

APPENDIX A: Additional Results

Table A1. Aggregate Income by Owner Type

		Interest		Dividends		Gross Proceeds/Redemptions		Other income	
		Total (Billions USD)	Share	Total (Billions USD)	Share	Total (Billions USD)	Share	Total (billions USD)	Share
Matched	Partnership	3.03	23.0%	13.23	46.4%	90.83	33.0%	132.22	63.5%
	Individual	2.80	21.2%	6.83	23.9%	49.51	18.0%	26.02	12.5%
	C Corporation	1.11	8.4%	0.80	2.8%	8.78	3.2%	11.10	5.3%
	Tax exempt entity	0.30	2.3%	0.29	1.0%	2.76	1.0%	4.38	2.1%
	Trust	0.14	1.1%	1.00	3.5%	3.36	1.2%	6.74	3.2%
	Foreign corporation	0.11	0.8%	0.03	0.1%	2.64	0.9%	0.55	0.2%
	S corporation	0.03	0.2%	0.10	0.3%	0.36	0.1%	1.03	0.4%
Unmatched	Missing TIN	3.27	24.8%	3.52	12.3%	91.66	33.3%	13.31	6.4%
	Unmatched entity	0.75	5.6%	0.39	1.3%	11.70	4.2%	4.06	1.9%
	Ambiguous match	1.07	8.1%	1.71	6.0%	6.72	2.4%	2.29	1.1%
	Unmatched TIN	0.48	3.6%	0.51	1.8%	5.29	1.9%	6.16	2.9%
	Unmatched individual	0.05	0.4%	0.05	0.1%	0.82	0.3%	0.11	<0.1%

Note: This Table reports owner shares weighted by income. We observe that most income types are similarly or slightly more concentrated in partnerships, compared to the wealth shares in Table 3. See Table 3 of the main text and the surrounding text for details on the construction of the estimates and group definitions.

Table A2 – Tax Year 2018 Totals: Haven versus Non-Haven Countries

	All	Havens	Non-Havens
No. of accounts	4,566,774	612,406	3,954,368
Total Account Balance (Billions USD)	3,981.8	1,939.8	2,042.0
Total Interest (Billions USD)	13.2	4.7	8.5
Total Dividends (Billions USD)	28.5	18.0	10.5
Total Gross Proceeds/Redemptions (Billions USD)	274.5	132.7	141.8
Total Other Income (billions USD)	208.0	173.7	34.3

Note: This table decomposes the total number of accounts and financial aggregates into totals for haven and non-haven countries. We observe that 14% of accounts and 49% of wealth is located in haven countries.

Table A3 – Total Values by Country Type – TY 2018

	Country type			
	Haven	Non haven	Non-haven – prior EOI	Non-haven – no prior EOI
No. of accounts	612,406	3,954,216	3,222,800	731,416
Total Account Balance (Trillions USD)	1.9	2.0	1.7	.2
Total Interest (Billions USD)	4.6	8.5	5.9	2.5
Total Dividends (Billions USD)	17.9	10.5	9.4	1.1
Total Gross Proceeds/Redemptions (Billions USD)	132.6	141.8	107.0	34.7
Total Other Income (Billions USD)	173.7	34.2	14.6	19.5

Note: This table reports the same statistics as Table A2 and further decomposes non-haven countries based on whether that country's government agreed to exchange information with the US prior to FATCA, under TIEAs and/or DTCs, which we label Prior Exchange of Information (EOI). The full set of country groups is as follows: Haven countries: Anguilla, Antigua and Barbuda, Aruba, Bahamas, Bahrain, Belize, Bermuda, Caribbean Netherlands, Cayman Islands, Cook Islands, Curacao, Cyprus, Dominica, Grenada, Guernsey, Hong Kong, Isle of Man, Jersey, Liechtenstein, Luxembourg, Malta, Mauritius, Monaco, Panama, Saint Kitts and Nevis, Saint Lucia, Singapore, Sint Maarten, Switzerland, Turks and Caicos Islands, United States, and British Virgin Islands. Non haven and prior EOI countries: Australia, Belgium, Canada, China, Czech Republic, Denmark, Estonia, Finland, France, Germany, Hungary, Iceland, India, Indonesia, Italy, Japan, Latvia, Lithuania, Mexico, New Zealand, Norway, Poland, Portugal, Slovakia, Slovenia, Republic of Korea, Spain, Sweden, Trinidad and Tobago, and the United Kingdom. All other countries are non-haven, without prior EOI.

Table A4 – Joint Distribution of Owner Type and Country – TY 2018

Panel A. Number of Accounts by Owner Type and Country

		Country type				Row Total
		Haven	Non Haven	NH – EOI	NH – No EOI	
Owner type	Individuals	403,909	2,262,500	1,705,589	556,911	2,666,409
	Partnership	42,592	19,669	15,024	4,645	62,261
	Other entity	22,241	31,548	24,609	6,939	53,789
	Unknown	143,383	1,639,170	1,477,358	161,812	1,782,553
Column Total		612,125	3,952,887	3,222,580	7,303,07	4,565,012

Panel B. Total Account Balance by Owner Type and Country (Billions USD)

		Country type				Row Total
		Haven	Non Haven	NH – EOI	NH – No EOI	
Owner type	Individuals	412	212	145	67	625
	Partnership	1,001	287	228	58	1,288
	Other entity	291	255	210	44	546
	Unknown	231	1,265	1,196	68	1,497
Column Total		1,937	2,020	1,781	239	3,957

Note: This table decomposes the total number of accounts (Panel A) and total wealth (Panel B) according to the country groups from Table A3 and the owner type categories from Figure 1. To facilitate the computation of shares, we also include row and column totals in each panel.

Table A5. Rates of Return in Selected Sub-Populations: Additional Specifications

	Sample restriction						
	All	Non-haven country	Haven country	Individual owners	Partnership owners	Other entity owners	Unmatched owners
Total reported wealth (billions USD)	3,982	2,042	1,940	626	1,292	279	1,510
Share of wealth with reported interest	33.3%	44.0%	22.0%	27.3%	16.9%	36.9%	43.9%
Quasi-rate of return: interest only	1.0%	0.9%	1.1%	1.7%	1.4%	0.7%	0.8%
Share of wealth with reported interest or dividends	37.7%	51.3%	23.4%	33.6%	19.8%	37.7%	49.9%
Quasi-rate of return: interest + dividends only	2.8%	1.8%	5.0%	4.6%	6.4%	1.1%	1.6%
Share of wealth with reported int., div., or gross proceeds	44.1%	55.0%	32.5%	36.5%	33.3%	38.7%	53.0%
Quasi-rate of return: int. + div. + 0.14*gross proceeds	4.6%	3.5%	6.5%	7.4%	6.7%	2.6%	3.5%
Share of wealth with reported int., div., G.P., or other income	59.2%	57.2%	61.5%	45.5%	67.2%	44.5%	56.3%
Quasi-rate of return: int. + div. + 0.14*(GP + other)	4.6%	3.7%	5.5%	7.2%	5.5%	2.7%	3.7%

Note: This table reports alternative specifications to Table 4 in the main text. The second set of estimates, where the quasi-rate of return includes interest and dividends only, is identical to Table 4. The other rows use alternative definitions of the rate of return. We observe that the higher rate of return for partnerships and individuals, and the higher return in havens than non-havens, both obtain for essentially any definition of the quasi-rate of return.

