Income-Shifting by U.S. Multinational Corporations

Amy Dunbar (University of Connecticut), with Ted Black (IRS Research, Applied Analytics and Statistics), Andrew Duxbury (James Madison University), and Thomas Schultz (Western Michigan University)[1]

I. Introduction

We explore the use of new foreign tax information reporting (FTIR) of payments to U.S. multinational corporations to estimate tax compliant income-shifting. The new FTIR data reports payments made by foreign-source payors to U.S. residents. This research is part of a larger effort to estimate the offshore component of the corporation income tax gap. The OECD's (Organization for Economic Cooperation and Development) Base Erosion and Profit Shifting (BEPS) project and "country-by-country" (CbC) reporting recommendations address concerns regarding the ability of multinational enterprises (MNEs) to engage in tax avoidance by exploiting gaps in international tax rules. In particular, global tax authorities find it challenging to fairly and consistently measure the economic value created by a multinational corporation (MNC) in a given country because transfer pricing and cost allocation strategies can artificially shift the recognition of profits to relatively low-tax or no-tax jurisdictions. We suggest that FTIR data provide a third-party source of tax data that will make Internal Revenue Service (IRS) operations more effective by identifying compliant corporate taxpayers.

We first estimate income-shifting by U.S. MNCs to/from their controlled foreign corporations (CFCs), using an N-state income-shifting model that computes after-tax rates of return for each MNC and CFC. We do this by equalizing the rates of return between the MNC and CFCs within the MNC group by adjusting profits, thereby removing the income shift (IS), and then aggregate the IS across all entities within the N jurisdictions each year to provide income-shifting estimates. We use data from the tax returns: the MNC's Form 1120 and the CFC's Form 5471 (*Information Return of U.S. Persons With Respect to Certain Foreign Corporations*) e-filed with the IRS from 2007–2015.[2]

Next, we use FTIR data that include information provided by payors to withholding agents in reporting countries[3] as a source to verify offshore income/deductions reported by U.S. MNCs and their related CFCs. The FTIR data are available from 1999 through 2012. We use data from 2007–2012 to match with e-file data. Thus, although we estimate income-shifting for 2007–2015, we can address compliance only for 2007–2012.

The FTIR data report the payor and recipient if the beneficial owner is a U.S. taxpayer. The data come from the foreign tax administration offices of approximately 30 countries, of which two-thirds designate whether the recipient is a corporation. We identify CFC payors that make payments to a corporate recipient who is also a member of the CFC's MNC group. The result of the payor and recipient matching is an intercompany matched dataset that allows us to compare Federal income tax return data and foreign information return data.

Finally, we develop two compliance measures to support inferences about compliant taxpayers, in contrast to most previous tax compliance research that relies upon evidence about noncompliance. These measures are based on two of the traditional components of the tax gap: filing compliance and reporting compliance. The

[1] We thank our discussant, Tim Dowd, Congressional Joint Committee on Taxation staff, for helpful comments. We also thank Mary-Helen Risler, IRS-RAAS, Charles Christian, Professor Emeritus, Arizona State University, and William Trautman, IRS LB&I, for their help. This paper results from work conducted under IRS contract # TIRNE-11-E-00052. The opinions expressed reflect those of the authors and not of the IRS.

[2] Treasury Regulations § 301.6011-5(a)(1) and (f) provide that a corporation must e-file its Form 1120 if it is required to file at least 250 returns during the calendar year ending with or within its taxable year and it reports "total assets at the end of the corporation's taxable year that equal or exceed $10 million on Schedule L of their Form 1120." The IRS tax return database contains tax return information for electronically filed business tax returns since 2006.

[3] Foreign information return data are provided by the source country to the residence country regarding various categories of income. See http://www.oecd.org/tax.

exploratory results provide global tax authorities with a basis for evaluating income-shifting risks and effectively deploying audit resources.

"Income-shifting," referred to as Base Erosion and Profit Shifting (BEPS)[4] by the OECD, has been studied since at least the early 1970s. These early studies generally documented intrafirm trading patterns that suggested tax motivated income-shifting (Jenkins and Wright (1975), Kopits (1976)). Since that time, studies have generally focused on estimating the amount of income-shifting, its sensitivity to changes in tax rates, and the influence of tax havens.

All studies struggle to determine the true amount of income shifted because "unshifted" profit or loss is unobservable, and therefore they use various methods to approximate unshifted profit. Several studies have attempted to approximate unshifted profit by calculating the amount of profit that should accrue to a firm based on labor, capital, and productivity, the method first used in Hines and Rice (1994). In estimating tax-motivated income-shifting, Hines and Rice (1994) add a tax incentive variable to the production function to estimate its sensitivity to tax rates. A few studies used a variation of this method by examining profit shocks including Dharmapala and Riedel (2013) and Markle (2015).

This stream of research also focuses on calculating the profit elasticity to tax rate changes. Most of these studies followed the main tests used in Hines and Rice (1994) and assumed a linear form. Dowd, Landefeld, and Moore (2017) challenge this linear assumption and find that prior studies overestimated the elasticity at higher tax rates and significantly underestimated elasticity at lower tax rates, suggesting that firms shift additional profits to low-tax jurisdictions when the benefits are at their greatest.

Christian and Schultz (2005) use a different approach based on the theoretical work by Rousslang (1997), which assumes the same marginal after-tax return on assets (ROA) across jurisdictions. Deviations from this norm may be tax-motivated. Using U.S. tax return data from 2001, they estimate that from $58 to $111 billion of income was shifted out of the United States. They identify several limitations in their study, including the use of a two-state model, which combines all foreign jurisdictions into one when comparing foreign and domestic ROAs.

Collins, Kemsley, and Lang (1998) use a two-state model to estimate the effect of the foreign tax rate and foreign tax credit position on foreign sales using publicly available data. Klassen and LaPlante (2012) use the same methodology as Collins, *et al.* (1998), but replace annual measures of tax and profitability with 5-year measures. They estimate firms with foreign tax rates lower than the U.S. rate increased income shifted out of the U.S. by approximately $10 billion over their sample period.

One of the major changes in U.S. tax law during the span of the literature is the advent of the "Check the Box" (CTB) regulations, which allowed MNCs to create a "disregarded" entity (DE) that was recognized as a corporation in the foreign jurisdiction but was disregarded as such by the United States. CTB broadened the ways in which U.S. multinationals could shift income among their foreign subsidiaries by enabling passive income (fixed, determinable, annual, or periodic (FDAP) income) to be paid from high-tax subsidiaries to low-tax subsidiaries while avoiding Subpart F, if structured properly. Several studies have examined the extent to which firms engage in additional income-shifting using DEs. Grubert (2012) uses a CTB indicator variable and finds that it contributed a 1-to-2-percentage-point decline in foreign effective tax rates. Altshuler and Grubert (2005) find similar results and estimate the total foreign tax reduction achieved to be approximately $7 billion by 2002.

We add to this literature in several ways. We first extend the two-state income-shifting analysis presented by Christian and Schultz (C&S) at the 2005 IRS Research Conference to an N-state shifting analysis. We then compare estimates of the income shifted from MNCs to: (1) CFCs in information reporting versus nonreporting countries; (2) CAP-owned CFCs versus non-CAP CFCs in any foreign country; and (3) CFCs that own a DE versus non-DE CFCs.

[4] We use the terms "income" and "profit" interchangeably.

Finally, we create two foreign tax information reporting compliance (FTIRC) measures based on a comparison of FTIR and IRS data. The first identifies the percentage of FTIR recipients relative to matched e-file recipients. The second compares information reporting data with reported income tax return data to reflect the correspondence of what CFCs report to the IRS vis-à-vis what the CFCs report to the tax administration in the source country. We use this second measure to infer compliant offshore income-shifting for the FTIR countries.

II. Income-Shifting Analysis

A. Income-Shifting Framework: N-State Model

Christian and Schultz (2005) posit convergence to a single capital factor price across two jurisdictions. That is, a corporation will allocate capital resources such that the marginal after-tax rate of return on capital investment (ATROC) will be equal between any two jurisdictions.[5] Our current research uses an N-state income-shifting model (ISM) that estimates the income shift variable, IS, to/from the foreign jurisdictions and the United States. Appendix A shows the derivation of the N-state model.

For example, if N = 3, the income-shifting framework posits convergence to a single capital factor price, ATROC, across three jurisdictions. If there is equivalence among all three jurisdictions in ATROC when adjusting for shifting, then:

$$\frac{[Y_d + IS_T] * (1-t_d)}{K_d} = \frac{[Y_1 - IS_1] * (1-t_1)}{K_1} = \frac{[Y_2 - IS_2] * (1-t_2)}{K_2}$$

where:

Y_d = pretax domestic income

Y_1 = pretax foreign jurisdiction 1 income

Y_2 = pretax foreign jurisdiction 2 income

K_d = domestic capital stock

K_1 = foreign jurisdiction 1 capital stock

K_2 = foreign jurisdiction 2 capital stock

t_d = tax rate domestic (Td/Yd) where T = level of taxes

t_1 = tax rate foreign jurisdiction 1

t_2 = tax rate foreign jurisdiction 2

IS_T = total income shift [IST = IS1 + IS2]

IS_1 = income shift jurisdiction 1

IS_2 = income shift jurisdiction 2.

The ISM computes ATROC for each MNC and CFC, then adjusts for the income shift (IS) to equalize the ATROCs, and finally aggregates the IS across all firms within N jurisdictions within a year to provide income-shifting estimates.

We cannot necessarily infer noncompliant income-shifting from the observation of unequal ATROCs. A higher-than-normal CFC ATROC, which reflects profit maximizing behavior in the ISM, corresponds to income being shifted from the U.S. parent to the CFC. However, a higher-than-normal ATROC can also reflect effective supply chain management, political risks, regional macroeconomic conditions, or other factors that raise the variability of returns in a given jurisdiction. Thus, a higher-than-normal ATROC may not be indicative of income-shifting that involves tax-motivated compliance risks.

[5] Rousslang (1997) uses a before-tax rate of return measure. We consider only income taxes when computing ATROC. Consistent with Christian and Schultz (2005), we use average rate of return as a proxy for the unobservable marginal rate of return.

While the ISM does not provide definitive income-shifting compliance inferences, it does provide estimates of the aggregate level of profit shifting offshore. The actual rate of return realized by some firms will be higher or lower than what the model predicts in a given jurisdiction, but on average and in the aggregate, the model's estimate of the total level of income shifted is likely to approximate what is actually realized.

B. MNC/CFC Observations

The dataset used for the ISM is comprised of tax returns filed by U.S. corporations as they appear in the e-file and often includes reporting error and data quality exceptions. CFC information returns, in particular, have numerous data quality exceptions that limit the number of usable CFC records for estimation of the ISM. The approach applied in this analysis is to delete CFC records with severe data quality exceptions. Consequently, a large number of CFC records are excluded from the analysis,[6] but enough records remain to provide reasonable inferences about compliant income-shifting.

1. MNC Industry Distribution

We identify 20,337 MNCs with 859,530 CFCs over 2007–2015. We selected only Forms 5471 that were Category 4 filers: U.S. corporations that have control of a foreign corporation (CFC) at some time during the tax year.[7] The final sample consists of 14,428 MNCs with 359,515 CFCs.

Table 1 shows the industry distribution of the parent company MNC over the sample period (for every other year beginning with 2007), based on the NAICS code on Form 1120. We do not have a balanced panel, so the unique EIN column represents the total number of unique MNCs that filed at least one Form 5471.

The industry distribution is similar for the original sample relative to the final sample. The manufacturing industry includes 26.5 percent of the MNCs, and controls 38.7 percent of the CFCs, whereas the professional services industry includes 24.8 percent of the MNCs, but controls only 12.2 percent of the CFCs.

2. Determination of Income-Shifting Database

Table 2 reports the determination of the income-shifting database, showing the number of unique MNCs, the total CFCs across the years, and CFCs within each year. We delete observations for problems with country identification, capital or income measures, tax rate measures, and rate of return measures. The income-shifting variables are defined in Appendix C. For the ISM, we determine the MNC's capital (K_d), income (Y_d), and taxes (T_d) and the CFC's country-aggregated K_f, Y_f, and T_f.

Country. The ISM solves for the equalizing ATROC between the U.S. parent and its CFC on a country-by-country basis. Thus, we must be able to identify the CFC's country. Each Form 5471 reports the country of incorporation (COI) in Box 1(c) and the principal place of business (PPB) in Box 1(e). We define our country variable CNTRY as PPB, or if PPB is missing, we use COI. If COI is also missing, or CNTRY is "US," the observation is deleted. Table 2 begins with the deletion of 1,782 CFC observations (68 MNCs) for which we could not determine the country and 1,826 CFC observations (24 MNCs) that listed the country as the United States. As a result, 92 MNCs no longer were in the database in any year. We delete only the observation with the exception; thus if an observation is deleted because of a CFC variable, the MNC that has other CFCs will continue to be included in the database.

