INTERNATIONAL MONETARY FUND

Quantifying the Revenue Yields from Tax Administration Reforms

Hassan Adan, Jean-Marc B. Atsebi, Nikolay Gueorguiev, Jiro Honda, and Manabu Nose

WP/23/231

2023
NOV

WORKING PAPER

 WP/23/231

IMF Working Paper
Fiscal Affairs Department

Quantifying the Revenue Yields from Tax Administration Reforms

Prepared by Hassan Adan, Jean-Marc Atsebi, Nikolay Gueorguiev, Jiro Honda, and Manabu Nose*

November 2023

ABSTRACT: Despite the criticality of tax administration (TA) reforms in enhancing domestic revenue mobilization, few studies have attempted to quantify the revenue impact of such reforms. This paper fills this gap by estimating the revenue yields associated with various tax administration capabilities, based on the International Survey on Tax Administration (ISORA), the Tax Administration Diagnostic Assessment Tool (TADAT), and TA reform episodes datasets (identified by Akitoby et al., 2020). It uses a Hausman-Taylor cross-country panel regression and an event study for specific TA reform episodes. Our results (using the ISORA data) show that an increase in the overall strength of TA from the 40th percentile to the 60th percentile is associated with an increase in tax revenue by 1.8 pp. of GDP (with a 95 percent confidence range of 0.5–2.6 pp. of GDP). Similarly, the event-study assessment shows that sustained TA reforms led to an increase in tax revenues between 2 to 3 pp. of GDP, in line with the experience in three country cases (Jamaica, Rwanda, and Senegal). Also, the revenue yields are increasing over time to more than 3 pp. of GDP after the 6th year following a comprehensive reform. The analysis also highlights the significant impact of specific measures including: i) strengthening compliance risks management, ii) enhancing public accountability, iii) establishing Large Taxpayer Offices (LTO), iv) strengthening accountability and transparency, and v) enhancing timely filing of tax declarations.

RECOMMENDED CITATION: Adan, H., Atsebi, J., Gueorguiev, N., Honda, J., Nose, M. (2023). Quantifying the Revenue Yields from Tax Administration Reforms. IMF Working Paper, WP/23/231. International Monetary Fund, Washington, DC.

JEL Classification Numbers:	C1, E62, H2, H83, O23
Keywords:	Tax Revenues Yields; Tax Administration, ISORA, TADAT.
Author's E-Mail Address:	HAdan@IMF.org, JAtsebi@IMF.org, NGueorguiev@IMF.org, JHonda@IMF.org, MNose@IMF.org

* The authors are grateful to Elena D'Agosto, Yongquan Cao, Seydou Coulibaly, Elizabeth Gavin, Jan Gottschalk, Anamaria Maftei, Andrew Masters, Andrew Okello, Paulo Paz, and Miguel Pecho for their insightful comments and the useful discussions. They also thank the participants of the IMF's Fiscal Affairs Department seminars. We also thank Dhruv Gandhi for excellent research assistance, and Dánea Trejo Carcamo and Victor Maradiaga for editorial assistance. The views expressed here are those of the authors and not necessarily those of the IMF, its Executive Board, or management.

WORKING PAPERS

Quantifying the Revenue Yields from Tax Administration Reforms

Prepared by Hassan Adan, Jean-Marc Atsebi, Nikolay Gueorguiev, Jiro Honda, and Manabu Nose

* The authors are grateful to Elena D'Agosto, Yongquan Cao, Seydou Coulibaly, Elizabeth Gavin, Jan Gottschalk, Anamaria Maftei, Andrew Masters, Andrew Okello, Paulo Paz, and Miguel Pecho for their insightful comments and the useful discussions. They also thank the participants of the IMF's Fiscal Affairs Department seminars. We also thank Dhruv Gandhi for excellent research assistance, and Dánea Trejo Carcamo and Victor Maradiaga for editorial assistance. The views expressed here are those of the authors and not necessarily those of the IMF, its Executive Board, or management.

