

THE *ICTD* GOVERNMENT REVENUE DATASET (GRD)

USER GUIDE & FAQs

Introduction:

The ICTD GRD was developed due to the weaknesses in terms of transparency, coverage and accuracy of existing data on government revenues. It combines data from a number of different international sources under a standard classification system, leading to demonstrable gains in both coverage and accuracy compared to any other single source. The GRD is constructed and presented in a transparent manner that allows users to track each figure back to its original source. This guide aims to assist users to understand the construction of the dataset and how best to make use of it. More detailed information on the construction of the dataset can be found in [Prichard *et al.* \(2014\)](#).¹

Data sources:

The GRD combines data from the following sources:²

- OECD *Revenue Statistics*
- OECD *Latin American Tax Statistics*
- IMF *Government Finance Statistics* (GFS)
- IMF Article IV Staff Reports
- CEPALSTAT *Revenue Statistics in Latin America*

For a large number of countries, data is available from more than one source. One of the key strengths of the GRD is that the authors have examined data from each underlying source and hand-picked the best source based on a number of criteria.³ In general, data that

- combines a long and consistent time series and high levels of disaggregation of sub-categories of revenue;
- accounts for natural resource revenues (whether classed as *tax* or *non-tax*) as distinct from other types of tax revenue
- extracts social contributions from tax

is preferred to data that does not. Furthermore, in the ‘*Merged*’ dataset, General Government data is preferred to Central or Budgetary Central Government, all else being equal;

There is, however, no strict preference ordering that can be applied to all countries. Each case has been considered in isolation, in order to ensure that the best data is included. The

¹ Prichard, W., Goodall, A. and Cobham, A. (2014) ‘*The ICTD Government Revenue Dataset*’, ICTD Working Paper 19, Brighton: International Centre for Tax and Development.

² In a very limited number of cases, data from individual countries’ public finance records are used as a supplementary source. Where applicable, this is noted in the GRD. A variety of additional sources were also consulted in constructing the data, but are not employed in the final dataset.

³ Data from all of those underlying sources – as well as noted on data choices where relevant - is available on request, in order to increase transparency.

inconsistent nature of revenue data across sources means that applying a one-size-fits-all rule or preference order would lead to a dataset with significant gaps or jumps.

The Dataset: General info

- The GRD is available from www.ictd.ac as a Microsoft Excel Spreadsheet (.xls), Open Document Spreadsheet (.ods) or Stata data file (.dta).
- The ‘*CentralGeneralMerged*’ file selects the best data for each country from either ‘Central’ or ‘General’ government, placing it into one ‘Merged’ sheet.
 - The ‘*CentralGeneralMergedFull*’ file includes an option for users to alter the choice between Central and General for each country.
- Those files suffixed by ‘Full’ are significantly larger in size and in general run slower on most computers, but include all the original formulae for expressing the data as % of GDP.
- Descriptions of all sheets in each file are included under ‘Variables & Info’. This includes information of flags that are included in the dataset to identify potentially problematic data, and it is essential that users exclude problematic data appropriately – and report such exclusions alongside econometric results.
- Notes on, for example, the choice of source or the calculations made using original source data for each country are included in the underlying *CentralGovernment* and *GeneralGovernment* files.
- The country codes, region codes and income codes included in the GRD allow for easy merging with other major sources, such as the World Development Indicators, for econometric analysis.
- The GRD is updated periodically, based on (i) changes or updates to the underlying data and (ii) feedback from users who are able to provide expertise or guidance with regard to particular countries or regions. In keeping with the wider aims of transparency and openness, users are encouraged to contact ICTD with feedback and suggestions where they believe the data to be questionable.

Frequently Asked Questions

1. *Should analysis employ data from the merged, general government or central government dataset?*

There is no universal answer to this question, as it will depend on the nature of the research question. That said, for studying total revenue collection most users are likely to

use the *merged* dataset, which employs general government data where available, while inserting central government data if, and only if, there is evidence that subnational revenue collection is limited. By contrast, relying exclusively on central government data will dramatically underestimate total revenue collection in fiscally decentralized states, while general government data is only available for a limited group of countries. While relying on the merged dataset implies a modest underestimation of revenue collection in countries for which general government data is not available, this minor estimation will generally be consistent over time, and generally preferable to the major risks introduced by relying exclusively on central or general government data.

