Turning on the Light: A New Assessment of Measurement Error in International Tax Data*

Rosanne Altshuler
Rutgers University

Lysle Boller
Penn Wharton Budget Model

Juan Carlos Suárez Serrato
Stanford GSB & NBER

September 2024

Abstract

A central issue in international taxation is the extent to which multinational corporations shift profits to low-tax jurisdictions. Policy analysis is obscured by difficulties in quantifying foreign earnings using existing datasets, which can be affected by measurement error. This paper sheds new light on this issue by examining a key source of measurement error in administrative international tax data: aggregation error that leads to double counting of foreign income and distortions in foreign tax rate calculations. We link data from tax filings and public disclosures to construct a firm-level proxy that uses "book-tax" differences to quantify the extent to which commonly-used aggregation techniques may result in double counting. A comparison of book and tax data allows us to proxy for levels of aggregation error in tax data and reveals an increasing trend over time consistent with larger measurement error. We show that applying a simple correction significantly harmonizes measures of foreign income and tax rates across firms' book and tax filings and resolves a systematic relationship between book-tax differences and the size of MNCs' foreign affiliate networks. Finally, we reexamine estimates from prior literature after correcting for aggregation error and find that their conclusions are generally robust to this correction.
Keywords: international taxation, profit shifting, measurement error, double counting
JEL Codes: D22, H25, H26, H32

This research was conducted as part of a Joint Statistical Research Program with the Internal Revenue Service (IRS). Any views expressed are those of the authors and not those of the IRS. The statistics reported in this paper have been reviewed and cleared for disclosure by the IRS.

*We are very grateful for insightful comments from Jennifer Blouin, Robert Moffitt, Leslie Robinson, and the Statistics of Income unit of the Internal Revenue Service. Special thanks to Chloe Gagin, Nuria McGrath, Melanie Patrick, Ralph Rector, Michael Weber, and Jason Wenrich for assistance with data access. Clare Doyle provided excellent research assistance. All errors remain our own.

1 Introduction

Tax competition and corporate tax avoidance have become central areas of focus for researchers and policymakers over the past several decades as the global economy has become increasingly integrated. Researchers have explored a variety of topics related to quantifying the effects of tax avoidance (e.g., Hines and Rice, 1994; Clausing, 2016*b*; Bilicka, 2019). Yet it is surprisingly difficult to reliably nail down even basic descriptive information regarding key elements of multinational corporate activity. For example, estimates of US tax revenue lost to profit shifting can differ by an order of magnitude, ranging from $10 billion to over $111 billion (Clausing, 2016*a*; Blouin and Robinson, 2023). Globally, researchers place estimates of profits shifted to low-tax jurisdictions between $616 billion and $1,076 billion (Clausing, 2016*a*; Tørsløv, Wier and Zucman, 2018). Recent surveys such as Dyreng and Hanlon (2021) underscore the breadth of this literature, emphasizing that overall profit shifting magnitudes remain largely unknown.

These discrepancies may be, in part, caused by measurement error that can arise when aggregating and allocating foreign earnings in commonly-used datasets of multinational activity. Altshuler and Grubert (2006) first highlighted this issue, and recent work by Blouin and Robinson (2023) has underscored the extent of measurement error in international tax data. While there are many possible sources of measurement error, this paper focuses on the specific problem of aggregation error, which can occur when the same earnings appear on the books of multiple affiliates within a multinational corporation (MNC). For example, if a foreign affiliate reports dividend income from a related foreign affiliate, summing the earnings of all affiliates within an MNC will introduce double counting because the same earnings appear on the books of both affiliates. Aggregation error can thus result in upwardly-biased estimates of foreign earnings, which in turn lead to biased estimates of profit shifting and foreign tax rates.

As MNCs have become increasingly complex, with potentially hundreds of affiliates distributed globally, this form of measurement error has likely increased. The growth in measurement error over time undercuts the reliability of analyses that rely on temporal variation to measure changes in corporate activity. This paper takes a closer look at the potential for aggregation error when using data from tax filings by foreign affiliates of US MNCs. By comparing aggregated income data from tax filings against consolidated foreign income data from public filings, we show that adjustments that correct for aggregation error in tax data appear to be quite

promising. Our research suggests that a simple correction can yield more reliable information on average foreign tax rates and income earned by foreign affiliates of US MNCs, providing higher quality information for the increasing number of researchers using IRS tax data to help evaluate and improve the US system for taxing international income.

Tax data have a number of advantages over data compiled from firms' public filings (e.g., in Compustat) or survey data from the Bureau of Economic Analysis (BEA), as we discuss in Section 2. First, Compustat data and BEA surveys typically focus only on large US MNCs. Tax data, on the other hand, are comprehensive, and samples compiled by the IRS contain information about smaller US firms that are active abroad. Second, IRS data contain information regarding the entity types that US corporations elect for the tax treatment of their foreign affiliates, which in turn reveals information about tax planning structures that one cannot observe in other data sources. Finally, measurement of corporate income tax can be more reliable in IRS data when compared with data from corporations' public filings. This is particularly true in certain industries, such as oil & gas, where firms have discretion over whether they categorize payments to foreign governments as income taxes on their books.

The insights gleaned from tax data, however, may be severely undercut by the inability of researchers to correctly measure foreign income and other outcomes. In Section 3, we examine aggregate differences in foreign earnings and reported tax payments by matching a sample of publicly-traded firms to their IRS tax filings. We then construct a novel, firm-level proxy for tax data aggregation error by examining differences between aggregated tax filings and consolidated financial statements that should not be subject to such error. This proxy allows us to examine the performance of a simple correction for aggregation error.

Our paper builds on the work of Blouin and Robinson (2023, henceforth B&R), who focus on the case of BEA multinational survey data. B&R show that failing to account for equity income between related affiliates in this data can lead to measurement errors, including double counting of income. B&R show that, because of the BEA's uniform standards on how firms report income, foreign earnings in the data can be aggregated without introducing measurement errors using a simple correction. In particular, they introduce a concept they call "Adjusted Pretax Income" that removes income from equity investments when computing aggregate foreign earnings. Our paper, in contrast, focuses on the use of firm-level data from IRS tax filings. While BEA survey data are designed to measure where economic activity takes place, IRS data instead measure

where income is reported. Given both this difference in measurement concept, as well as the fact that IRS reporting requirements for foreign affiliates are less standardized than the BEA's, it is an open question whether a similar correction may be effective at mitigating aggregation error in IRS data. Though B&R examine this question using aggregated IRS data, they do not have access to firm-level tax data. Our paper further illuminates this important measurement issue by using firm-level data.

Encouragingly, we find that this correction significantly harmonizes measures of foreign income and tax rates across firms' book and tax filings. Adjusting for aggregation error leads to a 30% reduction in the magnitude of foreign earnings as measured in tax data in 2016. Moreover, the correction resolves the systematic relationship we observe between book-tax differences and the size of an MNC's foreign affiliate network, which we ascribe to aggregation error. The fact that we find no evidence of this relationship after correction, nor evidence of changes in post-correction residual book-tax differences over time, indicates that this adjustment can greatly help researchers take advantage of the benefits of tax data by mitigating concerns related to aggregation error.

With the knowledge that this correction appears to remove systematic aggregation error in tax data, we turn to reexamining results from prior literature in Section 4. Specifically, we provide estimates of the tax semi-elasticity of foreign earnings, i.e., how the allocation of foreign earnings by US MNCs across countries correlates with average and statutory foreign tax rates. We first reproduce these semi-elasticities using the methodology of Dowd, Landefeld and Moore (2017, henceforth DLM), and then replicate them after correcting for aggregation error. Our correction-adjusted estimates yield elasticities with similar magnitudes. Furthermore, we find that DLM's general conclusion, that foreign earnings are more sensitive to tax rate changes in haven jurisdictions, is robust to aggregation error. We discuss this result in contrast to B&R, who perform a similar exercise and find that MNCs in correctly-adjusted BEA data do not demonstrate greater sensitivity to tax rate changes in haven jurisdictions.

This paper proceeds as follows. Section 2 discusses the advantages of tax data relative to other common data sources that are used to measure the activity of US MNCs. Section 3 describes how we construct a matched sample of MNCs that appear in both tax data as well as public filings compiled by Compustat. We use this sample to examine differences in how foreign earnings and income taxes are reported, to measure the extent of aggregation error that may

be generated by naive aggregation of foreign affiliates' tax data, and to assess the performance of a simple correction that aims to remove aggregation error from tax data. Section 4 presents our replications of tax semi-elasticity estimates from earlier studies. We also provide corrected estimates of average foreign effective tax rates by jurisdiction. Section 5 concludes.

2 Data Sources for Measuring Activity of US Multinationals

This section provides an overview of several datasets that researchers have used to study US MNC activity. We review the main issues encountered in these data sources. We then discuss the extent to which they may be affected by measurement issues. Table 1 briefly summarizes the main features and drawbacks of each data source.

Table 1: Data Sources for US MNC Activity

	Features	Drawbacks
Public Financial Statements	• Publicly available • Cleaned data available from providers such as Compustat	• No private firms • Limited foreign data
BEA DII	• No aggregation error	• Indirect ownership chains not observed • No foreign corp. tax data • No smaller firms
BEA AMNE	• Indirect ownership chains observed • Disaggregated info. on foreign affiliates • Foreign corp. tax data	• Possible aggregation error • No smaller firms • Does not specify foreign affiliate type (CFC vs. FDE)
SOI Tax Data	• Includes private firms • Indirect ownership chains observed • Disaggregated info. on foreign affiliates • Specifies foreign affiliate type (CFC vs. FDE)	• Possible aggregation error • No smaller foreign affiliates pre-2004

Notes: BEA refers to the Bureau of Economic Analysis. DII refers to the Direct Investment Income series. AMNE refers to the Activities of US Multinational Enterprises series. SOI refers to the IRS Statistics of Income division. Financial statement data is often accessed by researchers through proprietary databases such as Compustat. While this table presents all four datasets for comparison, our analyses only use financial statement data from Compustat and SOI data.

2.1 Public Filings

Publicly-traded firms in the United States are required to file financial statements with the Securities and Exchange Commission (SEC), which are then made available to the public. Throughout this paper, we refer to such filings as "book" data. Various services, such as Compustat, compile this information and provide it to researchers. Although public filings are a useful source of information for a large set of firms, they are not comprehensive—private firms, for example, are not covered by the database. Furthermore, filings are typically provided in a consolidated form, which heavily limits the availability of disaggregated information about MNCs' foreign operations. Some firms do, however, provide limited information about foreign income and foreign tax expenses.

2.2 BEA Multinational Data

The BEA provides two different sources of data on US MNCs. The first, "Direct Investment Abroad," forms the basis for the US's balance of payments data. The second, "Activities of US Multinational Enterprises," provides more disaggregated detail for a set of large US MNCs. Though our paper does not use BEA data, these datasets are very common in the literature (e.g., Hines and Rice, 1994; Blouin, Krull and Robinson, 2012; Clausing, 2016*a*; Albertus, 2023).

