# Methodological note: Corporate tax abuse, State of Tax Justice 2020

November 11, 2020

## 1 Data

The methodology exploits country by country reporting (CBCR) datasets which include information on MNCs which only became available recently and which is of heretofore unprecedented quality. The dataset was provided thanks to a CBCR regulation which stems from OECD Base Erosion and Profit Shifting (BEPS) Action 13 on CBCR and which requires all large MNCs to report how much tax they pay in individual countries, including tax havens. The regulation impacts MNCs with consolidated group revenues of at least EUR 750 million, headquartered in any country which has adopted the CBCR regulation. As the main data source for our analysis, we use the 2016 OECD CBCR data for large MNCs published by OECD for numerous headquarter countries in July 2020. In addition, instead of the 2016 data published by the OECD, we use 2017 US CBCR data published by the US Internal Revenue Service in December 2019; unlike the 2016 data, which included only 70% of the MNCs, this dataset is complete. Importantly, existing research compared these US CBCR data with other sources (Clausing, 2020b; Garcia-Bernardo et al., forthcoming) and established a good correlation between various types of data sources. Moreover, the CBCR data is outstanding in several dimensions.

First, one of the most obvious advantages of CBCR data over other data sources is its much more substantial country coverage. This is especially relevant for low- and middle-income countries and for selected parts of the world. For example, US CBCR data includes information on taxes and profits for 25 African countries while the frequently used data from the Bureau of Economic Analysis of the United States Department of Commerce only covers three. CBCR data includes data on large MNCs' profits and tax payments in, for example, up to 141 (United States) and 163 (India) jurisdictions in the full data set—93 and 134 jurisdictions respectively for the data set limited to firms with positive profits (the two data sets are discussed below). The exceptional data coverage provided by CBCR data thus enables us to collect evidence of profit shifting for many countries with low and middle per capita incomes. And this country coverage is one reason why UNODC and UNCTAD (2020) proposed to use this CBCR data for the Sustainable Development Goals indicator of illicit financial flows, likely in a similar way that we implement the profit misalignment method outlined in Section 2 (Cobham & Janský, 2020).

Second, CBCR ensures that profits and taxes are defined consistently with the concepts of corporate profits and taxes. By contrast, this is not the case with e.g. Bureau of Economic Analysis data where profits are imputed from a combination of net profits, intra-group dividends, interest paid, and other variables, as recently discussed by Blouin and Robinson (2020), Clausing (2020a), Clausing (2020b), Garcia-Bernardo et al. (forthcoming). Consequently, CBCR data excludes double-counting in revenue and likely in profit (with the exception of stateless entities dropped from our analysis and intercompany dividends, for which companies have neither instructions nor incentives). Furthermore, CBCR data is based on tax accounting and thus reflects how much MNCs in fact pay in taxes, rather than on financial accounting, which is the basis for most other datasets including Orbis and and which has been shown to underestimate profit shifting (Bilicka, 2019). Since CBCR data offers the best available information on MNCs' tax payments for many countries, it thus provides us with the first such dataset suitable for a high-quality cross-country comparison (for example, until now various proxies for profits were used, e.g. by Haberly and Wójcik (2015), Bolwijn et al. (2018) or Damgaard et al. (2019)).

Third, CBCR data are provided in two separate data sets, for all large MNCs ("All Sub-Groups") as well as for those large MNCs that have positive profits and so not losses ("Sub-Groups with Positive Profit"). To estimate semi-elasticity of profits to tax rate differentials, we prefer to use the data set for MNCs that have positive profits only, at the expense of a decrease in country coverage (Table **??**). By using the data with positive profits only, we avoid offsetting firms with losses and firms with profits and we can thus estimate ETRs more precisely. By contrast, data set which include both profits and losses likely understate profits (since losses are included) and overstate ETRs (since taxes are paid by companies earning profits, typically, though losses are also included in the denominator). We use this data set with all large MNCs for the misalignment method as well as to reallocate profits in the second step for all methods since for these purposes we prefer to have information on real economic activities of MNCs regardless of whether these MNCs are profit- or loss-making. It is also more suitable for comparison with other datasets (e.g. from the Bureau of Economic Analysis). Furthermore, unfortunately both data sets might be affected by a practice where MNCs prefer to report losses in countries with high taxes while locating their profits in countries with low taxes.

