

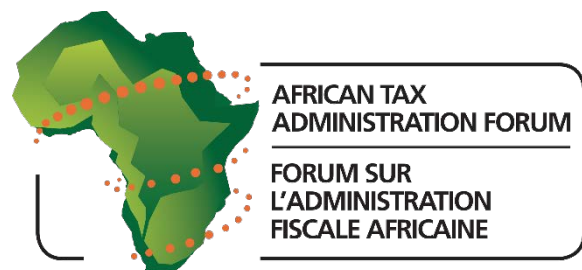
WIDER Working Paper 2024/13

# Tax revenue data in Africa: the Government Revenue Dataset and African Tax Outlook in comparison

Frankie Mbuyamba<sup>1</sup> and Kyle McNabb<sup>2</sup>

March 2024

In partnership with



**Abstract:** This paper compares two important sources of tax revenue statistics for African countries, namely the Africa Tax Administration Forum’s African Tax Outlook and the United Nations University World Institute for Development Economics Research Government Revenue Dataset. We consider the background, construction, sources, and user bases of each dataset before comparing the scope and attempting to understand where commonalities and differences lie. The bulk of the paper focuses on a quantitative comparison of ten key variables across the two datasets. We find that, although in principle the same concepts are captured across both, there are myriad differences in the values of key revenue variables between the two sources. We look in depth at three causes of such differences, namely the treatment of withholding taxes, revenue from the Southern African Customs Union revenue sharing agreement, and natural resource revenues.

**Key words:** data, development, tax, revenue statistics, Africa

**JEL classification:** H20, H30, F63, O11, O23

**Acknowledgements:** This study is an outcome of the collaboration between the African Tax Administration Forum (ATAF) and the United Nations University World Institute for Development Economics Research (UNU-WIDER).

The views expressed herein are those of the author(s) and do not necessarily represent the views of ATAF, the United Nations University, UNU-WIDER, or those of their Executive Directors or their affiliated organizations.

---

<sup>1</sup> African Tax Administration Forum, Pretoria, South Africa; <sup>2</sup> ODI, London, UK, corresponding author email [k.mcnabb@odi.org.uk](mailto:k.mcnabb@odi.org.uk)

This study has been prepared within the UNU-WIDER project [GRD – Government Revenue Dataset](#). It is part of UNU-WIDER’s [Domestic Revenue Mobilization \(DRM\)](#) programme, which is financed through specific contributions by the Norwegian Agency for Development Cooperation (Norad).

Copyright © UNU-WIDER 2024

UNU-WIDER employs a fair use policy for reasonable reproduction of UNU-WIDER copyrighted content—such as the reproduction of a table or a figure, and/or text not exceeding 400 words—with due acknowledgement of the original source, without requiring explicit permission from the copyright holder.

Information and requests: [publications@wider.unu.edu](mailto:publications@wider.unu.edu)

ISSN 1798-7237 ISBN 978-92-9267-471-7

<https://doi.org/10.35188/UNU-WIDER/2024/471-7>

Typescript prepared by Lesley Ellen.

United Nations University World Institute for Development Economics Research provides economic analysis and policy advice with the aim of promoting sustainable and equitable development. The Institute began operations in 1985 in Helsinki, Finland, as the first research and training centre of the United Nations University. Today it is a unique blend of think tank, research institute, and UN agency—providing a range of services from policy advice to governments as well as freely available original research.

The Institute is funded through income from an endowment fund with additional contributions to its work programme from Finland and Sweden, as well as earmarked contributions for specific projects from a variety of donors.

Katajanokanlaituri 6 B, 00160 Helsinki, Finland

The views expressed in this paper are those of the author(s), and do not necessarily reflect the views of the Institute or the United Nations University, nor the programme/project donors.

## 1 Introduction

The recent focus on domestic resource mobilization in low- and middle-income countries (LMICs) has led to a plethora of research that attempts to better understand the causes—and consequences—of revenue effort across countries. To carry out such work in a meaningful manner, one requires, naturally, good quality, cross-country data. For African countries the availability of cross-country data has markedly improved over recent years. The UNU-WIDER Government Revenue Dataset (GRD),<sup>1</sup> for example, has made great strides in making comparable cross-country revenue data available to researchers in a transparent and accessible manner. Similarly, and more recently, the OECD’s Revenue Statistics in Africa have, year by year, improved in coverage and consistency. Another recent advent is the African Tax Administration Forum (ATAF) African Tax Outlook (ATO). The ATO, which has grown to cover 37 countries as of 2023, strives to present a set of indicators that are meaningful to key user bases, such as the tax administrations of African countries.

This paper compares the UNU-WIDER GRD and ATAF ATO.<sup>2</sup> In Section 2, we discuss the background, construction, sources, and user bases of each dataset before attempting to map the scope of each dataset and understanding where commonalities and differences lie. This includes matching variables that should capture the same concepts across both datasets. In Section 4 we carry out a quantitative comparison of ten key variables that were matched across the two databases. We find that, although in principle the same concepts are captured across both, there are myriad differences in the values of key revenue variables between the two sources. Section 5 examines three causes of such differences in depth, namely the treatment of withholding taxes, revenue from the Southern African Customs Union (SACU) revenue sharing agreement, and natural resource revenues. We discuss the issues and provide guidance for users seeking to make better use of the two databases. Section 6 concludes.

## 2 Background, construction, and sources

### 2.1 Government Revenue Dataset (GRD)

#### 2.1.1 Background

The UNU-WIDER GRD—originally constructed by researchers at the ICTD—was launched in 2014. The initial construction of the GRD is described in detail in Prichard et al. (2014) and the data has since been described in depth in a series of further articles and technical notes that pertain to (i) specific changes made in the 2017 version (McNabb 2017); (ii) source selection (McNabb et al. 2021); (iii) variable description (Oppel et al. 2021); and (iv) country-specific notes (McNabb and Oppel 2021).

A major motivating factor behind the construction—and the continued production—of the GRD was/is to provide a dataset of comparable cross-country revenue statistics for a large sample of countries over a long time period. Prior to its construction few attempts had been made to

---

<sup>1</sup> Originally launched by and hosted at the International Centre for Tax and Development (ICTD).

<sup>2</sup> Where appropriate we also draw from other sources such as the International Monetary Fund (IMF) or Organisation for Economic Cooperation and Development (OECD) to provide additional context.

harmonize or understand the differences and similarities between revenue data from different providers (e.g., the IMF, OECD, CEPALSTAT, etc.), which led to a situation whereby questions over the quality (and coverage) of publicly available comparable, cross-country revenue data for LMICs were pervasive. Since its launch almost a decade ago in 2015, the GRD has been employed in countless research articles, donor reports, and technical documents. Following the release of the GRD, both the IMF and OECD launched similar efforts in the form of the World Revenue Longitudinal Dataset (WoRLD) and Global Revenue Statistics Database (GRSD) respectively. There has also, however, been a marked improvement in both the coverage and quality of the underlying data used in the GRD, largely owing to the efforts of the OECD to support the enhanced production of comparable revenue statistics in Latin America, Africa, and the Asia-Pacific region.

### 2.1.2 Construction and sources

The GRD includes data on tax and other government revenue streams across countries. It aims to give as detailed a picture of revenue collection as possible, and presents data according to the following breakdown:<sup>3</sup>

- Total revenue
  - Total taxes
    - direct taxes and subcomponents
    - indirect taxes and subcomponents
  - Non-tax revenue
  - Social contributions
  - Grants

The data is available in percentage of gross domestic product (GDP) or in nominal local currency. The data is presented at both the general and central government levels, with a ‘merged’ dataset presenting the ‘best’ of the two: typically, general government data is preferred as it paints a fuller picture of total government revenue, but this is often missing for LMICs.

The GRD project does not involve the collection of any ‘raw’ revenue data itself from countries. Rather, all of the data contained within it is publicly available through portals such as the IMF’s Government Revenue Statistics (GFS) and OECD’s Revenue Statistics contained in documents such as the IMF Article IV Staff Reports or, occasionally, from individual country sources (e.g., documentation on the websites of revenue authorities, statistics bureaus, or ministries of finance). However, one major challenge is that these sources are often not directly comparable with one another. The GRD attempts to overcome this by meticulously ensuring that the data is presented as consistently as possible across countries.

Often, revenue data from two sources will paint a slightly different story, either in terms of (i) level or (ii) composition. By *level* we mean that one source might report a higher or lower revenue figure than another, and by *composition* we mean that, even if two sources report comparable totals,

---

<sup>3</sup> A more detailed description of the exact variables contained can be found in Oppel et al. (2021).

different statistical guidelines and interpretations might mean that revenue streams are reported under different headings. Consider a brief comparison between the OECD’s Revenue Statistics and the IMF’s GFS, the two main sources that fill the GRD. According to its *Revenue Statistics Interpretative Guide*, the OECD reports social security contributions under ‘total taxes’ (and as a subcomponent of taxation) whilst, according to the *Government Finance Statistics Manual (GFSM) 2014*, the IMF does not, and instead reports social security contributions as a separate item under total revenue but not as a tax. Setting aside the question of which is a preferable indicator for understanding total tax collections across countries, this sort of inconsistency across sources makes comparisons somewhat difficult and an analyst seeking to compare the tax GDP ratio for a given country/year between the two sources will need to proceed with caution. At the very least any merging of ‘total taxes’ data from those two sources for a given country would not be possible as they capture quite different subsets of revenues.<sup>4</sup>

If comparing tax ratios between two sources, a further inconsistency can arise through the denominator—namely nominal GDP—which may also differ across sources. A larger (smaller) GDP figure leads to a lower (higher) tax ratio, *ceteris paribus*. Thus it is essential to compare tax ratios across sources using the same underlying denominator. The GRD incorporates data from all sources in nominal local currency and expresses this as a percentage of a ‘common’ GDP figure, which is taken from the most recent edition of the IMF’s World Economic Outlook.

