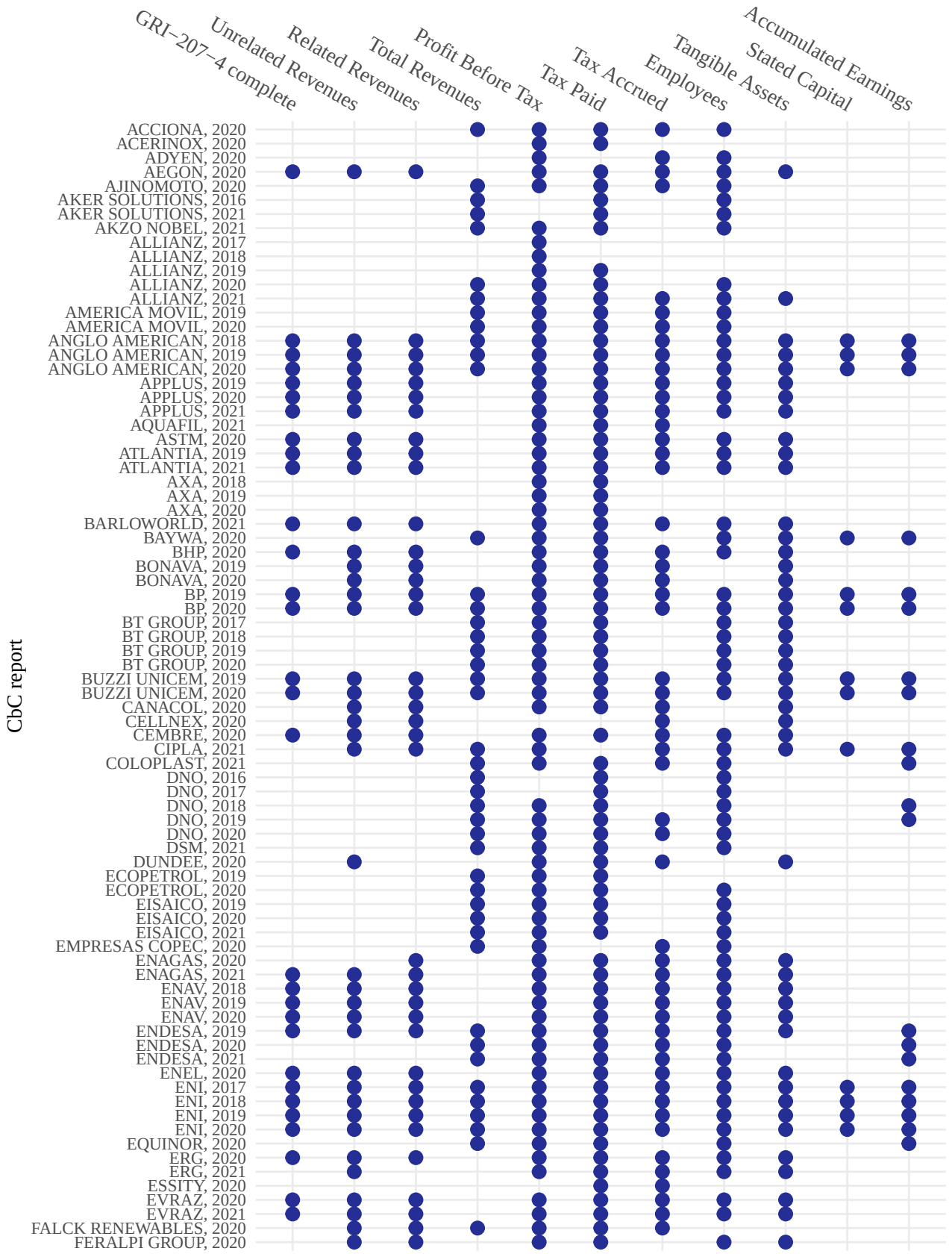