Direct Investment Abroad

The BEA conducts mandatory, confidential quarterly surveys that collect information on investment positions and transactions with directly-owned foreign affiliates of US investors.[1] Research that examines foreign earnings of US firms has used the "Direct Investment Income" (DII) measure contained in this data (e.g., Zucman, 2014, 2015). One advantage of DII data is that researchers do not introduce aggregation error related to double counting when using the data to examine foreign earnings. On the other hand, on a country-by-country basis, DII may not reflect the true geographic source of earnings. This is due to the fact that income generated by indirectly-held foreign affiliates will only be observed at the level of a directly-owned affiliate.

[1]The BEA DII data do not include indirectly-owned, or "pass-through," foreign affiliates. See Section 2.3 below for more detail on foreign affiliate classifications.

Activities of US Multinational Enterprises

The BEA also conducts annual surveys that collect detailed information on the activity of foreign affiliates directly or indirectly owned by US MNCs. This data is compiled into annual reports titled "Activities of US Multinationals" (AMNE). Unlike the DII data, AMNE data also provides measures of corporate income tax, which allows researchers to estimate taxes paid and average tax rates by jurisdiction. The AMNE data are the primary source used to measure foreign earnings in B&R.

As with IRS data, aggregate foreign earnings reported by the BEA may double count income that is included on the income statement of multiple affiliates. One key advantage of these data is that the BEA requires reporting entities to adhere to a strict accounting standard. B&R demonstrate that this reporting standard allows them to systematically remove aggregation error by adjusting aggregate foreign income by removing income from equity investments, which are separately reported by the BEA. They compare the BEA's stricter reporting requirements with IRS reporting standards for foreign income and suggest that the possibility for discretion in tax accounting implies that it may not be possible to fully correct for such measurement error in tax data.

2.3 Administrative Tax Data

The Statistics of Income division of the IRS produces two sets of corporate data that have been used by researchers to study US MNCs. These datasets provide data disclosed in tax returns for a large sample of US corporations, which includes information about the activities of both US parent companies as well as their foreign affiliates. The sample includes both public and private corporations, as well as relatively small firms that are not captured in BEA surveys. We refer to these data as "tax" data.

SOI Corporate Sample The first dataset, the IRS Statistics of Income (SOI) Corporate Sample, is an annual stratified sample of US corporations that SOI uses to produce publicly available aggregated business income statistics.[2] The SOI Corporate Sample provides information from unaudited tax returns for approximately 100,000 US corporations annually, and has been used in the business tax literature to study domestic firms (Yagan, 2015; Zwick and Mahon,

[2]Statistics are available at IRS (2022*b*), and the sampling procedure is described in IRS (2011).

2017). Our data focus on C corporations that were sampled between 1992 and 2016. This paper partially relies on Form 1120, the US Corporate Income Tax Return, as well as information provided in Schedule M-3, a Form 1120 attachment that reconciles firms' tax filings with their financial statements, and which is required as part of the tax return for corporations with total assets of at least $10 million (IRS, 2019).

SOI International Business Tax Sample The second dataset, which provides information related to foreign affiliates of US corporations, is used by SOI to publish aggregate statistics for international business taxes (IRS, 2022*a*). There are two types of foreign affiliates from the perspective of the IRS. The first type, which the IRS calls a "Controlled Foreign Corporation" (CFC), is viewed as an incorporated entity for tax purposes. The second type, which the IRS calls a "Foreign Disregarded Entity" (FDE), is treated as a pass-through entity for tax purposes, similar to a partnership. The IRS collects information about CFCs and FDEs via informational returns that are attached to the US parent's tax return (Forms 5471 and 8858, respectively).

Unlike the SOI Corporate Sample, which is constructed annually, information about CFCs is only collected in even years. Information about FDEs is collected less frequently, for four of the years in our sample period (2006, 2008, 2012, and 2016). Prior to 2004, this dataset only included information for foreign affiliates of large MNCs with greater than $500 million in assets. Starting in 2004, the sampling procedure became much broader.

Like the BEA's AMNE data, this SOI dataset includes information about both directly- and indirectly-owned foreign affiliates. One key difference from BEA's AMNE data, however, is that the IRS reveals whether entities are CFCs or FDEs. The entity type of a foreign affiliate can provide important information to researchers, who have used them to identify well-known profit shifting and tax minimization strategies (Altshuler, Boller and Suárez Serrato, 2024).

Foreign earnings in IRS data may be incorrectly aggregated in a manner similar to the BEA's AMNE dataset. This can occur if there is indirect ownership of one foreign entity by another. For example, foreign earnings that are distributed as dividends to the parent affiliate may appear as foreign earnings on each affiliate's informational return.

However, the SOI also provides information that can be used to correct for this type of error in a manner analogous to B&R's adjustment of BEA foreign earnings. In particular, Schedule M of Forms 5471 and 8858 reports transactions between related foreign affiliates. These forms

separately report dividend transactions, and SOI began publishing aggregate related dividends in 2010 as part of their publicly available international business statistics to address concerns around double counting. In the next section, we seek to address whether this information can provide a satisfactory adjustment to guard against aggregation error.[3]

3 Assessing Aggregation Error in Foreign Earnings and Taxes

Aggregation error potentially poses a serious problem for researchers: if SOI data overstate foreign earnings, estimates from research examining foreign earnings and foreign tax rates in these data may be biased. Worse, to the extent this error is growing over time, as one might suspect if MNCs' ownership structures are also becoming more complex, standard economic approaches to control for such error (e.g., unit fixed effects) might only exacerbate the problem (Bound, Brown, Duncan and Rodgers, 1994; Bound and Krueger, 1991).[4]

Below, we use data from Compustat and SOI to measure the extent to which tax data may suffer from aggregation error and examine whether adjusting foreign earnings to remove related dividend income, as suggested by B&R, appears to sufficiently correct for this error. First, we construct a sample of US MNCs that appear both in SOI and Compustat. Compustat reports foreign income and foreign tax figures on a consolidated basis for a set of MNCs.[5] The consolidated nature of Compustat's reporting implies that it should not be contaminated by aggregation error. This provides a baseline comparison that can be used to measure aggregation error in the tax data for a common set of firms, avoiding the possibility that aggregate differences may be due to differences in sample composition.

3.1 Sample Construction

Figure 1 provides an overview of the sample construction process. The top line in the left panel shows the annual sample size for US MNCs contained in the SOI data, which ranges between 7,597 and 9,157. US MNCs are defined as US companies that file Form 1120 as well as at least

[3]B&R perform this adjustment with aggregate country-level tax data, rather than firm-level data. Based on a comparison with BEA data, however, they question whether this adjustment is sufficient to fully correct for aggregation error.

[4]See, e.g., Blouin and Krull (2014) for evidence that multinational ownership structures have increased in complexity over time.

[5]The Compustat sample typically includes large, publicly-traded firms.

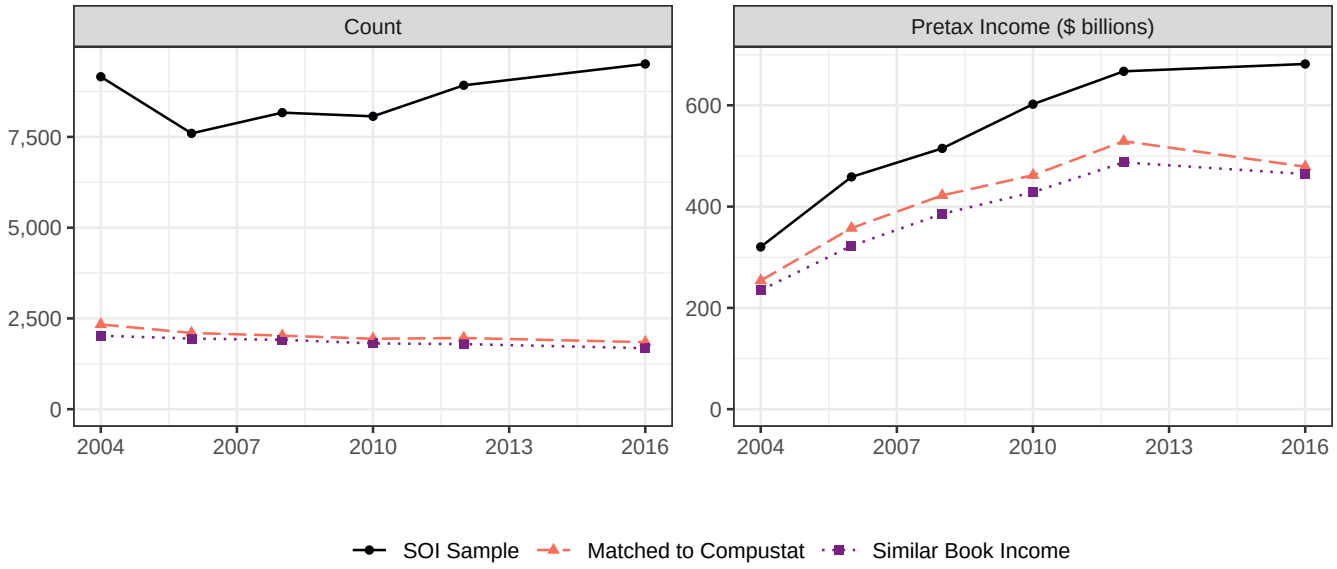
one Form 5471 for a foreign affiliate. SOI includes a firm identifier (EIN) that can be used to link the firm's corresponding data from Compustat. The size of this linked sample, shown in the orange dashed line in the left panel of Figure 1, ranges between 1,845 and 2,334. The fact that the matched sample is smaller than the full SOI sample is not surprising given that Compustat does not include information about privately-held firms.

To ensure that the match is of high quality, we use information from Form 1120, Schedule M-3, which is required for large US MNCs to reconcile book and tax financial information. Firms must report book earnings on their M-3 that in theory should be the same as those disclosed in their public filings. These earnings should be directly comparable to the net income figure reported in Compustat (variable *ni*); therefore, we drop firms from the matched sample that report M-3 net income with a different sign from Compustat net income, as well as firms that report M-3 net income not within 1% of the Compustat net income figure. The size of the final linked sample, which ranges from 1,680 to 2,023, is shown on the purple dotted line of the left panel of Figure 1. The figure demonstrates that only a small share of firms are dropped when using Schedule M-3 to match on net income, which is encouraging and indicative that the comparison is of high quality.[6]

The right panel of Figure 1 compares aggregate pretax income for each step of the sample construction. A much smaller proportion of foreign earnings are dropped relative to the reduction in sample size, confirming that we are dropping relatively small US MNCs.

[6]Figure A.1 provides an overview of the sample construction process starting with the full Compustat sample. An additional filter is applied to remove Compustat observations that report missing pretax foreign income. As shown in the left panel, a much smaller portion of Compustat MNCs are dropped in the match, which is unsurprising given that SOI is the more comprehensive database. In 2016, for example, the matched sample includes 68.8% of Compustat MNCs. In the right panel, the sample is presented in terms of aggregate foreign earnings as reported in Compustat, with the matched sample including 84% of the aggregate foreign earnings reported in Compustat in 2016.