Table A6 – Partnership account owners, assets and income by industry

	N account owners	Account Balance (Billions)	Interest (Billions)	Dividends (Billions)	Gross Proceeds (Billions)	Other (Billions)
Total						
Financial Services	15,754	910.00	1.90	7.58	58.40	102.50
Other finance	4,255	247.80	0.50	2.63	14.64	12.34
Other industries	1,647	13.27	0.04	0.09	0.50	1.30
Unknown	5,045	130.20	0.64	2.93	17.40	16.61
Non-haven						
Financial Services	4,081	204.00	1.04	0.91	22.40	8.00
Other finance	1,480	54.80	0.30	0.70	11.60	1.74
Other industries	1,155	5.87	0.01	0.02	0.33	0.14
Unknown	2,088	27.20	0.43	0.11	4.70	3.31
Haven						
Financial Services	11,673	706.00	0.86	6.67	36.00	94.50
Other finance	2,775	193.00	0.20	1.93	3.04	10.60
Other industries	492	7.40	0.03	0.08	0.17	1.16
Unknown	2,957	103.00	0.21	2.82	12.70	13.30

Note: This table shows asset and income information for partnerships for year 2018 reported on FATCA Form 8966 by industry and whether the account was held in a haven or non-haven country. It is categorized by the largest industry groups in each country category. The table shows the total number of 2018 partnerships reported on Form 8966 for each industry, total assets held and total income by source.

Table A7 – Allocation of Partnership Foreign Assets to Taxable Owners

	share of owners raw	share of owners assigned	share of assets raw	share of assets assigned
Individual	88.6%	90.1%	43%	54%
Unclassifiable	1.7%	0.0%	20%	0%
Foreign				
Individual/Entity	0.1%	0.1%	10%	13%
Tax Exempt	0.9%	0.9%	9%	11%
Trust	8.0%	8.2%	9%	11%
Foreign Corporation	0.2%	0.2%	5%	7%
Corporation	0.6%	0.6%	4%	4%

Note: Table A7 reports the shares of pass-through owners and allocated assets by owner type, the values underlying Figure 3.

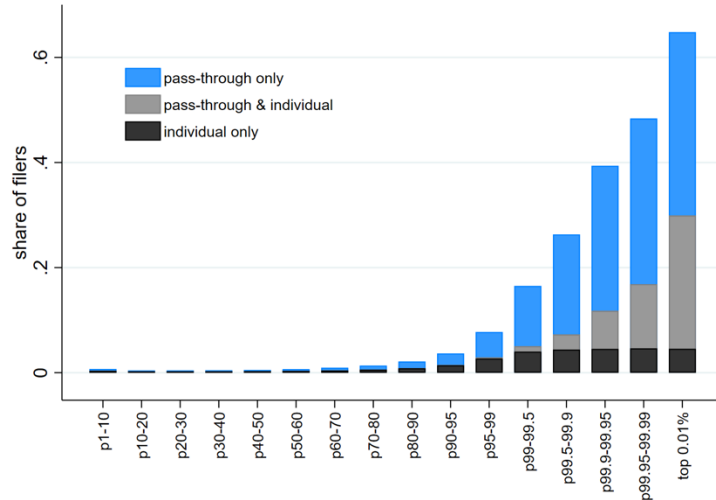
Table A8 – Allocated foreign assets by position in income distribution (2018)

Position in AGI Dist.	Haven Individual (Billion USD)	Haven Pass-through (Billion USD)	Non-haven Individual (Billion USD)	Non-haven Pass-through (Billion USD)	Total (Billion USD)
losses	12.4	8.8	5.9	2.8	29.9
p2-p10	41.2	2.7	19.5	2.5	65.9
p10-p20	5.3	2.8	8.3	0.4	16.8
p20-p30	11.6	6.6	3.7	4.5	26.5
p30-p40	9.1	7.6	3.4	0.9	21
p40-p50	2.5	4.7	4.3	2.2	13.7
p50-p60	17.1	2.9	10.6	0.8	31.5
p60-p70	2.8	8.3	6.8	1.3	19.1
p70-p80	5.5	5.6	9.1	1.8	21.9
p80-p90	7.4	10.9	16.9	6.5	41.7
p90-p95	31	5.6	12.1	6.9	55.5
p95-p99	19.4	13	24.2	4.7	61.3
p99-					
p99.5	12.8	9.2	11.5	27.4	60.8
p99.5-					
p99.9	71.3	37.3	19	11.8	139.4
p99.9-					
p99.99	67.7	77.7	17.5	22.2	185.1
p99.99-					
p100	66.5	192	12.1	66	336.7

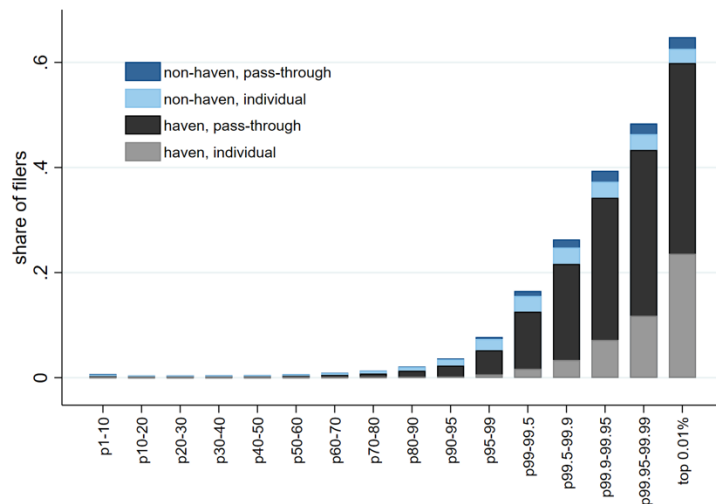
Note: Table A8 shows the asset values underlying-Figure 5.