These two examples highlight the GRD’s approach to overcoming common challenges when comparing cross-country revenue data. Many further—smaller—inconsistencies do appear and these are dealt with on a case-by-case basis and captured in country-specific notes in the GRD files themselves.

### *2.1.3 Key user bases*

The main user base of the GRD is the research community; the dataset presents a rich source of information for anyone seeking to make cross-country, regional, or global comparisons of tax and revenue ratios. It is also a valuable source of data for research which requires a variable that captures tax revenue in any cross-country econometric analysis. The GRD also forms the backbone of the revenue components presented in the World Bank’s Revenue Dashboard, a rich source of information on tax systems globally (World Bank 2023).

## **2.2 African Tax Outlook (ATO)**

### *2.2.1 Background*

In the past decade ATAF initiated the development of programmes that aim at conducting data-driven and evidence-based research to inform member services and domestic resource mobilization. The ATO programme emerged based on the need for reliable tax data from tax administrations in the region. The ATO is an ATAF flagship African publication that provides descriptive and analytical studies on tax issues, and brings together valuable, practical, and relevant information on participating countries.

The ATO has seen its number of participating members grow from 15 countries at the inception in 2016 to a total of 37 countries today. Guided by the ATAF vision and the ATO statement of not leaving any country behind, ATAF collaborates with all African countries in building strong,

---

<sup>4</sup> The GRD overcomes this particular problem by presenting total tax revenues that are both inclusive and exclusive of social security contributions, allowing the user to make the most informed comparisons.

effective, and efficient tax systems through reliable statistics to come up with efficient, effective, and sustainable policy and administrative decisions on taxation. However, these statistics are not always readily available, even within the tax administrations themselves. Knowing about this reality and pushed by the desire to see African tax administrations easily accessing quality information on taxation, the ATO aims to build a solid framework of meaningful indicators that will assist countries to compare, assess, and ultimately improve their tax administrations and revenue performance. For comparison purposes all monetary-related data is available in both local currency and in purchasing power parity (PPP) US dollars, whilst the databank includes nominal and real GDP figures that allow users to express the data as ratios, should they wish.

Amongst its objectives the ATO publication seeks to raise awareness of tax policy and administration amongst the heads and decision makers of its member countries and to promote revenue administration performance measurement. The main focal points in this work are the heads of research/statistics/revenue management who decided on and agreed to the critical indicators used for the ATO publication. Additionally, awareness was raised on the essence of evidence-based policy recommendations and, therefore, the significance of data collection and management within a revenue authority. This makes ATO a source of reliable information that serves as an African, and potentially global, reference as the process encourages full participation and ownership.

### *2.2.2 Construction and sources*

ATAF collects data from its members by using the ATO online data portal. The collected data is accessible directly from the ATO online data portal by all focal points. The impact of the ATO data has gone beyond African borders, which has raised the need for an ATAF databank to give public access for comparison and research purposes. As a result of the increased interest in using the ATO data, along with the ATAF flagship publication, it has become a valuable worldwide source of knowledge for African tax administrations, tax experts in policy-making, development partners, and researchers alike.

The ATO publication assesses and compares participating countries against themes on various categories, and the data is collected annually through the ATAF online data portal and conducted on the following categories:

- **Tax rates:** these provide and describe the rate (usually expressed as a percentage) of taxes collected from a taxpayer directly or indirectly by tax authorities, and include consumption, personal and corporate income, and presumptive taxes;
- **Tax base:** these indicators provide insights into the size and the structure of the economy. They include GDP, the level of final consumption and national debt, the population, and the total number of taxpayers;
- **Tax revenue:** these variables establish the composition of total net tax revenue, which is defined as the revenues collected from taxes on income and profits, taxes levied on goods and services, on wages, on the ownership and transfer of property, and other taxes;
- **Non-tax revenue:** this is the income that government earns from sources other than taxes. There are large sources of non-tax revenue, for instance, when people consume services offered by the government such as electricity, telecommunication and water, income from natural resource earnings/revenue, dividends, interests, fines, fees, and licences;

- **Tax administration:** this covers the organisational profile, cost of revenue collection, customs enforcement and declaration, tax returns, tax arrears and conducting tax audits, and investigation in participating countries;
- **Employee demographics:** this area describes the characteristics of employees in revenue administration. It gives the number of employees and the distribution of employees in relation to gender, education level, and main functions in the revenue administration; and
- **Tax compliance:** the indicators under this theme seek to establish the extent to which revenue and tax administrations reach out to taxpayers through services such as taxpayer education and campaigns.

### 2.2.3 Key user bases

The ATO database and publication are guiding tools which the African tax administrations and all other users are encouraged to utilize in their day-to-day work. The indicators are relevant to African tax authorities, researchers, and all other tax practitioners in understanding the tax landscape on the continent as well as in implementing tax reforms and policies.

## 3 Data mapping across the GRD and ATO

The variables that capture different revenue streams in the ATO (themes 3A, 3B, and 2) and GRD are as shown in Table 1. As a starting point Table 2 takes the ATO data structure and attempts to match comparable GRD variables, whilst Table 3 does the same exercise, taking the GRD as a starting point.

One primary difference in the way the data is presented across the two datasets arises under consumption taxes, such as value-added tax (VAT) or excise duty. The ATO separately reports consumption taxes on imports (e.g., 12. VAT on imports; 14. Excises on imports) and domestic sales (1. Domestic VAT revenue; 3 Excise tax revenue), whilst the GRD reports the sum of consumption taxes on domestic sales and imports under one heading (L1b. VAT and L2 Excises). The GRD, however, follows the GFSM practice of allocating all revenues collected under VAT & excise duty—be they from imports or local sales—under taxes on goods and services.

The ATO also presents data on VAT refunded, and thus, in order to arrive at a comparable VAT figure with the GRD's figure (L1b), it is necessary to sum (1-4+12).

Other notable differences between the datasets are as follows. The ATO separately presents data on presumptive tax collections, whilst the GRD does not. The ATO also presents data on withholding taxes (WHTs). In the ATO, withholding taxes are levied in terms of the Income Tax Act, wherein their rates and reporting obligations are stipulated. Withholding tax may be levied on royalties, interest, goods at importation, or other similar services.

Table 1: ATO and GRD variables<sup>5</sup>

	ATAF ATO	UNU-WIDER GRD <sup>6</sup>
<b>Theme 3A: Tax Revenue</b>	25. Revenue from excise on fuels	
1. Domestic VAT tax revenue	26. Revenue from excise on other goods	A. Total revenue (D+O+P+G); (B+C)
2. Sales tax revenue	27. Revenue from environmental taxes	B. Total resource revenue
3. Excise tax revenue	28. Energy taxes	C. Total non-resource revenue
4. VAT refunded	29. Transport taxes	D. Taxes (G+H)
5. Domestic consumptions tax revenue (1+2+3-4)	30. Pollution taxes	E. Resource taxes
6. Personal income tax revenue	31. Resource taxes	F. Non-resource taxes
7. Corporate income tax revenue	32. Carbon tax	G. Direct taxes (H + I + J)
8. Withholding tax revenue	33. Fossil fuel tax	H. Taxes on income, profits & capital gains (H1 + H2)
9. Presumptive tax revenue	<b>Theme 3B: Non-Tax Revenue</b>	H1. PIT
10. Other domestic tax revenue (specify)	34 Revenue from oil	H2. CIT
11. Total domestic tax revenue (5+6+7+8+9+10)	35. Revenue from other main natural resources (diamond, iron, crude oil, coffee...)	I. Taxes on payroll & workforce
12. VAT on imports	36. Revenue from fees, fines, penalties, and licenses	J. Property taxes
13. Import duty	37. Revenue from royalties	K. Indirect taxes (L + M + N)
14. Excises on imports	38. Non-tax revenue other than social security	L. Taxes on goods and services (L1 + L2)
15. Other customs taxes	39. Total non-tax (34+35+36+37+38)	L1. General taxes on goods & services
16. Customs tax revenue (12+13+14+15)	40. Contributions to the public pension system	L1b. VAT
17. Total tax revenue (11+16)	41. Contributions to the public health system	L2. Excise
18. Revenue from agriculture, forestry and fishing	42. Contributions to unemployment insurance system	M. Taxes on international trade
19. Revenue from mining and quarrying	43. Contributions to other social security systems	M1. Imports
20. Revenue from secondary sector (industry)	44. Total social security contributions (40+41+42+43)	M2. Exports
21. Revenue from tertiary sector (services)		N. Other taxes
22. Domestic revenue from large taxpayers	<b>Theme 2: Tax Base</b>	O. Non-tax revenue (O1 + O2)
23. Revenue from excise on tobacco	45. Total official development assistance (ODA) grants	O1. Resource non-tax revenue
24. Revenue from excise on alcohol		O2. Non-resource non-tax revenue
		P. Social contributions
		Q. Grants

Source: authors' elaboration from UNU-WIDER (2023) and ATAF (2023).

<sup>5</sup> The number (ATO) and letter (GRD) codes have been added for ease of comparison.

<sup>6</sup> NB this is an abridged list; the full set of variables can be found in the GRD (UNU-WIDER 2023).