Figure 1: SOI MNCs vs. Matched Sample



Notes: This figure provides an overview of the sample construction process. The "SOI Sample" contains all US MNCs from the SOI data sample. Note that prior to 2004, SOI statistics only included information for foreign affiliates of large MNCs with greater than $500 million in assets. Starting in 2004, the sampling procedure became much broader. The "Matched to Compustat" sample contains SOI Sample MNCs matched to Compustat via unique firm identifier (EIN). The final linked sample, "Similar Book Income," was created by dropping Compustat-matched MNCs whose M-3 net income, as reported on IRS Form 1120, Schedule M-3, has a different sign than or is not within 1% of the firm's Compustat net income. The left panel displays sample sizes in terms of number of MNCs, while the right panel shows the samples in terms of pretax foreign income. These panels showcase Compustat's lack of coverage for private MNCs, but underscore that the match quality of MNCs that are able to be linked is high.

3.2 Comparison of Aggregate Foreign Income and Tax

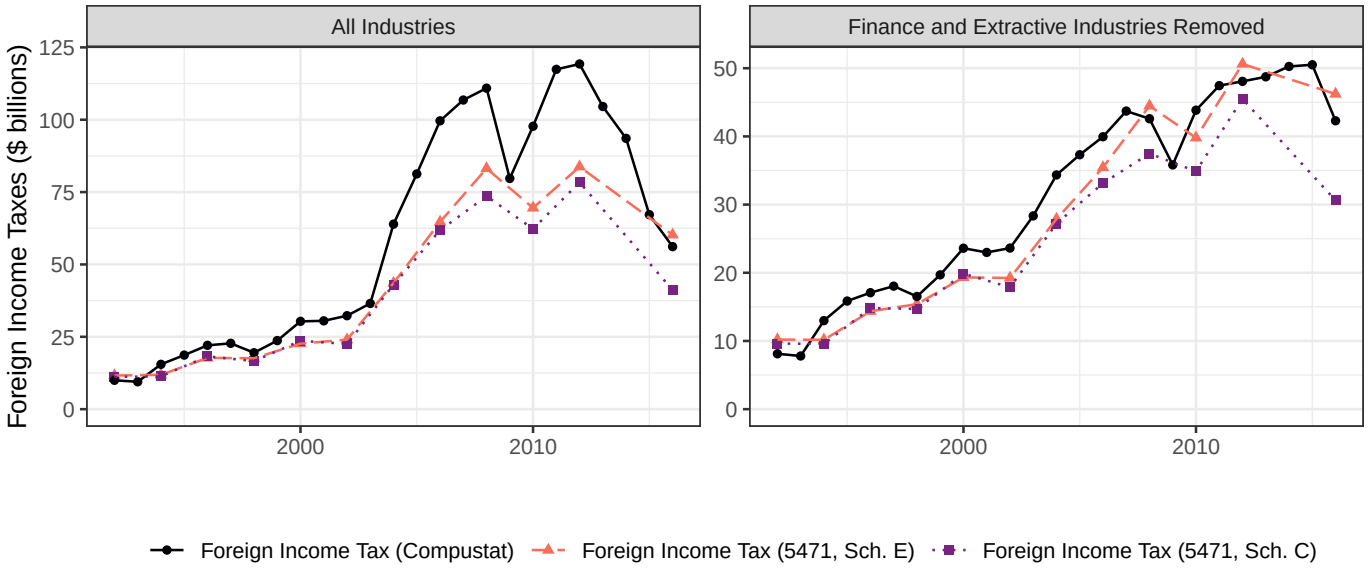
Having constructed a sample of MNCs for which both public filings and tax filings are observed, we now examine aggregate foreign earnings and tax outcomes over time within the sample.

Figure 2 shows three aggregate measures of foreign income tax. The solid line plots aggregate foreign tax from Compustat. The dotted lines provide two different estimates from tax data: the series represented by purple dots are computed from Form 5471, Schedule C (Income Statement), and the series represented by orange dashes are computed from Form 5471, Schedule E (Income, War Profits, and Excess Profits Taxes Paid or Accrued). We use Schedule E as the preferred measure. The left panel contains aggregates computed from MNCs in all industries. The right panel excludes MNCs classified as operating in the Financial, Utilities, Mining, Agriculture, or Oil & Gas sectors, which we refer to as "finance and extractive industries."[7]

Examining the full sample, it appears that Compustat tends to report much larger estimates

[7]This restriction is also implemented in Dowd, Landefeld and Moore (2017).

Figure 2: Comparison of Aggregate Foreign Tax Measures



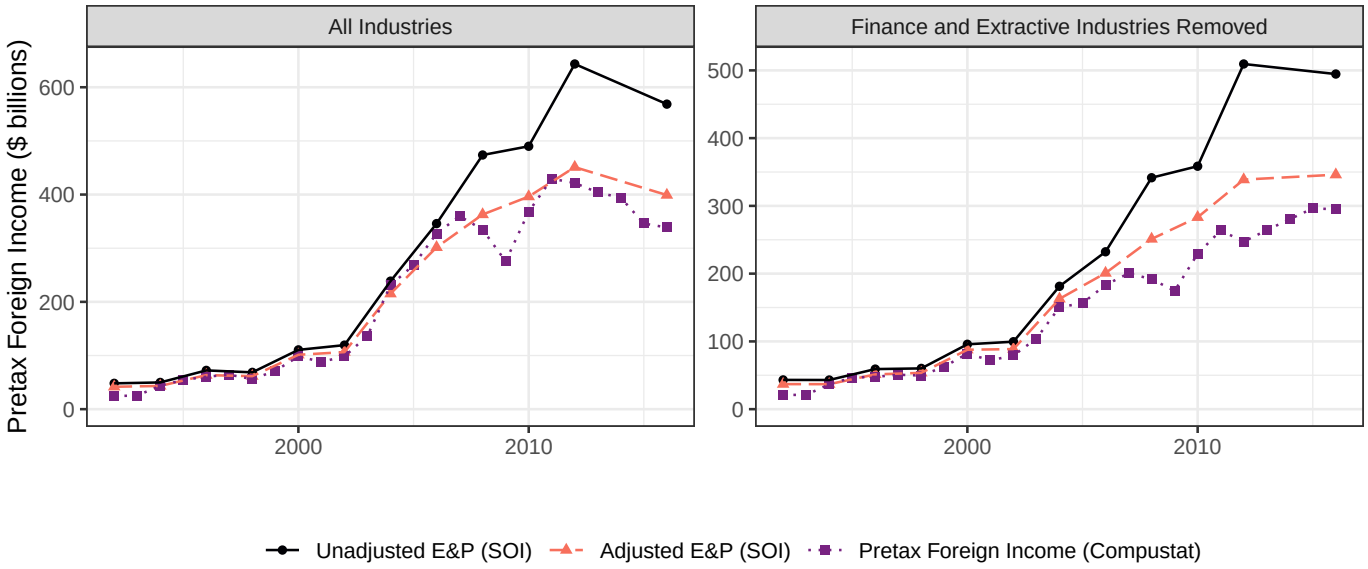
Notes: This figure shows three measures of aggregate foreign income tax as respectively reported by Compustat; IRS Form 5471, Schedule E; and IRS Form 5471, Schedule C. The left panel contains aggregates computed from MNCs in all industries, while the right panel excludes MNCs operating in Financial, Utilities, Mining, Agriculture, or Oil & Gas sectors. The panels show that Compustat tends to report larger estimates of foreign income tax than is contained in tax filings; however, this difference is lessened when excluding the industries mentioned above. We use Schedule E as the preferred measure.

of foreign income tax than what is contained in tax filings. This difference is markedly lower when excluding the industries selected above, as the MNCs in these industries tend to face different tax rules and may employ different foreign operating structures. One key reason for this discrepancy appears to be related to how extractive industries operate foreign concessions. These projects are often structured so that the foreign state receives a share of revenue or profits, and firms have discretion over whether to report these profit-sharing arrangements as income tax in their books. It does not appear that they have the same amount of discretion when they disclose information about foreign taxes in their returns.

Even after removing this set of industries, Compustat tends to report larger estimates of foreign corporate income tax than what firms disclose in their tax returns. This suggests that aggregation error is likely not a significant problem when computing a US MNC's foreign tax bill from its Form 5471 filings.

Figure 3 shows three aggregate measures of foreign pretax earnings for the matched sample. In this figure, the solid line plots represents unadjusted SOI earnings and profits (E&P) computed from Form 5471, Schedule H; the orange dashed line plots this data after adjusting to remove

Figure 3: Comparison of Aggregate Foreign Income (Pretax)



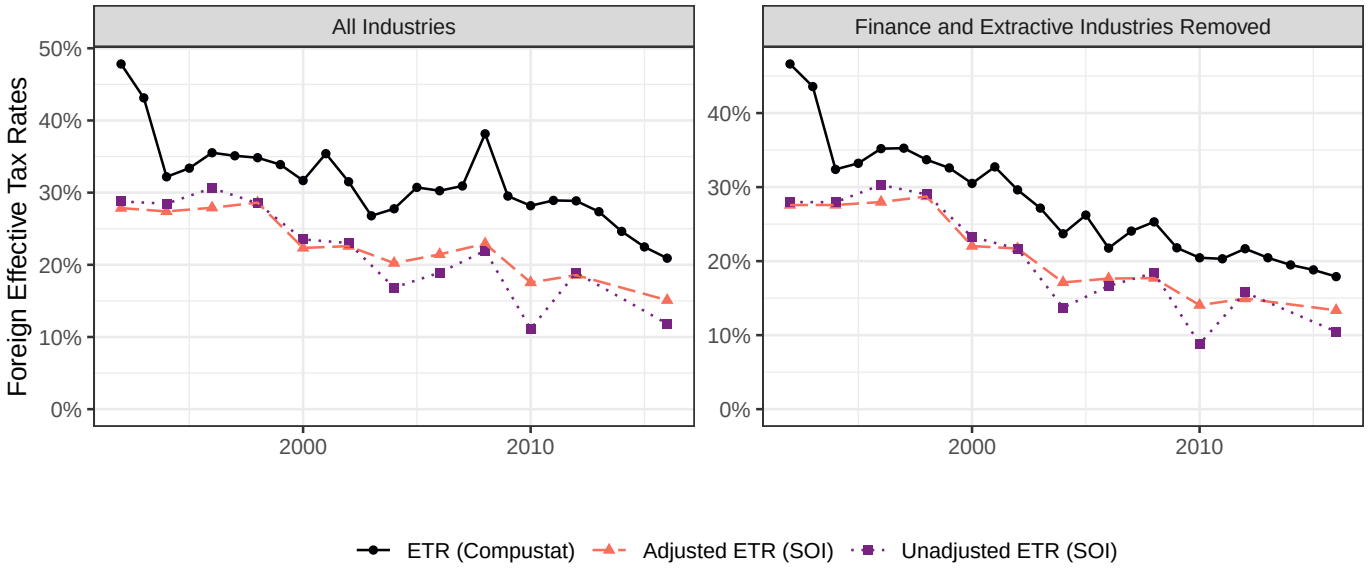
Notes: This figure shows three respective measures of aggregate pretax foreign income: unadjusted E&P computed from IRS Form 5471, Schedule H; Schedule H E&P, adjusted to remove dividends received from related affiliates as reported on Schedule M; and aggregate foreign pretax income as reported by Compustat. Both SOI measures include foreign taxes, as reported in Schedule E, added back to represent pretax earnings. As in Figure 2, the left panel contains aggregates computed from MNCs in all industries and the right panel excludes MNCs operating in Financial, Utilities, Mining, Agriculture, or Oil & Gas. The panels indicate that the E&P adjustment proposed by B&R is at least partially effective.

dividends received from related affiliates, as reported on Schedule M; and the purple dotted line plots aggregate foreign pretax income from Compustat. For the SOI data, we add back foreign taxes as reported on Schedule E so that both series represent pretax earnings. The left and right panels are inclusive and exclusive of the industries described above, respectively.