Figure A1: Share of taxpayers with a foreign account reported by FATCA by percentile – Distribution of Total Positive Income (TPI), TY2018

Panel A: Share of taxpayers holding accounts directly and/or through pass-throughs



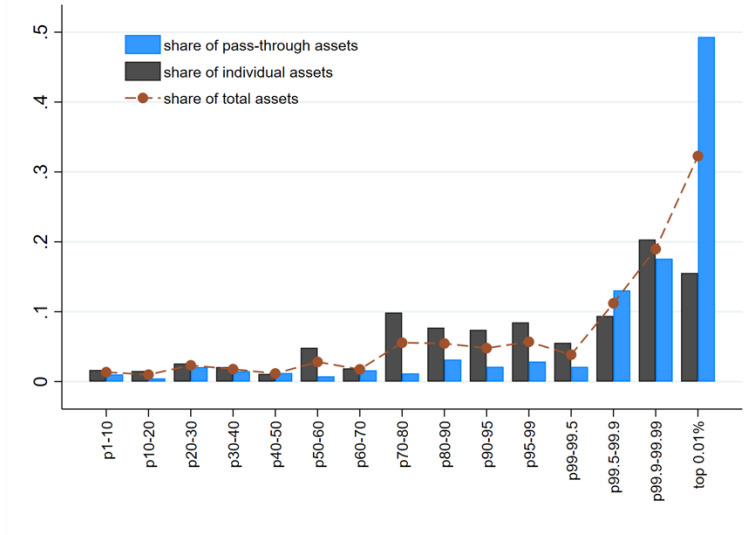
Panel B: By Haven or non-haven



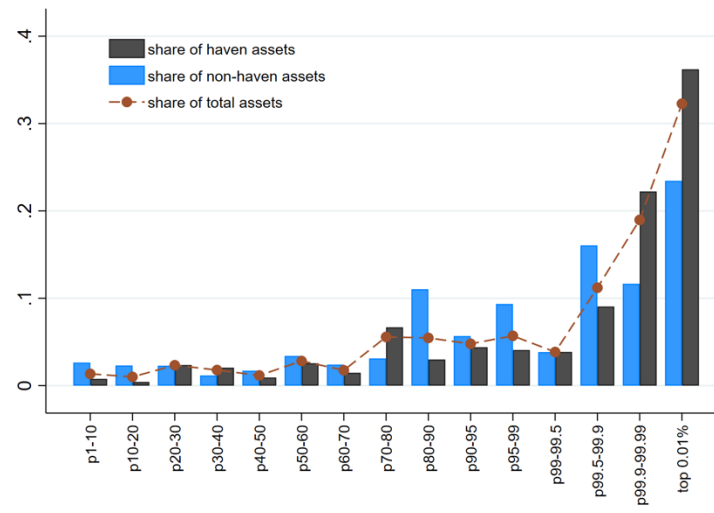
Note: This figure replicates Figure 4, but shows the share of taxpayers in each quantile of the distribution of total positive income (TPI) that receives a FATCA report (Form 8966) indicating ownership of a foreign account. The share is calculated as the total number of individuals in a quantile listed as an account owner on an 8966 over the total number of tax returns in that quantile. The black bars (individual) represent the share of individual account owners that own a foreign account directly. The blue bars (partnership K-1) represent the share of individual shareholders of partnerships that own foreign accounts. Explicitly, for partnership foreign account owners, we link the shareholders to the partnership through form K-1 and the blue bars represent individual shareholders of these partnerships. Panel A shows foreign account ownership shares by centile of the distribution. Over 20% of those in the top 1% are foreign account owners, with about 14% holding accounts through pass-through entities. Panel B zooms in on the top 1% of the income distribution, presenting the shares holding foreign accounts by the top 0.9-0.1% separately, then decomposes the top 0.1% into the 0.09-0.05%, the 0.05-0.01% and the top 0.01%. Foreign account ownership rates increase throughout the top 1%, with approximately 66% of those in the top 0.01% being FATCA account owners, just over half of which hold accounts through pass-through entities.

Figure A2: Share of assets owned by position in the individual income distribution – Distribution of Total Positive Income (TPI), TY2018

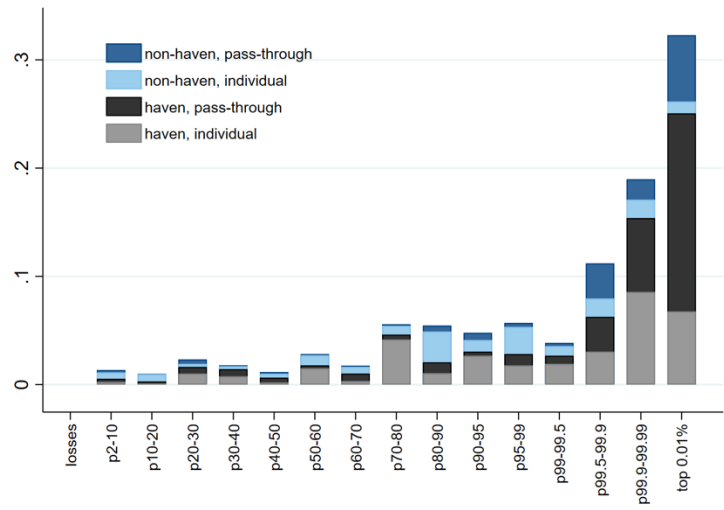
Panel A: Distribution of assets held directly and held through pass-through entities



Panel B: Distribution of assets held in havens and held in non-havens

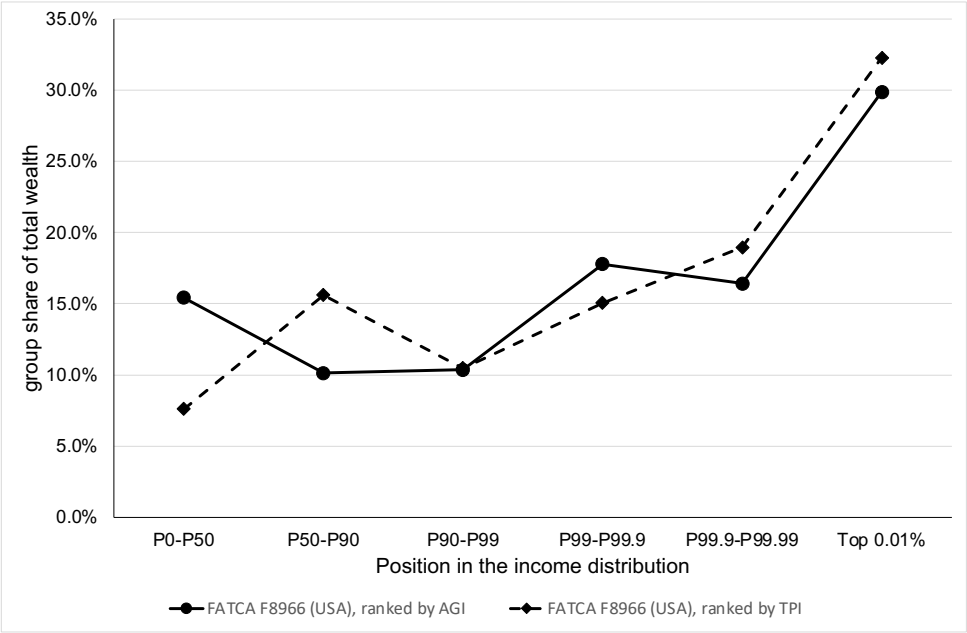


Panel C: Distribution of total assets held directly and through pass-through entities, in havens and non-havens