Capital (K) and Pretax Income (Y). K_d is measured as the sum of buildings and other depreciable assets, depletable assets, land, intangibles, and other assets as reported on Form 1120, Schedule L, Lines 10 through 14, for the domestic jurisdiction. K_f is the sum of Form 5471, Schedule F, Lines 8 through 12, for the foreign jurisdictions.

---

[6] The deletion of CFC records from the ISM database due to data quality exceptions creates a fundamental systemic bias in the income shifting estimates. Deleted CFC records constitute nearly one-half of all positive offshore income (Y_f) of CFCs worldwide.

[7] Control for this purpose means the U.S. corporation owns stock possessing: (1) more than 50 percent of the total combined voting power of all classes of stock of the foreign corporation entitled to vote; or (2) more than 50 percent of the total value of shares of all classes of stock of the foreign corporation. As category 4 filers, the Form 5471 must report intercompany transactions on the Schedule M, *Transactions Between Controlled Foreign Corporation and Shareholders or Other Related Persons*.

TABLE 1. MNC/CFC Industry by Year

Industry	Unique		Year					Total	
	EIN	Percent	2007	2009	2011	2013	2015	5471s	Percent
Agriculture, Forestry, Fishing and Hunting	48	0.3%	52	62	99	92	73	750	0.2%
Mining, Quarrying, and Oil and Gas Extraction	220	1.5%	476	439	463	423	339	3,848	1.1%
Utilities	38	0.3%	113	101	103	68	77	833	0.2%
Construction	105	0.7%	313	314	398	394	375	3,097	0.9%
Manufacturing	3,830	26.5%	15,811	14,439	15,779	15,860	15,062	139,242	38.7%
Wholesale Trade	1,218	8.4%	1,952	2,211	2,768	2,628	2,863	22,515	6.3%
Retail Trade	377	2.6%	562	625	695	800	830	6,194	1.7%
Transportation and Warehousing	206	1.4%	505	503	537	609	639	4,992	1.4%
Information	1,204	8.3%	1,842	2,030	2,340	2,534	2,694	20,895	5.8%
Finance and Insurance	424	2.9%	769	717	837	1,188	1,266	9,062	2.5%
Real Estate and Rental and Leasing	155	1.1%	642	684	689	584	584	5,705	1.6%
Professional, Scientific, and Technical Services	3,573	24.8%	4,151	4,437	4,759	5,255	5,768	43,758	12.2%
Management of Companies and Enterprises	2,343	16.2%	8,354	7,977	9,267	10,326	10,893	83,906	23.3%
Administrative and Support and Waste Mgmt	219	1.5%	553	428	750	840	977	6,368	1.8%
Educational Services	55	0.4%	99	132	158	139	110	1,147	0.3%
Health Care and Social Assistance	115	0.8%	103	95	106	75	98	850	0.2%
Arts, Entertainment, and Recreation	71	0.5%	208	233	208	224	245	2,036	0.6%
Accommodation and Food Services	68	0.5%	158	208	177	176	198	1,598	0.4%
Other Services (except Public Administration)	107	0.7%	123	201	282	304	157	1,796	0.5%
Other	52	0.4%	383	164	20	44	39	923	0.3%
Total	**14,428**		**37,169**	**36,000**	**40,435**	**42,563**	**43,287**	**359,515**	

NOTES: Industry based on NAICS (SCHK_BUS_ACTY_CD). See https://www.census.gov/eos/www/naics/.
Agriculture, Forestry, Fishing and Hunting: 111100–123456;
Mining, Quarrying, and Oil and Gas Extraction: 211110–213112; 21230–22100;
Utilities: 221000–221300;
Construction 233110–308700;
Manufacturing: 33400- 3390, 311110–412278;
Wholesale Trade: 421100–439100;
Retail Trade: 45411. 441110–454390, 455110;
Transportation and Warehousing: 481000–493120;
Information: 51112–51900, 511110–519190;
Finance and Insurance: 52312–52590, 521112–529990;
Real Estate and Rental and Leasing: 531100–537900;
Professional, Scientific, and Technical Services: 54119–54519, 541110–547100;
Management of Companies and Enterprises: 551110–55112;
Administrative and Support and Waste Management and Remediation Services: 561100–562211, 562910;
Educational Services: 611000-611710;
Health Care and Social Assistance: 621111–651112;
Arts, Entertainment, and Recreation: 711100–713910;
Accommodation and Food Services: 72215, 721110–722515;
Other Services (except Public Administration: 81121, 811110–818210;
Public Administration: ≥ 920000.

Values for K_d and K_f are required to be greater than zero for ATROC to be defined. Thus, we delete observations when $K_d <= 0$ or $K_f <= 0$. Most of these deletions result from zero assets in the MNC or CFC; thus the entity may be dormant.

Y_d is measured as taxable income reported on Form 1120, Line 30, less foreign-source dividends reported on Form 1120, Schedule C, lines 13-15. Foreign pretax income, Y_f, is measured as the current earnings and profits reported on Forms 5471, Schedule H, Line 5(d), plus foreign tax reported on Forms 5471, Schedule E, Line 8(d).

Because the effective tax rate is not defined when $Y = 0$, we delete observations where Y_d or Y_f is 0. We delete 434,129 CFC observations (855,922–421,793) and 5,636 MNCs (20,245–14,609).

Tax (T) and Effective Tax Rate (ETR). We next delete observations related to tax expense. The tax variable, T, represents the tax liability associated with an entity's taxable income reported in an individual jurisdiction. The value of T impacts the income-shifting estimates because ATROC is an after-tax concept.

The domestic tax variable, T_d, is equal to the sum of total tax reported on Form 1120, line 31, plus the foreign tax credit reported on Form 1120, Schedule J, Line 6(a), multiplied by the ratio of domestic pretax income, Y_d, to taxable income reported on Form 1120, Line 30.[8] The foreign tax variable, T_f, is defined as the income, war profits and excess profits taxes paid or accrued as reported on Forms 5471, Schedule E, Line 8(d). We delete observations where T<0. Because the returns are original, not amended, returns; we cannot explain negative tax expense.

TABLE 2. Determination of Income-Shifting Dataset: MNCs/CFCs, 2007–2015

Item	Total MNCs	Total CFCs	2007	2008	2009	2010	2011	2012	2013	2014	2015
MNCs	**20,337**		**7,053**	**7,575**	**8,151**	**9,119**	**9,872**	**10,426**	**11,103**	**11,972**	**12,571**
CFCs		**859,530**	**89,381**	**92,188**	**91,482**	**92,734**	**96,304**	**97,366**	**98,104**	**99,846**	**102,125**
Dropped obs.											
Missing CNTRY	68	1,782	355	266	217	185	169	141	138	142	169
CNTRY="US"	24	1,826	305	254	196	185	197	203	171	162	153
Remaining	**20,245**	**855,922**	**88,721**	**91,668**	**91,069**	**92,364**	**95,938**	**97,022**	**97,795**	**99,542**	**101,803**
$K_d = 0$	1,311	60,599	8,635	8,647	7,539	5,542	6,013	6,252	6,109	6,022	5,840
$K_d < 0$	38	4,720	617	854	676	350	615	482	573	269	284
$K_f = 0$	3,699	346,568	35,128	36,840	37,432	39,085	39,079	39,157	38,829	40,053	40,965
$K_f < 0$	19	4,398	451	432	361	422	506	481	514	562	669
$Y_d = 0$	493	13,635	1,015	971	1,012	1,794	1,739	1,861	1,676	1,739	1,828
$Y_f = 0$	76	4,209	491	495	412	436	499	461	478	482	455
Remaining	**14,609**	**421,793**	**42,384**	**43,429**	**43,637**	**44,735**	**47,487**	**48,328**	**49,616**	**50,415**	**51,762**
$t_d < 0$	55	29,133	2,002	2,606	3,899	3,400	3,534	2,954	3,054	3,412	4,272
$t_f < 0$	79	16,770	1,590	1,715	2,074	1,856	1,815	1,826	1,976	1,879	2,039
$tr_d > 1$	3	291	61	15	46	22	9	28	62	22	26
$tr_f > 1$	23	10,405	993	1,032	1,035	1,063	1,137	1,217	1,275	1,290	1,363
tr_d, $tr_f < -1$	21	5,667	569	587	571	577	557	663	686	682	775
Remaining	**14,428**	**359,527**[1]	**37,169**	**37,474**	**36,012**	**37,817**	**40,435**	**41,640**	**42,563**	**43,130**	**43,287**
MNCs by year			4,820	5,079	5,271	5,718	6,195	6,492	6,900	7,336	7,628
$atROA_d > 1$	205	3,724	419	583	504	412	323	308	413	398	364
$atROA_f > 1$	1,075	89,053	9,355	9,620	8,256	9,317	10,104	10,138	10,465	10,833	10,965
$atROA_d < -1$	1,623	9,197	915	906	1,022	828	838	1,016	1,067	1,259	1,346
$atROA_f < -1$	1,432	39,454	3,722	4,414	4,142	3,890	4,302	4,497	4,640	4,875	4,972
Remaining	**10,093**	**218,099**	**22,758**	**21,951**	**22,088**	**23,370**	**24,868**	**25,681**	**25,978**	**25,765**	**25,640**
MNCs by year			3,500	3,550	3,742	4,033	4,291	4,391	4,541	4,679	4,766

[1]The sample is further reduced to 359,515 CFCs by a single MNC's 12 CFCs that produced outlier income-shifting estimates despite truncating ATROC.

The domestic effective tax rate is then computed as domestic tax divided by domestic pretax income, or $ETR_d = T_d / Y_d$. The foreign tax rate is computed as foreign tax divided by foreign pretax income, or $ETR_f = T_f / Y_f$. We also delete observations where computed ETR is greater than 100 percent or less than -100 percent. After

[8] Total tax reported on Form 1120, Line 30, is net of the foreign tax credit (see Form 1120, Schedule J, Line 10), so the foreign tax credit is added back to Total Tax in calculating the domestic tax, T_d.

these deletions, the ISM database includes 14,428 MNCs and 359,527 CFCs. The sample is further reduced to 359,515 by a single MNC's 12 CFCs that produced outlier income-shifting estimates despite adjusting ATROC as described below.

After-Tax Return on Capital (ATROC). The ATROC ratio for an MNC in the U.S. or a foreign jurisdiction is defined as (Y-T)/K. We truncate observations where the ATROC is greater than 100 percent or less than -100 percent. An extremely large shifting estimate results when the foreign ATROC is high because of high income in the numerator and low assets in the denominator. The high foreign ATROC may be the result of an investment in a foreign country that has been expensed for tax purposes, but the investment is yielding large returns. If we dropped observations corresponding to ATROCs outside the range, the sample would be reduced to 10,093 MNCs and 218,099 CFCs. We choose to truncate rather than drop to maintain the sample size.

3. Descriptive Statistics

Table 3 reports the mean and median of the variables for each year. As stated before, we do not have a balanced panel of MNCs, even before the deletion process. The final database has 14,428 MNCs, but the number of MNCs ranges from 4,820 in 2007 to 7,628 in 2015. The number of CFCs ranges from 37,169 in 2007 to 43,287 in 2015. The CFC variables are reported both as aggregates across CFCs at the MNC level and at the CFC level. For example, in 2007, 4,820 MNCs invested on average $349 million across all their foreign CFCs, for a total of $1.68 billion. Each CFC has an average of $45 million in capital, for a total of $1.68 billion (37,169 CFCs * $45M).

After-Tax Return on Capital (ATROC). The ATROC ratio for an MNC in the U.S. or a foreign jurisdiction is defined as (Y–T)/K. We truncate observations where the ATROC is greater than 100 percent or less than -100 percent. An extremely large shifting estimate results when the foreign ATROC is high because of high income in the numerator and low assets in the denominator. The high foreign ATROC may be the result of an investment in a foreign country that has been expensed for tax purposes, but the investment is yielding large returns. If we dropped observations corresponding to ATROCs outside the range, the sample would be reduced to 10,093 MNCs and 218,099 CFCs. We choose to truncate rather than drop to maintain the sample size.