Notably, unadjusted E&P appears to diverge from the other measures over time. The Compustat and adjusted SOI foreign income measures are more closely aligned, although SOI seems to report a slightly higher figure on average. This provides suggestive evidence that the correction for aggregation error removes the bulk of disagreement between data sources. Figure A.2 provides a similar figure, but with aggregate post-tax earnings (instead of pretax), demonstrating a similar pattern.

Finally, Figure 4 presents a comparison of average annual foreign effective tax rates (ETRs) from Compustat and two SOI samples, one adjusted for aggregation error, the other unadjusted. Compared to the adjusted series, the average ETR computed from the unadjusted time series is downward-biased, as would be expected given that it overestimates the denominator. There is still considerable residual disagreement, however, when comparing adjusted ETRs from SOI with

Figure 4: Comparison of Aggregate Foreign Effective Tax Rates



Notes: This figure presents a comparison of average foreign effective tax rates from Compustat, SOI (adjusted for aggregation error), and SOI (unadjusted). These tax rates are based on pretax income as reported in Figure 3 and foreign tax as reported in Figure 2. The left panel contains aggregates computed from MNCs in all industries and the right panel excludes MNCs operating in Financial, Utilities, Mining, Agriculture, or Oil & Gas. The panels demonstrate that lack of adjustment to the SOI data creates downward bias in the tax rate estimates. Alternately, they also show that comparatively higher taxes and lower foreign income in the Compustat data lead to larger Compustat ETR estimates than adjusted or unadjusted SOI.

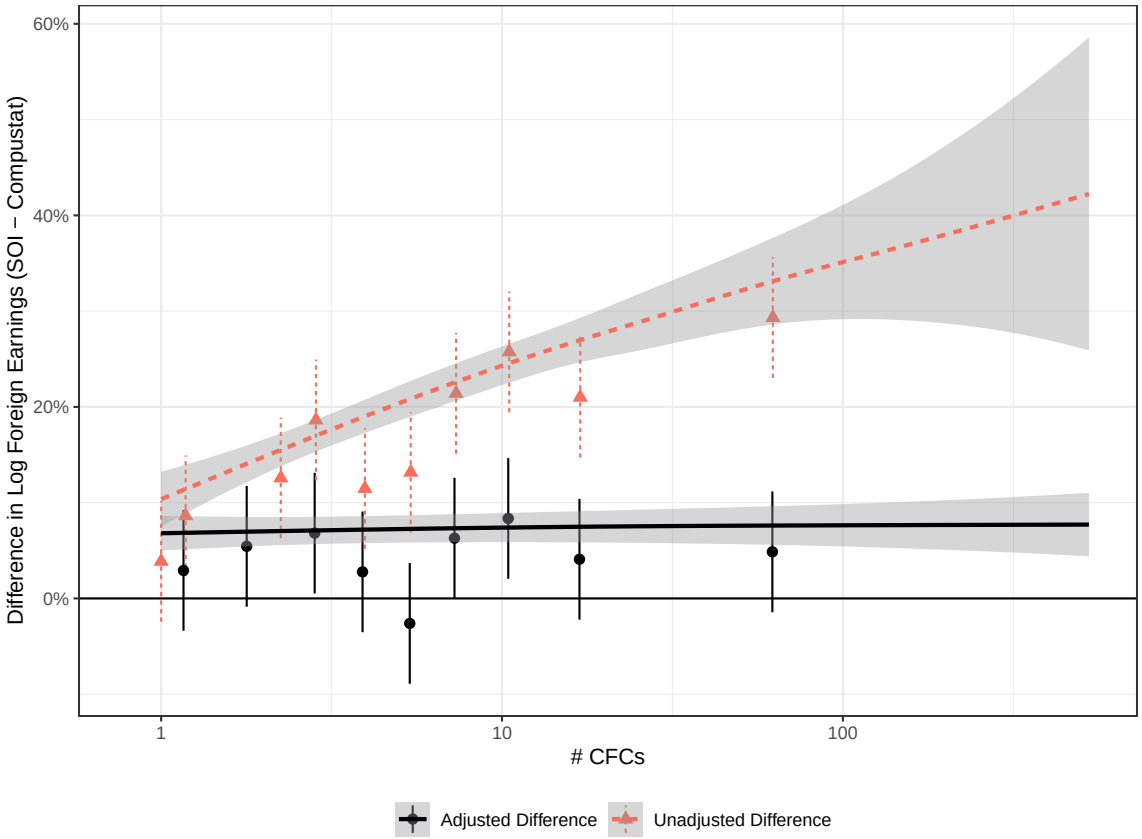
those computed from Compustat data. Compustat estimates are much higher throughout the sample period. This is a result of its higher tax estimates (Figure 2) and lower foreign earnings estimates (Figure 3).

Which source is more reliable? We do not have a full explanation for the remaining aggregate differences, but the Compustat series appear to be much noisier than the aggregate series computed from tax data. Finance and extractive industries distort aggregate rates upward in both Compustat and SOI data, but the distortion is much more evident in Compustat data, generating what seem to be unrealistically high average rates in certain periods. This suggests that reporting of foreign taxes and income in public filings may be less consistent and comparable over time relative to tax data.

3.3 Measurement Error and MNC Corporate Complexity

In this section, we show that aggregation error is closely related to the size of MNCs' foreign networks. More complex foreign networks create more potential for double counting, as the same dividends may be reported on the books of a potentially large number of foreign affiliates. Once

Figure 5: SOI/Compustat Foreign Earnings Differential



Notes: This figure presents the results of two cubic spline regressions where the dependent variable is the difference between log SOI foreign E&P and log Compustat foreign income. The orange line represents the estimates of a regression using unadjusted SOI E&P, while the black line uses adjusted SOI E&P. The orange triangles and black circles represent respective binscatter estimates produced for MNCs in 10 quantiles of foreign affiliate network size, as measured by the number of CFCs in the SOI data. The upward trends in the orange line and binscatter estimates indicate that aggregation error increases with the size of an MNC's foreign affiliate network. Alternately, the flat spline of the adjusted E&P regression indicates that residual differences between SOI and Compustat estimates are unrelated to a firm's foreign affiliate network size.

foreign earnings are adjusted to remove related dividends, however, there is no evidence that residual book-tax differences are systematically related to the complexity of an MNC's affiliate structure. Furthermore, we show that these adjustments result in residual measurement error that appears to be roughly constant over time. These results are encouraging for researchers, as they suggest that standard techniques to control for time-invariant heterogeneity and measurement error may be valid.

The aggregate trends shown above suggest that aggregation error primarily distorts estimates of foreign earnings, not foreign taxes. Given that the primary mechanism behind aggregation error relates to the double counting of foreign earnings that are distributed up through an MNC's foreign ownership network, we would expect aggregation error to increase with the size of an

MNC's foreign affiliate network. Figure 5 confirms that this is the case. The orange line and surrounding ribbon plot the estimates of a cubic spline regression. The dependent variable is the difference between log unadjusted foreign E&P as reported in SOI and log foreign income as reported in Compustat. The orange triangles are binscatter estimates produced for MNCs in 10 quantiles of foreign affiliate network size, as measured by the number of CFCs in the tax data. Note that there is a clear upward trend, indicating that aggregation error is closely related to MNCs' foreign network size.

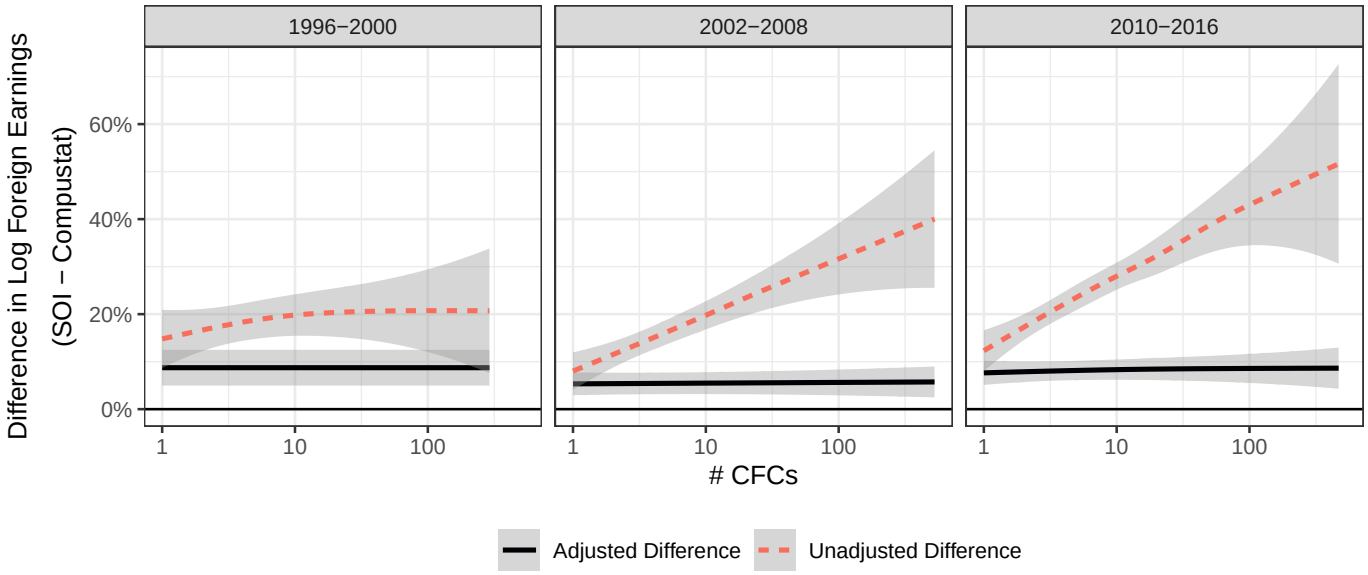
The black line and circles plot the same statistics computed for a sensitivity where the dependent variable uses SOI E&P, adjusted to remove dividend income. The spline is flat, suggesting that the residual differences between SOI and Compustat estimates are unrelated to the complexity of a firm's foreign network. If this adjustment were imperfect, and missed a portion of double-counted profits, we would expect this slope to perhaps be attenuated but not flat.

Figure 6 plots spline estimates for the same sample, but split into three distinct periods. This confirms that aggregation error has been increasing over time, as reflected by the increasing slope of the orange splines. Adjusted E&P, however, remains flat and roughly constant at the same level. This again suggests that residual measurement error has not been increasing over time, and instead might be due to other reporting differences between tax and public filings.