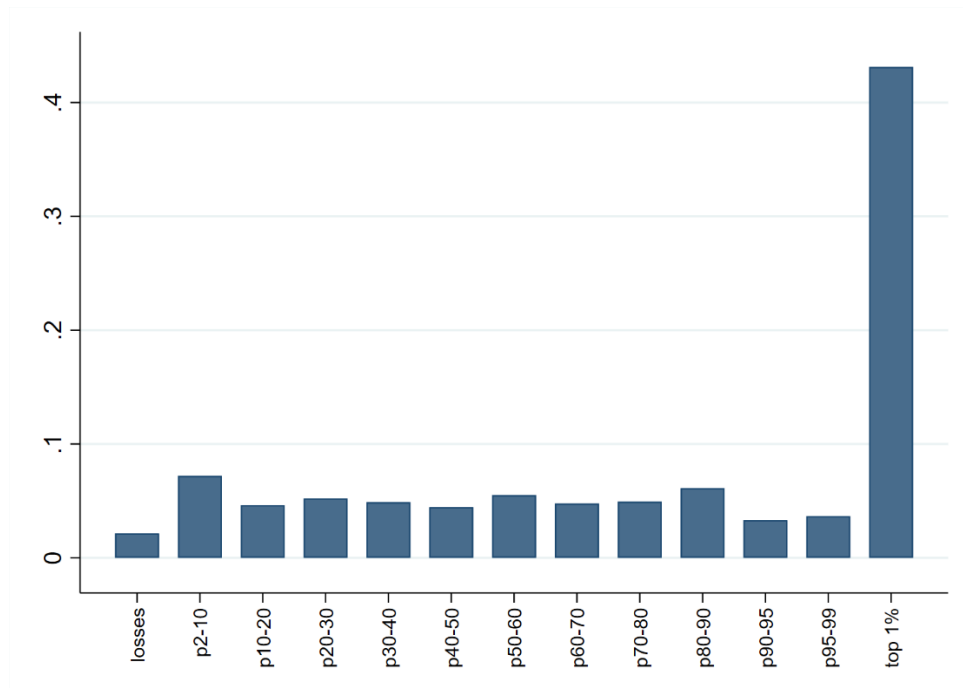
Note: This figure replicates Figure 5 but shows the share of assets held across the distribution of total positive income (TPI). Panel A looks at the distribution of all assets held in directly by individuals and the distribution of those held through pass-through entities (partnerships and S-corporations), separately. The red series plots total shares across both types. Shares are defined as the assets held by individual (pass-through) account owners in each percentile over the total amount of assets held by individual (pass-through) account owners. Total assets (red series) are very concentrated, with about 28% being held by the top 0.01%. The distribution of assets held through pass-through entities (blue bars), which are extremely concentrated, with over 45% being held by the top 0.01%. Directly held assets (black bars) are also concentrated, but less so, with about 12% being held by the top 0.01%. Panel B looks at the distribution of all assets held in havens and the distribution of those held in non-havens, separately. About 33% of assets in havens are held by those in the top 0.01%. Panel C decomposes the distribution of total assets into amounts held in havens and non-havens, directly and indirectly – 28% of total assets are held by the top 0.01%, the vast majority of which are held in havens (grey and black bars), particularly in through pass-through entities (black). Of non-haven assets held at the very top (light and dark blue bars), the majority is held through pass-through entities (dark blue).

Figure A3 – Comparison of Overall Wealth Shares Ranking by Total Positive Income (TPI) and Adjusted Gross Income (AGI)



Note: This Figure plots the share of total wealth belonging to individuals at different parts of the income distribution when we rank by AGI (as in Figure 5 of the main text) versus when we rank by Total Positive Income (as in Figure A2). We observe that ranking by TPI increases the concentration and especially decreases the ownership share in the bottom 50% of the distribution.

Figure A4 – Robustness of Asset Distribution to FATCA Exemption Threshold



Note: Foreign Financial Institutions (FFIs) are only required to report on account holders with aggregate asset values over \$50,000. Figure A4 shows an ad hoc adjustment to the distribution of foreign assets to assess the potential importance of this threshold. We arbitrarily assign \$40,000 in foreign assets to 10% of all households in the bottom 90% of the AGI distribution in 2018 (i.e. assume 10% of households in the bottom 90% have foreign accounts that do not appear in our data because they are slightly below the FATCA reporting threshold). The distribution of foreign assets across the AGI distribution is re-estimated. The level and share of assets going to the bottom 90% substantially increases, but the profile remains qualitatively similar; foreign assets would remain highly concentrated at the very top of the income distribution.

Appendix B: Cleaning Procedures and Statistics on Reporting Quality for Form 8966 Data

B1. The Form 8966

Form 8966 is submitted to the IRS by either foreign governments or by foreign financial institutions (FFIs) directly, depending on the model of the Intergovernmental Agreement (IGA) that the country has with the United States. See the main text for further details on IGAs. The contents of the form and requirements for what is reportable are generally the same under both IGA models, except that FFIs in Model 1 countries (about 88% of participating countries) were not required to report owner TINs for accounts existing prior to FATCA in the first years of reporting, as discussed in the main text.

The Form 8966 comprises the following five parts:

- **Part I, “Identification of Filer,”** contains information on the FFI or other entity (such as a Sponsoring Entity) that is reporting a U.S. owned account or ownership interest on the form. In addition to the name and address of the filer, this section contains a 2-digit filer category code describing what type of entity is filing the form, and the filing entity’s Global Intermediary Identification Number (GIIN).
- **Part II, “Account Holder or Payee Information,”** contains identifying information about the account holder, including whether the holder is an individual or an entity, what type of entity the holder is if it is an entity, the location of the account holder (address, city, state, and zip code), and the owner TIN. For the majority of 8966s, this section contains the identifying information of the account owner, who is also the account holder. Researchers worked with de-identified records throughout the research process, e.g. a dataset containing masked owner TINs and excluding the name of account holders.
- **Part III, “Owner Information,”** contains identifying information for the account owner when the account owner is different from the account holder. Generally, this part of the form will be populated in relatively infrequent situations, when a U.S. person (the “account owner”) owns an account indirectly through certain kinds of foreign entities (the “account holder”). Specifically, filling out this section is required for substantial U.S. owners of a passive non-financial foreign entity (NFFE), substantial U.S. owners of a Direct Reporting NFFE, substantial U.S. owners of a Sponsored Direct Reporting NFFE, and specified U.S. persons owning certain equity or debt interests in an owner-documented foreign financial institution (ODFFI).
- **Part IV, “Financial Information,”** contains information on the account balance and income for the reportable account in a given tax year. In addition to the account number (masked in the research dataset), currency code, and an indicator for whether the account was closed during that year, the financial account information comprises 5 items: account balance, interest, dividends, gross proceeds/redemptions, and other income.
- **Part V, “Pooled Reporting Type,”** pertains to instances where account information is pooled for reporting purposes. This is used for reporting on recalcitrant accounts (where the account holder is not compliant in providing identifying information to the FFIs). Part V is blank when a given Form 8966 pertains to a specific account owned by a US person, i.e. the accounts in our main analysis sample. When a Form 8966 contains a pooled report with information in Part V, Parts II through IV are blank.