Contents

I. Introduction 4

II. Estimating the Impacts of Overall Strength of Tax Administration 6
A. Data 6
B. Stylized Facts 7
C. Empirical Methodology 11
D. Results 13
1. Overall Strength of Tax Administration 13
2. Specific Components of Tax Administration: Multivariate Models 18

III. Estimating the Impacts Based on Reform Episodes 19
A. Data and Stylized Facts 19
B. Empirical Methodology 21
C. Results 24
1. Difference-in-Differences in a Staggered-Adoption Design 24
2. Synthetic Difference-in-Differences with Common Support 25
3. Controlled Interrupted Time-series Analysis Case Studies 27
a. Jamaica – Comprehensive TA Reforms 27
b. Rwanda – Comprehensive TA Reforms (Less LTO and Segmentation) 28
c. Senegal – Comprehensive TA Reforms (Less Improving Compliance) 29

IV. Concluding Remarks 30

Annex I. Data Descriptions and Sources 32

Annex II. Descriptive Statistics and Stylized Facts 33

Annex III. Results for Multivariate Models 35

Annex IV: The Impact of Tax Administration Reforms in the Post-GFC 37

References 38

BOX
1. The New IMF FAD Tax Administration - Assessment and Yield Tool 18

FIGURES
1. Scatter Plot Between ISORA Indexes Across Vintages (2017 vs. 2014) 8
2. Comparisons of TA Strength, Practices, and Characteristics Across Income Group Levels 9
3. Tax Revenues and ISORA and TADAT Aggregate Indexes 10
4. Scatter Plot Between ISORA and TADAT Aggregate Indexes 11
5. Estimated Revenue Yields from an Increase of the Strength of Tax Administration 14
6. Tax-to-GDP Ratio by Country Group 21
7. Group Time ATT (Yield in %) 24
8. Duration ATT (Yield in %) 25

9. Country Group vs. Control Group 26
10. CITSA Jamaica vs. Controls Average 28
11. CITSA Rwanda vs. Controls Average 29
12. CITSA Senegal vs. Controls Average 30

TABLES

1. Estimated Revenue Yields from an Increase in TA Strength from the 40th to the 60th Percentile 14
2. Baseline Results, Overall, Strength, ISORA 16
3. Baseline Results, Overall, Strength, TADAT 17
4. TA Reforms in EMs and LIDCs After the Global Financial Crisis 20
5. Aggregated ATE 25

ANNEX TABLES

AI.1. Data Descriptions and Sources 32
AII.1. Summary Statistics 33
AII.2. Correlations Between ISORA and TADAT Indexes Across Years 33
AII.3. Correlations among ISORA and TADAT Indexes, and between ISORA and TADAT Indexes 34
AIII.1. Multivariate Models, ISORA 35
AIII.2. Multivariate Models, TADAT 36

ANNEX FIGURES

AIV.1. Revenue Ratio Increase vs. Revenue Ratio prior Year 37
AIV.2. Pre- vs. Post- GFC Revenue/GDP Ratio 37

I. Introduction

1. **Can we quantify the impact of tax administration (TA) reforms? This paper tackles this seemingly straight-forward but challenging question through a set of empirical approaches.** Given the need to enhance domestic revenues for a country's development (particularly in developing countries), many countries have been implementing tax policy and administration reforms. Critical as these tax administration reform measures are, the quantification of their benefits involves significant challenges. To address these challenges, we propose a methodology to measure the strength of tax administration and its practices, although the data on tax administration is quite limited for analyses over an extended period. We also employ an empirical methodology that accounts for the potential endogeneity of the strength of tax administration and its practices to tax revenues.

2. **Very few empirical studies have explored the impact of TA reforms.**[1] Ebeke et al. (2016), Crivelli (2019), and Chang et al. (2020) conducted cross-country analyses on the establishment of Large Taxpayers Office (LTO), Semi-Autonomous Revenue Agency, the strength of tax administration, and tax performance. Further, there has been an emerging microeconomic literature that identifies the effectiveness of TA reforms using an experimental design in a single developing country context, related to the use of third-party information (Pomeranz, 2015; Carrilo et al., 2017) and organizational reforms in strengthening tax administration (Basri et al., 2021).[2] However, in a cross-country context, a more robust empirical framework (that accounts for slow-moving features of TA indexes and the endogeneity related to a country's TA status) is needed. A detailed examination of specific reform episodes is also needed to better identify the revenue gains and predict the impact of TA reforms over time (i.e., how long after the reform the effect starts materializing).

3. **To fill the gap in the literature, we employ two different methodologies to estimate the yields from TA measures while addressing the shortcomings of earlier studies.** The use of the different methodologies aims at cross-checking the results, which may lead to complementary findings.