Table 2: Variable mapping ATO – GRD

<b>Theme 3A: Tax Revenue</b>	
1. Domestic VAT tax revenue	Part of L1b VAT
2. Sales tax revenue	Not captured
3. Excise tax revenue	Part of L2. Excise
4. VAT refunded	Part of L1b VAT
5. Domestic consumptions tax revenue (1+2+3-4)	Not captured
6. Personal income tax revenue	H1. PIT
7. Corporate income tax revenue	H2. CIT
8. Withholding tax revenue	Not captured
9. Presumptive tax revenue	Not captured
10. Other domestic tax revenue (specify)	Not captured
11. Total domestic tax revenue (5+6+7+8+9+10)	Not captured
12. VAT on imports	Part of L1b VAT
13. Import duty	M1
14. Excises on imports	Part of L2. Excise
15. Other customs taxes	Not captured
16. Customs tax revenue (12+13+14+15)	Not captured
17. Total tax revenue (11+16)	D. Taxes
18. Revenue from agriculture, forestry and fishing	Not captured
19. Revenue from mining and quarrying	Not captured
20. Revenue from secondary sector (industry)	Not captured
21. Revenue from tertiary sector (services)	Not captured
22. Domestic revenue from large taxpayers	Not captured
23. Revenue from excise on tobacco	Part of L2. Excise
24. Revenue from excise on alcohol	Part of L2. Excise
25. Revenue from excise on fuels	Part of L2. Excise
26. Revenue from excise on other goods	Part of L2. Excise
27. Revenue from environmental taxes	Not captured
28. Energy taxes	Not captured
29. Transport taxes	Not captured
30. Pollution taxes	Not captured
31. Resource taxes	E. Resource taxes
32. Carbon tax	Not captured
33. Fossil fuel tax	Not captured
<b>Theme 3B: Non-Tax Revenue</b>	
34. Revenue from oil	Part of O1. Resource non-tax revenue
35. Revenue from other main natural resources (diamond, iron, crude oil, coffee...)	Part of O1. Resource non-tax revenue
36. Revenue from fees, fines, penalties, and licenses	Part of O2. Non-resource tax revenue
37. Revenue from royalties	Part of O1. Resource non-tax revenue
38. Non-tax revenue other than social security	Not captured
39. Total non-tax (34+35+36+37+38)	O. Non-tax revenue
40. Contributions to the public pension system	Not captured
41. Contributions to the public health system	Not captured
42. Contributions to unemployment insurance system	Not captured
43. Contributions to other social security systems	Not captured
44. Total social security contributions (40+41+42+43)	P. Social contributions
<b>Theme 2: Tax Base</b>	
45. Total ODA Grants	Q. Grants

Note: key: **Green** = Match; **Amber** = Partial match; **Red** = Not matched.

Source: authors' elaboration from UNU-WIDER (2023) and ATAF(2023).

Table 3: Variable mapping GRD – ATO

UNU-WIDER GRD	Corresponding ATO variable
A. Total revenue (D+O+P+G); (B+C)	<b>Captured: 17+ 23 + 28 + 29</b>
B. Total resource revenue	Partially captured: 18 + 19 + 21
C. Total non-resource revenue	Partially captured: (17 + 23 + 28 + 29) - (18 + 19 + 21)
D. Taxes (G+H)	<b>Captured: 17</b>
E. Resource taxes	<b>Captured: 31</b>
F. Non-resource taxes	Not captured
G. Direct taxes (H + I + J)	<b>Captured: 6 + 7 + 8 + 9 (+ 10?)</b>
H. Taxes on Income, profits & capital gains (H1 + H2)	<b>Captured: 6 + 7</b>
H1. PIT	6
H2. CIT	7
I. Taxes on payroll & workforce	Not captured
J. Property taxes	Not captured
K. Indirect taxes (L + M + N)	<b>Captured: 5 + 16</b>
L. Taxes on goods and services (L1 + L2)	
L1. General taxes on goods & services	
L1b. VAT	<b>Captured: (1-4) + 12</b>
L2. Excise	<b>Captured: 3 + 14</b>
M. Taxes on international trade	Partially captured: 13
M1. Imports	<b>Captured: 13</b>
M2. Exports	Not captured
N. Other taxes	Not captured
O. Non-tax revenue (O1 + O2)	<b>Captured: 17</b>
O1. Resource non-tax revenue	Not captured
O2. Non-resource non-tax revenue	Not captured
P. Social contributions	<b>Captured: 28</b>
Q. Grants	<b>Captured: 29</b>

Note: code: **Green** = Match; **Amber** = Partial match; **Red** = Not matched.

Source: authors' elaboration from UNU-WIDER (2023) and ATAF(2023).

#### 4 Data comparisons: GRD and ATO

In this section we compare the GRD and ATO indicators according to the mapping in Tables 2 and 3. Where the mapping has identified that concepts are matched across both datasets, this represents a situation whereby we should expect the figures to match. We compare ten indicators, namely: total revenue (excluding grants); total taxes; personal income tax (PIT); corporate income tax (CIT); VAT; excise duty; customs duty (imports); non-tax revenue; social contributions; and grants.

In order to ensure comparability, a number of steps are followed:

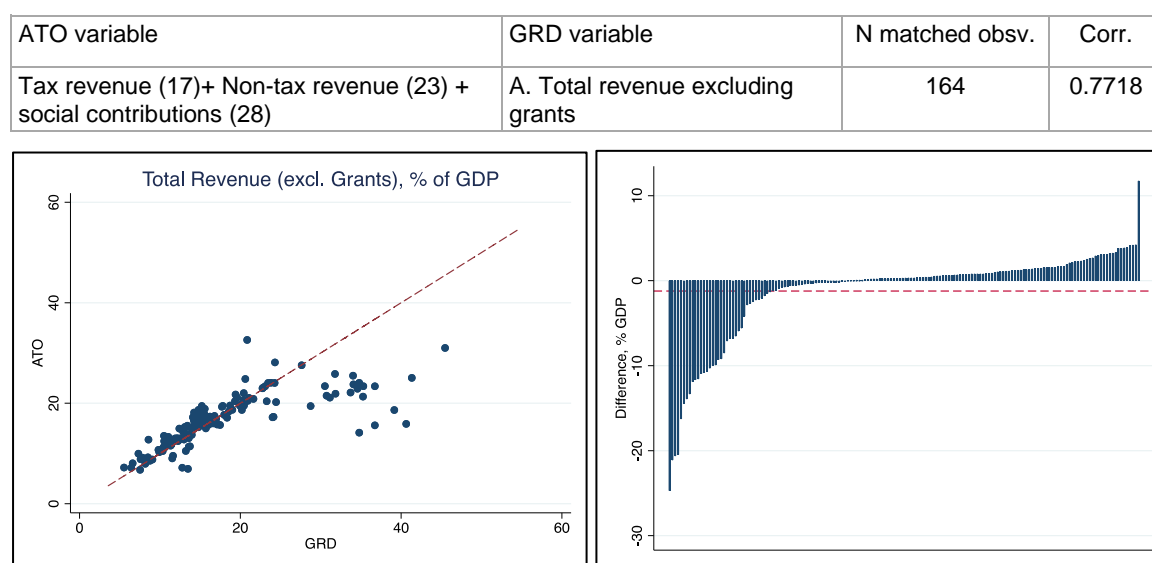
1. All data is firstly accessed in nominal local currency units (LCU).
2. All LCU data is then converted to % of GDP, using the nominal GDP series in current LCU from the IMF's World Economic Outlook April 2023 edition.<sup>7</sup>

Having followed these steps, the expectation is that the same concept in both datasets should be identical, or at least very similar. Any differences that emerge will, then, likely be due to a difference in concepts incorporated or a difference in measurement. For each variable we provide the number of common observations, the correlation coefficient, a scatter chart that plots all of the common observations, and a bar chart showing the difference between the ATO and GRD. This is calculated as the ATO value minus the GRD value; so a negative value for the difference implies that the GRD has a higher value for that indicator. The red dashed line indicates the average difference.

#### 4.1 Total revenue

The first variable we compare is total government revenue, excluding grants. This is constructed as the sum of tax revenue, non-tax revenue, and social contributions in both datasets. We exclude grants from the total revenue calculation due to the very small sample size for that variable (see Section 4.10 below).

Figure 1: Total revenue ATO and GRD



Source: authors' calculations from UNU-WIDER (2023) and ATAF (2023).

We see from the left-hand panel that the two series are quite closely correlated (corr. = 0.77), with many observations clustered around the 45-degree line. However, there are a number of outlying observations where the GRD includes a significantly higher figure than the ATO. This is also reflected in the right-hand panel. Whilst the average difference in the series is -1.22% of GDP, there are a large number of observations where the GRD is more than 5% higher than the ATO.

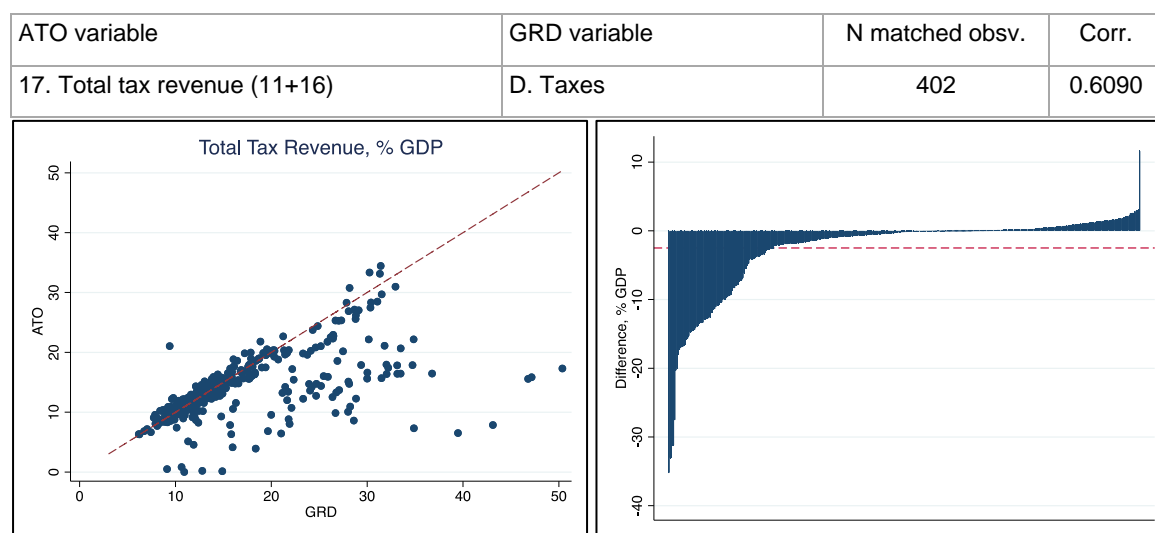
<sup>7</sup> The ATO dataset also contains nominal GDP figures; in reality, the choice to utilize the IMF data over this data is irrelevant for our comparisons, with the important point being that the same figure is used to express the nominal revenue figures for both datasets.