Even though the slope is flat, the black lines in both Figures 5 and 6 do appear to be significantly positive, indicating that SOI foreign earnings are on average about 8% larger than the figures reported in Compustat. While understanding the differences between book and tax data is notoriously difficult, the test we implement in this section provides a meaningful validation of the correction proposed by B&R for two reasons. First, though there are some existing level differences even after correction, these differences do not grow with increased potential for double counting, as measured by the number of related affiliates. Second, these results are robust across different time periods over which foreign tax planning by US MNCs became much more prevalent, as shown in Figure 6.

Table 2 provides a regression formulation of the graphical analysis shown above that estimates

Figure 6: SOI/Compustat Foreign Earnings Differential by Period



Notes: This figure plots the same spline estimates as Figure 5 above, split into three time periods. The increase in the slope of the orange line over the three periods further confirms that aggregation error in the unadjusted SOI data has been increasing over time. However, the roughly constant black line indicates that residual measurement error has not been increasing.

the following specification,

$$\log(\text{Foreign Pretax E\&P}_{it}^{f,\text{SOI}}) - \log(\text{Foreign Pretax Income}_{it}^{f,\text{Compustat}}) = \beta \log(N_{it}^{\text{CFC}}) + \mu_t + u_i. \tag{1}$$

Again, the dependent variable is the difference between log foreign E&P as reported in SOI and log foreign earnings as reported in Compustat. Odd columns use unadjusted E&P and even columns introduce the double counting adjustment. The independent variable is the log of the number of CFCs in a firm-year. For the unadjusted difference, the coefficients on this variable are positive, with large t-statistics. The adjusted coefficients, however, are close to zero with narrow confidence intervals. Columns (3), (4), (7), and (8) include the year fixed effects μ_t, and Columns (5) through (8) remove the industries excluded in the aggregate plots shown in the previous section.

4 Measurement Error and Tax Elasticities of Foreign Income

In the previous section, we show that introducing a simple correction to remove related dividends appears to remove systematic aggregation error in the measurement of foreign earnings in tax

Table 2: Measurement Error and MNC Foreign Affiliate Network Size

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(Intercept)	0.095***	0.082**			0.081***	0.079***		
	(0.027)	(0.027)			(0.016)	(0.016)		
Log # CFCs	0.043***	−0.001	0.043***	−0.012	0.047***	0.005	0.050***	0.000
	(0.009)	(0.009)	(0.009)	(0.009)	(0.005)	(0.005)	(0.006)	(0.005)
Num. Obs.	9993	9993	9993	9993	9040	9040	9040	9040
FE: year	-	-	X	X	-	-	X	X
Remove Outliers	-	-	-	-	X	X	X	X
Adjusted Earnings	-	X	-	X	-	X	-	X

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Notes: This table displays coefficients from regressions of the form 1. The dependent variable is the difference between log SOI foreign E&P and log Compustat foreign income. Odd columns use unadjusted SOI E&P and even columns use adjusted SOI E&P. Columns (3), (4), (7), and (8) add year fixed effects, and Columns (5) through (8) remove MNCs operating in Financial, Utilities, Mining, Agriculture, or Oil & Gas. *, **, and *** denote statistical significance at the 10, 5, and 1% level.

data. In this section, we follow extensive work in this literature (Dharmapala and Riedel, 2013; Heckemeyer and Overesch, 2013) to recompute tax semi-elasticities of foreign earnings following the methodology in Dowd, Landefeld and Moore (2017, henceforth DLM). DLM use unadjusted earnings and a simple regression framework to examine how sensitive the allocation of MNC earnings are to average and statutory tax rates by country.

First, we attempt to replicate their results without any alterations using unadjusted earnings. DLM show that foreign earnings are more sensitive to tax rates in jurisdictions typically classified as tax havens. We construct a sample that produces very similar sample statistics to those of DLM, and obtain similar regression results with their aggregation process. We then modify their approach to utilize adjusted foreign earnings that remove aggregation error. Although this adjustment does result in noticeably different estimates, their general findings are robust to this measurement error.

Notably, our finding that DLM's results are robust to measurement error is different than the takeaway from B&R, who use BEA data to demonstrate that, once aggregation error is taken into account, MNCs do not appear to be more sensitive to rate changes in haven jurisdictions. One reason for this could be specification differences—B&R run country-level, rather than firm-level, regressions. We examine a similar specification that continues to show larger elasticities in tax haven jurisdictions. This apparent contradiction can be resolved by understanding how foreign

earnings are geographically classified in BEA vs. tax data.

4.1 Sample Construction

DLM provide a clear description of how they construct their sample. This allows us to create a similar analysis sample with data from Form 5471. They start with a sample of CFCs with positive foreign earnings for MNCs that do not operate in extractive industries, finance, insurance, or utilities. As shown in Section 3, this appears to be a reasonable selection criterion given that tax reporting can be quite different for these industries. DLM require that at least 10 CFCs operate in countries that are included in the analysis.[8] They also require that CFCs report positive amounts for wages and tangible assets.[9]

Table A.1 contains three iterations of relevant sample statistics for our analysis. First, we reproduce DLM's sample statistics as published in Table 1 of their paper. Second, we replicate these statistics using SOI data. There are some small differences between the samples—ours is slightly smaller (90,746 vs. 96,959 in their sample), which may be due to differences in how they classify industries.[10] Overall, however, the summary statistics are very close.

Finally, we produce a similar sample using adjusted foreign earnings.[11] Unsurprisingly, average profits are lower. The sample size is also somewhat smaller due to the fact that foreign earnings for a small part of the sample become negative once the adjustment is introduced.

4.2 Corrected Measurements of Tax Semi-Elasticities of Foreign Earnings

Adjusted Estimates from DLM

We use the samples described above to estimate the regression,

[8] This meets standard criteria for disclosure rules from the SOI.

[9] Assets are reported on the balance sheet disclosed on Form 5471, Schedule F. Tangible assets are calculated as the sum of end-year figures for buildings and other depreciable assets, depletable assets, and land (Sch. F, lines 8a, 9a, and 10 on the 12-2007 revision of the form). Although DLM indicate that depreciation is removed, we obtain similar summary statistics only when computing tangible assets before removing accumulated depreciation and depletion and proceed with the analysis accordingly. Wages are computed from the income statement disclosed in Schedule C (line 10 on the 12-2007 revision, "Compensation not deducted elsewhere").

[10] DLM do not provide the exact industry codes that they exclude in their analysis.

[11] We make two adjustments to DLM's measurement of foreign earnings. They use pretax income disclosed in the income statement (Schedule C). We instead use foreign earnings and profits (E&P), disclosed on Schedule H, and remove related dividends disclosed on Schedule M. B&R suggest that E&P is reported in a more consistent manner than the earnings figure reported on Schedule C.

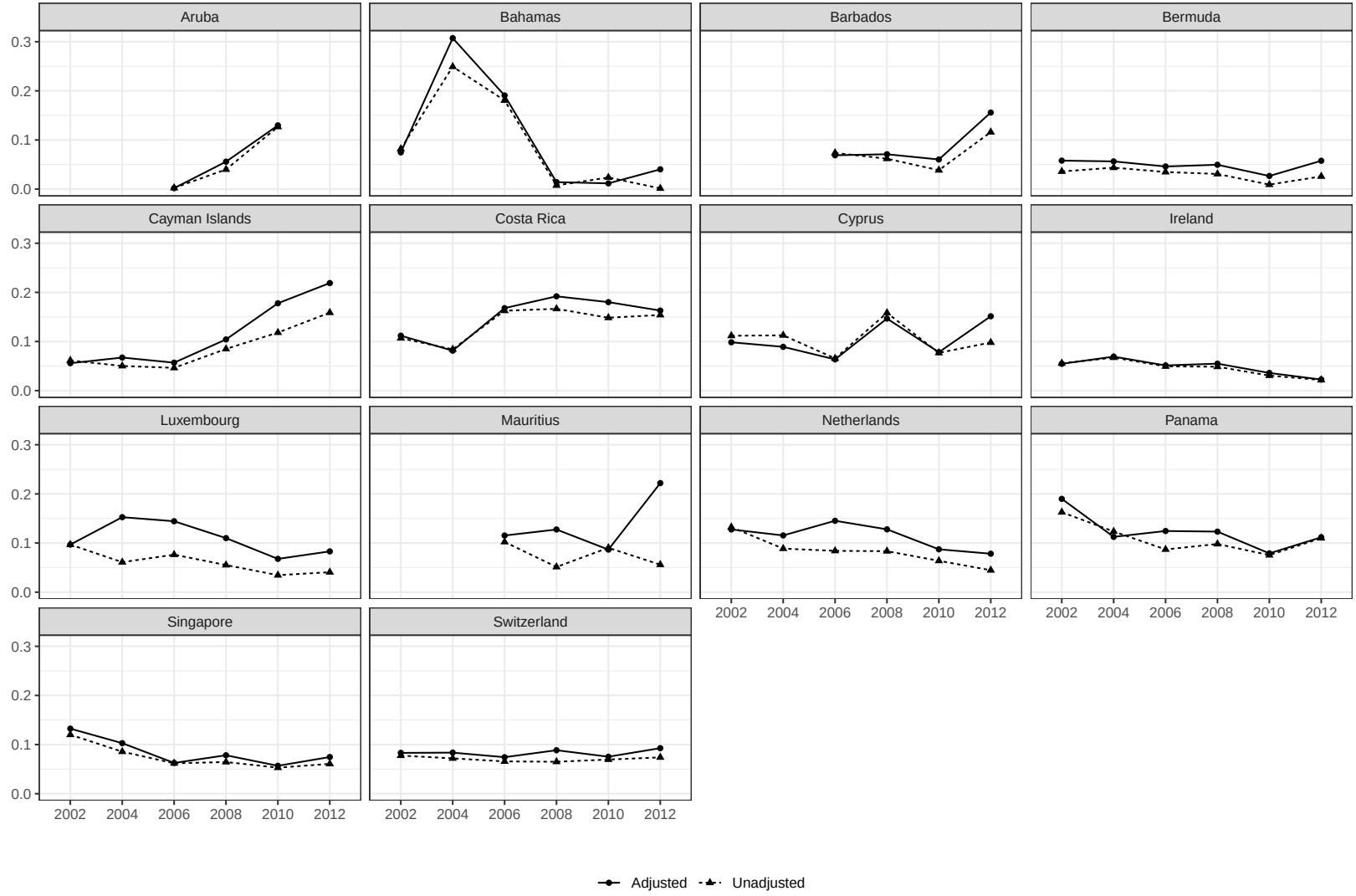
$$\log \pi_{ict} = \alpha + \beta_1 \log K_{ict} + \beta_2 \log L_{ict} + \beta_3(1 - \tau_{ct}) + \beta_4 X_{ct} + \mu_t + \psi_g + \epsilon_{it}, \tag{2}$$

where π_{it} are foreign earnings following the notation of DLM, K_{it} are tangible assets, L_{it} is the wage bill, and τ_{ct} is either the average tax rate faced by CFCs incorporated in jurisdiction c or the statutory rate for country c in period t. X_{ct} is a vector of country-level controls that includes a second-order polynomial of GDP per capita and population. μ_t is a year fixed effect and ψ_g is an MNC fixed effect. As in DLM, β_3 is the parameter of interest. This parameter captures how sensitive MNC profit allocations are to tax rates. In practice, when using average tax rates, τ_{ct} is computed using a jackknife procedure to remove the effect of firm i on the average rate.