While the substantial country coverage as well as the other advantages of CBCR data open new avenues for research, at least two challenges associated with the new data source remain (and we summarise them alongside the above discussed advantages in Table 1). First, unfortunately a certain extent of double counting in profit due to intercompany dividends is likely inevitable as MNCs are instructed not to double count intercompany dividends in revenue but not so explicitly in profit. This potential double counting has been explored recently for US data by Horst and Curatolo (2020). Significantly, there do not seem to be incentives for double counting profits in tax havens by MNCs (since they know this data is to be used for assessing transfer pricing risk). Furthermore, we exclude stateless income, another potential source of double counting. Consequently, the US CBCR data produce totals reassuringly consistent with other data sources Clausing (2020b), Garcia-Bernardo et al. (forthcoming).

| Selected advantages |
|---|
| Includes data on large MNCs' profits and tax payments in around 100 jurisdictions for at least five headquarter countries |
| Does not include double counting in revenue and likely not in profit. |
| Enables to use data on large MNCs and those with positive profit only (the latter estimates ETRs more precisely). |
| **Selected disadvantages** |
| Might include some double counting in profit due to intercompany dividends or stateless entities (which we drop). |
| Includes a sample of large MNCs for 2016 for some countries in aggregated and anonymised form (which we address). |

**Table 1.** Summary of selected advantages and disadvantages of the CBCR data

Second, data from 2016, i.e. the only year for which we have data for most headquarter countries, does not provide complete coverage of all large MNCs for all countries. 2017 is thus the first year with complete data available, though only the United States has published its data thus far (OECD is to publish data from 2017 for other headquarter countries likely only in July 2021). In this paper, we therefore use both of the two years of data to make use of the best available data possible. While the CBCR regulation has been implemented by approximately 100 countries so far, only some of them agreed to share their data publicly in aggregated and anonymised form; moreover, some have chosen to aggregate data to a far greater extent than others. Of the 137 jurisdictions included in the Inclusive Framework on BEPS, 58 jurisdictions accepted collected CBCR, 35 received data from 20+ MNCs, 26 jurisdictions shared a limited amount of data, and only 10 jurisdictions disaggregated the data on more than 60 countries, with the United States and India (137 and 163 jurisdictions, respectively) leading the way (Table **??**). In the end, we kept data for 12 countries with information on at least 30 jurisdictions.

Since the aggregation of profits and taxes hides profit shifting, we retained in our main sample the 11 countries which reported on at least 7 offshore financial centres (Fig. 1).

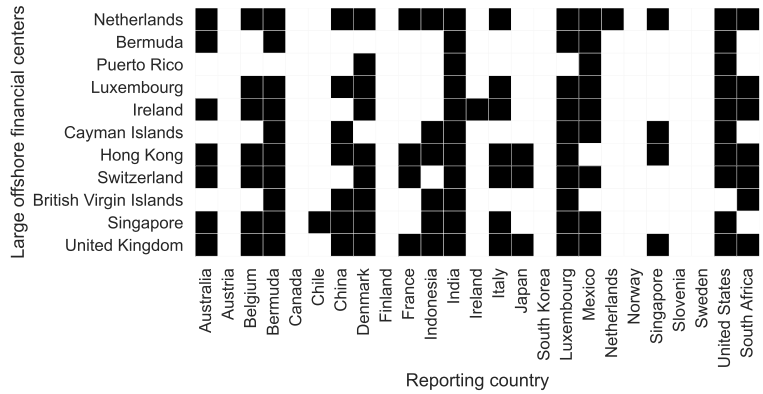


**Figure 1.** Country availability. Black squares mark the availability of data on large offshore financial centres for each reporting country.

We focus on the remaining challenges posed by this data in section 3, where we empirically deal with three additional issues: the lack of completeness in the data of reporting countries, the varying combinations of countries in the aggregated country categories and the lack of reporting by some countries. Other limitations of the CBCR data (e.g revenues unavailable according to the location of the final customer) are discussed by the OECD, which published the data with an "Important disclaimer regarding the limitations of the country-by-country report statistics", and by Garcia-Bernardo et al. (forthcoming) and Clausing (2020b).