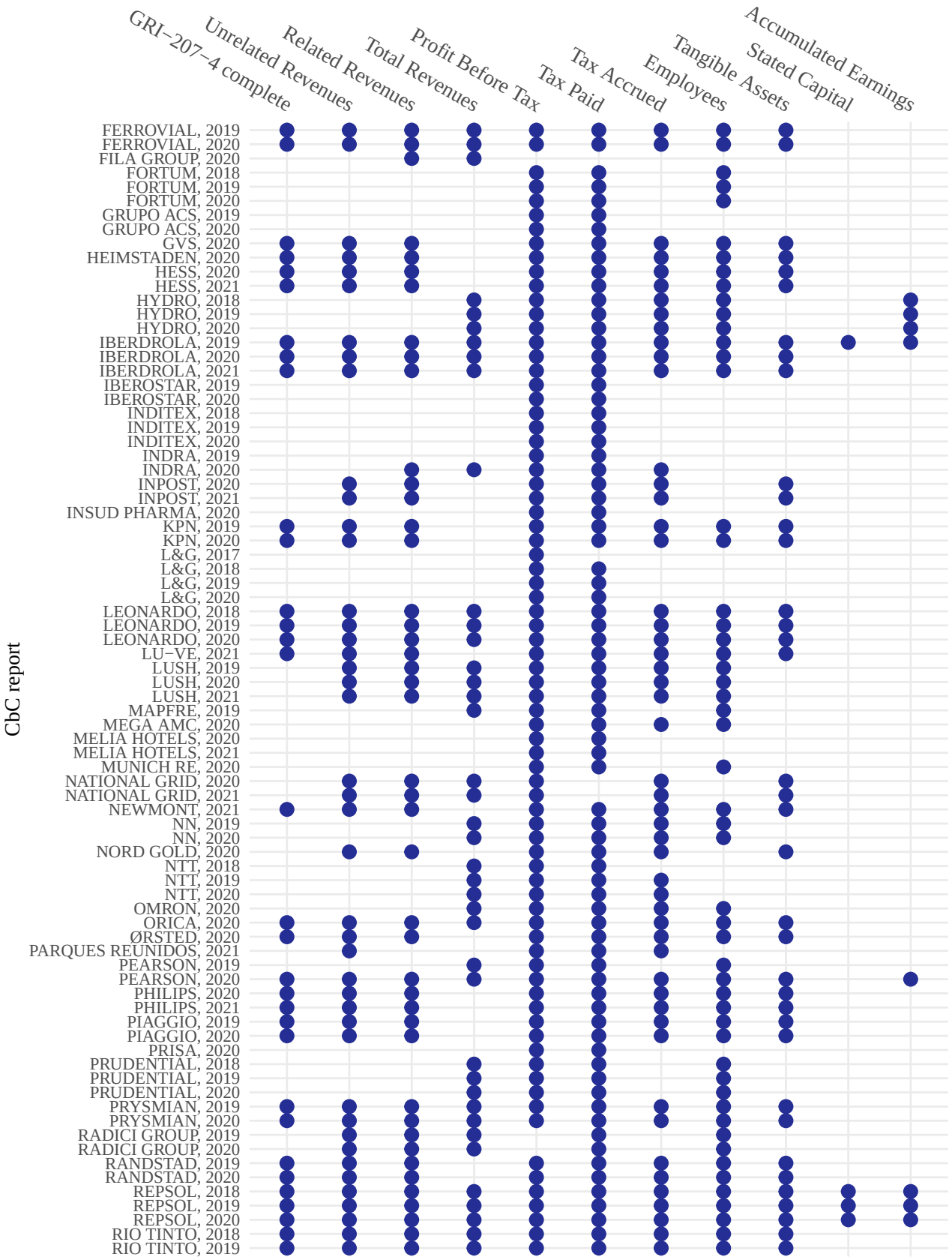