2. *Should analysis include or exclusive social contributions from total revenue and total tax revenue?*

Conceptually, the answer is ambiguous and highly dependent on the research question: social contributions are not the same as other taxes, as they are contributions toward a specific area of public spending, but they nonetheless represent an important area of government involvement. Pragmatically, however, most users are advised to rely on revenue and tax figures exclusive of social contributions, owing to problems of completeness and comparability for social contributions figures. Specifically, many countries do not consistently report social contributions, leading to more extensive missing data, while what is reported in the ‘social contributions’ category does not appear to be entirely consistent across countries, particularly for developing countries and when relying on IMF Article IV reports, thus raising concerns that including social contributions in the analysis could produce misleading results.

3. *If I am interested in resource revenues, should I use the “total resource revenue” variable, or an alternative strategy to measure resource dependence?*

We would ideally be able to produce a figure for “total resource revenue” for every country, which would capture all resource revenues, whether those revenues are recorded as tax or non-tax revenue in official government accounts. However, in practice, this is simply not possible: Relatively few countries report a figure for “total resource revenue”, with some additional countries identifying the resource components of total tax revenue, or of total non-tax revenue, but not both. Current international initiatives to strengthen the collection of resource revenue data are correspondingly extremely valuable.

In the absence of such data, users may nonetheless wish to include a proxy for resource wealth in econometric analyses. In these cases the best option available to most users is likely to be to calculate “total non tax revenue” as the difference between “total revenue” and “total non-resource tax revenue”. This variable will thus include all types of non-tax revenue, including resource revenue, but acts as a useful proxy for resource revenue

because resource revenue explains the vast majority of the variation in “total non tax revenue” across countries. See discussion in Prichard et al. (2014).⁴

Note that the “total non-tax revenue” variable described here is calculated for users at the far right of the Excel file containing the ICTD GRD, but is distinct from the variable “composite non-tax revenue” in the dataset. The latter generally includes only the revenue that states themselves record as non-tax revenue, while the former also includes any resource revenue that states record as tax revenue.

4. *Why do the subcomponents of tax revenue not sum to the respective totals?*

- (i) Total tax revenue is generally equal to the sum of the sub-components of tax revenue. However, the total tax figure exceeds the sum of the sub-components in cases where some revenue was not allocated to any specific category in the underlying source, which may occur for a variety of reasons. Thus, for example, Total Taxes on Income, Profits and Capital Gains is sometimes larger than the sum of CIT + PIT, as some sources allocate some revenue to “other income taxes”, which cannot be allocated to either personal or corporate income taxes.
- (ii) Occasionally, the subcomponents of revenues are slightly greater than or smaller than the reported totals in the underlying source. The GRD, as a rule, does not modify data from underlying sources, thus retaining these minor discrepancies in order to ensure transparency.⁵ If users wish to correct such data for their own analyses, they are free to do so, though changes of this kind to the data should be carefully recorded for any analysis.
- (iii) For many EU countries, the data in the GRD comes from the OECD Revenue Statistics. The total tax figure in such cases is *inclusive* of ‘Customs duties collected for the EU’, reflecting the ‘Total Tax Revenue’ figure in the OECD. However, this item is not allocated to any subcomponent, meaning that the sum of subcomponents will often be marginally lower than the total tax figure.
- (iv) Users should note that detailed descriptions of all variables are included in the ‘Variables & Info’ sheet of the .xls and .ods files. Only limited descriptions have been included in the variable labels of the .dta files.

⁴ Prichard, Wilson, Paola Salardi, and Paul Segal (2014), Taxation, Non-Tax Revenue and Democracy: New Evidence Using New Cross-Country Data, *ICTD Working Paper 23*, Brighton: Institute of Development Studies.

⁵ The only exception being where the surrounding data all sums and there is a glaringly obvious error. This would be noted in the notes column of the underlying file.