This warrants further investigation below. In particular, an examination of the total tax, social contributions, and non-tax elements of total revenue should help to shed light on this fairly stark difference.

## 4.2 Total tax revenue

Total tax revenue is the variable for which we observe the most ‘common’ data points across sources, at 402. The correlation between the two sources is, however, lower than for total revenue at just 0.61.

Figure 2: Total tax revenue ATO and GRD



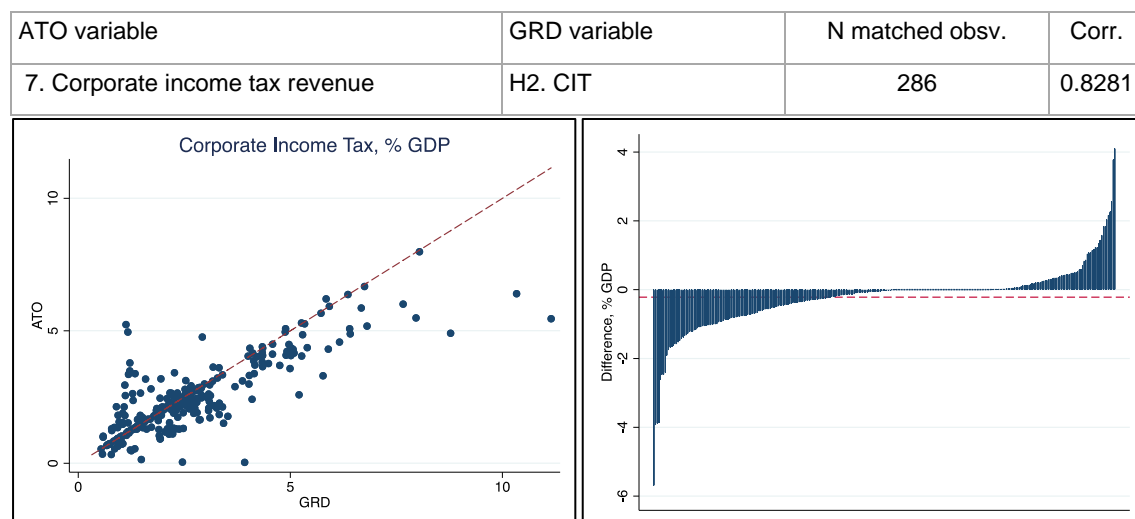
Source: authors' calculations from UNU-WIDER (2023) and ATAF (2023).

We see a similar picture emerge as for total revenue: in a large number of observations, the GRD captures significantly higher levels of tax revenue. The average difference is -2.49% of GDP. In some cases the difference is very large, as much as 20% or 30% of GDP. Discrepancies of this magnitude suggest that there are, for some countries, structural differences in the definition of the underlying concepts.

## 4.3 Corporate income tax (CIT)

There are 286 common observations for CIT, with a correlation coefficient of 0.83. As can be observed, again, there are a large number of observations where there is a high degree of similarity between the two datasets. This is reflected by the low average difference of just -0.22% of GDP. However, again, some large differences emerge at the tails.

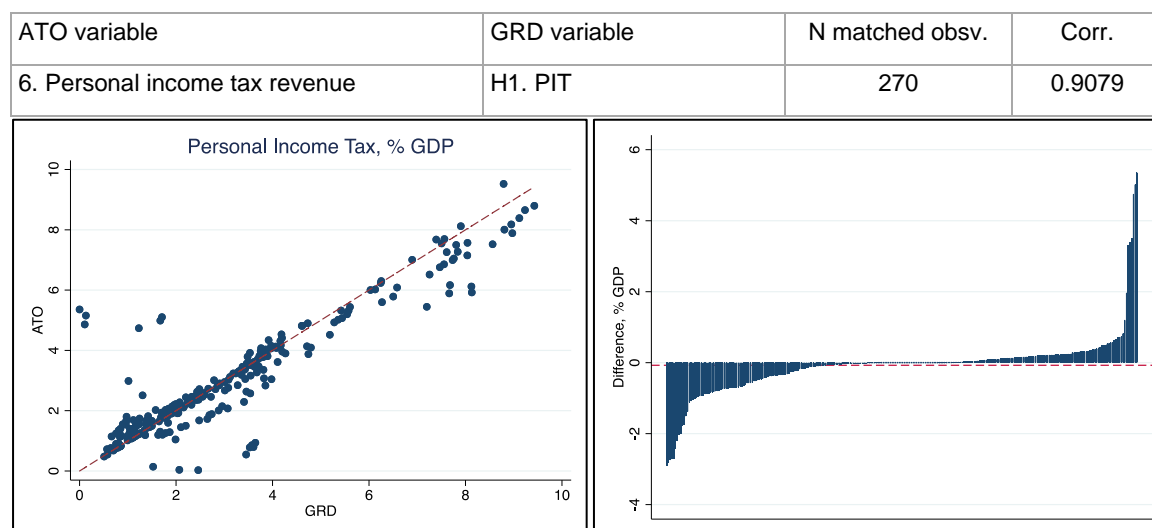
Figure 3: Corporate income tax ATO and GRD



Source: authors' calculations from UNU-WIDER'(2023) and ATAF (2023).

#### 4.4 Personal income tax (PIT)

Figure 4: Personal income tax ATO and GRD



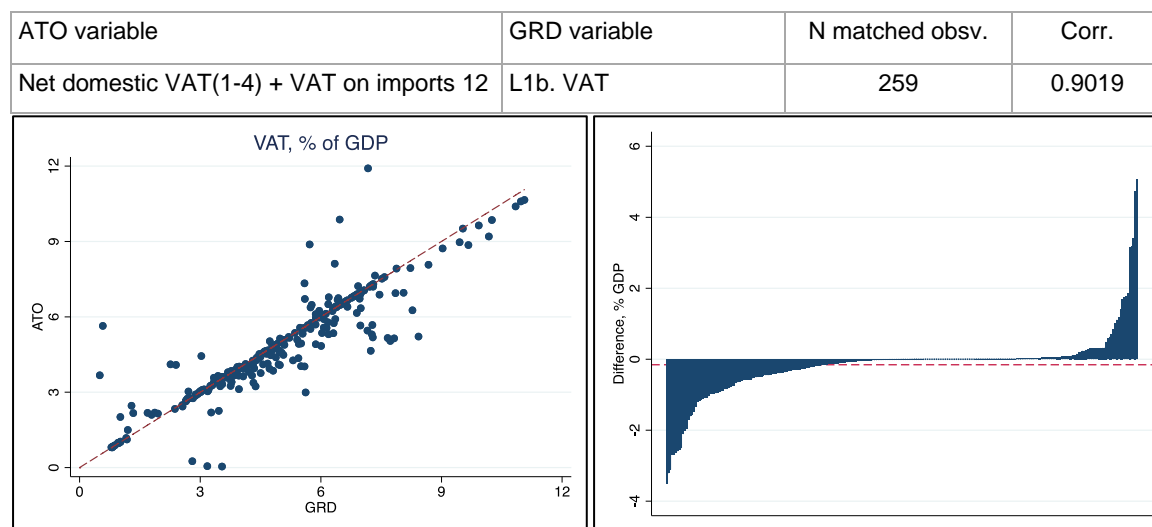
Source: authors' calculations from UNU-WIDER (2023) and ATAF (2023).

Personal income tax is one of the variables where we observe the highest degree of similarity across the two datasets. Across 270 common observations, the correlation coefficient is 0.91, whilst the average difference is just -0.07% of GDP. Again there are a number of observations at the tails, where there are quite large differences between the datasets.

#### 4.5 Value-added tax (VAT)

The diagnostics for VAT are fairly similar to those for PIT. A correlation coefficient of 0.90 over 259 observations and an average difference of just -0.16% of GDP suggest that the two datasets broadly capture the same concepts for VAT.