After the introduction of Form 8966 in TY2014, a few changes to the form were made in TY2015. The following table describes these changes, with line number references to the TY2015 version of the Form 8966:

Form Section	TY2014	TY2015-present
Part I, Line 1b: Filer Category Code	Absent	Present
Part II, Line 1b: Indicate whether account holder is an individual or entity	Absent	Present
Part II, Line 5: entity type	Includes Direct Reporting NFFE	Excludes Direct Reporting NFFE
Part IV, Line 3b: Check if account closed during the year	Absent	Present

Following Tax Year 2015, the contents of the form have not changed. We discard TY2014 data because very few accounts were reported for TY2014, so changes to the form over time are irrelevant for our analysis.

B2. Data Cleaning Procedures

The following describes the data cleaning process for the dataset of contents of Forms 8966 used in our analysis. The data were extracted from the IRS database and contains all fields of the Form 8966, as well variables generated when processing the forms. We note that all data cleaning procedures outlined here are done purely for statistical purposes for this paper and do not represent how the IRS processes Form 8966 records internally.

Owner Information

We first create a new set of variables for the beneficial owner of the account (i.e., TIN, owner’s country code and country name). Based on how FFIs are instructed to fill in the form (see above), we pull owner information from Part III of the form when Part III is populated, and otherwise from Part II. In some cases, the beneficial owner TIN reported in Part II or III is missing or is in an invalid format (e.g. a single-digit number); we create a variable that encodes a missing or invalid TIN.

The type of owner for an account, either “Individuals” or “Entity,” is reported in Part II, Line 1b. If “Entity” is selected, Part II, Line 5 is filled out to indicate the type of entity, “US Person Entity,” “Passive NFFE with US Owners,” “Direct Reporting NFFE,” “FFI with US Owners,” and “Nonparticipating FFI.” We combined these five entity categories, along with the “Individual,” to create a categorical variable for owner type. In characterizing owner types in the main text, we generally defer to the owner type variable created during the form match process (see below). Information about the type of owner as reported on the Form 8966 is nevertheless useful for cleaning data and e.g. in disambiguating form matches.

Filer Information

The main variable of interest to us from the Filer information in Part I of the form is the country in which the account is located, but we use other RFI information in data cleaning to check for duplicate records and jointly owned accounts, and to check the validity of very large dollar values (see below). Since most filer information is split between RFI variables and Sponsoring Entity (SPS) variables, these variables are combined to create a single filer variable for all identifying information. Box 1b in Part I of the form contains a 2-digit filer category code with eleven categories. A categorical variable is created by combining

the RFI and SPS filer category codes. Similarly, RFI and SPS GIINs are combined into one variable representing the GIIN of the financial institution where the assets are located.

Currency Conversions

The financial information in Part IV of the form can be reported in a local currency or in U.S. Dollars (USD). All such fields have a corresponding variable containing the three-letter currency code for the currency in which the field is denominated. The raw data contained versions of the account value variables that had already been converted to USD by a team at the IRS. However, to examine the validity of extremely large dollar values and clean data it was useful to convert the raw, local-currency version of these variables to USD ourselves using the same set of exchange rates used by the internal team. The end result is three sets of variables for Part IV of the form: the IRS converted amount, our converted amount, and the amount as reported by the FFI or country.

We generally defer to the IRS converted amount, with two exceptions. First, in some cases, the variable for an unconverted amount is present while the internally converted USD amount is missing. In such cases, we replace the missing internally converted amount with our converted amount. Second, we use originally reported currency amounts to screen for outliers and data issues related to mis-reported currency codes by the filer, for instance when a filer reports an amount in local currency but mistakenly indicates the amount is in USD. When one USD is equivalent to an extremely large nominal amount of local currency, such misreporting can create egregious outliers, as discussed further below.

Sample Restrictions

Our sample restrictions focus on information reported in Part IV (Financial Information) because the key statistics about offshore wealth and related income are drawn from this part of the form. In order to avoid deleting observations that might be useful in other contexts, we create a categorical variable to indicate inclusion in the main analysis sample or, if the observation is excluded from the main analysis sample, to encode the reason why.

We first exclude records that are not useful for our analysis. Specifically, these exclusions are as follows:

- All records related to recalcitrant accounts are removed from the preferred sample (414,450 observations), as these are usually pooled for reporting purposes by FFIs.
- A “Record Status Code” is generated automatically upon submission of each form by filers. Observations with a record data status code of “Bad Data-Void,” “Bad Data-Void Error, or “Bad Data- Record Void Requested” are removed from the sample (2,934,393 observations). Such records contain little to no actual data.
- Observations where the Form 8966 owner type is missing, or the owner type is a Nonparticipating FFI (which also suggests that what is being reported involves recalcitrant accounts), are removed (256,102 observations)
- Finally, we exclude from our sample a large number of observations (6,697,618) in which no financial information reported for any income field or account balance.

The last of these restrictions excludes by far the most observations. We elected to drop these records because our goal in this paper is to characterize the offshore wealth and associated income that is reported by FFIs under FATCA, and in these records no wealth or income is reported. Based on the

instructions for the Form 8966, it is unclear why this information is so frequently missing. The most informative clue we observed is that, in 95% of the records with no financial information, the RFI fills in the account balance field as the number “0” rather than not entering any information at all – meanwhile, the income fields typically contain no information rather than the number 0. These clues suggest that the vast majority of accounts with no information are dormant accounts containing no assets. We consulted with experts on the data who generally agreed with this assessment. Still, some of these accounts with no financial information could be of interest from a tax enforcement or other perspective. We defer a fuller examination of these accounts to other work.

There are also a number of records in which account balance is reported but many or all of the other financial variables are missing or zero. We keep these records in the data and discuss missing income information further below and in the main text.

Following the initial restrictions above, we carefully clean the financial information variables to ensure that the key summary statistics we report in the paper based on these variables are accurate. Our general principle is to assume that FFIs report correctly unless we see a strong reason to believe otherwise. Still, there are a number of highly implausible records that are clearly due to some kind of reporting error. Including these records in the sample would bias our key statistics based on these financial variables.