Capital Stock (K). For all MNCs, the mean (median) of K_d is $1,653 ($68) million in 2007 and $1,734 ($36) million in 2015. The values of K_f aggregated at the MNC level represent the offshore investment across CFCs for an MNC, decreasing from a mean (median) of $349 ($5) in 2007 to $286 ($2) in 2015. At the CFC level, the mean (median) is $45 ($1) in 2007 and $50 ($1) in 2015.

Pretax Income (Y). For all MNCs, the mean (median) of Y_d is $96 ($0) million in 2007 and $76 ($0) in 2015. The zero medians indicate that losses are reported in the domestic jurisdiction by one-half of the MNCs. The values of Y_f aggregated at the MNC level represent the offshore pretax income across CFCs for an MNC, decreasing from a mean (median) of $47 ($1) in 2007 to $69 ($0) in 2015. For all CFCs, the mean (median) of Y_f is $6 ($0) million in 2007 and $12 ($0) in 2015. The zero median indicates that losses are also reported on average by the CFCs.

Tax (T) and Effective Tax Rate (ETR). For all MNCs, the mean (median) values for T_d is $37 ($0) with a corresponding ETR of 0.11 (0.00) in 2007. The values in 2015 are $30 ($0) with an ETR of 0.13 (0.00). For all CFCs, the mean (median) value of T_f is $2 ($0) in 2007 with a corresponding ETR of 0.14 (0.07). The values in 2015 are $2 ($0) with an ETR of 0.12 (0.01).

After-Tax Return on Capital (ATROC). The ATROC ratio for an MNC in the United States and a CFC in a foreign jurisdiction is defined as (Y–T)/K. Table 3 reports the untruncated results. The mean (median) value of $ATROC_d$ before truncation is -15,906 percent (1 percent) and after truncation is -5 percent (1 percent) (not tabulated). For CFCs, the mean (median) value of $ATROC_f$ before truncation is 36,921 percent (18 percent) and after truncation is 24 percent (18 percent). In 2015, the mean (median) value of $ATROC_d$ before truncation is -78,451 percent (-1 percent) and after truncation is -16 percent (-1 percent). For CFCs, the mean (median) value of $ATROC_f$ before truncation is 812,374 percent (15 percent) and after truncation is 20 percent (15 percent).

TABLE 3. Descriptive Statistics for Input Variables and Key Ratios: Domestic vs. Foreign

[$ amounts in millions]

Variable	2007			2008			2009		
	N	Mean	Median	N	Mean	Median	N	Mean	Median
K_d	4,820	1,652.86	67.72	5,079	1,559.20	61.71	5,270	1,492.62	51.58
K_f-MNC	4,820	348.86	5.05	5,079	279.91	4.15	5,270	312.81	3.79
K_f-CFC	37,169	45.24	1.35	37,474	37.94	1.25	36,000	45.79	1.29
Y_d	4,820	96.22	0.31	5,079	61.49	-0.20	5,270	35.64	-0.38
Y_f-MNC	4,820	46.84	0.83	5,079	0.60	0.53	5,270	56.32	0.32
Y_f-CFC	37,169	6.07	0.39	37,474	0.08	0.34	36,000	8.25	0.24
T_d	4,820	36.76	0.00	5,079	27.65	0.00	5,270	20.71	0.00
T_f-MNC	4,820	16.06	0.18	5,079	17.91	0.13	5,270	10.74	0.07
T_f-CFC	37,169	2.08	0.03	37,474	2.43	0.03	36,000	1.57	0.02
ETR_d	4,820	0.11	0.00	5,079	0.10	0.00	5,270	0.09	0.00
ETR_f	37,169	0.14	0.03	37,474	0.13	0.02	36,000	0.12	0.00
$atROC_d$	4,820	-159.06	0.01	5,079	-18.24	-0.01	5,270	-8.47	-0.02
$atROC_f$	37,169	369.21	0.18	37,474	12,912.55	0.18	36,000	-1,179.36	0.13

Variable	2010			2011			2012		
	N	Mean	Median	N	Mean	Median	N	Mean	Median
K_d	5,718	1,683.84	58.01	6,195	1,562.84	56.98	6,492	1,638.74	49.37
K_f-MNC	5,718	312.95	3.23	6,195	300.96	3.11	6,492	309.63	2.57
K_f-CFC	37,817	47.32	1.32	40,435	46.11	1.30	41,640	48.27	1.32
Y_d	5,718	64.47	-0.05	6,195	57.05	-0.05	6,492	70.73	-0.01
Y_f-MNC	5,718	99.87	0.42	6,195	80.24	0.34	6,492	80.92	0.27
Y_f-CFC	37,817	15.10	0.33	40,435	12.29	0.32	41,640	12.62	0.30
T_d	5,718	27.97	0.00	6,195	26.64	0.00	6,492	29.69	0.00
T_f-MNC	5,718	14.40	0.08	6,195	15.24	0.07	6,492	13.33	0.06
T_f-CFC	37,817	2.18	0.03	40,435	2.33	0.03	41,640	2.08	0.03
ETR_d	5,718	0.14	0.00	6,195	0.14	0.00	6,492	0.15	0.00
ETR_f	37,817	0.13	0.02	40,435	0.13	0.02	41,640	0.13	0.03
$atROC_d$	5,718	-3.46	0.00	6,195	-109.58	0.00	6,492	-10.90	0.00
$atROC_f$	37,817	27,250.51	0.16	40,435	1,776.09	0.16	41,640	294.56	0.15

Variable	2013			2014			2015		
	N	Mean	Median	N	Mean	Median	N	Mean	Median
K_d	6,900	1,612.36	44.99	7,336	1,693.10	39.83	7,628	1,734.38	36.27
K_f-MNC	6,900	302.18	2.20	7,336	287.44	1.81	7,628	285.50	1.60
K_f-CFC	42,563	48.99	1.31	43,130	48.89	1.22	43,287	50.31	1.13
Y_d	6,900	73.43	-0.04	7,336	86.68	-0.08	7,628	75.57	-0.16
Y_f-MNC	6,900	70.85	0.23	7,336	84.71	0.18	7,628	68.86	0.13
Y_f-CFC	42,563	11.49	0.29	43,130	14.41	0.28	43,287	12.13	0.24
T_d	6,900	29.95	0.00	7,336	33.61	0.00	7,628	30.42	0.00
T_f-MNC	6,900	12.16	0.05	7,336	31.54	0.04	7,628	9.24	0.03
T_f-CFC	42,563	1.97	0.03	43,130	5.36	0.03	43,287	1.63	0.03
ETR_d	6,900	0.15	0.00	7,336	0.14	0.00	7,628	0.13	0.00
ETR_f	42,563	0.13	0.02	43,130	0.13	0.02	43,287	0.12	0.01
$atROC_d$	6,900	-14.56	0.00	7,336	-481.44	-0.01	7,628	-784.51	-0.01
$atROC_f$	42,563	4,836.65	0.15	43,130	1,324.64	0.15	43,287	8,123.74	0.15

C. *Income-Shifting Results*

We estimate income-shifting by MNC by country. We then compute the aggregate income-shifting estimates, IS, for each tax year (see Table 4). IS is an estimated $176 billion in 2007 and rises to $274 billion in 2015. The current specification of the N-state ISM does not estimate the amount of cross-subsidiary shifting between countries. Rather, it attributes all income-shifting to transactions directly between the United States and each jurisdiction.

We examine the income-shifting results by segregating the final database by three attributes: (1) whether the CFC is operating in a reporting or nonreporting country; (2) whether the CFC is controlled by an MNC that has a DE in its corporate structure; and (3) whether the CFC is controlled by an MNC that is a CAP taxpayer.

1. Reporting vs Nonreporting Country Analysis

We report income-shifting measures grouped by CFCs operating in reporting and nonreporting countries.[9] Table 4, Panel A, shows the number of CFCs operating in a reporting versus nonreporting country in each year and their corresponding IS, computed at the MNC level as the aggregate of outbound shifting to its CFCs in reporting versus nonreporting countries. Approximately two-thirds of the CFCs operate in a reporting country (227,853/359,515), but account for only 18 percent of the IS ($368,919/$1,995,737).

TABLE 4. Income-Shifting (IS) Estimates

[$ amounts in millions]

Panel A: IS by CFCs Operating in Reporting/Nonreporting Countries

Year	All CFCs		Reporting		Not Reporting		Not Identified	
	N	Total IS	N	IS	N	IS	N	IS
2007	37,169	$176,355	23,790	$24,878	13,107	$146,910	272	$4,567
2008	37,474	202,984	24,016	33,569	13,258	161,201	200	8,213
2009	36,000	171,306	23,011	60,603	12,871	106,227	118	4,475
2010	37,817	252,807	24,149	53,090	13,580	194,397	88	5,320
2011	40,435	255,712	25,742	66,457	14,576	183,168	117	6,086
2012	41,640	210,140	26,311	17,925	15,210	185,485	119	6,731
2013	42,563	220,053	26,658	19,587	15,784	194,359	121	6,107
2014	43,130	232,109	26,926	45,787	16,080	183,461	124	2,861
2015	43,287	274,271	27,250	47,021	15,916	226,963	121	288
Total	**359,515**	**$1,995,737**	**227,853**	**$368,919**	**130,382**	**$1,582,171**	**1,280**	**$44,648**

MNCs can have CFCs in both reporting and nonreporting countries. The ISM determines IS country-by-country, summing the variables for multiple CFCs in a single country. When the Form 5471 provides an unrecognized country code (possible data error), we treat the country as "Not Identified." N is the number of CFCs that remain in the final income-shifting dataset in each year (see Table 2).

Panel B: IS by CFCs With and Without a Disregarded Entity

Year	All CFCs		CFCs with DEs		CFC no DEs		Not Identified	
	N	IS	N	IS	N	IS	N	IS
2007	37,169	176,355	1,790	110,669	35,025	65,419	354	268
2008	37,474	202,984	1,885	141,030	35,192	61,618	397	337
2009	36,000	171,306	1,890	117,232	33,729	53,875	381	198
2010	37,817	252,807	2,157	194,604	35,647	58,212	13	(9)
2011	40,435	255,712	2,324	178,800	38,102	76,915	9	(3)
2012	41,640	210,140	2,566	174,186	39,071	35,952	3	3
2013	42,563	220,053	2,744	187,534	39,816	32,518	3	1
2014	43,130	232,109	2,821	177,018	40,308	55,652	1	(561)
2015	43,287	274,271	3,000	217,515	40,286	56,705	1	51
Total	**359,515**	**1,995,737**	**21,177**	**1,498,588**	**337,176**	**496,865**	**1,162**	**284**

[9] When Form 5471 provides an unrecognized country code (possible data error), we treat the country as "Not Identified."

2. DE/Non-DE Analysis

A relatively easy way to reduce foreign taxes is to shift income from a high-tax foreign entity to a sister low-tax foreign entity through interest or royalty payments. However, this approach was generally ineffective before the "Check the Box" (CTB) regulations became effective in 1997. Because the income earned by the low-tax foreign entity was considered Subpart F income and immediately taxed in the United States, worldwide tax savings were not realized.[10] The strategic use of CTB can eliminate the recognition of Subpart F income on such transactions. By "checking the box" on related tax forms, the two entities are treated as part of the foreign parent CFC for U.S. tax purposes. Thus the *intercompany* payments become *intracompany* transactions, which typically are not subject to Subpart F.

This strategy takes advantage of a mismatch of entity classification between the U.S. and the foreign jurisdictions. The OECD in its "Action Plan on Base Erosion and Profit Shifting" specifically references "country rules that allow taxpayers to choose the tax treatment of certain domestic and foreign entities could facilitate hybrid mismatches" as a way to achieve unintended tax results.[11] Because CTB can reduce the foreign tax rate, it also increases the incentive for MNCs to shift the recognition of what would otherwise be U.S. income to foreign locations.

We assign DE status to the MNC if any related CFC checks the Form 5471, Schedule G box for the question: "During the tax year, did the foreign corporation own any foreign entities that were disregarded as entities separate from their owners under Regulations sections 301.7701-2 and 301.7701-3."

Table 4, Panel B, shows the number of CFCs with at least one disregarded entity and their corresponding IS. Only 6 percent (21,177/359,515) of the CFCs have at least one disregarded entity, but they have 75 percent ($1,498,588/$1,995,737) of the IS.

3. CAP/Non-CAP Analysis

The Compliance Assurance Process (CAP) is an IRS initiative designed to resolve uncertain tax positions in real-time. This research examines whether 221 firms selected for the CAP program have lower shifting measures compared with a sample of 280 non-CAP firms matched on income, assets, and debt equity ratio, using tax return data.