Table 2 shows the summary statistics of the CBCR data with positive profits for the countries in our sample. Furthermore, we remove observations with an effective tax rate (ETR, computed as the ratio of corporate income tax to gross income ) larger than 50% or smaller than 0% (in this case using accrued taxes). We also remove observations when reported profits are below USD 1 billion. This eliminates outliers and allows for a more efficient estimation of the tax semi-elasticity. Furthermore, Table 2 distinguishes between domestic and foreign activities of MNCs - domestic one are those in the reporting (i.e. headquarter) countries, while foreign ones are those in all other countries (i.e. except for the domestic one). For most countries domestic profits and activities are higher than foreign ones. The exceptions are mainly Bermuda and Luxembourg, which are often considered tax havens, as well as Belgium. The observed balance between domestic and foreign activities provides a useful guidance for when we estimate missing data in Section 3.

## 2 Misalignment model

We estimate a profit misalignment method, which typically starts from a given relationship between real profit ($p$) and a combination of labour (measured using wages and employees), capital (often approximated with tangible assets) and revenue. Profit misalignment is then calculated as the difference between reported profits ($\pi$) and theoretical profits ($p$). In our version of this method we allocate 25% of the weight to employees, 25% of the weight to wages, and 50% of the weight to unrelated party revenues.

$$\frac{\hat{p_i}}{\sum_i \hat{p_i}} = R_i \cdot \sum_i \pi_i \tag{1}$$

Importantly, since MNCs can report zero or negative profits in a country in order to avoid taxes, we use the data on all sub-groups instead of the data on sub-groups with positive profit, which is used to

| Reporting | Partner | Firms profits>0 | Profits US$bn | Tax accrued US$bn | Tax paid US$bn | Employees Thousands | Revenue US$bn | Assets US$bn | ETR (%) accrued | ETR (%) cash |
|---|---|---|---|---|---|---|---|---|---|---|
| Australia | Domestic | 94 | 69.8 | 13.9 | 10.0 | 949.9 | 365.2 | 340.7 | 19.9% | 14.3% |
| | Foreign | 758 | 26.5 | 3.1 | 2.7 | 335.3 | 124.2 | 92.9 | 11.6% | 10.1% |
| Belgium | Domestic | 43 | 18.7 | 1.0 | 0.8 | 146.0 | 88.4 | 90.8 | 5.4% | 4.1% |
| | Foreign | 52 | 82.2 | 7.1 | 8.5 | 499.3 | 167.5 | 101.2 | 8.7% | 10.3% |
| Bermuda | Domestic | 26 | 12.7 | 0.0 | 0.0 | 2.1 | 10.4 | 9.1 | 0.2% | 0.1% |
| | Foreign | 26 | 88.7 | 9.4 | 12.8 | 530.3 | 520.5 | 620.3 | 10.6% | 14.5% |
| China | Domestic | 77 | 391.8 | 65.5 | 77.9 | 11543.8 | 2785.2 | 5471.2 | 16.7% | 19.9% |
| | Foreign | 905 | 57.3 | 5.3 | 5.2 | 355.6 | 376.7 | 336.1 | 9.2% | 9.1% |
| Denmark | Domestic | 35 | 17.0 | 2.4 | 2.5 | 120.5 | 44.7 | 24.0 | 13.9% | 14.5% |
| | Foreign | 39 | 10.0 | 0.7 | 0.8 | 783.4 | 74.9 | 40.5 | 6.6% | 7.7% |
| India | Domestic | 146 | 74.8 | 18.4 | 22.2 | 3891.3 | 564.2 | 864.0 | 24.6% | 29.7% |
| | Foreign | N/A | 15.0 | 2.3 | 5.9 | 548.5 | 130.5 | 121.8 | 15.5% | 39.6% |
| Italy | Domestic | 104 | 48.3 | 6.6 | 7.4 | 630.5 | 340.7 | 209.1 | 13.6% | 15.4% |
| | Foreign | 130 | 44.9 | 5.7 | 6.4 | 612.8 | 263.0 | 148.2 | 12.6% | 14.3% |
| Luxembourg | Domestic | 52 | 8.2 | 0.1 | 0.1 | 8.8 | 8.6 | 18.4 | 1.4% | 0.9% |
| | Foreign | 119 | 34.3 | 2.4 | 2.8 | 1142.9 | 372.1 | 128.6 | 6.9% | 8.1% |
| Mexico | Domestic | 60 | 26.5 | 6.4 | 6.8 | 1228.6 | 139.3 | 100.6 | 23.9% | 25.5% |
| | Foreign | 334 | 9.8 | 2.3 | 2.3 | 340.9 | 114.2 | 79.0 | 23.1% | 23.4% |
| United States | Domestic | 1094 | 1310.5 | 257.9 | 209.6 | 19601.7 | 9426.7 | 4880.5 | 19.7% | 16.0% |
| | Foreign | 1548 | 873.6 | 102.5 | 100.5 | 10972.0 | 3338.4 | 1722.2 | 11.7% | 11.5% |
| South Africa | Domestic | 34 | 16.5 | 1.7 | 2.4 | 604.3 | 69.5 | 77.0 | 10.1% | 14.6% |
| | Foreign | 574 | 5.2 | 1.2 | 0.9 | 316.9 | 43.7 | 41.1 | 23.8% | 18.1% |