After these restrictions, the data contain 23,738,383 observations, which break down by tax year as follows:

Tax Year	2014	2015	2016	2017	2018	2019
N. Obs.	1,210,434	2,559,273	4,595,466	5,058,423	5,351,336	4,963,451

Cleaning Financial Variables

One of our goals is to provide the most accurate account possible of the wealth and income reported on Forms 8966. To do this, we carefully screened the financial variables in Part IV of the form, as outliers due to mis-reporting can substantially skew distributional statistics of the kind we seek to calculate.

To begin with, accounts where the account balance is greater than \$100 billion, or any of the income fields are greater than \$10 billion, are removed (399 observations). Inspecting each of these records by hand, we found that all of them appear to be erroneous.

We then excluded observations with suspiciously small income amounts relative to their account balance. Observations where the account balance is greater than \$1 billion, at least one income variable is not missing, and the sum of the income variables is less than \$10 million are removed (386 observations removed).

The above restrictions excluded the most obviously problematic records for tabulating total account balances, but we found that some erroneous outliers remained. We next carefully examined total account balances *by country and year* to identify suspicious observations that warranted further examination. We found that this was a useful way to screen for problematic reporting errors because such errors usually generated inconsistencies within countries across tax years (e.g., a huge spike in a single year in a given

country). Upon finding such an inconsistency, we isolated the records responsible for the inconsistency by breaking down the totals further. In some cases, the anomaly was driven by a very small number of accounts in a given country (e.g., a single account over \$10 billion at an FFI that managed no other accounts over \$1 billion in any other year). In other cases, a particular FFI would appear to report a larger number of accounts inconsistently across time. For example, an FFI might report a few hundred accounts in 2015, 2017 and 2018 but several thousand accounts in 2016. From conversations with experts, one reason we believe we might observe this is if the FFI reports on *all* of the accounts it maintains rather than the accounts owned by U.S. persons. In other cases, the anomaly would be driven by single FFI whose account values spiked massively in a single year – e.g. from a few million to several billion dollars – for a large number of accounts, for unknown reasons. In both of these cases, we typically excluded all accounts or accounts over \$10 million at the problematic FFI. In the latter case, we also checked whether the anomaly could be attributable to a misreported currency code, in which case we would correct the currency code and keep the records in the analysis sample. In yet another case, the anomaly occurred because account balances contained repeated digits (e.g., a string of 9’s) that indicate that they are place holders rather than actual account balances; we excluded such records.

In total, we removed 78,560 observations due to these issues, leaving 23,659,823 observations in the sample, which break down by year as follows:

Tax Year	2014	2015	2016	2017	2018	2019
N. Obs.	1,191,647	2,551,608	4,579,049	5,052,811	5,333,362	4,951,346

Of these, 20,805 were exclusions of a very small number of accounts and 68,528 were broader exclusions implemented at the FFI level. We found obvious currency conversion issues causing anomalies in 33,866 records and corrected them. We are unable to provide further details due to privacy concerns and the data protection provisions of IGA agreements. We note that one implication of these exclusions is that the total wealth figures likely represent a lower bound for the actual total offshore wealth and income of US persons. If some problematic records are still included, this could bias the totals upwards rather than downwards, but we believe that our careful screening ensures any upward bias from this issue is likely minimal.

We next turned to cleaning the financial income variables: interest, dividends, gross proceeds/redemptions, and other income. Observations with over \$10 million in interest income and the interest amount is greater than half of the account balance, or have \$10 million in interest income and the account balance is missing, are dropped from the preferred sample (1,402 observations removed). All observations with over \$1 billion in dividend income are suspicious, as all of these observations have low or no account balance (50 observations removed). For many observations where Gross Proceeds makes up a considerable amount of the total account balance, it appears that the account is new and therefore not suspicious. Accordingly, the restriction for gross proceed amounts is 1.5 times the total account balance. If gross proceed income is greater than \$10 million, and gross proceeds are 1.5 time the account balance, the observations is removed (9,044 observations removed). Accounts with over \$1 billion in gross proceeds and missing account balance are removed (484 observations removed). The restriction of 1.5 time the account balance is also used with other income to discard suspicious observations. If other income is \$1 billion and 1.5 times greater than the account balance, the observation is removed (126

observations removed). Additionally, if other income is greater than \$1 billion and account balance is missing, the observation is removed (52 observations removed).

Following our use of the decomposition of account values by country and year, we followed the same cleaning procedures for the other financial information fields. In other words, we broke down the totals of these fields by country and year, and reviewed this decomposition to identify further anomalies. From this procedure we removed another 17,771 observations, leaving 23,642,052 observations in the sample, which break down by year as follows:

Tax Year	2014	2015	2016	2017	2018	2019
N. Obs.	1,190,809	2,549,525	4,574,280	5,048,854	5,329,536	4,949,048

Going forward, we disregard data from 2014 and 2019. Tax year 2014 contain far fewer records, as we can see in the tabulations above. Tax year 2019 data were incomplete as of the data pull upon which we built our analysis sample, and this is reflected by a drop in the number of observations from 2018 to 2019.

Duplicate Records

The raw data also contain a sizable number of duplicate records, which we turn to next.

We first create a data quality code variable to rank observations by data quality for use in cleaning duplicates. The highest level of data quality are observations that have a “Good Data” Record Data Status code, a positive total amount of income, and a positive account balance (6,717,526 observations). The second highest data quality rank are observations that have a “Good Data” Record Data Status code, but have either a missing, 0, or negative account balance or total income, with the last of these being the most common by far (13,002,262 observations). The next rank are observations that have a “Bad Data-Correction Requested” Record Data Status Code (10,133,629 observations). The records that have this data status code and survive our other cleaning procedures typically contain useful financial information; our understanding is that the requested corrections alluded to in the code pertain to missing owner information (e.g. owner TINs) or parts of the Form 8966 that we do not use in our analysis. The remainder (3,198,766 observations) are ranked as the lowest level of data quality.

To check for duplicate records, we first counted the number of observations per each unique tax year/RFI GIIN/account number combination. We find that 67% of observations are unique to each account number/RFI GIIN/tax year combination, while 33% have more than one observation. We created a “non-duplicate flag” at this point in the data, and assume that the 67% of records with a unique account number, RFI, and tax year are not duplicates. For the remaining 33%, we need to do more work because not all of these records are duplicates. In particular we could observe the same account number, RFI GIIN and tax year for multiple records in two common situations: jointly owned accounts, and accounts where the (masked) account number we observe represents not an actual account number but a placeholder value (e.g. for some types of reportable accounts an account number may not exist or be known to the filer). The next several steps represent our best attempt at separating out the true duplicate records from these other types of situations.