Figure 5: Value-added tax ATO and GRD

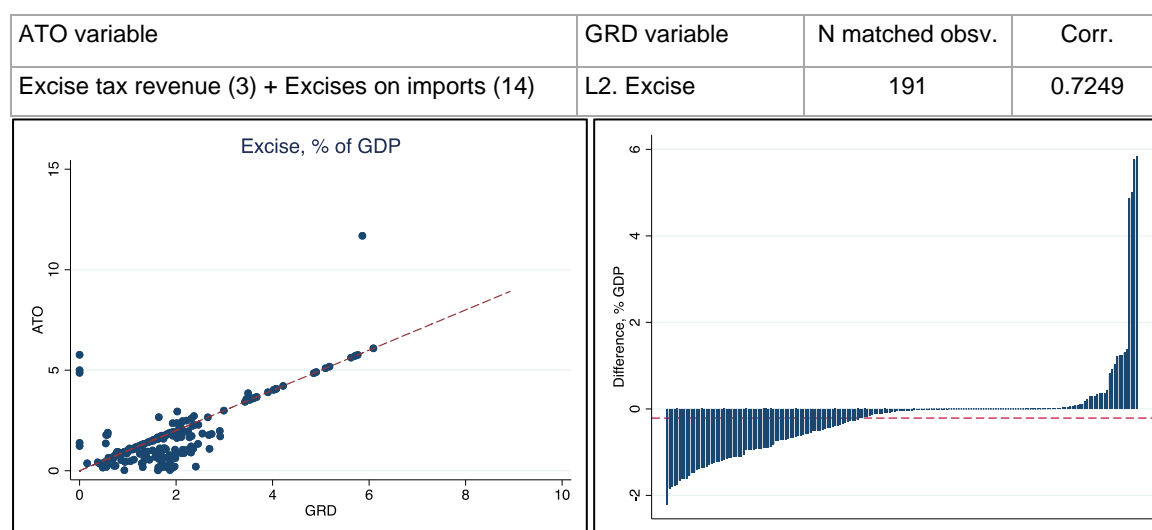


Source: authors' calculations from UNU-WIDER (2023) and ATAF (2023).

#### 4.6 Excise duty

Across 191 common observations the correlation coefficient for excise duty is 0.73. The average difference is small, at just -0.21% of GDP, although again some observations at the tails are very different.

Figure 6: Excise duty ATO and GRD

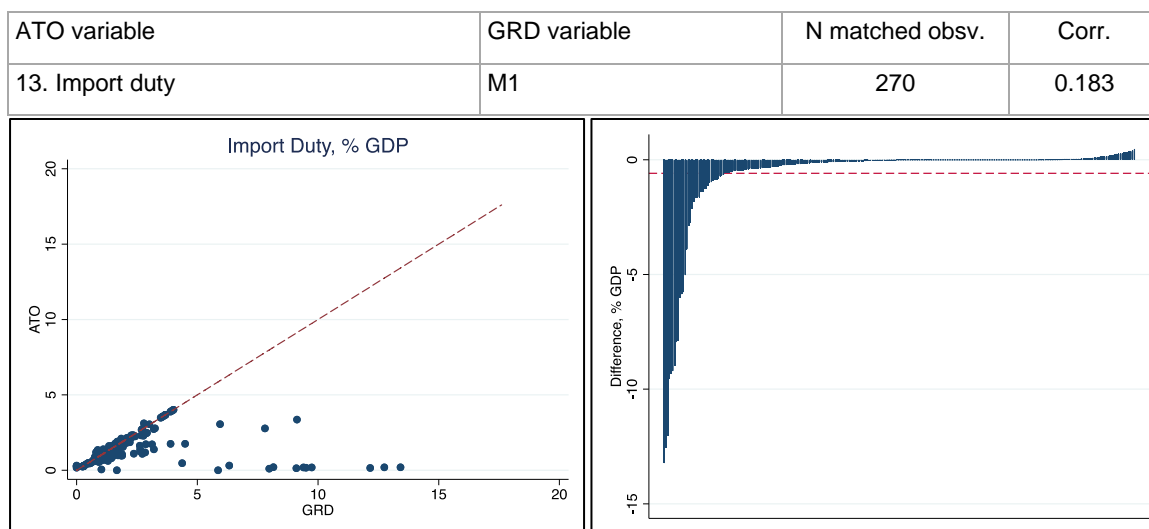


Source: authors' calculations from UNU-WIDER (2023) and ATAF (2023).

#### 4.7 Import duty

Import duty stands out as the one tax subcomponent where the two datasets are most dissimilar. Specifically, across 270 observations, the correlation coefficient is just 0.18, whilst the average difference between the two estimates is -0.6% of GDP. However, at the extreme of the distribution, some estimates of import duty from the GRD are more than 10% of GDP lower than those from the ATO.

Figure 7: Import duty ATO and GRD

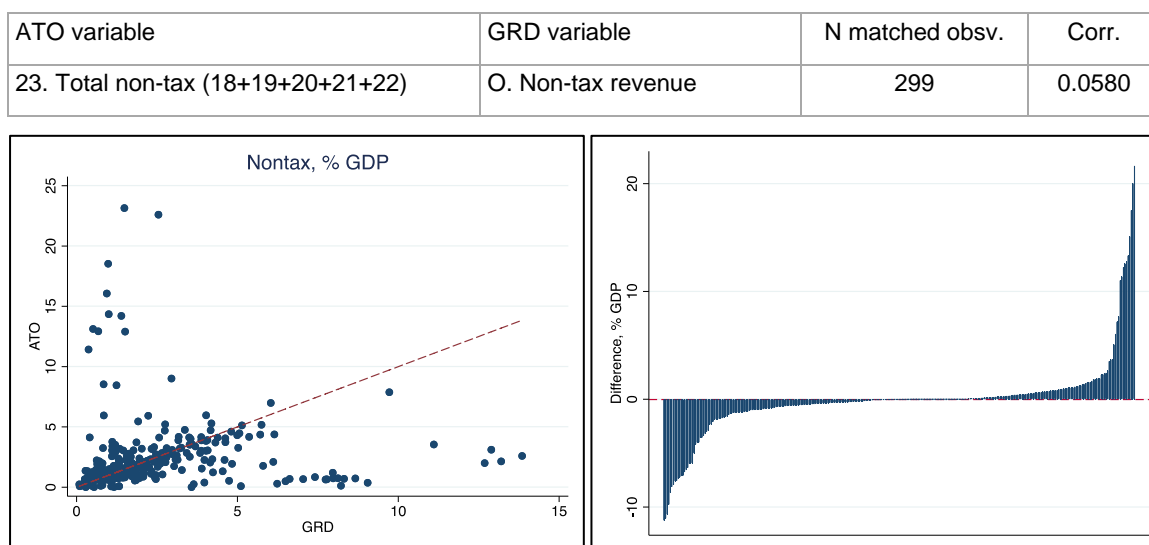


Source: authors' calculations from UNU-WIDER (2023) and ATAF (2023).

#### 4.8 Non-tax revenue

Turning to non-tax revenue, there is a very low correlation across the two datasets ( $\text{corr.} = 0.06$ ), with large differences observed at the tails (many observations where GRD is  $\pm 5\%$  of GDP compared to the ATO). The average difference is very low, at around  $-0.035\%$ , but this is essentially a result of 'balance' between extreme values at either end of the distribution.

Figure 8: Non-tax revenue ATO and GRD



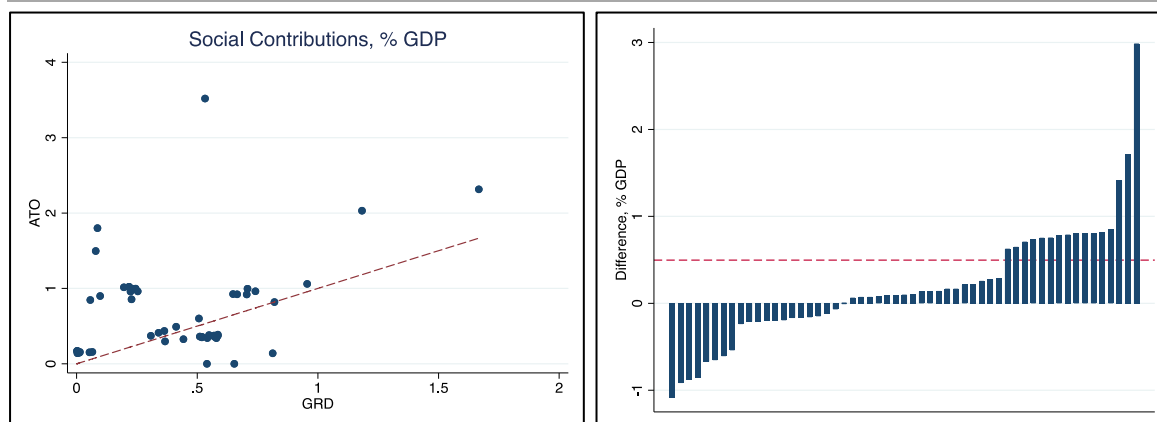
Source: authors' calculations from UNU-WIDER (2023) and ATAF (2023).

#### 4.9 Social contributions

Only a small number of common observations (50) were identified for social contributions and we find a low correlation coefficient of 0.32. The average difference is 0.5% of GDP. Notably, this is the only indicator where the average difference is above zero.

Figure 9: Social contributions ATO and GRD

ATO variable	GRD variable	N matched obsv.	Corr.
28. Total social security contributions (24+25+26+27+28)	P. Social contributions	50 <sup>8</sup>	0.3207

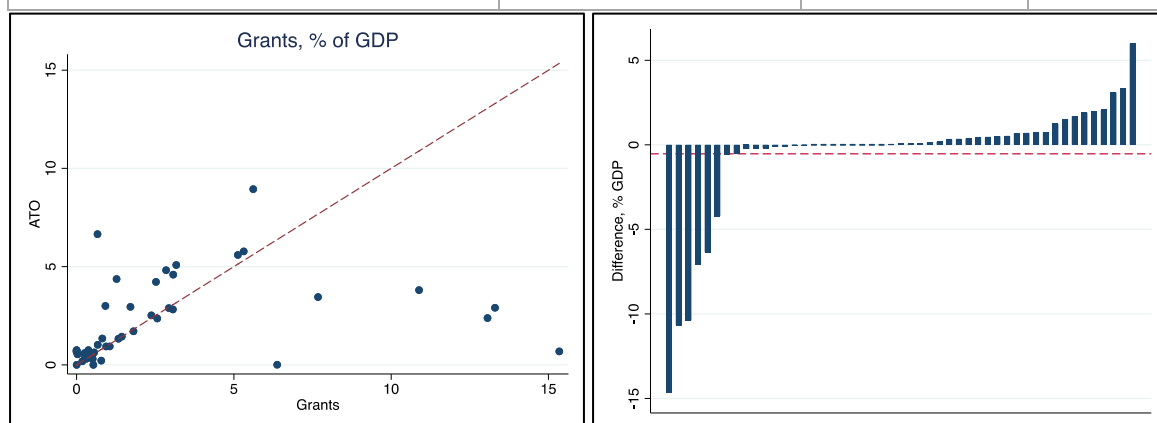


Source: authors' calculations from UNU-WIDER (2023) and ATAF (2023).