**Table 2.** Summary statistics for the 11 countries of the sample. Domestic indicate the financial reporting of MNCs in the reporting (i.e. headquarter) countries, while foreign in all other countries (i.e. except for the domestic, or headquarter, one). Note that since we are using "Sub-Groups with positive profits", the number of firms included in the domestic section can be lower than the number of firms reporting on foreign operations.

calculate tax semi-elasticities. The ETRs (used to calculate tax revenue losses) are still calculated from the data on sub-groups with positive profit. For observations which were available in the data on all sub-groups but not in the data on sub-groups with positive profit we used the average country ETR if available and the statutory corporate income tax rate otherwise. Since the misalignment method is not affected by outliers (e.g. small countries) we keep all observations in the sample; in the regression models, we dropped observations with profits below $1 billion or extreme ETRs.

Profit shifting is calculated as the difference between booked profits and estimated profits:

$$\hat{S}_i = \pi_i - \hat{p}_i \tag{2}$$

In this case $\sum \hat{S}_i = 0$ and $\Delta P_i = \hat{S}_i$. However, we add one extra constraint. The profit misalignment of all foreign observations (pairs of reporting and investment countries where reporting and investment countries differ) with a tax rate higher than 25% was set to zero since we assumed that an MNC would not shift profits to a country with a tax rate over 25%. This corrects for extreme outliers such as high profits of Bermudian companies in Peru and high profits of MNCs in resource-rich countries.

## 3 Estimating missing data

The most important limitation of studies on profit shifting has been a lack of data completeness. While the availability of CBCR data constitutes a significant step forward and partially corrects this issue, as discussed in section 1, three specific limitations remain to be addressed. The first limitation concerns the lack of completeness in the data of reporting countries. We address this limitation by comparing the number of companies in Orbis, a frequently used database covering over 300 million public and private firms worldwide, with the number of companies observed in CBCR (Table 3). While the number of countries observed is similar to the number of companies expected for most countries, we observe large

| Country | # Expected (Orbis) | # Observed (CBCR) | Ratio | Country | # Expected (Orbis) | # Observed (CBCR) | Ratio |
|---|---|---|---|---|---|---|---|
| China | 583 | 82 | 7.11 | US (2017) | 1501 | 1548 | 1.03 |
| Denmark | 69 | 39 | 1.77 | AUS | 111 | 110 | 1.01 |
| Bermuda | 60 | 39 | 1.54 | India | 158 | 165 | 0.96 |
| Singapore | 48 | 32 | 1.50 | Norway | 55 | 60 | 0.92 |
| US (2016) | 1,501 | 1,101 | 1.36 | Chile | 29 | 32 | 0.91 |
| South Africa | 58 | 44 | 1.32 | Finland | 48 | 54 | 0.89 |
| Japan | 891 | 715 | 1.25 | Netherlands | 136 | 155 | 0.88 |
| Italy | 151 | 130 | 1.16 | Australia | 64 | 73 | 0.88 |
| Indonesia | 22 | 19 | 1.16 | Belgium | 45 | 54 | 0.83 |
| France | 206 | 180 | 1.14 | Poland | 24 | 29 | 0.83 |
| Canada | 179 | 160 | 1.12 | Slovenia | 5 | 7 | 0.71 |
| Korea | 205 | 185 | 1.11 | Mexico | 40 | 74 | 0.54 |
| Sweden | 95 | 88 | 1.08 | Luxembourg | 30 | 120 | 0.25 |
| Ireland | 47 | 45 | 1.04 | | | | |