- **In the first part, this paper applies an empirical strategy using cross-country panel data with slow-moving indicators of TA strength that accounts for the endogeneity of the TA reform impact.** Specifically, the Hausman-Taylor cross-country panel regression with endogenous time-invariant variable (Chatelain and Ralf, 2021) estimates the contribution of characteristics of tax agency to domestic tax revenues, controlling for other key determinants (macroeconomic, structural, and tax policy changes) as identified in the past literature.
- **In the second part, an event-study design is used to quantify the yields from past TA reform episodes in Low-Income Developing Countries (LIDCs) and Emerging Markets (EMs) with sustained revenue mobilization efforts in the aftermath of the global financial crisis (GFC) (Akitoby et al., 2020).** This analysis employs three recent extensions to the canonical difference-in-differences (DID) to

[1] Cross-country studies focused on macroeconomic conditions (Tanzi, 1992; Gupta, 2007; IMF, 2018), tax policy (Dabla-Norris et al., 2020; Ngoma and Krsic, 2017; Akitoby et al., 2020), and governance or socio-political institutions (IMF, 2019; Morrow et al., 2022) as key determinants of tax revenues.

[2] On corporation taxation in Indonesia, Basri et al. (2021) examined the impact of more "intensive" tax administration, through the creation of medium-sized taxpayer offices and increasing the revenue staffing-to-taxpayers ratio on actual tax filings and payments. Compared with the elasticity of taxable income with respect to the corporate tax rate, they found that improved tax administration creates significant returns in generating revenues (equivalent to raising top CIT rates on all firms by eight percentage points).

estimate the revenue yields from TA reforms in different groups, while providing insights on the evolution of the yields in the years following the reforms.

4. **Our findings are threefold**.

- **First, improvements in the overall strength of the TA are associated with a significant increase in tax revenues.** Specifically, we find that an increase in the strength of the tax administration that would move a country from the 40th percentile to the 60th percentile in the distribution of IMF member countries by this indicator is associated with an increase in tax revenues by up to 1.4 percentage points (pp.) of GDP for all countries and 1.8 pp. of GDP for EMs and LIDCs.[3] These point estimates of revenue yields are broadly consistent with the estimates in our event study of specific TA reforms after the GFC (i.e., additional revenues between 2 to close to 3 pp. of GDP).[4]
- **Second, while significant TA reforms may lead to some initial gains, it could take time for the full results to be realized.** The event-study analysis reveals that the revenue yields are increasing over time to more than 3 pp. of GDP after the 6th year following a far-reaching and comprehensive reform across several practices of tax administration.
- **Third, several specific TA measures are strongly correlated among themselves and associated with better tax performance.** Using ISORA, our results suggest that compliance risk management improvements have the highest revenue yields. In addition, public accountability and LTO also play a more significant role in increasing revenue in EMs and LIDCs. While using TADAT, we find that accountability and transparency, and timely filing of tax declarations matter the most for raising revenues. We also find high correlations among ISORA and TADAT sub-indexes. This suggests that different TA functions support and reinforce each other.

5. **This paper broadens the analysis by Chang et al. (2020) and provides a comprehensive database in benchmarking a country's strength of TA and its practices against comparators across multiple dimensions and estimating revenue yields.** From an operational perspective, the comprehensive database and analysis, as well as an operational tool developed in parallel (Box 1), help to identify gaps in TA and systematically estimate the contribution of past or future TA reforms for IMF member countries.

6. **This paper is organized as follows:** The first part analyzes the role of TA strength and practices on tax revenues using the ISORA and TADAT datasets (see below for datasets description). The second part studies the impact of TA reforms in the post-GFC period. In each part, Section A presents the data and stylized facts, Section B discusses the empirical strategy, and Section C presents the findings. Lastly, the paper concludes and provides some policy recommendations.

[3] In our sample, 17 countries achieved an improvement of their ISORA's OS index that is higher than an amount equivalent to the change from the 60th percentile to the 40th percentile during the period 2014-17. Those with the highest increase are Grenada, St. Vincent and Grenadines, Uruguay, Greece, Ghana, Poland, and Myanmar.

[4] Given the limited availability of data, the analyses with ISORA and TADAT data do not necessarily cover the periods of the reform episodes in section III (Jamaica, Rwanda, and Senegal). Based on the episodes, we assume that the reforms described in section III would broadly correspond to an increase in TA strength equivalent to the change from 40th percentile to 60th percentile.

II. Estimating the Impacts of Overall Strength of Tax Administration

A. Data

7. **We use the International Survey on Revenue Administration (ISORA) and the IMF's Tax Administration Diagnostic Assessment Tool (TADAT) datasets to measure countries' overall strength of the TA, as well as the strength across the different characteristics and practices of the TA.** The ISORA surveys countries' TA features through both numerical and categorical survey questions (with 982 data points) on a wide spectrum of practices and institutional features agreed on by four partner international organizations.[5] The data is collected through a voluntary self-assessment, where countries respond to an online questionnaire. It covers 154 countries, including 36 Advanced Economies (AEs), 71 EMs, and 46 LIDCs, and encompasses four vintages.[6] However, we only use the first two vintages for our analysis, ISORA 2016 and ISORA 2018, given the major revision of the survey for the subsequent vintages.[7] The ISORA and TADAT data are increasingly being used in depth-research on TA (see e.g., Dabla-Norris et al., 2020).