As of the end of 2015, there were 221 firms that have participated in CAP, which started in 2005 with 16 firms. Of the 221 firms, 36 have left the CAP program. Beginning in 2012, the CAP program moved from an invitation-only program to an application program. Initially, taxpayers were invited by the IRS to apply to the CAP pilot program. The IRS required applicants to have assets greater than $10 million, to have audited financial statements, not to be in litigation with the IRS or any federal or state agency, and to display a general willingness to be transparent and cooperative with the IRS.

Table 5, Panel A, reports the aggregate income-shifting estimates, the sums of IS for each tax year for CAP and non-CAP MNCs. Although we have a matched sample of CAP and non-CAP taxpayers, the Table 5, Panel A, results cannot tell us if CAP MNCs shift less income without normalization of the IS measure.

To be able to determine whether the CAP attribute is correlated with IS, we determine the percentage of income shifted for CAP and non-CAP MNCs. We retain the CAP/non-CAP observations where $Y_d > 0$. We normalize IS by year by dividing by pretax income + income shift (Y_d + IS) for each year. For example, if for one year, cumulative Y_d = $500 and IS = $300, then normalized NIS = 300/(500 + 300) = 0.375.[12] Table 5, Panel B contains the results. For example, in 2007, CAP firms shifted 14 percent of income relative to 21 percent by non-CAP firms. CAP MNCs consistently shift less income on a percentage basis than non-CAP MNCs, suggesting firms that participate in CAP appear to be more compliant taxpayers.

[10] Prior to CTB, firms often minimized foreign taxes by using strategies similar to "Commissionaire" structures that limit profits in high-tax jurisdictions, leaving excess profits in low-tax jurisdictions. While there are other legal aspects to these structures, and other structures firms used, they typically earned a small profit. These structures often had undesired business consequences such as VAT.

[11] OECD (2013).

[12] Alternatively, we could have normalized at the MNC level, but adding percents is problematic.

TABLE 5. Normalized IS from CFCs Controlled by a CAP-MNC vs. Non-CAP-MNCs
[$ amounts in millions]

Panel A: IS to CFCs Controlled by a CAP-MNC or Non-CAP-MNC

Year	CAP		Non-CAP	
	N	IS	N	IS
2007	2,794	25,416	5,386	40,563
2008	2,343	20,507	5,121	44,633
2009	2,695	21,772	4,929	40,430
2010	2,781	30,347	5,329	71,558
2011	2,488	24,947	5,193	70,636
2012	2,583	13,440	5,300	53,470
2013	2,666	23,432	5,346	59,491
2014	2,576	15,973	4,922	57,268
2015	2,576	41,779	4,679	48,865
Total	**23,502**	**217,611**	**46,205**	**486,912**

NOTE: 221 CAP firms were matched with 280 non-CAP firms using income, assets, and debt equity ratio from tax return data.

Panel B: Normalized IS to CFCs Controlled by a CAP-MNC or Non-CAP-MNC

Year	CAP					Non-CAP				
	N					N				
	MNCs	CFCs	IS	Yd	NIS	MNCs	CFCs	IS	Yd	NIS
2007	92	1,529	19,576	118,785	0.141	126	2,923	32,471	119,056	0.214
2008	76	998	12,604	81,065	0.135	115	2,544	37,409	103,731	0.265
2009	73	1,084	15,696	78,268	0.167	101	2,278	25,413	84,750	0.231
2010	88	1,356	24,473	106,205	0.187	109	2,716	61,397	106,321	0.366
2011	83	1,248	21,468	80,342	0.211	111	2,884	64,719	103,570	0.385
2012	89	1,388	14,887	101,455	0.128	109	3,049	46,879	110,736	0.297
2013	92	1,510	20,733	122,747	0.145	109	3,124	53,609	126,083	0.298
2014	94	1,468	13,899	139,405	0.091	103	2,831	50,848	149,238	0.254
2015	94	1,395	35,731	131,757	0.213	93	2,630	40,942	132,409	0.236
Total	**781**	**11,976**	**179,066**	**960,028**		**976**	**24,979**	**413,685**	**1,035,894**	

NOTE: We retain the CAP/non-CAP observations where Y_d >0. We normalize IS at the MNC level by dividing by pretax income + income shift (Y_d + IS) for each MNC. For example, if for one MNC in 2007, Y_d = $500 and IS = $300, normalized NISp = 300/(500 + 300) = .375. To avoid adding percents, we sum Y_d within year for CAPs and non-CAPs. In 2007, NIS = 19,576/(118,785 + 19,576) = .141. With the exception of 2007, CAPs consistently shift less income on a percentage basis than non-CAPs.

III. Foreign Tax Information Reporting (FTIR) Data Analysis

A. FTIR Records

This section introduces the use of FTIR data for corporate tax compliance research and explores how this data can be used to draw tax compliance inferences about income-shifting by CFCs. The information shared with the United States is collected by the host countries as part of their regular tax administration processes.

IRS research has demonstrated that tax compliance among individuals improves when they are subject to accurate information reporting by third parties. This study examines whether corporate taxpayers accurately report information to the IRS that is consistent with the information reported to the offshore tax administration authorities. The FTIR data analyzed in this study consists of records that report payments of income sourced in the foreign jurisdiction when the beneficial owner of the reported income is a U.S. taxpayer.[13] The OECD describes the data as follows:

[13] Although there is no legal definition of "beneficial owner" for treaty purposes (http://www.oecd.org/tax/treaties/47643872.pdf), we treat the term as synonymous with "tax resident." To obtain treaty benefits, many U.S. treaty partners require the IRS to certify that the person claiming treaty benefits is a resident of the U.S. for Federal tax purposes. The IRS provides this residency certification on Form 6166, a letter of U.S. residency certification. The IRS procedure for requesting a certificate of residency (Form 6166) is the submission of Form 8802, *Application for United States Residency Certification*. [https://www.irs.gov/individuals/international-taxpayers/form-6166-certification-of-u-s-tax-residency]

> Information which … is typically bulk information comprising many individual cases of the same type, usually consisting of payments from and tax withheld in the supplying State, where such information is available periodically under that State's own system … {… *With the use and application of this information*} compliance is generally improved and fraud can be detected which otherwise would not have come to light. {*emphasis added*}[14]

The FTIR data has approximately 20 million records for the period 2007–2012 from the reporting countries. Each record has the payor name, the income type and amount paid, and the recipient name and type: (1) corporate; (2) individual; (3) unknown (where the recipient type is reported blank or is unknown); and (4) other (a mix of up to 20 other minor recipient categories). Payors are generally business entities operating in the source countries. We focus on payors that are affiliates of U.S. MNCs, including CFCs, DEs, and foreign corporate partnerships (FCPs).[15] The payments made to related corporate recipients are intercompany transactions. Thus, the payments should be reported as expense items on the payor's Form 5471 and as income items on the MNC's return.

Corporate-recipient records account for less than 7 percent of the 20 million records, but include almost half of the $791 trillion income reported during 2007–2012. Of the reporting countries, about two-thirds of the countries report the type of recipient, so we do not capture all corporate recipients.[16]

Table 6 reports the income types and total amounts paid to corporations versus all recipient types over 2007–2012. We focus on payments to corporations of fixed, determinable, annual, or periodic (FDAP) income—specifically interest, royalties, dividends, and capital gains.[17] The FDAP income accounts for 80 percent of the total income reported in the corporate recipient records.

TABLE 6. Corporate Recipient Income for 2007–2012

[$ amounts in millions]

Item	Corporate	All Recipients
FDAP Income		
Dividends	$175,745	$284,712
Interest	62,598	153,852
Royalties	68,489	114,287
Capital Gains	1,881	4,127
Non-FDAP Income		
Other income	46,618	158,743
Income from immovable property	966	3,175
Business profits	14,160	14,925
Missing income type	15,865	57,705
Total Income	**$386,322**	**$791,526**

NOTE: Fixed, determinable, annual, periodic (FDAP) income includes interest, royalties, dividends, and capital gains. The data includes eight other income categories that we classify as "other income."

In next section, we explain how we match recipients to the tax return data, and then identify intercompany transactions. We assume that U.S. beneficial owners of the income sourced offshore will identify themselves to the source country tax administration, usually to claim treaty preferences for tax withholding rates;[18] therefore, we should be able to match to U.S. tax records.

---

[14] OECD and Council of Europe (2011).

[15] CFCs file Form 5471; DEs appear in Form 8858 and FCPs in Form 8865.

[16] Assuming the same distribution of corporate vs. individual records (where the recipient type is known) for the records with an unknown recipient type, corporate records would constitute 70-80 percent of the total income reported.

[17] [http://www.irs.gov/Individuals/International-Taxpayers/Fixed,-Determinable,-Annual,-Periodical-(FDAP)-Income]

[18] The income recipient will claim U.S. tax residence using Form 6166 through submission of Form 8802, the application for U.S. tax residency certification. [https://www.irs.gov/individuals/international-taxpayers/form-6166-certification-of-u-s-tax-residency]

B. Matching Information Records to Tax Return Records

Comparing the FTIR information data to comparable data in the corporate e-file database requires identification of the corporate payor and recipient entities. The FTIR data does not provide a global taxpayer identification number (TIN).[19] Hence, we use quantitative linguistics to match names,[20] which involves identifying all possible entities and affiliates of consolidated corporate taxpayers that file Form 1120 tax returns with the IRS. Appendix B describes the matching process. The entities and affiliates in the e-file are defined as the participating entity list (PEL).[21] We refer to any entity name found in the e-file as a "PELname."

1. Results of Matching Process

Matched Recipients. The recipient matches to an e-file corporate entity are reported in Table 7, Panel A. The recipient matching results provide information about the amount of corporate entity income sourced offshore. Although we match only 29 percent of the recipients, these certain matches account for 69 percent of the income. Our focus in this paper is on the matched payors and their reporting of the payments as expenses on Form 5471.

Matched Payors. The payor matches are summarized in Table 7, Panel B, which shows that out of 47,076 unique payors that make payments to corporate recipients, 7,911 (17 percent) are related payors, which means that if the payor is a CFC, we should be able to match the payment to the payor's Form 5471. Certainly, payors may be affiliates of corporate taxpayers from different countries that pay income to recipients where the beneficial owner is a U.S. (corporate) taxpayer. For example, a payor could be a subsidiary of a French corporation operating in the UK, paying a U.S. entity. But some of the payors will be affiliates of U.S. MNCs (related payors), and those payors will file a U.S. tax return (Form 5471, Form 8858, or Form 8865).

These matched payors account for 55 percent of the income paid to corporate recipients. The related payors make payments to U.S. beneficial owners that average $27M ($212,000/7,911). The unrelated payor remits an average $4.5M to U.S. beneficial owners. Thus, the non-U.S. affiliated payors make lower payments to U.S. corporate recipients.[22]

Of the $212M payments from related payors, 66 percent was paid to recipients that were identified with certainty (not tabulated). We know, however, that the payments were sent to U.S. beneficial owners, whether or not we could match the recipients with certainty. Thus, the payments should be reported on the payor's tax return. If the payor is a CFC, we should be able to match the payment to the expense on the Form 5471. We use the results of the recipient process to help us identify related payors.

Table 7, Panel C. shows the entity form of the related payors. We focus on the $129B that was paid by a CFC that should be reported as an expense on a Form 5471. In particular, 80 percent of royalty payments were made by CFCs ($33,573/$42,011). DEs paid relatively more dividends (42 percent) and interest payments (46 percent); possibly reflecting the use of treaty preferences by the DE entity.

[19] If a source country TIN is reported, it may not be clear if it is a VAT number, a legal codification number (such as for incorporation), or some other designation.

[20] A virtual TIN was developed for this purpose by the IRS. The process involves the solution of a large combinatorial problem that mathematically identifies every possible combination and ordering of name-words for each payor and recipient reported in the information reporting data. See Appendix B.

[21] The PEL is a list of all business entities that are capitalized in the corporate taxpayers' commercial ventures and includes domestic and foreign corporate subsidiaries with both majority and minority positions, domestic and foreign partnerships, disregarded entities, hybrids and reverse hybrids.

[22] Recipients that may be U.S. corporations for unrelated payors may be indirectly related through a foreign parent. That is, the payor has a foreign parent, where the foreign parent is the owner of the recipient entity that is a U.S. corporation. In this instance, there is an indirect relationship between the payor and recipient that turns upon the role of the common foreign parent.