Figure A.10: Variables collected per report - ordered by reporting completeness.



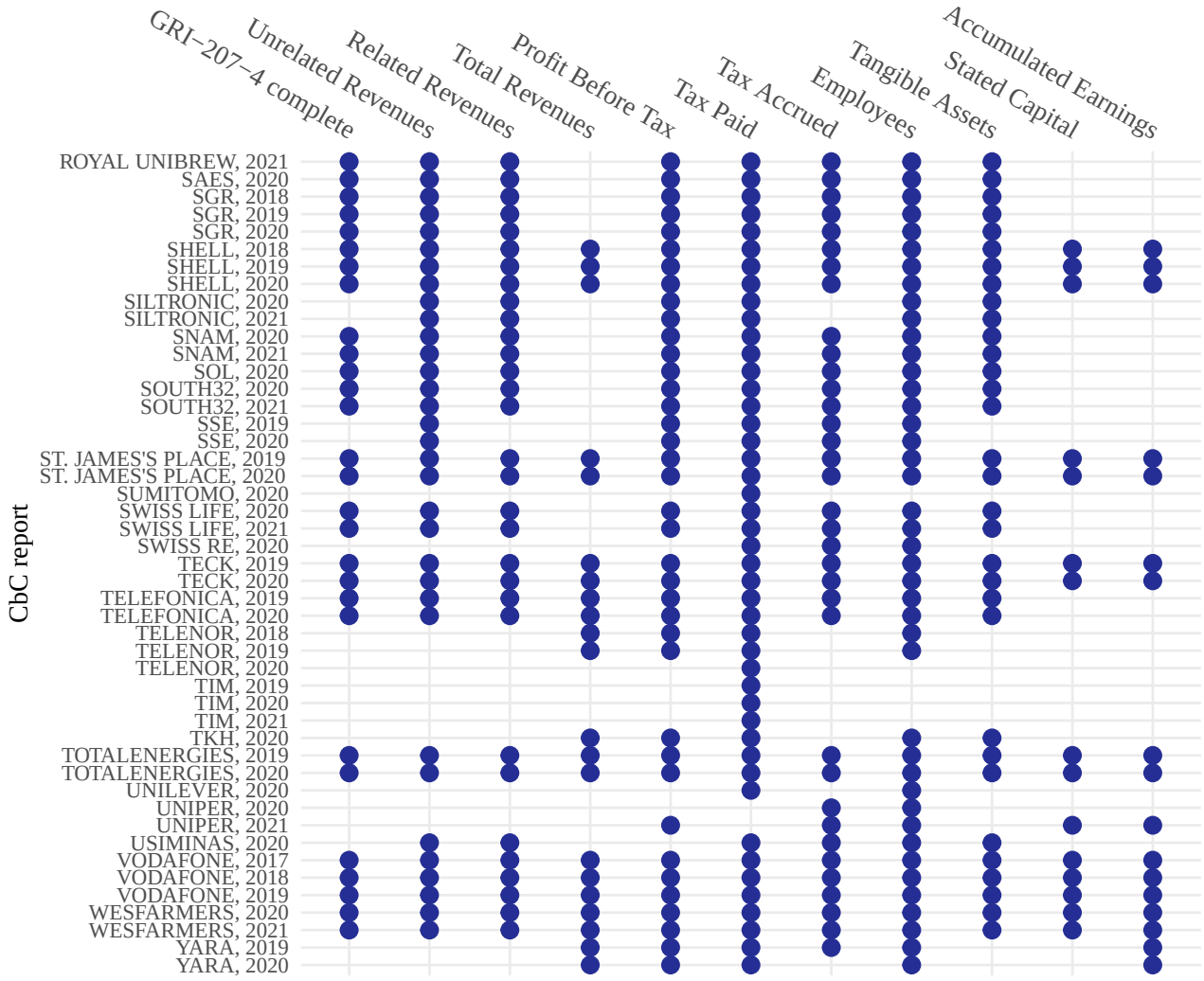
Notes: The GRI−207−4 complete column indicates whether the report complies with the standard, meaning all variables but Stated Capital, Accumulated Earnings and Total Revenues are disclosed.

Figure A.11: Variables collected per report (part 1 of 3).



Notes: The GRI−207−4 complete column indicates whether the report complies with the standard, meaning all variables but Stated Capital, Accumulated Earnings and Total Revenues are disclosed.

Figure A.12: Variables collected per report (part 2 of 3).



Notes: The GRI−207−4 complete column indicates whether the report complies with the standard, meaning all variables but Stated Capital, Accumulated Earnings and Total Revenues are disclosed.

Figure A.13: Variables collected per report (part 3 of 3).

Table A.6: Number of reports, across the years, with data on each variable, part 1 of 2.

Year	Number of Reports	Tax Paid	Profit Before Tax	Employees	Tax Accrued	Unrelated Revenues	Tangible Assets
2016	2	2	0	2	0	0	0
2017	6	4	5	4	2	2	3
2018	20	19	20	15	10	9	10
2019	50	50	48	39	33	28	26
2020	97	91	86	73	69	55	55
2021	33	30	31	27	26	22	21
Total reports	208	196	190	160	140	116	115
Number of MNC	111	104	101	86	81	65	64

Table A.7: Number of reports, across the years, with data on each variable, part 2 of 2.

Year	Related Revenues	Total Revenues	Accumulated Earnings	Stated Capital	GRI-207 complete	OECD complete
2016	0	2	0	0	0	0
2017	2	4	2	2	2	2
2018	9	13	7	5	9	5
2019	27	32	15	11	24	11
2020	56	43	16	11	41	10
2021	20	12	5	3	15	1
Total reports	114	106	45	32	91	29
Number of MNC	64	51	22	15	49	12

Notes: The 'GRI-207' column indicates the number of reports that comply with that standard.
The 'Complete reports' column indicates the number of reports that have data for all variables the database aims to collect.