8. **Using the ISORA data, we closely follow Chang et al. (2020) to construct an operational strength index of TA and seven sub-indexes measuring the strength of specific TA practices and characteristics.** The seven sub-indexes cover administrative and operational practices and structural foundations (laws, regulations, and policies) of the TA. They encompass: i) compliance risk management (CRM) approach, ii) the use of third-party data, iii) digitalization of services, iv) service orientation, v) public accountability, vi) the degree of autonomy, and vii) the establishment of LTO or large taxpayer program. In addition, they are compiled as the equally weighted average of the specific questions related to each sub-index.[8,9] We also compile an overall operational strength (OS) index as the equally weighted average of the seven sub-indexes, which measures the strength of the TA across practices and characteristics.

9. **The TADAT data provides information on the relative strengths and weaknesses of key components of a country's TA by focusing on the performance of the major national taxes.**[10] Starting in 2013, it has covered 86 countries, with 11 repeated assessments in the pre-pandemic period (i.e., between

[5] The ISORA dataset is the collaboration of the Inter-American Center of Tax Administrations (CIAT), the Fiscal Affairs Department (FAD) of the IMF, the Intra-European Organisation of Tax Administrations (IOTA), and the Organisation for Economic Co-operation and Development (OECD). More information about the survey, questionnaires, ISORA data and related publications can be found following this link: https://data.rafit.org/?sk=BA91013D-3261-42F8-A931-A829A78CB1EC.

[6] The initial ISORA dataset covers 159 countries, however due to some missing values, we cover 154 countries in our analysis.

[7] The five vintages include ISORA 2016, ISORA 2018, ISORA 2020, ISORA 2021 and ISORA 2022 covering respectively, the years 2014–15, 2016–17, 2018–19, 2020, and 2021. Until ISORA 2021, the survey was run biennially. Following the ISORA 2018, a major revision of the questionnaire has resulted in a discontinuity of many questions, a smaller set of questions asked annually, and additional questions to be asked every four years. This major revision limits us from using the ISORA 2020 and 2021 vintages for our analysis. See Crandall et al. (2021) for a comprehensive description of the results of the ISORA 2018.

[8] For more details on the computation of the ISORA sub-indexes, please refer to Annex I in Chang et al. (2020) or module 3 of the developed FAD operational tool TA-AYT that provide the different questions used to construct the indexes.

[9] Other data from the ISORA is also used. This includes: i) the ratio of tax staff over the labor force; and ii) the active taxpayers over the labor force, measuring the human resource capacity and taxpayers' base of the TA, respectively.

[10] TADAT dataset is not publicly available for several countries. Detailed information about the TADAT dataset and methodology can be found following this link: https://www.tadat.org/home#overview.

2013 and 2019).[11] TADAT assessment focuses on nine key performance outcome areas (POAs) capturing the most critical TA functions, processes, and institutions, including: i) integrity of the registered taxpayer base, ii) effective risk management, iii) supporting voluntary compliance, iv) timely filling of declarations, v) timely payment of taxes, vi) accurate reporting in declarations, vii) effective dispute resolution, viii) efficient revenue management, and ix) accountability and transparency.[12] These nine POAs are compiled using the 32 high-level indicators (HLIs), which are in their turn constructed using 55 measurement dimensions (MDs), following the TADAT aggregation M1 and M2 methods to construct the 32 HLIs.[13] We also compile an overall performance assessment report (PAR) index as the equally weighted average of the nine POAs, which measures the strength of the TA across all practices and characteristics.

10. **Other data for empirical analyses include the tax-to-GDP ratio and standard macroeconomic controls.** As a dependent variable, the tax-to-GDP ratio (excluding trade taxes and social security contributions) is used, which is compiled from multiple sources. We exclude trade taxes and social security contributions to account for the volatility of commodity prices and exclude resource revenues, and because other agencies than the TA may collect those revenues. We also use several control variables coming from different sources. They are discussed in the methodology section, and their data descriptions and sources are presented in Annex I.