TABLE 7. Matching Results: 2007–2012

Panel A: Payments Received by Matched U.S. Corporate Recipients

Item	Certain	Uncertain	Total
N (unique)	63,921	159,311	223,232
Payments ($M)	$265,649	$117,322	$382,971
Percent of Recipients	28.6%	71.4%	
Percent of Income	69.4%	30.6%	

NOTE: The recipient matching process found multiple potential matches for 223,232 unique recipients, but there were recipients that could not be matched at all. For example, the name on the data record was simply random characters. When there was a possible match, we retained the match with the highest matching score. The certainty of the match is estimated to be 95 percent; thus 5 percent of the "certain" entity matches are not meaningful. These certain and uncertain matches account for $382,971/$386,322 (99 percent) of the payments (in $M) to U.S. corporate recipients.

Panel B: Matched Payor Payments ($M) to All U.S. Corporate Recipients

Item	Related	Unrelated	Total
N (unique payors)	7,911	39,165	47,076
Dividends	$99,910	$75,835	$175,745
Interest	36,388	26,210	62,598
Royalties	42,011	26,478	68,489
Capital Gains	182	1,699	1,881
Other	33,510	44,100	77,610
Total	**$212,001**	**$174,321**	**$386,322**
Percent of Payors	16.8%	83.2%	
Percent of Payments	54.9%	45.1%	

NOTE: Payors that are related to a U.S. MNC account for 17 percent of the payors and 55 percent of the payments to corporate recipients. The related payors make payments to U.S. beneficial owners that average $27M ($212,001/7,911) over the 2007–2012. The unrelated payors remit an average $4.5M to U.S. beneficial owners. Thus, the non-U.S. affiliated payors pay less to U.S. corporate recipients.

Panel C: Related Payor Payments ($M), by Type of Payor Entity

Item	Total	CFC	DE	FCP
Dividends	$99,910	$55,986	$42,383	$1,541
Interest	36,388	17,837	16,767	1,783
Royalties	42,011	33,573	7,628	809
Capital Gains	182	182		
Other	33,510	21,796	9,593	2,121
Total	**$212,001**	**$129,374**	**$76,372**	**$6,255**

2. CFC e-File Details

This section discusses the e-file part of the matching process and uses royalty expense to illustrate. Although we focus on royalty payments, we provide other detail from Form 5471 as background information. Table 8, Panel A, shows the total amounts of the following Form 5471 line items for all CFCs in the reporting and non-reporting countries for 2007–2012:

- CFC income (per books) from Schedule C, line 21;
- Assets (end of period) from Schedule F line 22b;
- Royalty income from Schedule C, line 6b;
- Royalty expense from Schedule C, 11b; and
- Intangible assets (end of period) from Schedule F line 11a+11b+11c.

Relative percentages of global totals for each of these line items is presented in Panel B. For example, in 2007, the reporting countries account for 51 percent of total CFC income, 59 percent of assets, 24 percent of the royalty income, 42 percent of the royalty expense, and 60 percent of intangible assets.

TABLE 8. CFC Form 5471 by Reporting/Nonreporting Countries: 2007–2012
[$ amounts in millions]

Panel A: Form 5471 Line Items by CFCs in Reporting vs. Nonreporting Countries

Category	Item	2007	2008	2009	2010	2011	2012
Reporting	**N**	**41,387**	**43,298**	**43,137**	**43,441**	**45,150**	**45,969**
	CFC income	1,048,722	1,153,273	958,128	1,066,877	1,349,711	1,139,497
	Total assets	8,774,534	8,775,613	6,829,747	7,861,438	8,374,122	8,744,304
	Royalty income	4,753	9,802	9,455	10,780	13,357	11,506
	Royalty expense	15,985	23,245	20,603	20,900	34,417	24,591
	Intangibles	293,831	281,047	325,136	341,586	413,109	405,353
Nonreporting	**N**	**25,055**	**26,099**	**26,559**	**27,557**	**29,144**	**30,063**
	CFC income	988,070	1,104,552	1,078,703	1,346,182	1,304,061	1,541,212
	Total assets	6,136,057	6,400,989	6,840,861	7,558,854	8,282,967	9,531,800
	Royalty income	14,992	26,338	23,841	21,974	37,952	36,974
	Royalty expense	18,713	31,613	30,616	36,006	39,824	47,101
	Intangibles	194,833	202,081	227,007	256,374	310,830	376,609
Other	**N**	**401**	**303**	**162**	**147**	**182**	**194**
	CFC income	21,516	19,985	12,865	14,335	22,254	18,188
	Total assets	56,181	44,739	29,917	33,816	49,730	71,374
	Royalty income	52	1	0	1	53	34
	Royalty expense	3,561	40	13	11	4	7
	Intangibles	1,264	829	322	523	550	1,246
Total	**N**	**66,843**	**69,700**	**69,858**	**71,145**	**74,476**	**76,226**
	CFC income	2,058,308	2,277,811	2,049,696	2,427,394	2,676,025	2,698,897
	Total assets	14,966,771	15,221,341	13,700,525	15,454,107	16,706,819	18,347,478
	Royalty income	19,797	36,141	33,297	32,754	51,362	48,515
	Royalty expense	38,259	54,898	51,233	56,917	74,245	71,699
	Intangibles	489,928	483,957	552,466	598,484	724,489	783,209

Panel B: Reporting Countries' Share of Total CFC Line Items

Category	Item	2007	2008	2009	2010	2011	2012
Reporting	N	62%	62%	62%	61%	61%	60%
	CFC income	51%	51%	47%	44%	50%	42%
	Total assets	59%	58%	50%	51%	50%	48%
	Royalty income	24%	27%	28%	33%	26%	24%
	Royalty expense	42%	42%	40%	37%	46%	34%
	Intangibles	60%	58%	59%	57%	57%	52%

IV. Tax Compliance Inferences From Matched FTIR Data

A. FTIRC Measures

We create two FTIRC measures that are consistent with conventions used in the estimation of the tax gap. The IRS tax gap estimation methodology defines two components of the tax gap:

- **nonfiling gap**—the tax not paid on time by those who do not file tax returns on time; and
- **underreporting gap**—the tax misreported on timely filed returns.

The recipient filing measure (FTIRC1) provides a means to quantify the nonfiling gap, albeit for that portion of the filers that identify themselves as U.S. beneficial owners to a foreign tax administration (FTA). We assume that if the corporate recipient entity reported in the FTIR data takes on the burden of doing so, then that entity should also be reported in the corporate e-file.

The expense reporting ratio (FTIRC2) is characterized as quantifying elements of income underreporting. That is, FTIRC2 is the ratio of payments reported to the FTA relative to the expense reported to the IRS. IRS research has shown that accurate third-party information reporting is a significant determinant of individual income tax reporting compliance. Thus, when we draw compliance inferences using FTIRC2, we assume that accurate third-party reporting of offshore corporate tax positions should also be consistent with corporate tax compliance, especially for offshore income-shifting.

Drawing inferences about compliant tax positions is a conceptual departure from most antecedent tax compliance research that uses evidence from noncompliant taxpayers to draw inferences about tax compliance in the population. We draw on the CAP/non-CAP income-shifting comparisons (see section II-C3) that suggests CAP taxpayers have more tax compliant positions vis-à-vis the positions of corporate taxpayers that are included in the non-CAP sample; but more importantly, CAP firms are very different (i.e., have tax positions that are more compliant) from the population of corporate taxpayers in general.

The two FTIR compliance measures reflect how accurately CFC recipient entities: (1) file returns with the IRS (when the recipient entity is a U.S. tax resident); and (2) report FDAP payments as an expense on Form 5471, Schedule M, Related Transactions Statement, relative to what the CFC reports to the FTA. Thus, we compare and contrast what the payor reports to the foreign tax administration with what the payor reports to the IRS. Then, using the FTIRC2 measure that quantifies the comparisons of the FTA tax data to the IRS tax data, inferences are drawn about compliant income-shifting.

1. FTIRC1 Measure—Recipient Filing

Measurement: We first examine whether the recipient entities of FDAP file a U.S. tax return. The foreign payor provides the FTA information similar to what U.S. payors of FDAP provide to the IRS using Form 1042-S: the recipient, amount, and type of payments. We rely on the virtual TIN from the matching process described in Part III in which recipients are either matched with certainty or do not have a meaningful match to an entity in the corporate e-file.[23] The match logic supporting the virtual TIN assumes that if the recipient has identified itself as a U.S. tax resident (an entity with a beneficial owner that is a U.S. corporate taxpayer) to the FTA, the recipient will also identify itself to the IRS in the corporate e-file. Thus, unmatched recipients suggest potential filing noncompliance: the recipient is either not identifying itself to the IRS or is not identifying itself correctly to the FTA. For example, the recipient may notify the FTA that it is a U.S. tax resident (or there is a U.S. corporate beneficial owner) to claim beneficial tax treaty withholding tax rates, or the payor is providing poor recipient identification to the tax authorities offshore.

FTIRC1 is the ratio of the matched recipients to all recipients for each related payor, weighted by the FDAP payments. Thus, FTIRC1 is the percent of the payments reported to the FTA for recipients matched to the corporate e-file. For example, suppose a related payor has 10 recipients in the FTIR data, and each recipient receives a payment of $100. Six of the recipients are matched to the corporate e-file, and four are unmatched. The FTIRC1 ratio is 60 percent (600/1,000).

As shown in Table 9, the average CAP FTIRC1 measure is 88 percent, meaning 88 percent of all FDAP payments were made to recipients identified in the e-file. The non-CAP average is 75 percent. The difference between the FTIRC1 measure for CAPs and non-CAPs is significant at the 95 percent level of certainty. Thus, CAP-related payors report more matched recipients relative to non-CAP-related payors, suggesting that CAP recipient entities have better filing compliance with the IRS.

[23] Type II errors are also possible if the matching process failed to identify a recipient that does exist in the e-file.

TABLE 9. FTIRC1

Item	CAP	Non-CAP	All Others	Total
N	224	418	3,460	4,102
FTRIC1	0.883	0.754	0.670	0.690

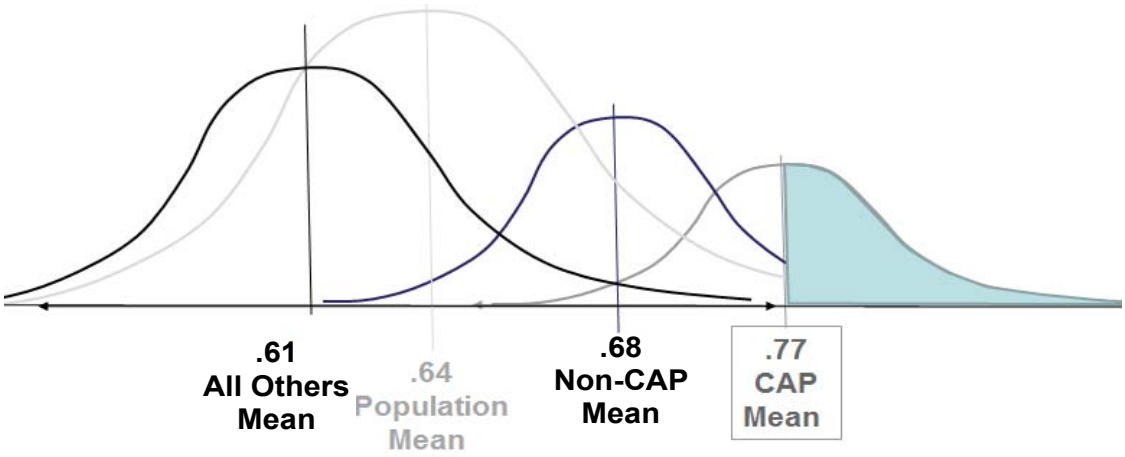
NOTES: FTIRC1 is the ratio of the matched recipients to all recipients for each related payor, weighted by the FDAP payments. Thus, FTIRC1 is the percent of the payments reported to the FTA for recipients matched to the corporate e-file. As shown in Table 9, the average CAP FTIRC1 measure is 88 percent, meaning 88 percent of all FDAP payments were made to recipients identified in the e-file. The non-CAP average is 75 percent. The difference between the FTIRC1 measure for CAPs and non-CAPs is significant at the 95 percent level of certainty. Thus, CAP-related payors report more matched recipients relative to non-CAP-related payors, suggesting that CAP recipient entities have better filing compliance with the IRS.

This table is based on an earlier FTIR dataset for the years 2007–2010. The "All Others" are Large Business & International (LB&I) taxpayers that were not in the CAP/non-CAP sample. We will update the table using the dataset used for the other tables.

We infer that when a CFC's suppliers are U.S. beneficial owners filing with the IRS, the CFC is engaged in intercompany financial transactions that are reported to the IRS. Recipients that claim to be U.S tax residents (are identified as U.S. tax residents) to the FTA in the source country but are not identified in the e-file merit further review.