#### 4.10 Grants

Figure 10: Grants ATO and GRD

ATO variable	GRD variable	N matched obsv.	Corr.
29. Total ODA/grants	Q. Grants	49	0.3409



Source: authors' calculations from UNU-WIDER (2023) and ATAF (2023).

Finally, we observe a similar result for grants; a correlation of 0.34 from just 49 common observations. The average difference is -0.53% of GDP, with some extreme negative values being observed.

#### 4.11 Conclusions from the comparison

1. Compared to the ATO, it appears that the GRD reports higher figures on average (as denoted by negative bars in the 'difference' graphs) across almost all indicators. This

<sup>8</sup> We restrict the comparison here to observations where a non-zero value is reported in both the GRD and ATO.



suggests that more revenue streams are being accounted for in the underlying GRD source data (normally OECD or IMF).

2. Whilst total revenue shows some major differences (both positive and negative across the two datasets), these differences are not explained by systematically low or high values of any one component part (i.e. taxes, social contributions, or non-tax revenues). There appear to be myriad causes of the disparities observed.
3. The left-skewness of the import duties comparison is quite stark and warrants further investigation.

## 5 Further investigations and user recommendations

The preceding analysis has shown that there are significant differences between the revenues captured in the two datasets and there are many variables that, according to their definitions, should capture the same revenues but do not. The specific reasons for these disparities are too numerous to discuss in depth. However, in this section we focus on three key areas where the source of the disparity was clear and systematic. The discussion for each area should provide sufficient guidance for users seeking to navigate the two databases and employ data on income tax, SACU countries, or natural resource revenues.

### 5.1 Components of income tax

One common source of inconsistency between the GRD and ATO that was identifiable following an examination of underlying data lay with the construction of the various income tax components.

In the GRD taxes on income, profits, and capital gains (**H**) is most often comprised of the sum of corporate (**H2**) and PITs (**H1**). In some rare cases there is a residual amount incorporated in the total but not allocated to either of H1 or H2; the GRD does not explicitly report such revenues as a separate category, but they are calculable by  $[H - (H1 + H2)]$ .<sup>9</sup> Meanwhile the ATO reports significant sums of revenue under WHTs for many observations. This occurs when reporting countries do not allocate the withheld amounts to either of PIT or CIT, as might be the case when reporting to GFS standards, for example. However, the sum of PIT, CIT, and WHT in the ATO should, in theory, be more or less equal to the total value for category **H** in the GRD. We describe two cases where there were large differences, as outlined in Section 4, between the datasets for either PIT or CIT.

Namibia is one country where, in GRD, an income tax ‘residual’ exists. We show the subcomponents of income taxes in Namibia across both datasets for the years 2016–21 in Figure 11. A couple of insights emerge:

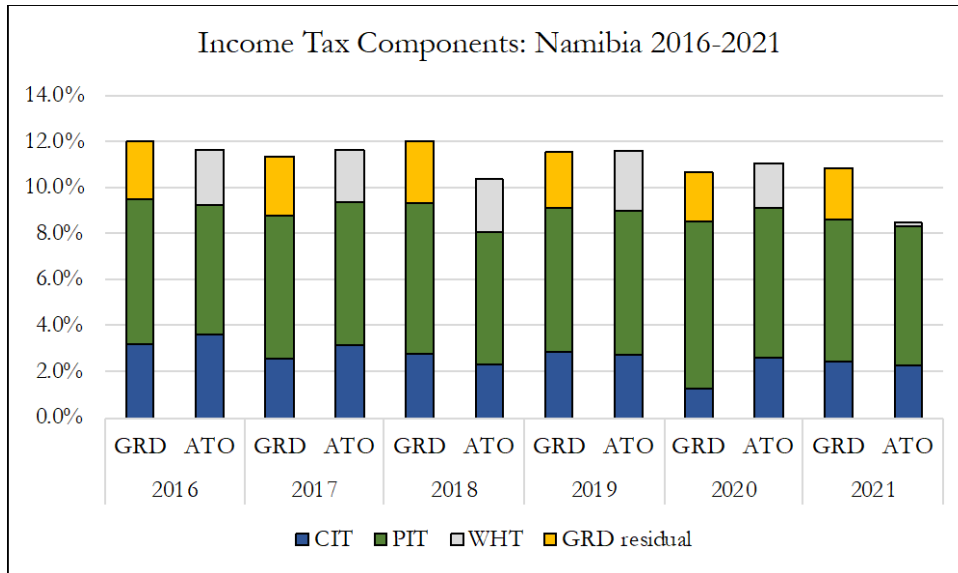
- (i) For some of the years between 2016 and 2020, the total of income taxes in the GRD (including any residual amounts) is roughly equal to the sum of PIT, CIT, and WHT in the ATO. This is true for 2016, 2017, 2019, and 2020.
- (ii) For other years there appear to be structural differences in the amounts recorded. For example, in 2018, the absolute amounts for each of PIT, CIT, and WHT in the ATO

---

<sup>9</sup> Such revenues also appear in the underlying IMF or OECD data as categories 11132 and 1300 respectively.

are all lower than the amounts contained in the GRD. In 2021, whilst the amounts of PIT and CIT in the GRD are fairly close to those in the ATO, there is an almost total collapse in revenue recorded under WHT in the ATO to just 0.2% of GDP, having stood at at least 2% of GDP in the prior years.

Figure 11: Income tax components: Namibia 2016–21

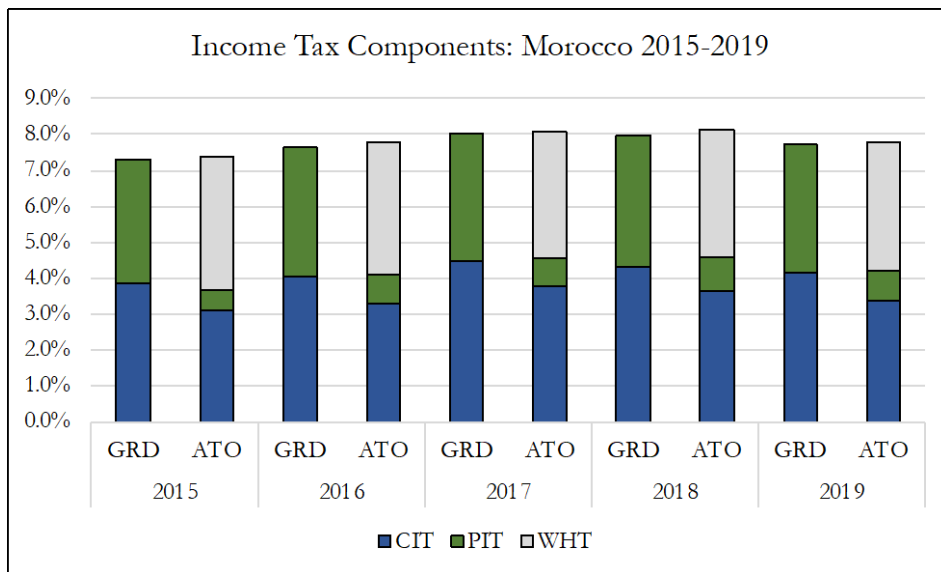


Source: authors' illustration based on GRD and ATO.

A second case is that of Morocco. In this case the total of income taxes (**H**) in the GRD is always equal to the sum of PIT and CIT (and thus no residual values exist). However, Morocco reports substantial amounts of WHT in the ATO. Comparing the two datasets alongside one another, as in Figure 12, it is clear that the WHT revenue is made up largely—but not completely—of PIT.

To classify most of PIT as WHT is not totally incorrect; pay-as-you-earn, as operates in Morocco (and in many other countries), is a form of WHT. However, this example highlights a potential drawback for analysts who wish to compare tax subcomponents across countries within the ATO. Whilst some countries allocate WHT to the appropriate subcomponent (such as PIT), others do not.

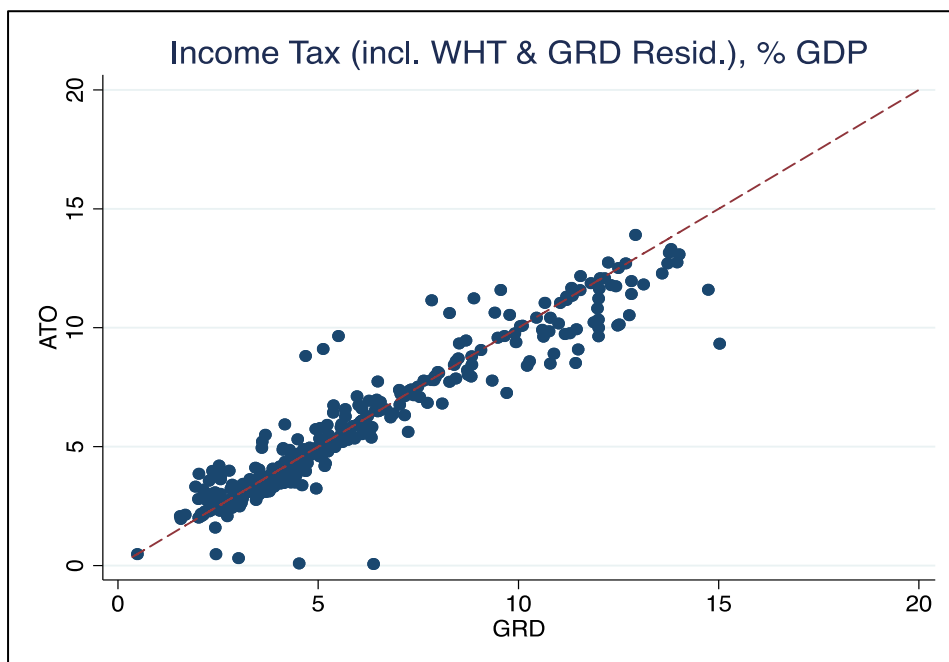
Figure 12: Income tax components: Morocco 2015–19



Source: authors' illustration based on GRD and ATO.