**Table 3.** Number of companies expected (according to Orbis) versus observed in the CBCR data

differences in the case of some countries. We therefore multiplied all reported financial information by a ratio listed in Table 3 in case that ratio was above one, with the exception of two countries described below.

We address the lack of completeness in the data of two reporting countries, the United States and China, in a specific way. In the case of the United States, we expected 1,501 companies according to Orbis. Instead, we find 1,101 companies in the 2016 data. This is due to a lack of completeness of 2016 data in the United States (Garcia-Bernardo et al., forthcoming). US Internal Revenue Service data for 2017 indicate that we should observe approximately 1,575 companies–1,548 with profits in at least one jurisdiction. In order to correct for this disparity we use US data for 2017. In China, instead of the expected 583, only 82 companies reported satisfactory data to the OECD. However, those 82 companies reported \$2.9 tn of sales domestically, and \$0.45 tn abroad; for comparison, the numbers for the United States in 2016 were \$7.8 tn domestically and \$3.4 tn abroad. This indicates that the data is not as erratic as it may appear. Lacking a better heuristic, we multiply the financial information for China by a conservative factor of two[1].

The second limitation concerns the combinations of countries in aggregated categories (e.g. Other Africa, Europe). The aggregation criterion is different for different countries. While India and South Africa do not seem to aggregate data, the United States aggregates countries with a low number of reporting MNCs. This is problematic as aggregation affects particularly low- and middle-income countries and low tax jurisdictions. For instance only three countries report information on Zambia and only two countries report on the Isle of Man. The other countries aggregate information on Zambia and the Isle of Man in larger categories such as Other Africa and Other Europe. If we decided to ignore these grouped data, we would be missing a significant part of the operations in those countries, leading to an underestimation of the extent of profit shifting.

We address theses biases by modelling the location of employees and sales for each pair of countries using the Histogram-based Gradient Boosting Regression Tree, a type of gradient boosting based on decision trees which frequently outperforms other machine learning algorithms while offering some interpretability on the most relevant variables (Friedman, 2001; Ke et al., 2017). Specifically, we use the Python implementation in scikit-learn (Pedregosa et al., 2011). Another of its advantages is that it offers native support for missing values, and as such is able to use the full available information without data

[1] We run a robustness test in which the data of China was not adjusted. This increases total profit shifting by 5%, especially towards China and Macao, and away from the United States and the United Kingdom.

imputation. We train the location of profits, employees and sales using variables from the gravity data set of CEPII, imports and exports from UN Comtrade, and foreign direct investment from the World Bank as well as from other sources. We obtain a median out-of-sample R-square of 0.49, 0.50 and 0.51 respectively for employees, sales and profits.

We use the model to estimate the total number of employees and unrelated party sales for each pair of countries in the world. For reporting countries, we then adjust the estimated values so their sum corresponds to the aggregated sum in CBCR. Let us demonstrate using the following model scenario: French MNCs have 10,000 employees in Other America, and Other America comprises Paraguay and Suriname – we can establish this by checking which countries are missing from the CBCR data of France. If our model estimates 6,000 employees in Paraguay and 5,000 employees in Suriname, we multiply the employees of those countries by 10,000 and 11,000 respectively. For each country, we compare the sums of those estimated values with values observed in the CBCR data. We then use the lowest of the two ratios (estimated vs reported employees and sales) to adjust the profits shifted in order to correct for the combination of small countries in aggregated groups. While this step typically increases total shifted profits by approximately 30%, it is key with respect to accounting for missing data in countries underrepresented in the sample, i.e. typically low- and middle-income countries. Without this step, we would redistribute too few profits to those countries. Figure 2 shows the available information on CBCR, displaying how data coverage is especially worrisome in the case of low- and middle-income countries.