The compliance inference that is drawn using the FTIRC1 measure is illustrated in Figure 1. The sample means of FTIRC1 for CAP MNCs, matched non-CAP MNCs, and all other corporate taxpayers are significantly different, meaning it is unlikely these separate samples could be drawn from the population at random. Thus, the FTIRC1 measure reflects fundamental differences among these three groups.[24]

FIGURE 1. FTIRC1—Recipient Filing



CAPs, non-CAPs, and All Others have FTIRC1 means that are significantly different. Thus, the groups are not likely to be drawn from the same population. The inference for CAPs is that 77 percent of cross-border FDAP payments are with entities known to the IRS.

2. FTIRC2 Measure—Expense Reporting

The second measure is based on the assumption that accurate third-party reporting is coincident with corporate tax compliance—a relationship demonstrated for individual taxpayer compliance. FTIRC2 is computed for each matched payor for each year as the ratio of expense reported by the CFC to the FTA, relative to the expense reported by the CFC to the IRS. The FTA numerator uses all recipient royalties and interest reported in the FTIR data, whether or not paid to a related party or even matched to an e-file recipient.[25] The IRS

[24] If the population means are not different, then the inference process must be applied for each entity as we do when we compute FTIRC2. That is, each CFC's FTIRC value is compared to the mean for that group, using a 1-tail test.

[25] An alternative measure for FTIRC2 can be computed using data for the FDAP reported to the FTA that is ascribed with certainty to entities in the e-file. This FTIRC2 measure would be lower.

denominator includes royalty and interest expense reported to the IRS on the payor's Form 5471 [Schedule C, *Income Statement* for royalty expense (line 11b) and interest expense (line 12)].[26] If the CFC paid all of its royalties and interest to U.S. beneficial owners, the ratio of expense reported to the FTA should equal the expense reported to the IRS, and FTIRC2 would be equal to 1. CFCs, however, also make payments to entities that are not U.S. beneficial owners, in which case the FTIRC2 will be less than 1.[27]

We evaluate FTIRC2 on a group basis, comparing the payor's ratio to the group's mean ratio computed across years. The group is based on industry and excludes CFCs that own a DE. The FTIRC2 measure could be affected if the CFC makes payments to a related DE. Such payments would be reported to the FTA, but not reported on the Form 5471 because DE payments are eliminated as intracompany transactions.[28]

The average FTIRC2 for royalty expense reported by CAP MNCs and non-CAP MNCs is shown in Table 10, Panel A. For example, the average FTIRC2 for CAPs is 25 percent in 2010. Thus, on average, one quarter of the CFCs' royalty remittances reported to the IRS are also reported to the FTA in the FTIR data. The corresponding FTIRC2 for non-CAP MNCs is 33 percent in 2010.

The average FTIRC2 for interest expense is shown in Table 10, Panel B. For example, the FTIRC2 for CAP MNCs is 1.4 percent in 2007. The corresponding FTIRC2 for non-CAP MNCs is 2.9 percent. Evidently, most CFC interest payments are not paid to U.S. tax residents because only a small percentage of all interest expense reported by CFCs to the IRS are also reported to the FTA in the FTIR data. Additional work is ongoing to develop the compliance measures further. Any results reported in this paper are preliminary.

TABLE 10. FTIRC2—Royalty and Interest Expense, 2007–2012

Panel A: Mean FTIRC2—Royalty Expense

Year	CAP	Non-CAP	All Others
2007	0.192	0.396	0.240
2008	0.240	0.444	0.280
2009	0.262	0.319	0.241
2010	0.248	0.326	0.290
2011	0.247	0.317	0.291
2012		0.402	0.271

Panel B: Mean FTIRC2—Interest Expense

Year	CAP	Non-CAP	All Others
2007	0.014	0.029	0.046
2008	0.011	0.033	0.040
2009	0.032	0.023	0.044
2010	0.020	0.017	0.034
2011	0.023	0.056	0.040
2012		0.009	0.036

NOTE: FTIRC2 is computed for each matched payor for each year as the ratio of payments reported by the CFC to the FTA, relative to the expense reported by the CFC to the IRS on Form 5471 [Schedule C, *Income Statement* for royalty expense (line 11b) and interest expense (line 12). If the CFC paid all its royalties and interest to U.S. beneficial owners, the ratio of expense reported to the FTA should equal the expense reported to the IRS and FTIRC2 would be 1. CFCs, however, also expense payments to entities that are not U.S. beneficial owners, in which case the FTIRC2 will be less than 1.

[26] In future work, we plan to also use Schedule M, *Transactions Between Related Entities*, for royalties paid (line 20, cols. b–f) and interest paid (line 22, cols. b-f).

[27] We determine the measure only for CFC payors filing Form 5471. We could also compute the measure for DE and FCP payors. However, neither the DE, Form 8858, nor the FCP, Form 8865, include a royalty expense line item detail in the reported income statement amounts. Future work could use expenses reported on Form 8858, Schedule M, *Transactions Between Foreign Disregarded Entity of a Foreign Tax Owner and the Filer or Other Related Entities*, and Form 8865, Schedule N, *Transactions Between Controlled Foreign Partnership and Partners or Other Related Entities.*

[28] If the FTIRC2 observations for CFCs with DEs were included with CFCs that do not have DEs, the income consolidation rules imply that the reported FTA expense>IRS expense. Thus, for CFCs with DEs, the permissible range of FTIRC2>1, whereas the range for CFCs without DEs is between 0 and 1. If these two distributions are combined, compliance inferences would be confounded because the two ranges are discrete.

B. Compliant IS

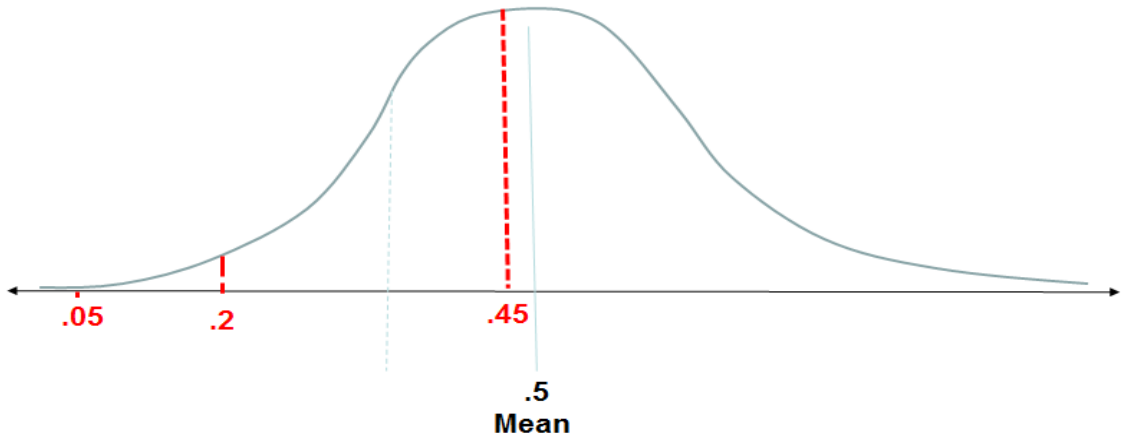
This section describes the process we use to draw compliance inferences about the estimated levels of income-shifting (IS) by CFCs. The ISM provides income-shifting estimates for each CFC, but does not identify whether the income-shifting estimate is compliant or noncompliant. The inferences that are drawn about the compliant portion of the estimated income-shifting levels rely upon the cumulative probability distribution of the CFCs' FTIRC2 measures. While the income-shifting estimates are determined for all CFCs in both reporting and nonreporting countries, the inferences about compliant income-shifting may be determined only for CFCs domiciled in reporting countries. We assume that the FDAP "footprint" depicted by the FTIR data provides an empirical proxy measure for the compliance of the estimated income-shifting.

We reject an inference of compliance when FTIRC2 (FTA expense/IRS expense) of the given CFC is significantly different from the ratio observed in the group of like CFCs (based on industry and DE). We use the FTIRC2 measures to compute a compliance adjustment factor (CAF) at the CFC level. If a CFC's FTIRC2 < mean FTIRC2 for the CFC's group, we reduce the estimated income-shifting amount by (1 – CAF) where CAF = FTIRC2/mean FTIRC2. CAF is first computed for royalties using the royalty FTIRC2 and then for interest expense using the interest FTIRC2. We then compute a weighted average of the two CAFs, where the weights are the relative share of (interest or royalty) expense in the total of interest expense and royalty expense in Schedule C. For example, suppose for a CFC, that Schedule C reports a royalty expense of $60 and an interest expense of $40, and that the royalty CAF measure is 0.4, with the interest CAF measure equal to 0.3. If so, then the overall CAF applied to the income-shifting estimate is 0.36 (= 0.6*0.4+0.4*0.3).

The CAF derivation relies upon the cumulative distribution of the FTIRC2 measure. For example, assume that a CFC has an IS of $100. Further assume the CFC's royalty FTIRC2 is 0.2 and for simplicity no interest expense is reported by the CFC to the IRS. Thus, what the CFC reported to the FTA is only 20 percent of what is reported to the IRS. Suppose this CFC is in a group of comparable CFCs with a mean FTIRC2 of 0.5. Thus, because the FTIEC2 is below the mean, the CAF is 0.2/0.5, or 40 percent.

The example is illustrated in Figure 2. When the FTIRC2 measure is less than the group's mean, the CFC has a ratio of FTA/IRS expense that is less than its peers. Continuing the previous example, the CFC's group reports royalty expense to the FTA that is on average 50 percent of what the group reports to the IRS. The CFC with royalty FTIRC2=0.2, however, reports to the FTA only 20 percent of the expense it reports to the IRS. Thus, this CFC is conducting transactions that are at variance from the group CFCs with the result that the compliance adjustment factor for this CFC based on FTIRC2 is 40 percent, which suggests that the CFCs reported financial transactions are only 40 percent of what is expected for like CFCs.

FIGURE 2. FTIRC2—Expense Reporting



NOTES:

FTIRC2 = .5. The group's average royalty expense reported to the IRS is 50 percent of what is reported in the source country to the FTA. ISp is deemed 100 percent compliant when FTIRC2 >=.5.

FTIRC2 = .45. The CFC's royalty expense reported to the IRS is 45 percent of what is reported in the source country. Because this CFC's FTIRC2 is below the mean, 90 percent (.45/.5) of the ISp is deemed to be compliant.

FTIRC2 = .2. 40 percent (.2/.5) of the ISp is deemed to be compliant.

FTIRC2 = .05. 10 percent (.05/.5) of the ISp is deemed to be compliant.

Table 11 presents the levels of estimated income-shifting that reflect compliant offshore positions of the CFC, where the CFC is in an information reporting country. When a CFC does not have an FTIR record, the mean value of the FTRIC2 measure is assigned to that CFC. The table reports the results for CAP MNCs, non-CAP MNCs, and all other CFCs. The differences are not significant for CAPs, non-CAP, and others. In general, we infer that just under half of the ISp is "compliant." We caution that these results are preliminary, and our research in ongoing.

TABLE 11. Compliant Income-Shifting Estimates in FTIR Countries Based on FTIRC2 for IS>0

Year	IS			Compliant IS			Compliance Adjustment Factor (Percent Compliant)		
	CAP	Non-CAP	Others	CAP	Non-CAP	Others	CAP	Non-CAP	Others
2007	11,085	14,852	48,029	4,842	6,611	21,106	0.437	0.445	0.439
2008	6,287	12,085	60,905	2,962	5,471	26,723	0.471	0.453	0.439
2009	11,275	12,874	55,309	5,338	5,894	24,789	0.473	0.458	0.448
2010	10,113	15,814	61,709	4,785	7,107	28,029	0.473	0.449	0.454
2011	11,613	18,572	68,353	5,326	8,287	29,870	0.459	0.446	0.437
2012	10,590	18,273	48,595	4,973	8,223	21,487	0.470	0.450	0.442

NOTE: We evaluate FTIRC2 on a group basis, comparing the payor's ratio to the group's mean ratio computed across years. The group is based on industry and DE. When a CFC does not have an FTIR record, the mean value of the FTRIC2 measure for that group is assigned to that CFC. We use the FTIRC2 measures to compute a compliance adjustment factor (CAF) at the CFC level. If a CFC's FTIRC2 < mean FTIRC2 for the CFC's group, we reduce the estimated income-shifting amount by (1-CAF) where CAF = FTIRC2/mean FTIRC2. CAF is first computed for royalties using the royalty FTIRC2 and then for interest expense using the interest FTIRC2. We then compute a weighted average of the two CAFs, where the weights are the relative share of (interest or royalty) expense in the total of interest expense and royalty expense in Schedule C. For example, suppose that for a CFC, Schedule C reports royalty expense is $60 and interest expense is $40, and that the royalty CAF measure is .4, with the interest CAF measure equal to .3. The overall CAF applied to the income-shifting estimate is .36 (60*.4+.40*.3).