Figure 7 shows average tax rates for the set of tax havens considered in DLM. After implementing the correction for aggregation error, these estimates generally increase. This is expected, given that aggregation error increases the denominator of ETR estimates, while leaving the numerator unchanged, creating downward bias.[12]

[12]We also include Netherlands in this figure given its importance as a domicile for international tax planning structures, although it is not classified as a tax haven for the DLM-style regressions for comparability. The Netherlands is classified as a haven by B&R.

Figure 7: Adjusted and Unadjusted Average Tax Rates for Selected Haven Jurisdictions



Notes: This figure presents adjusted and unadjusted average tax rates over time for the set of tax haven countries we consider in our DLM-style regression analysis (Equation 2), as well as the Netherlands.

We also run two modified specifications that are also presented by DLM. The first allows the intercept and β_3 to vary according to whether country c is classified as a tax haven,

$$\log \pi_{ict} = \alpha_h + \beta_1 \log K_{ict} + \beta_2 \log L_{ict} + \beta_{3,h}(1 - \tau_{ct}) + \beta_4 X_{ct} + \mu_t + \psi_g + \epsilon_{it}. \tag{3}$$

Note that the only difference between Equation 2 and Equation 3 is the addition of h subscripts on α and β_3. The second allows for a second-order polynomial in the keep rate,

$$\log \pi_{ict} = \alpha_h + \beta_1 \log K_{ict} + \beta_2 \log L_{ict} + \beta_3(1 - \tau_{ct}) + \gamma_3(1 - \tau_{ct})^2 + \beta_4 X_{ct} + \mu_t + \psi_g + \epsilon_{it}. \tag{4}$$

Existing literature interprets β_3 as a semi-elasticity. Here, we do not emphasize a causal interpretation, which would require conditional exogeneity of τ_{ct}.[13] Instead, we focus on whether these parameter estimates are sensitive to measurement error.

Table 3 shows estimates for the specification contained in Equation 2 using both unadjusted and adjusted earnings. Columns (1) and (3) present estimates using statutory rates, and Columns (2) and (4) present estimates using average rates. The coefficient for the keep rate in the adjusted regression is slightly larger, but close in magnitude to the unadjusted estimates.

Table 4 presents unadjusted and adjusted results for the specification contained in Equation 3, following the haven classification of Gravelle (2013). This classification includes Aruba, the Bahamas, Barbados, Bermuda, the Cayman Islands, Costa Rica, Cyprus, Ireland, Luxembourg, Mauritius, Panama, Singapore, and Switzerland. The coefficients on the interaction term are large and significant, which replicates the general findings of DLM. This suggests that foreign earnings are more sensitive to changes in tax rates within havens. Table A.2 contains an alternate specification contained in Equation 3 but defines havens as countries in the smallest decile of the tax rate distribution by year. Both definitions are used by DLM.

Finally, Table 5 presents unadjusted and adjusted results for the specification contained in Equation 4. As in DLM, the second-order keep rate coefficient is significant, and actually increases in magnitude for the adjusted results.

[13] One can imagine that countries that lower effective tax rates on foreign investment may also change other policy levers to attract capital, which would violate this assumption.

Table 3: Firm-Level Semi-Elasticity of Net Income to Corporate Tax Rates

	(1)	(2)	(3)	(4)
1-Rate	1.293***	0.740***	1.308***	0.836***
	(0.150)	(0.086)	(0.155)	(0.102)
Log capital	0.321***	0.320***	0.317***	0.317***
	(0.009)	(0.009)	(0.009)	(0.009)
Log wages	0.394***	0.393***	0.391***	0.390***
	(0.012)	(0.012)	(0.013)	(0.013)
GDP per capita	4.866***	2.432+	4.159**	2.291+
	(1.329)	(1.313)	(1.384)	(1.388)
GDP per capita sq.	53.218***	72.183***	46.753**	61.697***
	(15.239)	(15.468)	(15.108)	(15.662)
Population	0.001***	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Population sq.	0.000***	0.000**	0.000***	0.000**
	(0.000)	(0.000)	0.000***	0.000**
Sample	Unadjusted	Unadjusted	Adjusted	Adjusted
Num.Obs.	90,746	90,746	88,801	88,801
R2 Within	0.463	0.463	0.457	0.457
ETR	Statutory	Average	Statutory	Average

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Notes: This table displays coefficients from regressions under Equation 2. The results in this table correspond to Dowd, Landefeld and Moore (2017) Table 2. The dependent variable is log profits for MNC i in country c in period t, using unadjusted (Columns (1) - (2)) or adjusted (Columns (3) - (4)) SOI foreign E&P. The parameter of interest, 1-Rate, captures the sensitivity of MNC profit allocation to changes in tax rates. Odd columns present estimates using statutory tax rates and even columns present estimates using average rates. *, **, and *** denote statistical significance at the 10, 5, and 1% level.

Country-Level Regressions

B&R also estimate DLM-style regressions using corrected data from the BEA, and show that there is no evidence for heterogeneity once estimates of net income are corrected to adjust for aggregation error. One reason for this may be differences in specification. B&R estimate country-level regressions of the form,

$$\log \pi_{ct} = \alpha + \beta_1 \log K_{ct} + \beta_2 \log L_{ct} + \beta_3(1 - \tau_{ct}) + \beta_4 X_{ct} + \mu_t + \gamma_c + \epsilon_{ct}, \tag{5}$$

where the dependent variable is aggregate foreign earnings for foreign affiliate located in country c in year t. They present results inclusive and exclusive of country fixed effects γ_c.

Table 4: Firm-Level Semi-Elasticity of Net Income to Corporate Tax Rates: Using Gravelle (2013) Haven Definition

	(1)	(2)	(3)	(4)
Haven	−1.997***	−2.631**	−2.361***	−1.543*
	(0.410)	(0.971)	(0.397)	(0.728)
1-Rate	0.539**	0.578***	0.418*	0.605***
	(0.182)	(0.089)	(0.182)	(0.112)
Haven × 1 - Rate	2.657***	3.020**	3.144***	1.872*
	(0.519)	(1.054)	(0.502)	(0.804)
Log capital	0.319***	0.320***	0.316***	0.316***
	(0.009)	(0.009)	(0.009)	(0.009)
Log wages	0.395***	0.393***	0.392***	0.391***
	(0.012)	(0.012)	(0.013)	(0.013)
GDP per capita	3.566**	3.233*	2.626+	2.559+
	(1.318)	(1.318)	(1.361)	(1.382)
GDP per capita sq.	50.850***	48.653**	44.019**	47.561**
	(15.208)	(15.383)	(14.896)	(15.555)
Population	0.001***	0.001***	0.001**	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Population sq.	0.000**	0.000***	0.000**	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Sample	Unadjusted	Unadjusted	Adjusted	Adjusted
Num.Obs.	90,746	90,746	88,801	88,801
R2 Within	0.464	0.463	0.458	0.457
ETR	Statutory	Average	Statutory	Average

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Notes: This table displays coefficients from regressions under Equation 3. The results in this table correspond to Dowd, Landefeld and Moore (2017) Table 4. The dependent variable is log profits for MNC i in country c in period t, using unadjusted (Columns (1) - (2)) or adjusted (Columns (3) - (4)) SOI foreign E&P. The parameter of interest is an interaction term between 1-Rate, which captures the sensitivity of MNC profit allocation to changes in tax rates, and a dummy variable equal to one if country c is classified as a tax haven. Tax havens are classified as defined in Gravelle (2013), the same methodology used by DLM (Figure 7). Odd columns present estimates using statutory tax rates and even columns present estimates using average rates. *, **, and *** denote statistical significance at the 10, 5, and 1% level.

We aggregate the sample described in the previous section to the country level, and estimate similar regressions, presented in Tables 6 and 7 below. Even columns in Table 6 include the haven interaction term analogous to the specification contained in Equation 3, and the same columns in Table 7 include second-order polynomials in the tax rate analogous to specification contained in Equation 4.[14] B&R present results with respect to the tax rate (as opposed to the

[14]We include the same tax havens used by B&R—Bermuda, the Cayman Islands, Ireland, Luxembourg, the

Table 5: Non-Linear Firm-Level Semi-Elasticity of Net Income to Changes in Corporate Tax Rates

	(1)	(2)	(3)	(4)
1-Rate	−8.847***	−0.542+	−9.963***	−3.666**
	(1.944)	(0.291)	(1.943)	(1.150)
1-Rate sq.	6.764***	0.885***	7.521***	2.916***
	(1.320)	(0.211)	(1.316)	(0.751)
Log capital	0.320***	0.320***	0.317***	0.317***
	(0.009)	(0.009)	(0.009)	(0.009)
Log wages	0.394***	0.393***	0.391***	0.391***
	(0.012)	(0.012)	(0.013)	(0.013)
GDP per capita	4.512***	2.514+	3.755**	2.365+
	(1.336)	(1.311)	(1.388)	(1.389)
GDP per capita sq.	47.706**	67.960***	40.917**	57.132***
	(15.173)	(15.392)	(14.880)	(15.498)
Population	0.001***	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Population sq.	0.000**	0.000**	0.000**	0.000**
	(0.000)	(0.000)	(0.000)	(0.000)
Sample	Unadjusted	Unadjusted	Adjusted	Adjusted
Num.Obs.	90,746	90,746	88,801	88,801
R2 Within	0.464	0.463	0.458	0.457
ETR	Statutory	Average	Statutory	Average

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Notes: This table displays coefficients from regressions under Equation 4. The results in this table correspond to Dowd, Landefeld and Moore (2017) Table 5. The dependent variable is log profits for MNC i in country c in period t,using unadjusted (Columns (1) - (2)) or adjusted (Columns (3) - (4)) SOI foreign E&P. The parameter of interest, 1-Rate, is the semi-elasticity of sensitivity of MNC profit allocation to changes in tax rates. Odd columns present estimates using statutory tax rates and even columns present estimates using average rates. *, **, and *** denote statistical significance at the 10, 5, and 1% level.

keep rate in prior regressions). We do the same to facilitate comparison with their results.

In both tables, Columns (1) through (4) present estimates using unadjusted foreign earnings, and Columns (5) through (8) present estimates using adjusted foreign earnings. As in B&R, estimates are computed with and without country fixed effects. Table 6 demonstrates that, unlike in B&R, the coefficient on the haven interaction remains large and significant, although slightly attenuated. Results in Table 7 are mixed—the adjusted estimates appear to weaken the non-linear term. However, the non-linear term is not significant even in the unadjusted

Netherlands, Singapore, and Switzerland.

specification with country fixed effects. These results highlight the value of using firm-level data, which yield more robust results in all cases and which generally have more statistical power. Differences between our results and those of B&R at the country level may be due to several factors, including differences in how profits are allocated geographically across countries between the two datasets, and the inclusion of smaller MNCs that are not surveyed by the BEA.