**Figure 2.** Available information on CBCR. Colour denotes increasing GDP per capita. Countries with availability below 10% are annotated.

The third limitation concerns the lack of reporting by some countries including e.g. Germany, Spain, and the United Kingdom. This is partially addressed in the previous step, where financial information for all pairs of countries is estimated even for non-reporting countries. However, domestic information is important, especially for large countries. This is addressed by estimating the number of domestic employees and revenue for all non-reporting countries. We do so by using a linear model based on the number of expected companies in each country, its GDP, population, the ETRs and the total consolidated banking claims on an immediate counterparty basis (Table B4 of the BIS data) (R-square 0.91, 0.98 respectively for employees and sales. We only use this information to redistribute profits back to the home countries but not to calculate profit shifted. This is a conservative strategy since domestic profits of companies in non-reporting countries with low tax rates (e.g. the British Virgin Islands) do not counted towards the estimate.

Finally, we assess our results' sensitivity to the estimation of missing information. To do so, we train the models 1,000 times using bootstrapped samples of the data (i.e. the gradient boosting ensemble to address the second limitation and the linear regression to address the third limitation) and record the impact in our results. Since the sampling randomly removes information, samples without important dyads (e.g. USA–Netherlands, or China–Hong Kong) will be heavily affected. We thus offer a conservative strategy allowing us to partially understand how our results depend on methodological choices. In the end, we use the median value for our point estimates.

## References

Bilicka, K. A. (2019). Comparing UK Tax Returns of Foreign Multinationals to Matched Domestic Firms. *American Economic Review*, *109*(8), 2921–2953. https://doi.org/10.1257/aer.20180496

Blouin, J., & Robinson, L. A. (2020). *Double counting accounting: How much profit of multinational enterprises is really in tax havens?* (SSRN Scholarly Paper No. ID 3491451). Social Science Research Network. Rochester, NY. Retrieved December 5, 2019, from https://papers.ssrn.com/abstract=3491451

Bolwijn, R., Casella, B., & Rigo, D. (2018). Establishing the baseline: Estimating the fiscal contribution of multinational enterprises. *Transnational Corporations*, *25*(3), 111–142. https://doi.org/10.18356/de4a3b2c-en

Clausing, K. (2020a). Five Lessons on Profit Shifting from the US Country by Country Data. *Tax Notes Internationa and Tax Notes Federal*.

Clausing, K. A. (2020b). Profit Shifting Before and After the Tax Cuts and Jobs Act. *National Tax Journal*. https://doi.org/10.2139/ssrn.3274827

Cobham, A., & Janský, P. (2020). *Estimating Illicit Financial Flows: A Critical Guide to the Data, Methodologies, and Findings*. Oxford University Press. https://doi.org/10.1093/oso/9780198854418.001.0001

Damgaard, J., Elkjaer, T., & Johannesen, N. (2019). What Is Real and What Is Not in the Global FDI Network? *IMF Working Paper*, *2019*(274), 1–51.

Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, *29*(5), jstor 2699986, 1189–1232.

Garcia-Bernardo, J., Jansky, P., & Tørsløv, T. (forthcoming). Multinational Corporations and Tax Havens: Evidence from Country-by-Country Reporting. *International Tax and Public Finance*.

Haberly, D., & Wójcik, D. (2015). Tax Havens and the Production of Offshore FDI: An Empirical Analysis. *Journal of Economic Geography*, *15*(1), 75–101. https://doi.org/10.1093/jeg/lbu003

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett, Eds.). In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems 30*. Curran Associates, Inc. Retrieved October 5, 2020, from http://papers.nips.cc/paper/6907-lightgbm-a-highly-efficient-gradient-boosting-decision-tree.pdf

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

UNODC, & UNCTAD. (2020). *Conceptual Framework for the Statistical Measurement of Illicit Financial Flows*. https://www.unodc.org/documents/data-and-analysis/statistics/IFF/IFF_Conceptual_Framework_for_publication_15Oct.pdf