V. Conclusion

This exploratory research illustrates a potential use of FTIR data to identify compliant income-shifting. We first extend the prior income-shifting model enabling estimation of income-shifting on a CFC/country-by-country basis (N-state model). We then match the FTIR data to e-file data, which allows us to use third-party reporting to examine compliance. We construct two FTIR compliance measures that have foundations in tax compliance related to the filing gap and the income underreporting gap. Whether the FTIRC measures could be useful for tax gap estimation remains open to further study by using enforcement data. The initial specifications are designed to analyze taxpayer compliance and make use of the distinctions between CAP and non-CAP taxpayers to demonstrate the inference framework.

Evidence provided by the FTIRC measures confirms some but not all expectations about taxpayer compliance. First, the income-shifting results suggest that CAP taxpayers have less estimated income-shifting. Second, the recipient entity filing measure (FTIRC1) showed expected differences between CAP, non-CAP and all other filers. The last measure (FTIRC2), however, did not support the conclusion that CAP taxpayers are more compliant. Further study is needed to confirm the attributes used to create the groups.

Our data has measurement errors due to data quality problems in both the FTIR data and the corporate e-file data. Observations from the e-file that have data quality problems have been dropped. Had they not, the ISM solution would require notional adjustments in line items reported by the taxpayer in either the numerator or denominator to reach a reasonable ATROC. The most serious e-file data exceptions have dramatic impacts on the shifting estimates from the ISM. These dropped entities contain CFCs with asset valuation issues, especially for what we expect are intangible assets and may need further scrutiny to address the valuation reporting weaknesses.

We recognize the effects of the asset-valuation problem as being one of the determinants of the data exceptions in the shifting estimates. As noted, CFCs with about half of the offshore income (where CFC income is positive, i.e. excluding CFC NOLs) are affected by the asset-valuation issues. We surmise these asset-valuation issues are likely to have emerged from intangible assets being expensed during development and within cost sharing platforms. Since the income-shifting model depends upon accurate asset valuations, this measurement problem implies that there would be significant limitations for evaluating arms' length transfer pricing since

asset values are pivotal data inputs for transfer pricing computations. Hence, we conclude that intangible expensing creates statistical irregularities for tax administration that severely limit effective enforcement of the arms' length standard using the income-shifting concepts analyzed here, as well as for transfer pricing methods in general. This finding is consistent with conclusions of a National Academy of Sciences Workshop[29] that concluded intangible expensing created challenges for the accurate measurement of investment in financial statements and in the national income and product accounts (NIPA). These investment measures influence Wall Street financial decisions, and are a key data input used by the Federal Reserve.

Perhaps the largest hurdle we had to overcome in this research is creating a virtual TIN using computational linguistics to match FTIR data to e-file data. In the process of recipients confirming to the FTA that they are U.S. residents, the U.S. TIN (EIN for corporations) should be included in the shared FTIR data. The new country-by-country reporting initiative may resolve many, but not all, of the entity identification issues.

[29] National Research Council (2009).

References

Altshuler, R., and H. Grubert. 2005. The Three Parties in the Race to the Bottom: Host Governments, Home Governments, and Multinational Companies. *Florida Tax Journal* 7(3): 154–183.

Black, T. 2012. Quantification of Taxpayer Text Identifiers: ALPHA Function, IRS Office of Research, February, Washington, DC.

Christian, C., and T. Schultz. 2005. ROA-Based Estimates of Income-Shifting by U.S. Multinational Corporations. IRS Research Conference, 2005, Washington, DC.

Collins, J., D. Kemsley, and M. Lang. 1998. Cross-Jurisdictional Income-Shifting and Earnings Valuation. *Journal of Accounting Research* 36(2): 209–229.

Desai, M. A., C. F. Foley, and J. R. Hines Jr. 2006. The Demand for Tax Haven Operations. *Journal of Public Economics* 90(3): 513–531.

Dharmapala, D., and J. R. Hines Jr. 2009. Which Countries Become Tax Havens? *Journal of Public Economics* 93(9–10): 1058–1068.

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: 95–107.

Dyreng, S.D., and B.P. Lindsey. 2009. Using Financial Accounting Data To Examine the Effect of Foreign Operations Located in Tax Havens and Other Countries on U.S. Multinational Firms' Tax Rates. *Journal of Accounting Research* 47(5): 1283–1316.

Dyreng, S. D., and K. S. Markle. 2016. The Effect of Financial Constraints on Income-Shifting by U.S. Multinationals. *Accounting Review* 91(6): 1601–1627.

Grubert, H. G. 2012. Foreign Taxes and the Growing Share of U.S. Multinational Company Income Abroad: Profits, not sales, are being globalized. *National Tax Journal* 65(2): 247–275.

Grubert, H. G. 1998. Taxes and the Division of Foreign Operating Income Among Royalties, Interest, Dividends and Retained Earnings. *Journal of Public Economics* 68(2): 269–290.

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

Jenkins, G. P., and B. D. Wright. 1975. Taxation of Income of Multinational Corporations: The Case of the United States Petroleum Industry. *Review of Economics and Statistics*. 57(1):1–11.

Kopits, G. F. 1976. Taxation and Multinational Firm Behavior: A Critical Survey, *IMF Staff Papers* 23 (3):513–516.

Klassen, K., and S. Laplante. 2012. Are U.S. Multinational Corporations Becoming More Aggressive Income Shifters? *Journal of Accounting Research* 50 (5): 1245–1285.

Manning, C. D., P. Raghavan, and H. Schütze. 2008. *Introduction to Information Retrieval*, Cambridge University Press. [http://www-nlp.stanford.edu/IR-book/].

Markle, K. 2015. A Comparison of the Tax-Motivated Income-Shifting of Multinationals in Territorial and Worldwide Countries. *Contemporary Accounting Research* 22(1): 7–43.

National Research Council. 2009. Intangible Assets: Measuring and Enhancing Their Contribution to Corporate Value and Economic Growth. Washington, DC: The National Academies Press. https://doi.org/10.17226/12745.

OECD. 2013. Action Plan on Base Erosion and Profit Shifting, OECD Publishing. http://dx.doi.org/10.1787/9789264202719-en.

OECD and Council of Europe. 2011. The Multilateral Convention on Mutual Administrative Assistance in Tax Matters: Amended by the 2010 Protocol, OECD Publishing. http://dx/doi.org/10.1787/9789264115606-en.

Rousslang, D. 1997. International Income-Shifting by U.S. Multinational Corporations. *Applied Economics* 29 (71): 925–934.

Salton, G., A. Wong, and C. S. Yang. 1975. A Vector Space Model for Automatic Indexing, *Communications of the ACM*, vol. 18, nr. 11, pages 613–620. [http://www.cs.uiuc.edu/class/fa05/cs511/Spring05/other_papers/p613-salton.pdf].

Wilde, J., and R. Wilson. 2017. Perspectives on Corporate Tax Avoidance: Observations from the Past Decade. Working Paper.

Appendix A

Derivation of the N-State Income-Shifting Model

The income-shifting framework provides a methodology for quantifying the income-shifting activities of U.S. multinational corporations based on the theoretical work of Rousslang (1997). The framework uses a profit optimization model that assumes firms account for the benefits of income-shifting when making investment decisions and, consistent with economic theory, allocate capital so that marginal after-tax rates of return on assets are the same across all members of an affiliated group. Factor price equalization in a single factor context implies that capital is allocated among jurisdictions to realize the greatest overall return. Capital is reallocated from jurisdictions where the returns are lower and moved into jurisdictions where returns are higher, resulting in factor price equalization.[30]

Specifically, the original Christian and Schultz[31] income-shifting two-state solution yields an annual estimate of the direction and magnitude of income shifted between a firm's domestic and foreign operations as a function of the pretax income, tax, and long-term asset values reported in tax records. Because the two-state (domestic and foreign) solution aggregates the rate of return in all observed foreign jurisdictions, it does not distinguish between the amounts of income-shifting attributed to individual foreign tax jurisdictions. The IRS extended the two-state income-shifting framework to reflect a three-state solution (domestic, foreign #1, and foreign #2), so the amounts of income shifted between a firm's domestic operations and each of two foreign super-jurisdictions can be estimated.

The income-shifting framework posits convergence to a single capital factor price across jurisdictions.

If there is equivalence between all three jurisdictions in the rate of return after tax when adjusting for shifting, then:

$$\frac{[Y_d + IS_T] * (1-t_d)}{K_d} = \frac{[Y_1 - IS_1] * (1-t_1)}{K_1} = \frac{[Y_2 - IS_2] * (1-t_2)}{K_2},$$

where:

Y_d = pretax domestic income

Y_1 = pretax foreign jurisdiction 1 income

Y_2 = pretax foreign jurisdiction 2 income

K_d = domestic capital stock

K_1 = foreign jurisdiction 1 capital stock

K_2 = foreign jurisdiction 2 capital stock

t_d = tax rate domestic (T_d/Y_d) where T = level of taxes

t_1 = tax rate foreign jurisdiction 1

t_2 = tax rate foreign jurisdiction 2

IS_T = total income shift [$IS_T = IS_1 + IS_2$]

IS_1 = income shift jurisdiction 1

IS_2 = income shift jurisdiction 2.

[30] While the ISM does not provide compliance inferences for the tax transactions to shift income, it does provide estimates of the aggregate level of profits shifting offshore. The actual rate of return realized by some firms will be higher or lower than what the model predicts in a given jurisdiction, but on average and in the aggregate, the model's estimate of the total level of income shifted is likely to approximate what is actually realized. Firms expect higher-than-normal returns in foreign jurisdictions to compensate for the added risks. The ISM accounts for the risk in an underwriting framework and aggregates the total returns across all firms to provide accurate income shifting estimates.

[31] Christian, Charles and Thomas Schultz, "ROA-Based Estimates of Income Shifting by U.S. Multinational Corporations," IRS Research Conference, 2005, Washington, DC.

This paper extends the three-state model to an N-state model as derived below. Assuming factor-price equalization in a single factor framework, it follows that the capital price will converge across jurisdictions for each corporation.[32] That is, a corporation will allocate capital resources such that the rate of return to capital will be equal between any two jurisdictions (jurisdiction "i" and jurisdiction "j"). In general form, showing the effect of income shifted between the jurisdictions is as follows:

$$(Y_i - X_i) * \alpha_i = (Y_j - X_j) * \alpha_j \quad [1A]$$

where:

$$\alpha_i = (1 - t_i) / K_i \quad [1B]$$

$$\alpha_j = (1 - t_j) / K_j. \quad [1C]$$

Then, solving for the level of corporate income shifted in jurisdiction "i",

$$X_i = Y_i - (Y_j - X_j) * \alpha_j / \alpha_i. \quad [1D]$$

Now, including the domestic jurisdiction in the expression, the rate of return in the domestic jurisdiction is equal to the rate of return in any jurisdiction "j". That is, when accounting for the corporate income shifted offshore, and adjusting both the domestic and offshore jurisdictions for the corporate income shifted out of the domestic jurisdiction and into the offshore jurisdiction, we obtain:

$$(Y_d - X_d) * \alpha_d = (Y_j - X_j) * \alpha_j \quad [1E]$$

where:

$$\alpha_d = (1 - t_d) / K_d, \quad [1F]$$

and where the total income shifted out of the domestic jurisdiction (IS or Xd) is equal to the sum of the (negative) income shifted in each of the "N" offshore jurisdictions (and the domestic jurisdiction "d" is defined for j=1):

$$IS = X_d = - \sum_{j=2}^{N} X_j. \quad [1G]$$

The domestic pre-tax corporate income plus the corporate income shifted is equal to the terms on the right hand side of equation [2] below:

$$Y_d + \sum_{j=2}^{N} X_j = (Y_j - X_j) * \alpha_j / \alpha_d. \quad [2]$$

Solving equation [2] for the income-shifting term (Xj) provides the general form of the income-shifting in any jurisdiction "j" where X_1 is X_d (the total shifting from/to the domestic jurisdiction) and $X_1 = X_d = - \sum_{i=2}^{N} X_i$. Then,

$$X_2 = [(\alpha_2 + \Sigma_{i=3}^{N} \alpha_1,\alpha_2/\alpha_i) * Y_2 - \alpha_1 * Y_1 - \alpha_1 * \Sigma_{i=3}^{N} Y_i] / (\alpha_1 + \alpha_2 + \Sigma_{i=3}^{N} \alpha_1,\alpha_2/\alpha_i) \quad [3A]$$

and for any CFC in the jurisdiction "j":

$$X_j = Y_j - \alpha_2/\alpha_j * Y_2 + \alpha_2/\alpha_j * X_2, \quad [3B]$$

[32] We assume that the shifting of income from/to jurisdiction $\mathbf{X_i}$ and $\mathbf{X_j}$ will be reflexive; meaning that $\mathbf{X_i} = - \mathbf{X_j}$ such that the income shifted out of one jurisdiction must equal the income shifted into the other jurisdiction in the 2-state case. The N-state condition is found in expression 1G.

where the total income-shifting to/from the domestic jurisdiction (X_d=X_1) is:

$$X_1 = X_d = -\sum_{i=2}^{N} X_i \quad . \qquad [3C]$$

Equation [3C] shows the general form of the solution for the level of corporate income shifted into (or out of) any jurisdiction "j". The sign of the data in the solution for Xj reflects whether the income enters or leaves the jurisdiction. If the sign is negative, income is shifted into the jurisdiction. If the sign is positive, then income is shifted from this jurisdiction.