Table 6: Country-Level Semi-Elasticity of Net Income to Corporate Tax Rates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ETR	−2.833**	−1.173+	−2.706*	−2.644**	−2.313***	−1.581*	−2.445*	−2.414*
	(0.480)	(0.575)	(0.678)	(0.631)	(0.286)	(0.408)	(0.641)	(0.675)
Haven × ETR		−12.099*		−11.802+		−9.054**		−8.290*
		(3.020)		(4.872)		(1.948)		(2.420)
Log capital	0.480***	0.437**	0.370*	0.341*	0.426***	0.390***	0.160	0.123
	(0.067)	(0.084)	(0.128)	(0.129)	(0.044)	(0.043)	(0.171)	(0.165)
Log wages	0.549**	0.554**	0.358	0.294	0.604***	0.606***	0.358	0.360
	(0.098)	(0.110)	(0.258)	(0.268)	(0.056)	(0.068)	(0.234)	(0.248)
GDP per capita	−0.177+	−0.167+	−0.009	0.010	−0.158+	−0.135+	−0.179	−0.128
	(0.085)	(0.070)	(0.069)	(0.085)	(0.068)	(0.053)	(0.130)	(0.104)
GDP per capita sq.	0.024*	0.014	0.008	0.003	0.025**	0.012*	0.021*	0.012+
	(0.009)	(0.008)	(0.006)	(0.007)	(0.006)	(0.004)	(0.008)	(0.005)
Population	0.137	0.105	−0.277	0.227	0.169	0.199	−3.250	−2.974
	(0.104)	(0.071)	(2.183)	(1.919)	(0.153)	(0.138)	(1.623)	(1.546)
Population sq.	−0.012	−0.008	0.016	0.001	−0.014	−0.016	0.131	0.122
	(0.007)	(0.005)	(0.100)	(0.091)	(0.011)	(0.010)	(0.070)	(0.068)
Sample	Unadjusted	Unadjusted	Unadjusted	Unadjusted	Adjusted	Adjusted	Adjusted	Adjusted
Num. Obs.	269	269	269	269	269	269	269	269
Year FEs	X	X	X	X	X	X	X	X
Country FEs			X	X			X	X

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Notes: This table displays coefficients from regressions under Equation 5. The results in this table correspond to Blouin and Robinson (2023) Table 2, Panel A. The dependent variable is aggregate foreign earnings for a foreign affiliate located in country c in year t. The parameter of interest is effective foreign tax rate. Even columns include the same tax haven interaction term as presented in Equation 3 (Tables 4 and A.2). Columns (1) through (4) use unadjusted SOI foreign E&P; Columns (5) through (8) use adjusted E&P. Columns (3), (4), (7), and (8) also include country fixed effects.

Table 7: Non-Linear Country-Level Semi-Elasticity of Net Income to Corporate Tax Rates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ETR	−2.833**	−12.001**	−2.706*	−9.359+	−2.313***	−3.832*	−2.445*	−3.200+
	(0.480)	(2.781)	(0.678)	(4.519)	(0.286)	(1.360)	(0.641)	(1.470)
ETR Sq.		18.742*		12.321		1.760		0.759
		(4.817)		(7.329)		(1.689)		(1.159)
Log capital	0.480***	0.508***	0.370*	0.392*	0.426***	0.442***	0.160	0.153
	(0.067)	(0.071)	(0.128)	(0.136)	(0.044)	(0.041)	(0.171)	(0.175)
Log wages	0.549**	0.522**	0.358	0.278	0.604***	0.594***	0.358	0.368
	(0.098)	(0.096)	(0.258)	(0.271)	(0.056)	(0.054)	(0.234)	(0.245)
GDP per capita	−0.177+	−0.164+	−0.009	−0.084	−0.158+	−0.153+	−0.179	−0.188
	(0.085)	(0.079)	(0.069)	(0.069)	(0.068)	(0.067)	(0.130)	(0.119)
GDP per capita sq.	0.024*	0.019	0.008	0.009	0.025**	0.023*	0.021*	0.020*
	(0.009)	(0.010)	(0.006)	(0.005)	(0.006)	(0.006)	(0.008)	(0.007)
Population	0.137	0.079	−0.277	−0.278	0.169	0.194	−3.250	−3.442
	(0.104)	(0.094)	(2.183)	(1.913)	(0.153)	(0.151)	(1.623)	(1.767)
Population sq.	−0.012	−0.007	0.016	0.022	−0.014	−0.016	0.131	0.141
	(0.007)	(0.007)	(0.100)	(0.090)	(0.011)	(0.011)	(0.070)	(0.076)
Sample	Unadjusted	Unadjusted	Unadjusted	Unadjusted	Adjusted	Adjusted	Adjusted	Adjusted
Num. Obs.	269	269	269	269	269	269	269	269
Year FEs	X	X	X	X	X	X	X	X
Country FEs			X	X			X	X

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table displays coefficients from regressions under Equation 5. The results in this table correspond to Blouin and Robinson (2023) Table 2, Panel B. The dependent variable is aggregate foreign earnings for a foreign affiliate located in country c in year t. The parameter of interest is effective foreign tax rate. Even columns include the same tax haven interaction term as presented in Equation 4 (Table 5). Columns (1) through (4) use unadjusted SOI foreign E&P; Columns (5) through (8) use adjusted E&P. Columns (3), (4), (7), and (8) also include country fixed effects.

5 Conclusion

Increasing attention on tax avoidance and profit shifting has led to a recent wave of tax policies regulating the multinational activities of both individuals (e.g., Johannesen, Reck, Risch, Slemrod, Guyton and Langetieg, 2024) and corporations (e.g., Bilicka, Devereux and Güçeri, 2023). For researchers who use accounting and tax data to study these issues in the context of multinational firms, properly measuring tax avoidance is a primary concern. IRS tax data provide a wealth of information for researchers, but have the potential to generate significant measurement error if researchers do not properly aggregate foreign earnings.

This paper improves the general understanding of the benefits and measurement challenges inherent in using IRS tax data, and in particular evaluates one potential solution for tax data aggregation error. Using a novel proxy based on a comparison with consolidated measures from public filings, we show that a simple adjustment proposed by Blouin and Robinson (2023) significantly harmonizes book and tax measures of foreign income and tax rates. After applying this correction, we find no relationship between residual book-tax differences and MNC foreign affiliate network complexity, and no increase in residual book-tax differences over time. Finally, by replicating existing literature estimates of tax semi-elasticities across haven and non-haven jurisdictions, we show that these prior conclusions are generally robust to aggregation error.

The performance of this adjustment is encouraging for the use of IRS tax data in analyzing MNC activity. Ultimately, however, whether tax data is the most "correct" data source depends on the goal and estimand of the research. By shedding new light on the performance of this correction in IRS data, our results can help researchers and policymakers more precisely estimate tax data outcomes and more accurately interpret different measures of MNC activity across data sources.

References

Albertus, James F. 2023. "Income Shifting Out of the United States by Foreign Multinational Firms." *Available at SSRN 2691065*.

Altshuler, Rosanne, and Harry Grubert. 2006. "Governments and Multinational Corporations in the Race to the Bottom." *Tax Notes*, 110(8).

Altshuler, Rosanne, Lysle Boller, and Juan Carlos Suárez Serrato. 2024. "Tax Planning and Multinational Behavior." *Working Paper*.

Bilicka, Katarzyna Anna. 2019. "Comparing UK Tax Returns of Foreign Multinationals to Matched Domestic Firms." *American Economic Review*, 109(8): 2921–53.

Bilicka, Katarzyna, Michael Devereux, and Irem Güçeri. 2023. "Tax-Avoidance Networks and the Push for a "Historic" Global Tax Reform." *Tax Policy and the Economy*, 37: 57–108.

Blouin, Jennifer L., and Leslie Robinson. 2023. "Accounting for the Profits of Multinational Enterprises: Double Counting and Misattribution of Foreign Affiliate Income." SSRN October 1, 2023.

Blouin, Jennifer L., and Linda K Krull. 2014. "Does Organizational Form Affect Firms' Foreign Operations? The Role of "Check-the-Box" on Multinational Tax Planning." Vol. 107, 1–31, JSTOR.

Blouin, Jennifer L., Linda K. Krull, and Leslie A. Robinson. 2012. "Is U.S. Multinational Dividend Repatriation Policy Influenced by Reporting Incentives?" *The Accounting Review*, 87(5): 1463–1491. _eprint: https://publications.aaahq.org/accounting-review/article-pdf/87/5/1463/21222/accr-50193.pdf.

Bound, John, and Alan B Krueger. 1991. "The extent of measurement error in longitudinal earnings data: Do two wrongs make a right?" *Journal of labor economics*, 9(1): 1–24.

Bound, John, Charles Brown, Greg J Duncan, and Willard L Rodgers. 1994. "Evidence on the validity of cross-sectional and longitudinal labor market data." *Journal of Labor Economics*, 12(3): 345–368.

Clausing, Kimberly A. 2016*a*. "The Effect of Profit Shifting on the Corporate Tax Base in the United States and Beyond." *National Tax Journal*, 69(4): 905–934.

Clausing, Kimberly A. 2016*b*. "The Effect of Profit Shifting on the Corporate Tax Base in the United States and Beyond." *National Tax Journal*, 69(4): 905–934.

Dharmapala, Dhammika, and Nadine Riedel. 2013. "Earnings shocks and tax-motivated income-shifting: Evidence from European multinationals." *Journal of Public Economics*, 97(Supplement C): 95 – 107.

Dowd, Tim, Paul Landefeld, and Anne Moore. 2017. "Profit Shifting of U.S. Multinationals." *Journal of Public Economics*, 148: 1–13.

Dyreng, Scott, and Michelle Hanlon. 2021. "Tax avoidance and multinational firm behavior."

Gravelle, Jane G. 2013. "Tax Havens: International Tax Avoidance and Evasion." Congressional Research Service Web Site https://crsreports.congress.gov/product/details?prodcode=R40623 (Accessed: July 4, 2023).

Heckemeyer, Jost H., and Michael Overesch. 2013. "Multinationals? Profit Response to Tax Differentials: Effect Size and Shifting Channels."

Hines, James R., and Eric M. Rice. 1994. "Fiscal Paradise: Foreign Tax Havens and American Business." *The Quarterly Journal of Economics*, 109(1): 149–182.

IRS. 2011. "SOI Sampling Methodology and Data Limitations." U.S. Internal Revenue Service https://www.irs.gov/pub/irs-prior/p946--2002.pdf (Accessed: February 7, 2022).

IRS. 2019. "IRS Form 1120, Schedule M-3." U.S. Internal Revenue Service https://www.irs.gov/pub/irs-pdf/f1120sm3.pdf (Accessed: February 6, 2024).

IRS. 2022*a*. "SOI Tax Stats - Controlled Foreign Corporations ." U.S. Internal Revenue Service Web Site https://www.irs.gov/statistics/soi-tax-stats-controlled-foreign-corporations (Accessed: February 7, 2022).

IRS. 2022*b*. "SOI Tax Stats - Corporation Complete Report." U.S. Internal Revenue Service Web Site https://www.irs.gov/statistics/soi-tax-stats-corporation-complete-report Accessed: February 7, 2022).

Johannesen, Niels, Daniel Reck, Max Risch, Joel Slemrod, John Guyton, and Patrick Langetieg. 2024. "The Offshore World According to FATCA: New Evidence on the Foreign Wealth of US Households." *Tax Policy and the Economy*, 38: 61–99.