First, we assume if we observe multiple records with the same account number at the same RFI in the same tax year, *but the owner TINs are distinct*, then these do not represent duplicate records. Rather they

are likely to be distinct accounts with a placeholder/missing account number or jointly owned accounts. We handle jointly owned accounts in the next section. Based on this assumption, when we see the same RFI, account number, and *non-missing* owner TIN for multiple records in the same tax year, we keep exactly one of these records. We use the data quality code described above to decide which of the duplicate records to keep – for example, if the Record Data Status Code was “Bad Data, Correction Requested” and the FFI submitted a corrected record, this will ensure that we keep the corrected record. Empirically this happens quite often. In the case where multiple duplicates have the same data quality code, we keep the first record appearing in the dataset.

When owner TINs are not missing, adding the distinct TIN screen is likely all we need to screen out duplicate records. The observations without missing TINs that this procedure flags as duplicates are therefore removed from the sample (2,402,674 observations removed).

Potential duplicates with the same account number at the same RFI in the same year, but with missing owner TINs, are more difficult to screen. We note that in the case of missing owner TINs, keeping one record is sufficient as we have no hope of matching multiple owners, so the main challenge is to ensure that we do not exclude the records with missing TINs and placeholder account numbers. To do this, we screen for distinct account balances, as account balance is rarely missing and truly distinct accounts should have distinct account balances in the vast majority of cases. As such, we encode those records with the same masked account number at the same RFI but a missing owner TIN as a non-duplicate if the account balance is distinct. In the event that we observe multiple records with the same account number and RFI, missing owner TINs, and the same account balance, we regard these as duplicates, as before, we keep the first record with the highest data quality code (469,629 observations removed).

Joint Ownership

Generally, when an account is jointly owned, per the instructions for Form 8966, we expect to see an 8966 for each owner, with the *same account number* and *distinct owner information* on each of these 8966's. Additionally, RFIs are instructed to report the income and asset information in total for the account and file a different Form 8966 for each owner, rather than to divvy these up to various joint owners, so the information on Part IV should also be the *same in the case of joint ownership*.

In principle, we can therefore identify jointly held accounts by looking for cases where the same account number at the same RFI is reported as owned by distinct owner TINs. However, implementing this principle naively would cause us to inadvertently flag accounts with placeholder account numbers as jointly owned accounts. To deal with this issue, we obtained a list of the masked analogue of the 7 most commonly observed “placeholder account numbers” from the IRS (the most commonly observed of which is the masked analogue of the placeholder that RFIs are actually instructed to use when an account number does not exist). We replace the account number to missing when the account number takes one of these placeholder values, and we do not encode any of these as null accounts. Doing so helps us ensure that we do not mistakenly assume that accounts with these placeholder account numbers are jointly held by many owners. If the account number is not missing, the owner TINs are distinct, and the account balance matches another record with the same account number, we assume that these records represent jointly owned accounts.

We next need to devise a way of handling jointly held accounts that 1) avoids double counting the same wealth when we tabulate totals of financial variables, and 2) properly counts the number of owners of

offshore wealth (e.g. to count the fraction of individuals in a given part of the income distribution with an interest in an offshore account). To do this, we keep all of the records for jointly owned accounts (so that owner counts are correct), but we divide the financial variables proportionally among the owners (to avoid double counting dollars of income/wealth). For example, if two individuals own a joint account (by far the modal case), we replace the account balance and income variables by two, allocating half of the wealth and income to each owner. When true ownership is not an equal split among all joint owners, this allocation is conservative in most respects, as e.g. it could reduce the concentration of ownership of offshore wealth.

We next looked in detail at records with large numbers of owners, as these could correspond to errors that are not actually jointly held accounts. On inspection, it appeared that instances where the number of owners is greater than 25 are virtually all due to errors. For example, the distribution of account balances tilts sharply toward very small values above around 25 owners. If the number of owners is greater than 25, we keep the first observation for that account so that total wealth and the total number of accounts are unaffected, but the remaining records are removed from the sample (36,597 observations removed). Overall, 13.41 % of observations in the preferred sample are jointly owned.

Variables for Estimating Rates of Return

We create a few auxiliary variables in order to estimate rates of return. First, we create a series of indicator variables indicating the following: whether interest income is non-missing, whether interest or dividend income is non-missing, whether interest, dividends, or gross proceeds is non-missing, and whether any income is non-missing. We create an analogous set of auxiliary variables equal to the account balance if specified income variables are non-missing, and otherwise the variable is missing.

Unresolved Data Anomaly

While we managed to resolve most obvious data anomalies, we could not resolve one large spike in total wealth in a single country in 2016. Total wealth in that country spikes in 2016, drops in 2017, and then increases again in 2018; these spikes are large enough to matter for the time series of total wealth in the Form 8966 data. We were able to identify three RFIs in this country that were reporting an anomalously large number of accounts with very large total account balances in 2016, but no issue was found either in the currency conversion process or with a specific observation within these RFIs. Meanwhile, all RFIs report a sizable (if inconsistent) number of large accounts in every year, with many more large accounts in 2018 and 2016 than 2017. In other words, it could be that these RFIs reported a number of large accounts unnecessarily in 2016, but it could also be that these RFIs actually neglected to report on some large accounts in 2017. In the main sample, we left these accounts as is, but we report in Table B1 what the total account balances would be if we either 1) exclude all records from these RFIs in 2016 (middle column, our best proxy if 2016 was the problem year) or 2) if we replace total wealth at these RFIs in 2017 with the midpoint between TY2016 and TY2018 (third column, our best proxy if 2017 was the problem year).

Table B1 – Sensitivity analysis for problematic RFIs: Total Wealth Over Time with Various Specifications

SPECIFICATION			
TAX YEAR	Baseline Total Account Balance (Trillions USD)	Drop 2016 records at problematic RFIs (Trillions USD)	Replace 2017 records at problematic RFI with average wealth (Trillions USD)
2015	1.65	1.65	1.65
2016	3.65	2.65	3.65
2017	3.23	3.23	3.74
2018	3.98	3.98	3.98

B3. Linking Owners’ Tax Returns

We use standard fuzzy matching procedures to link F8966 owners to tax returns of individuals and entities on the basis of TINs and/or names. The initial output of this process is a nonexclusive categorization of every record as follows:

- Records linked to individuals via Forms 1040 or other records pertaining to individuals (e.g. Social Security records) are classified as **individuals**
- Records linked to partnerships (Forms 1065 and 1066) are categorized as **partnerships**
- Records linked to C-corporation tax returns (the form 1120 family excluding 1120-S and 1120-F) are categorized as **C corporations**
- Records linked to S-corporation tax returns (Form 1120-S) are categorized as **S corporations**
- Records linked to tax returns of foreign corporations (Form 1120-F) are categorized as **Foreign Corporations**
- Records linked to tax exempt entity returns (the Form 990 family) are categorized as **tax exempt**
- Records linked to tax returns of trusts (Form 1040, 1041A, Form 5227) are categorized as **trusts**

For the most part, these describe the owner types for the matched records, see e.g. Table 3 of the main text. A few records belong to more than one of the above categories. One set of cases this occurs is when the entity changed its classification, so we link the 8966 record to one type of entity return in one year and a different type in another year. In these cases we use the most recent tax return we observe to classify the entity. In another set of cases, the TIN matches a unique tax return but the entity name matches another tax return of a different type of entity; in this case we use the TIN match to classify the entity. Third, when we have an ambiguous match but the TIN matches another form filed by just one of the candidate entity types, we disambiguate the owner type on the basis of the other form. For example, a record that matches to a partnership return and a trust return will be classified as a partnership if the owner information also matches a Form 8804. Records that belong to multiple categories after our best attempts at disambiguation are assigned an **ambiguous match** owner type.