If we compare the two indicators as depicted in Figures 11 and 12 (including both the WHT (ATO) and residual income tax (GRD) components), we find a much stronger correlation than for either of PIT or CIT in isolation (0.95 from 353 observations). Figure 13 highlights.

Figure 13: Income tax (including WHT and GRD residual) as % of GDP



Source: authors' illustration based on GRD and ATO.

## 5.2 Countries belonging to SACU

A key source of the discrepancies uncovered in Section 4 lies with countries that belong to SACU, namely Botswana, Eswatini, Lesotho, Namibia, and South Africa. SACU’s pooled customs revenues are shared between its members according to a predetermined ‘revenue sharing formula’.

The revenue sharing formula comprises three components, as follows (SACU 2021):

- (i) **Customs component:** this is distributed based on the value of each country’s share of ‘intra-SACU’ imports;
- (ii) **Excise component:** 85% of excise duties collected within the SACU, distributed based on members’ share of total SACU GDP per capita; and
- (iii) **Development component:** the remaining 15% of excise duties are distributed based on the inverse of each member’s GDP per capita and weighted to favour the less-developed countries within SACU.

It should, then, be apparent that to compare customs and excise duty revenues of SACU countries with those in other parts of Africa may be very misleading. However, there are also potential pitfalls in even comparing *within* and *across* datasets for SACU countries, as different sources handle the issue in different ways. For the SACU countries, the GRD sources the revenue data from either IMF country reports or the GFS. A simple breakdown of this data for 2021 is shown in Table 4.

Table 4: SACU countries (GRD)

Country	Underlying source	Total revenue	Total tax	Trade	Non-tax	Grants
Botswana	IMF country report	30.32%	22.11%	6.78%	8.22%	0.00%
Lesotho	IMF GFS	48.44%	30.00%	4.76%	5.38%	13.06%
Namibia	IMF country report	30.43%	28.17%	8.11%	2.26%	0.00%
Eswatini	IMF GFS	25.65%	24.67%	9.10%	0.75%	0.22%

Note: all revenue % of GDP.

Source: authors’ illustration based on data from UNU-WIDER (2023).

For Lesotho, where the GRD draws from IMF data, some SACU revenue is allocated to trade tax, whilst the bulk is classified as a grant. However, the corresponding IMF country report lists all of this revenue as a ‘SACU transfer’, not allocating it to either grants or taxes on international trade. When we consider the same variables from the ATO, where available, we see the following picture emerge:

Table 5: SACU countries (ATO)

Country	Source	Total revenue	Total tax	Trade	Non-tax	Grants
Botswana	ATO	N/A	10.73%	2.58%	N/A	N/A
Lesotho	ATO	N/A	15.60%	2.24%	N/A	N/A
Namibia	ATO	N/A	30.75%	13.77%	N/A	N/A
Eswatini	ATO	N/A	12.72%	4.35%	N/A	N/A

Note: All revenue % of GDP.

Source: authors’ illustration based on data from ATAF (2023).

Clearly, there are large disparities between the two datasets. For Namibia SACU transfers are accounted for in the ATO data under ‘other customs revenue’, whilst for the other three countries, it appears that the SACU revenues are not fully accounted for, although the revenues that appear under customs (trade) are clearly, at least partially, comprised of revenue shared within SACU. Indeed, it is the SACU countries which account for the most disparate observations highlighted in Section 4.7.

For context we consider a third source of data for the same observations, this time from the OECD’s Revenue Statistics in Africa. Table 6 displays this data. The aggregate revenue figures are fairly close to those reported in the GRD, but the allocation across tax and non-tax is not mirrored in either the GRD or ATO. Large sums of revenue are allocated to non-tax revenue in the OECD’s classification and an examination of the country-specific tables shows that for Namibia, Lesotho, and Botswana there is a separate line for ‘SACU revenue’. However, for Eswatini, it exists as a residual subcomponent of non-tax revenue.

Table 6: SACU revenues, OECD data

Country	Source	Total revenue	Total tax	Trade	Non-tax	Grants
Botswana	OECD	31.8%	15.0%	0.0%	16.7%	0.0%
Lesotho	OECD	50.5%	20.7%	0.0%	25.2%	4.6%
Namibia	OECD	29.9%	19.7%	0.8%	10.2%	0.0%
Eswatini	OECD	29.3%	18.0%	0.0%	11.0%	0.3%

Note: all revenue % of GDP.

Source: authors’ illustration based on data from OECD (2023).

The overarching challenge with SACU revenues—as we hope is clear—is *where* to allocate them. Unfortunately, it does not appear that there is a solution that is at the same time straightforward, consistent, or ‘correct’.

One potential area for improvement—or at least consistency across datasets—might be to allocate the **customs component** to customs revenue (as, according to the revenue sharing formula, this is linked somewhat closely to SACU members’ import and export activity).<sup>10</sup> It is not clear whether this is currently the approach taken in the ATO, although the figures contained therein under customs revenues do appear within a plausible range for this approach. It is less obvious where the excise and development component should be allocated. Clearly, the development component is more akin to a grant or a non-tax revenue than an excise duty, although it is at least formulaically related to the latter.

Another area where improvement could be made is to at least ensure consistency *within* datasets. Neither the GRD nor ATO does this particularly well at this stage, e.g. for the GRD some of the SACU revenue appears as a grant for Lesotho, whilst in the ATO it seems that SACU revenues are only accounted for in full by Namibia.

The case of SACU is a fairly extreme example of where comparing tax to GDP ratios across countries can be misleading or problematic. Analysts who utilize data for this region should

---

<sup>10</sup> SACU (2021) notes that this is linked to *intra-SACU* trade. It is not clear whether or not, for example, a good imported from China to Lesotho that arrived at a port in South Africa would count toward intra-SACU trade.

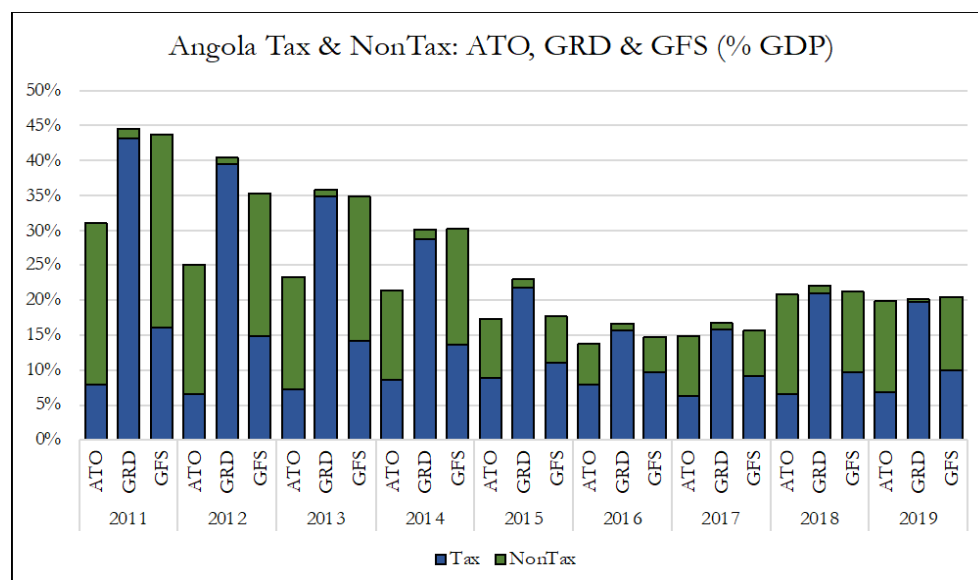
exercise extreme caution; Lesotho’s tax ratio of close to 50%, for example, is by no means grounded in the same economic roots as countries with a similarly high ratio (such as the Nordics).

### 5.3 The treatment of natural resource revenues

One common shortcoming of cross-country revenue data from, for example, the IMF’s GFS is that it does not often disaggregate between revenue which accrues from natural resource extraction and that which accrues from other non-resource sectors. The GRD and ATO take different approaches to dealing with natural resource revenue and this is a cause of some of the differences highlighted in Section 4. The ATO has separate categories for resource taxes and non-tax revenue from (i) oil, (ii) other natural resources, and (iii) royalties. The GRD has, per Table 1, a variable that captures total natural resource revenue, total natural resource tax, and total natural resource non-tax. These should, in theory, be broadly comparable, but a closer look at two cases shows that this is not necessarily the case. We consider Angola and Nigeria, two large oil-producing countries where resource revenues constitute a significant share of the government budget.

#### 5.3.1 Angola

Figure 14: Angola tax and non-tax: ATO, GRD, and GFS (% GDP)



Source: authors’ calculations from ATAF (2023), UNU-WIDER (2023), and IMF (2023b).