Tørsløv, Thomas R, Ludvig S Wier, and Gabriel Zucman. 2018. "The Missing Profits of Nations." National Bureau of Economic Research Working Paper 24701.

Yagan, Danny. 2015. "Capital Tax Reform and the Real Economy: The Effects of the 2003 Dividend Tax Cut." *The American Economic Review*, 105(12): 3531–3563.

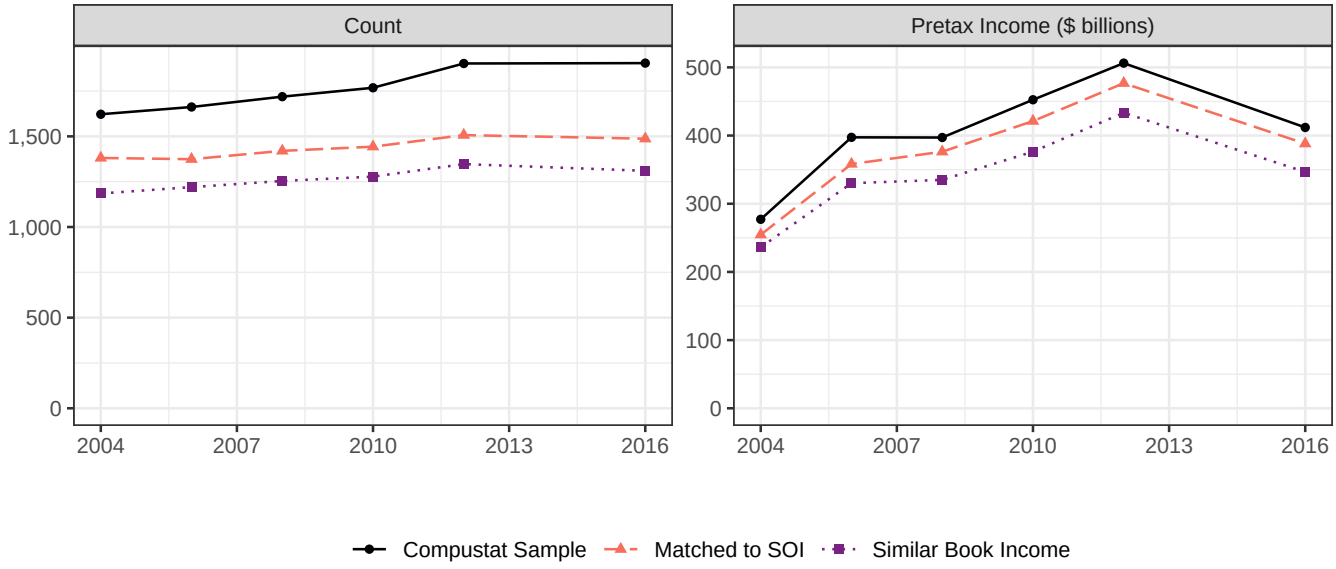
Zucman, Gabriel. 2014. "Taxing across Borders: Tracking Personal Wealth and Corporate Profits." *Journal of Economic Perspectives*, 28(4): 121–48.

Zucman, Gabriel. 2015. *The Hidden Wealth of Nations.* The University of Chicago Press.

Zwick, Eric, and James Mahon. 2017. "Tax Policy and Heterogeneous Investment Behavior." *American Economic Review*, 107(1): 217–48.

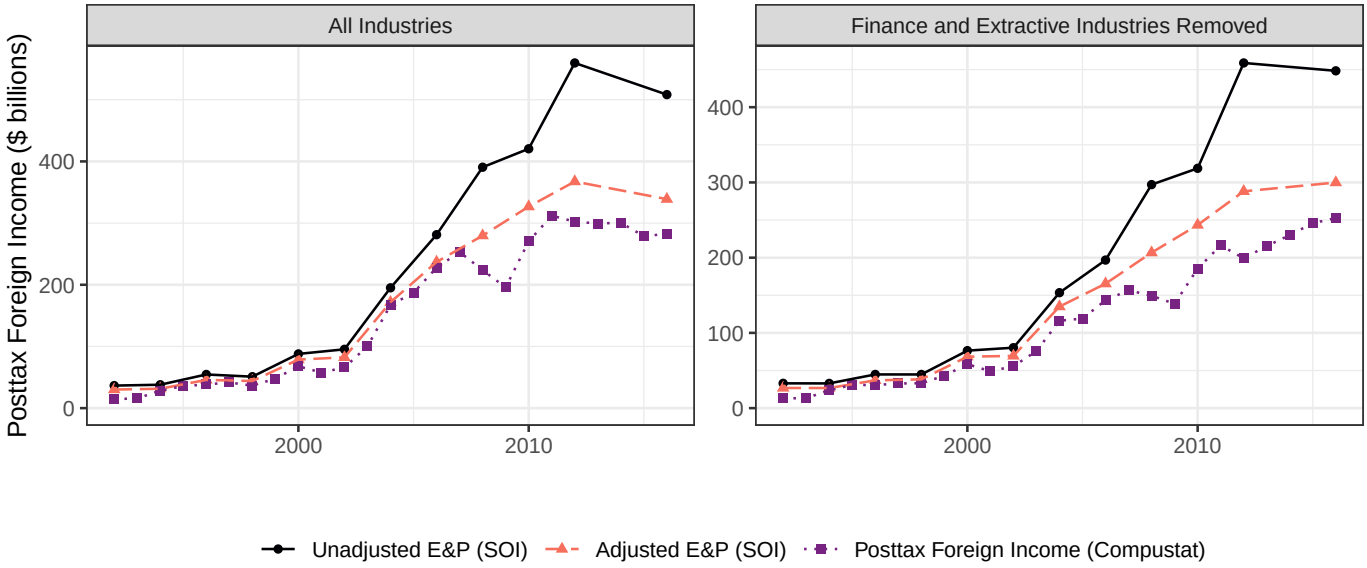
A Additional Tables and Figures

Figure A.1: Compustat MNCs vs. Matched Sample



Notes: This figure provides an overview of the sample construction process, following the same methodology as described in Figure 1 but starting with the Compustat sample rather than the SOI sample. Observations reporting missing pretax foreign income are removed from the initial Compustat sample. The left panel shows sample sizes in terms of number of MNCs, while the right panel shows sample sizes in terms of pretax foreign income. These panels show that the majority of Compustat MNCs were able to be linked to an SOI MNC, and that the match quality of linked MNCs is high.

Figure A.2: Comparison of Aggregate Foreign Income (Post-Tax)



Notes: This figure displays the same three aggregate foreign income measures as Figure 3 on a post-tax, rather than pretax, basis. It similarly underscores the effectiveness of B&R's adjustment for aggregation error in the SOI E&P data.

Table A.1: Sample Statistics

Sample	DLM Reproduction			Unadjusted			Adjusted		
	All	Nonhavens	Havens	All	Nonhavens	Havens	All	Nonhavens	Havens
Profits	14.41	9.01	57.19	15.25	9.42	76.31	13.33	7.99	69.60
Average Tax Rate	0.17	0.18	0.08	0.17	0.18	0.07	0.20	0.21	0.08
Statutory Tax Rate	0.29	0.30	0.18	0.29	0.30	0.18	0.29	0.30	0.18
Capital	27.37	24.40	50.91	27.15	23.25	68.05	26.82	22.83	68.93
Wages	7.75	7.18	12.32	8.08	7.37	15.49	8.01	7.25	16.05
GDP per capita	0.03	0.03	0.04	0.03	0.03	0.05	0.03	0.03	0.05
Population	150.24	168.54	5.14	154.50	168.82	4.60	156.43	170.83	4.58
2002	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.11	0.10
2004	0.16	0.16	0.15	0.16	0.16	0.15	0.16	0.16	0.15
2006	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17
2008	0.18	0.18	0.18	0.18	0.18	0.19	0.18	0.18	0.19
2010	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19
2012	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20
Num obs.	96,959	86,099	10,860	90,746	82,831	7,915	88,801	81,112	7,689

Notes: This table contains relevant sample statistics that correspond to Dowd, Landefeld and Moore (2017, henceforth DLM) Table 1. DLM use a sample of CFCs with positive foreign earnings for MNCs that report positive amounts for wages and tangible assets and do not operate in extractive industries, finance, insurance, or utilities. Statistics are shown for the entire sample, for CFCs located in non-haven countries, and for CFCs located in tax havens. Columns (1) - (3) show unadjusted sample statistics as reported in DLM Table 1. Columns (4) - (6) show our replication of the unadjusted sample statistics: while our sample is slightly smaller, the summary statistics are very similar between DLM's original and our replication. Finally, Columns (7) - (9) show sample statistics using adjusted foreign earnings data; whereas DLM use pretax income disclosed in IRS Form 5471, Schedule C, we instead use foreign E&P disclosed in Schedule H and remove related dividends disclosed in Schedule M. This results in lower average profits and a slightly smaller sample size post-adjustment.

Table A.2: Firm-Level Semi-Elasticity of Net Income to Corporate Tax Rates: Using Lowest Decile Haven Definition

	(1)	(2)	(3)	(4)
Haven	−1.871***	−6.361***	−2.706***	−6.844***
	(0.400)	(1.849)	(0.416)	(1.801)
1-Rate	0.759***	0.597***	0.735***	0.644***
	(0.154)	(0.087)	(0.155)	(0.104)
Haven × 1 - Rate	2.582***	6.932***	3.518***	7.499***
	(0.517)	(1.961)	(0.523)	(1.933)
Log capital	0.319***	0.319***	0.316***	0.316***
	(0.009)	(0.009)	(0.009)	(0.009)
Log wages	0.395***	0.393***	0.392***	0.391***
	(0.012)	(0.012)	(0.013)	(0.013)
GDP per capita	4.185**	3.158*	3.700**	2.124
	(1.318)	(1.319)	(1.377)	(1.387)
GDP per capita sq.	49.668***	56.189***	39.629**	60.933***
	(15.077)	(15.161)	(14.897)	(15.460)
Population	0.001***	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Population sq.	0.000***	0.000**	0.000**	0.000**
	(0.000)	(0.000)	(0.000)	(0.000)
Sample	Unadjusted	Unadjusted	Adjusted	Adjusted
Num.Obs.	90,746	90,746	88,801	88,801
R2 Within	0.464	0.463	0.458	0.457
ETR	Statutory	Average	Statutory	Average

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Notes: This table displays coefficients from regressions under Equation 3. The results in this table correspond to Dowd, Landefeld and Moore (2017) Table 4. The dependent variable is log profits for MNC i in country c in period t, using unadjusted (Columns (1) - (2)) or adjusted (Columns (3) - (4)) SOI foreign E&P. The parameter of interest is an interaction term between 1-Rate, which captures the sensitivity of MNC profit allocation to changes in tax rates, and a dummy variable equal to one if country c is classified as a tax haven. Tax havens are defined as the countries in the smallest decile of tax rate distribution by year. Odd columns present estimates using statutory tax rates and even columns present estimates using average rates. *, **, and *** denote statistical significance at the 10, 5, and 1% level.