Aggregating the income shifted across all jurisdictions, the total level of income-shifting (IS) in the domestic jurisdiction (restating equation 3C) is:

$$\text{IS: } X_d = -\sum_{i=2}^{N} X_i \quad . \qquad [4]$$

Appendix B

Matching Foreign Tax Information Reporting Data to the Corporate e-File Using Computational Linguistics

This appendix provides an overview of a process to match foreign tax information reporting (FTIR) records to IRS corporate tax return records. There is no global TIN available, and thus the matching of payor and recipients to a corporate tax return is done using name matching.

The corporate tax return names are drawn from the participating entity list (PEL). The PEL is a list of all business entities in the corporate taxpayers' commercial ventures and includes domestic and foreign corporate subsidiaries with both majority and minority positions, domestic and foreign partnerships, disregarded entities, hybrids and reverse hybrids. The tax return names and the information reporting names are subject to pre-process to recognize the data quality in the entity name and provide greater uniformity in the text presentation and text format. Two different areas of data quality affect the matching process: blank names and semblants. Blank recipient names and blank payor names are found in the information reporting records. When an information reporting record has a blank payor or recipient name, there is no match logic that can be applied, and the record is deleted.

Semblants are minor differences in entity names that reflect the inconsistent use of capitalization, punctuation, and word abbreviations. Semblants can also reflect misspellings of words, but only minor differences in capitalization, punctuation and abbreviations are included.

A. Recipient Matching Methods

Recipient names are matched to PEL names using a two-step process. The first step finds all possible match candidates and the second step assigns a rank or certainty to select the "best" matching candidate name. All possible candidate names are determined by examining every combination of the name-words in a name reported in the information reporting data and comparing them to every combination of name-words in the list of PELnames. The ranking or certainty of the "best" match, among all the PELname candidates for a name in the information reporting data, is computed using the vector scoring model.[33]

The number of combinations is vast and a direct solution, representing a match, is computed using an analytical approach from quantitative linguistics that develops numeralizations of name-words using integer mapping of character sets. Each word in the name from the information reporting data and PELname is converted using a numeralization into a unique (and invertible) number. These name-word numeralizations are input to a mathematical function to compute a very large integer.[34] This very large integer number is essentially unique to each name.[35] A PELname candidate is found for a recipient name using the numerical solution where the functional value for the word combinations of a PELname candidate is equal to the functional value for the word combinations of a name reported in the information reporting data. The existence of a solution implies that one or more name-words in the PELname candidate are common to the name-word(s) in the name reported in the FTIR data.

The PELname candidate name-word combinations for the name-word combinations of a name reported in the FTIR record identifies name-words that match for some of the words in the names but, not necessarily all the name-words. That is, some but not all of a PELname candidates' name-word may match only some of the name-words in a name in the information reporting data record. In other words, it is a partial match. The numeralizations of the name-words ensure that any ordering of a given combination of name-words can be found in the functional solution. For example, suppose a name is composed of three words (A B C with sepa-

[33] See Manning, *et al.* (2008).

[34] The functional value is the multiplicative product of the numbers corresponding to the name-words. Each name-word number represents the numeralizations of each name-word. The functional value is a very large number and numerical precision is needed for all digits to make comparisons for matching. The functional values have about 100 to 150 digits (or more), and keeping computational precision requires special software routines to increase the number of registers used in the computations as most software programming languages lose mathematical precision for numbers larger than 10^{30}.

[35] Black (2012).

rating blanks), denoted for simplicity where each word is a capital letter. The number of possible combinations is 2N - 1 with N=number of name-words in the name and here N=3 with 7 combinations: (1) ABC, (2) A B, (3) A C, (4) B C, (5) A, (6) B, (7) C.

The ordering of the name-words does not matter because the function (using the name-word numeralizations) ensures an identical solution for any word order; the function is a multiplicative product of the numeralizations and multiplicative products are transitive. For instance, if the numeralizations of the name-words are (with simulated numerical values): A=10, B=27 and C=39, then function value equals the multiplication product of the three name-word's numerical values A*B*C=10530. Further, the product is the same regardless of the name-word order:

$$A^*B^*C = A^*C^*B = B^*A^*C = B^*C^*A = C^*A^*B = C^*B^*A = 10530,$$

for every possible ordering of the three-word name. This numeralization feature of name-word order equivalence greatly reduces the computational problem, compared to text methods, such as regular expressions or boulean equivalence.[36] In addition, transforming the matching into a numerical domain vastly reduces the memory required by each processing instruction since each digit in the numerical value of the solution of the comparator function can attain 10 values, while a single character in text has 256 potential values (EPCDIC) for standard character sets and 512 or more for extended character sets, such as those using special characters found in foreign languages used in the information reporting data record reported by the source information reporting country's tax administration.

The resulting set of PELname candidates for each recipient name is then ranked to select the "best" PELname candidate for each recipient. A PELname candidate detected in the first step of the matching process has one or more name-words that are common with the name-words in the name from the information reporting data record. The vector space model is used to quantify an index similar to what is used in Internet search engines.[37] The index reflects both the similarity and the dissimilarity of the PELname candidate and the recipient name. In general, the index is inversely correlated with the frequency of use of the name-words. Therefore, matching a word like "corporation", which is very common in the information reporting data names and PELnames, has far less consequence than for an uncommon word in an affiliate or entity name like Arapahoe or Stryker.

The resulting "best" PELname candidate for each recipient is further assigned a certainty measure that provides a statistical inference about the meaningfulness of the "best" match. In other words, not all "best" matches are meaningful. Assume a recipient entity or affiliate (a PELname) of a U.S. consolidated corporate taxpayer that earns income in a source information reporting country is properly identified as such (i.e., a U.S. corporate recipient) to the tax administration in that information reporting country. If so, then that recipient entity likely will be reported in the corporate tax return filed by the U.S. consolidated corporate taxpayer. The distribution of all PELname candidates, using the vector score of the PELname candidates as a random variable, is used to define the certainty of the match. Certainty is defined as those PELname candidates with vector scores that are more than two standard deviations above the mean (upper one-tail test statistics).

B. Payor Matching Methods

The payor names are also matched to PEL names to identify intercompany transactions. PELname candidates for each payor in the information reporting data are first derived using a two-step process similar to that used to match the recipients. However, the PELname list is restricted to entities and affiliates reported on three forms: Form 5471 for CFCs, Form 8858 for disregarded entities (DEs), and Form 8865 for foreign corporate partnerships (FCPs). All other PELnames are excluded from the list of potential PELname candidates in the

[36] The number of name-word orderings is N! where N=number of name-words in the name reported in the information reporting data or the number of name-words in the PELname. Since the average number of name-words is about 5 in a name, the numeralization and computational function reduces the number of comparator operations by a factor of about 1/N!=1/120—not to mention the efficacy of the computer memory required for the numeralizations instead of text character comparators.

[37] Salton, *et al.* (1975).

matching process. In addition, the number of unique payor names is much smaller than the number of unique recipient names because each payor may report many recipients. However, the average number of name-words of payor names is much larger than the average recipient name. Consequently, the size of the combinatorial problem for payors is actually larger than the combinatorial universe for recipients.

Once the PELname candidates for each FTIR payor have been selected, an additional processing step is applied to determine the "best" PELname payor candidate that is a meaningful match, using information about associated entities of the payors.[38] We expect that the payors and at least some of the recipients of that payor will be affiliated with the same consolidated corporate tax group reported in the U.S. tax return.

This condition of the associated entities suggests that of all the payor PELname candidates found for a payor name, the "best" match that is meaningful is that payor PELname candidate with a group EIN that also matches one or more of the group EINs for the matched PELname for the corresponding recipient. In other words, a related payor will have recipients that are also part of the same consolidated corporate tax group, and the payor and recipients have the same consolidated group EIN reported in the corporate tax return database that is found in the matching process.

Each FTIR record reports both the payor and the recipient. The PELname candidate for the payor includes the group EIN, and that is then compared to each of the EINS of the (previously) best matched PELname recipients. The "best" payor PELname candidate has a group EIN (of the consolidated tax group) that is common to one or more of the "best" match recipient PELnames. This condition demonstrates the existence of intra-company financial transactions between affiliated entities within the consolidated corporate tax group. When the matched EINs are found, this is deemed to identify the "best" PELname match that is meaningful for the payor name, and we conclude that this payor/recipient record is an intercompany transaction.

[38] The distribution of vector scores (VS) for payors candidates lacks properties of normality that would allow drawing conventional statistical inferences about meaningfulness. The recipient VS distribution has normality and 95%+ certainty can be associated with the properties of the normal distribution. However, because most payors do not have U.S. tax nexus, the payor candidate VS distribution would not be expected to have the same qualities.

Appendix C
Income-Shifting Variables

K_d Domestic capital stock is measured as the sum of buildings and other depreciable assets, depletable assets, land, intangibles, and other assets as reported on Form 1120, Schedule L, Lines 10 through 14. K_d must be greater than zero to enable calculation of return on assets.

K_f Foreign capital stock is measured as the sum of buildings and other depreciable assets, depletable assets, land, intangibles, and other assets as reported on Forms 5471, Schedule F, Lines 8 through 12. K_f must be greater than zero in all foreign jurisdictions to enable calculation of return on assets.

Y_d Domestic pretax income is measured as taxable income reported on Form 1120, Line 30 (TXBL_INCOM), less foreign-source dividends reported on Form 1120, Schedule C (Line 13: SCHC_OTH_FRGN_DIV, Line 14: SCHC_CNTRL_FRGN_INCM, Line 15: SCHC_FRGN_DIV_GR_UP).

Y_f Foreign pretax income is measured as the current earnings and profits reported on Forms 5471, Schedule H, Line 5(d), plus foreign tax reported on Forms 5471, Schedule E, Line 8(d).

T The tax liability associated with an entity's taxable income reported in an individual jurisdiction. The domestic tax variable, Td, is equal to the sum of total tax reported on Form 1120, line 31, plus the foreign tax credit reported on Form 1120, Schedule J, Line 6(a), multiplied by the ratio of domestic pretax income, Y_d, to taxable income reported on Form 1120, Line 30.

T_f The foreign tax variable, T_f, is equal to the sum of total income, war profits and excess profits taxes paid or accrued as reported on Forms 5471, Schedule E, Line 8(d).

ETR td; tf The effective tax rate, ETR, is calculated as T/Y by taxpayer by jurisdiction by year. Prior to calculating the income-shifting estimates, we truncate t at the EIN-level as follows: when $t > 1$, then t is set to 1; when $t < 0$, then t is set 0. The domestic tax rate is computed as domestic tax divided by domestic pretax income, or $ETR_d = T_d / Y_d$. The foreign tax rate is computed as foreign tax divided by foreign pretax income, or $ETR_f = T_f / Y_f$.

CNTRY Two-character identifier used to indicate the foreign jurisdiction associated with each filed Form 5471. CNTRY is defined as the principal place of business (PPB) reported in Box 1(e) on Form 5471, or if PPB is missing, CNTRY is the country of incorporation (COI) reported in Box 1(c) when known.

fyear We define TXY as the first four digits of TX_PRD_END_DR, an 8 digit variable, e.g., 20121231, FYR is the next two digits, the ending month of the fiscal year. We then define FYEAR as TXY unless FYR <6; then FYEAR = TXY -1. Thus, FYEAR agrees with Compustat's FYEAR.