Figure 14 highlights the case of Angola, showing the breakdown between total tax revenue and total non-tax revenue across the ATO, GRD, and IMF GFS—the latter is included for additional context.<sup>11</sup> The underlying GRD data draws from the IMF’s Article IV reports. Two notable features emerge. Firstly, the sum of both components is relatively similar across all three sources in most recent years, but, between 2011 and 2014, the ATO figures are markedly lower than those from the IMF or GRD. Secondly, the breakdown of revenue between non-tax and tax is quite similar between the ATO and GFS, whereas the GRD allocates almost all revenue to tax, with only marginal amounts listed as non-tax. A closer look at the notes in the GRD shows that resource revenues—which are significant in Angola—are wholly allocated under tax in order to maintain historical consistency with earlier Article IV reports. However, this is also reflective of how the revenue is reported in more recent reports. The most recent country report for Angola (IMF

<sup>11</sup> The sample here is restricted to 2011–19 due to the more limited availability of GFS data.

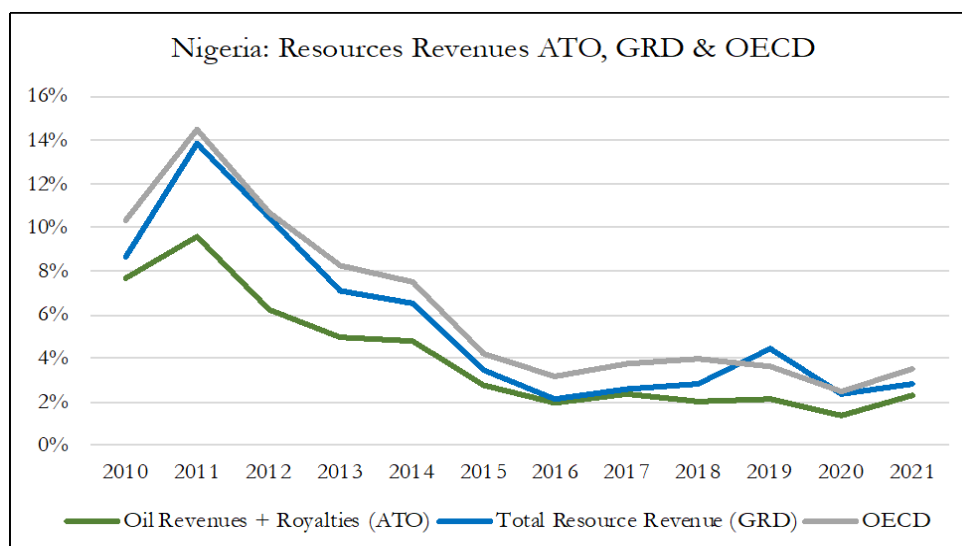
2023a: Table 2b) lists almost all of government revenue as tax, broken down by ‘oil’ and ‘non-oil’. However, the ATO and GFS appear to handle resource revenues differently, with significant sums—or potentially all—being reported as non-tax revenue. Neither dataset, however, lists any ‘resource tax’ revenue; no data is reported in the ATO, and the GFS does not contain the appropriate breakdown of revenues by resource or non-resource. Ideally, one would want to know tax and non-tax revenues broken down by resource or non-resource source. Both the GRD and ATO do strive to do this, although the incompleteness of the underlying data and inconsistency in reporting across countries mean that, in practice, there are few countries where this is done satisfactorily.

The comparison between the ATO and GFS provides some indication that the GRD approach of allocating all resource revenue in Angola as ‘tax’ is likely inappropriate.<sup>12</sup> However, the ATO and GFS still do not quite ‘agree’ on what might be the actual breakdown as fairly significant disparities in both the share of tax to non-tax and the absolute amount of revenue collected exist in many of the years depicted.

### 5.3.2 Nigeria

Another example of how the GRD and ATO report resource revenues differently can be found in the case of Nigeria. The ATO again reports no resource ‘tax’ but reports significant amounts of oil revenue and royalties as non-tax revenue. The GRD, meanwhile, only includes an aggregate figure for resource revenue, not reported as either tax or non-tax. For context we also incorporate a measure of total resource revenue from the OECD’s revenue statistics. This is computed as the sum of resource tax (petroleum profits tax plus income tax on gas exploration) and resource non-tax, (revenue from oil plus revenue from royalties). A comparison of these amounts is shown in Figure 15.

Figure 15: Nigeria: resource revenues ATO, GRD, and OECD



Source: authors’ illustration based on data from ATAF (2023), UNU-WIDER (2023), and OECD (2023).

The first thing to notice is that all three sources trend broadly in the same direction, whilst the GRD and OECD appear to often be closest in absolute terms. It turns out that the disparity between the ATO and the other series (it is more often than not lower) is due to resource tax

<sup>12</sup> However, it is reflective of what is contained in the underlying source data (IMF 2023a).

revenue that is unaccounted for in the ATO database. When we compare the ATO and OECD, we see that the resource non-tax figures match exactly. Thus, the resource *tax* component would appear to exist as a part of CIT in the ATO but is not separately reported as a resource tax.

In this case both the GRD and ATO have their limitations. The GRD does not allocate resource revenues to either of tax or non-tax, whilst the ATO only reports resource non-tax revenue, with resource tax revenue remaining bundled as a subcomponent of CIT. Given that data in the OECD Revenue Statistics in Africa and ATO often track each other quite closely (as evidenced by the matching resource non-tax figures), it would seem that there is significant scope to improve coverage in the ATO by more carefully disentangling resource tax revenue from other sources. The GRD could also consider changing its underlying source for Nigeria to the OECD, where the data appears more complete and better disaggregated.

## 6 Conclusion

This study compared, in depth, the revenue variables of the UNU-WIDER GRD and ATAF ATO. Having discussed the key elements related to the background, construction, and user bases of both datasets, we mapped the similarities and attempted to answer the question of whether comparable indicators on paper capture the same revenues in practice. Considering ten key indicators, we found substantial differences across the two datasets in both the magnitude of revenues and the manner in which revenues are classified into various subcomponents. Our investigation focused on three key areas where differences arise, namely the revenues classified in the ATO as withholding taxes, revenues from the SACU revenue sharing agreement, and the treatment of natural resource revenues. Across these areas we attempted to better understand the reasons why each dataset presents revenues in the manner it does, and we provided an in-depth discussion which we trust will be useful for users of the dataset.

One broad lesson that emerged through these investigations is that having more than one source of revenue data can often help the user to contextualize what is contained in any one source. Indeed, as we documented, it is occasionally useful to compare the data in the ATO or GRD with a third source (such as the OECD's Revenue Statistics or IMF's GFS) to truly gain a better understanding of what is being presented. Both datasets have some weaknesses and strengths and we aimed to highlight these to users in the hope that it can lead to a better understanding—and use—of either dataset for analysis.



## References

- ATAF (2023). 'African Tax Outlook Databank'. Pretoria: African Tax Administration Forum Publishing. Available at: <https://ato.ataftax.org/atafdatabank/> (accessed 3 March 2024).
- IMF (2023a). 'Angola: 2022 Article IV Consultation-Press Release; Staff Report; and Statement by the Executive Director for Angola'. IMF Country Report 2023/100. Washington DC: International Monetary Fund. Available at: <https://www.imf.org/en/Publications/CR/Issues/2023/03/06/Angola-2022-Article-IV-Consultation-Press-Release-Staff-Report-and-Statement-by-the-530604> (accessed 27 November 2023).
- IMF (2023b). 'Government Finance Statistics'. Washington DC: International Monetary Fund. Available at: <https://data.imf.org/?sk=a0867067-d23c-4ebc-ad23-d3b015045405> (accessed 28 November 2023).
- McNabb, K. (2017). 'Toward Closer Cohesion of International Tax Statistics: The ICTD/UNU-WIDER GRD 2017'. WIDER Working Paper 2017/184. Helsinki: UNU-WIDER. <https://doi.org/10.35188/UNU-WIDER/2017/410-0>
- McNabb, K., and A. Oppel (2021). Government Revenue Dataset (2021): Country Notes. WIDER Technical Note 2021/21. Helsinki: UNU-WIDER. <https://doi.org/10.35188/UNU-WIDER/WTN/2021-21>
- McNabb, K., A. Oppel, and D. Chachu (2021). 'Government Revenue Dataset (2021): Source Selection'. WIDER Technical Note 2021/10. Helsinki: UNU-WIDER. <https://doi.org/10.35188/UNU-WIDER/WTN/2021-10>
- Oppel, A., K. McNabb, and D. Chachu (2021). Government Revenue Dataset (2021): Variable Description'. WIDER Technical Note 2021/11. Helsinki: UNU-WIDER. <https://doi.org/10.35188/UNU-WIDER/WTN/2021-11>
- OECD (2023). *Revenue Statistics in Africa 2023*. Paris: OECD Publishing. <https://doi.org/10.1787/15bc5bc6-en-fr>.
- Prichard, W., A. Cobham, and A. Goodall (2014). 'The ICTD Government Revenue Dataset'. Working Paper 19. Brighton: Institute of Development Studies. <https://doi.org/10.2139/ssrn.2496442>
- SACU (2021). 'Factsheet: Understanding the SACU Revenue Sharing Arrangement'. Windhoek, Namibia : SACU Secretariat Communications Unit. Available at: <https://www.sacu.int/docs/brochures/2021/SACU-1st-Factsheet.pdf> (accessed 3 March 2024).
- UNU-WIDER (2023). 'Government Revenue Dataset'. Version 2023. <https://doi.org/10.35188/UNU-WIDER/GRD-2023>
- World Bank (2023). 'World Bank Revenue Dashboard'. Washington DC: World Bank. Available at : <https://www.worldbank.org/en/topic/taxes-and-government-revenue/brief/tax-revenue-dashboard> (accessed 28 November 2023).