5. *Why does revenue data in the GRD not match that in other sources?*

- (i) The GRD expresses all data in terms of a common underlying GDP figure, taken from the World Economic Outlook (WEO).⁶ This helps to ensure comparability of data across different sources, but might be higher, or lower, than tax ratios expressed in other sources, based on different GDP data.
- (ii) The underlying GDP figures taken from the 2015 version of the WEO reflect the adoption by many countries of SNA2008 and its European counterpart, ESA2010. In some cases, this has led to large increases in GDP (Nigeria, for example), which might cause users to find lower tax ratios than an older source, which does not express tax figures in terms of up-to-date GDP figures.
- (iii) Tax statistics in other sources may be inflated by revenues earned from natural resources. The GRD, where possible, presents both total revenue and total tax figures both inclusive and exclusive of resource revenues.

6. *Is it possible to estimate total subnational revenue by subtracting Central Government revenue from General Government revenue?*

Generally speaking the difference between General Government and Central Government revenue should be equal to total subnational revenue. However, any such calculations should be undertaken with extreme caution, and checked carefully, as definitional differences, or other unusual accounting practices, between central and general government sources could lead to over or under-estimating subnational revenue.

7. *Data that I am interested in using has been flagged as potentially problematic.*

- (i) The authors of the GRD have striven to alert users to cases where data might be potentially unreliable or misleading, through a series of ‘flags’. These are defined in each data file, and often notes relating to the need for the individual flags are included in the *CentralGovernment* and *GeneralGovernment* files. Users are advised to exclude any data that is flagged as *Prob1 – Prob4* from econometric analysis.
- (ii) Data that is flagged as *Treat with Caution (TWC)* may be used, but very carefully. For example, a cross-country regression that used such data might wish to report results both with and without the data flagged as *TWC*. In cases where results are reliant on the inclusion of such data, users would be wise to a) communicate this to their audience and/or b) exclude the data all together.

⁶ In cases where no data is available in the WEO, GDP figures are taken from the underlying source of the tax data

- (iii) More generally, because of well-known imperfections with the collection of government revenue data – and with GDP data used in some countries to calculate tax revenues – best practice for econometric analysis should always be to pursue extensive robustness testing, including sensitivity to the inclusion of data from individual, or small groups, of countries. Caution should also always be employed when comparing tax ratios across countries, as high, or rapidly rising, tax ratios in individual countries may reflect the underestimation and irregular rebasing of GDP.

8. *What are some of the limitations to the GRD?*

- (i) There are widely recognised concerns over the quality of government revenue data in existing sources. The GRD strives to present the ‘best’ available data consistently and to flag any noticeably problematic or questionable data. This already represents an improvement over any other source. However, the broader accuracy of the numbers contained in the GRD is only as good as the underlying data obtained from each source.
- (ii) In particular, data on natural resource revenues have not historically been collected consistently, or according to strictly enforced definitions. As such, the disaggregation between non-resource taxes and resource taxes is likely to be imperfect at the margin – though still better than the alternative of not making any distinction. There is potential for more precise data in the future, as more systematic data collection efforts are put in place – something toward which both the IMF and OECD are working. In a similar vein, no data on resource revenues is generally available for countries with small levels of resource revenue (generally less than 1% of GDP) and, as such, all revenue is treated as non-resource tax revenue in these cases.
- (iii) There are also potential inconsistencies across countries, or over time, in the disaggregation of tax revenue between taxes on international trade and taxes on goods and services. This reflects potential inconsistencies in the classification of taxes on goods and services collected at the border, by customs officials. While every effort has been made to ensure consistency, users should be aware of this potential issue in individual countries when working with these disaggregated categories.
- (iv) There may be cases where the data in the GRD appears slightly less complete (in terms of disaggregation) than that in, for example, the GFS. This would be as a result of using IMF Article IV data that, whilst not as detailed, allows the resource component of tax to be isolated.

The ICTD welcomes all feedback from users of the GRD. Queries should be directed to

Wilson Prichard (Research Director, ICTD) – wilson.prichard@utoronto.ca

Kyle McNabb (Research Assistant) - kyleianmcnabb@gmail.com