Apart from the ambiguous matches, we create a few other categories for records we cannot confidently link to an owner’s tax return. Some records do not match to the tax returns for any of the entity types described above, but the records to match the filer information for non-tax-return forms filed by

businesses or employers (Forms 940, 941, 943, 944, 945, 2290, or 720). In this case, the owner type is classified as an **unknown entity**. (Otherwise the owner type is classified based on the type of tax return it files). Some observations match a valid individual (TIN and/or name) in Social Security administrative records but no tax returns; these individuals are classified as **unknown individual**.

Records where the owner TIN was missing and we were unable to match to the owner’s tax return with confidence based on name are classified as **missing TIN**. Records where an owner TIN was present on the Form 8966 but matched none of these other data, and a name match was not possible, are classified as **Unmatched TIN**. We suspect many of these correspond to incorrect or invalid TINs.

B4. Missing Data

This section contains supplemental analysis of reporting quality issues in the main sample.

Table B2 reports the share of observations (accounts) in the main sample in 2018 for whom a given variable is missing, breaking observations down by country type. These rates of missing data are relatively stable over time; we focus on 2018 information because this data are most used in the main text. We observe that Missing TINs are especially concentrated in non-havens and in particular non-havens with prior automatic exchange of information. For the financial variables, the rate at which variables are missing does not vary systematically by owner type. Just over 5% of observations have no account balance but some reported income (recall that observations without account or income information are excluded from the sample, see above). Income information is far more frequently missing, with around 60% of observations having no reported interest and over 90% having no information for the other three income variables.

Table B2 - Share of accounts with missing information by country type (TY2018)

	Country type			
	Haven	Non-haven	NH - EOI	NH - No EOI
Missing Tin	20.8	38.8	43.3	19.0
Missing Account Balance	6.4	5.4	5.5	5.4
Missing Interest	59.6	61.5	62.7	56.1
Missing Dividends	93.1	92.1	91.2	96.4
Missing GP	93.6	93.6	93.8	92.6
Missing Other	89.1	96.7	97.1	95.1
Number of observations	612,406	3,954,216	3,222,800	731,416

In the previous table and in the main Table figures concerning missing data (e.g. Table 2), we count as missing instances where the variable in question was reported as the number zero and those where the reported dollar amount is blank. In Table B3 below, we separately tabulate the frequency of these two cases. We observe that in general, blank variables are more common than zeros for the income fields,

while account balances, in contrast, are more often zero than blank. This suggests that while many “missing” account balances might represent accounts that had no balance at the end of the year, the income variables are often missing because the FFI did not report this information rather than because the FFI indicated that the account received none of the given type of income.

We also report dollar-weighted figures for the income variables, i.e. the share of account balances where an income variable was blank or zero. Observations are zero interest income are noticeably less common for large accounts, so the dollar-weighted zero share is much smaller for interest, suggesting that when we observe zero interest, in many cases, this is because the account received no interest income. However, this pattern does not appear so strongly for other income variables, making it less clear what to make of zeros for these. Dividends are almost always blank rather than zero. Likewise, given that a large share of observations and dollars of wealth are associated with blank income fields, we find it plausible that usually, when these variables are missing, it occurs because the filer did not report income, not because there was no income to report.

In Figure B1, we examine how the likelihood of a blank or zero income variable varies by the account balance. In Panel A, we observe that the probability of a blank income variable is decreasing as we move to very large accounts, but the share blank remains high even for very large accounts. Even at the very top of the account balance distribution, about of accounts have no reported income information of any kind. The probability of observing zeros for income variables is lower. For interest, this probability drops sharply with the account balance. For other income fields, the probability is surprisingly flat by account balance – we might have expected larger accounts to be associated with riskier and/or more sophisticated investments that would generate dividend and gross proceeds income, making zeros less likely.

From these statistics and supplemental analysis of the blank versus zeros question, we concluded that we cannot be certain that when the number zero appears, this is an actual report of zero dollars of income/account balance rather than that the information is not reported by the filer. Likewise, blank income fields are far to prevalent even for very large accounts that almost certainly should receive some type of income for us to regard blank observations as zeros. As such, we included both cases in our missing data definition in the main text. We have discussed this issue with experts on the data, who indicated they were aware of the prevalence of missing income data but unaware of how we might further disambiguate true zero amounts from non-reported information.

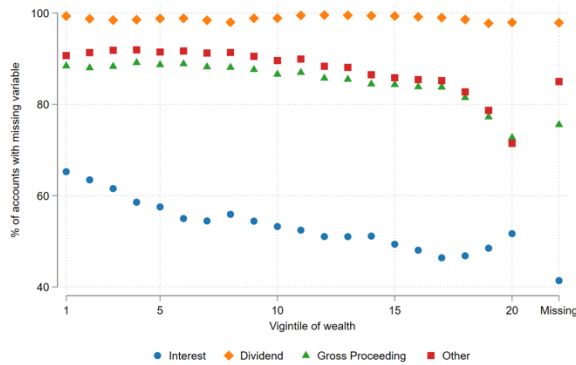
Table B3 – Blank and zero probabilities for Dollar-Valued Variables (TY2018)

A. Pr(variable is blank), in %					
	Account Balance	Interest	Dividends	Gross Proceeds	Other
Account weighted	1.1	53.2	98.8	84.9	87.6
Weighted by Acct Bal	--	62.9	99.3	78.0	70.5
B. Pr(variable is zero), in %					
	Account Balance	Interest	Dividends	Gross Proceeds	Other
Account weighted	4.5	14.8	0.1	8.9	8.2
Weighted by Acct Bal	--	3.9	0.0	7.8	10.0
C. Pr(variable is blank variable is blank or zero), in %					
	Account Balance	Interest	Dividends	Gross Proceeds	Other
Account weighted	19.3	78.3	99.9	90.5	91.4
Weighted by Acct Bal	--	94.1	100.0	90.9	87.5

Note: This Table reports the share of account (unweighted and weighted by account balance) with blank (panel A) or zero (panel B) in one of the main variables on all accounts and share of blank on all accounts with blank or zero value (panel C).

Figure B1. Probability of Blank versus Zero Income Variables by Quantiles of Account Balance

Panel A. Prob Income Variable is Blank



Panel B. Prob Income Variable is Zero