B. Stylized Facts

11. **This section presents several stylized facts on the TA strength, its practices, and characteristics using ISORA and TADAT datasets. It discusses the differences between ISORA and TADAT aggregate indexes and the correlations among indexes over time and between themselves.** It also provides comparisons of TA strength, practices, and characteristics across income group levels, as well as a foretaste of the relationship between TA strength and tax revenues. The summary statistics of the indexes constructed using ISORA and TADAT datasets are presented in Annex II. Our main findings are the following:

a) **The strength of TA generally changes only slowly over time, with some exceptions**. Looking at the changes of the ISORA and TADAT indexes over time (across vintages and repeated assessments, respectively), we show that TA strength, practices, and characteristics only slightly changed over the years—they are slow-moving variables (see illustration in Figure 1 for ISORA). It indicates that it takes time as well as strong and sustained efforts to improve TA strength. Table AII. shows that the correlations between ISORA indexes for 2014 and 2017 are elevated. For instance, eight out of the ten ISORA's indexes have a coefficient of correlation higher than 0.8.[14]

b) **The strength of TA, its practices, and its characteristics vary substantially with the levels of economic development**. The results of the comparison across income group levels are presented in

[11] This is a very dynamic database. The number of assessments has now increased (see the exact number as of today: https://www.tadat.org/home). However, since our focus is on the period 2010-2019, we do not include any of the assessments after 2020 (post-pandemic period).

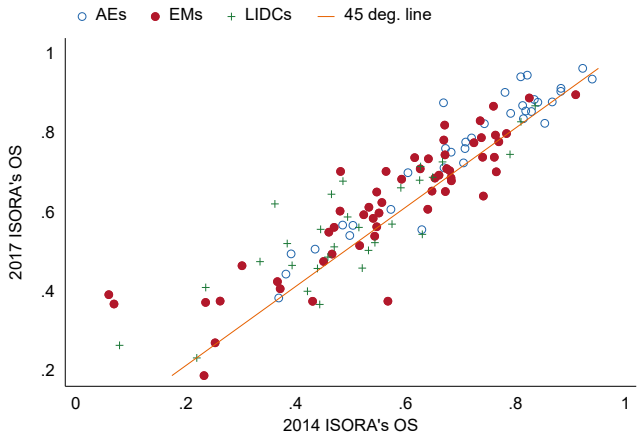
[12] TADAT's timely filing and timely payment scores factor in the credibility of the data provided: if regarded as not credible, administrations receive the lowest score. i.e., these are not only a measure of the actual timeliness of filing, but they are combined with an assessment of whether the administration's record keeping on filing and payment is reliable.

[13] Please follow this link for the description of the M1 and M2 methods: https://www.tadat.org/tadatAtAGlance.

[14] On exceptional basis, there were a few countries which demonstrated significant improvement. Those are: Grenada, St. Vincent and Grenadines, Uruguay, Greece, Ghana, Poland, and Myanmar.

Figure 2. The TA strength, as measured by the ISORA and TADAT indexes, is on average much stronger in AEs, while LIDCs have the weakest TA (Panel A). These findings point to the potential scope for LIDCs and EMs to further improve the strength of TA across all practices and characteristics.

Figure 1. Scatter Plot Between ISORA Indexes Across Vintages (2017 vs. 2014)



Sources: Authors' calculations using ISORA.

- **For ISORA, AEs have higher scores than in EMs and LIDCs, except for the establishment of LTO.**[15] EMs also have higher scores on average than LIDCs, except for autonomy and establishment of LTO, where they have almost similar scores. In EMs and LIDCs, the weaknesses are more accentuated for the use of third-party data and digitalization, and compliance risk management.
- **For TADAT, the differences in TA performance between AEs, EMs, and LIDCs are more pronounced.** AEs have higher scores across all practices and characteristics (in particular for timely filing of tax declarations and timely payments of taxes), which are also higher on average than in EMs and LIDC. Also, EMs have higher scores on average than LIDCs across all practices and characteristics. However, they have weaker scores for effective risk management, accurate reporting in declarations, timely filing of tax declarations, and efficient revenue management. LIDCs have the lowest scores on average across all practices and characteristics, especially for timely filing of tax declarations, efficient revenue management, accurate reporting in declarations, and effective risk management.
- **Not surprisingly, the strength of TA and tax revenues are positively correlated.** Specifically, we find that the correlation between ISORA and TADAT indexes and tax revenues (excluding trade taxes and social contributions) is high, although there is a high variance among peer group countries (Figure 3).

c) **There are significant heterogeneities by income groups.** AEs tend to collect more tax revenues with stronger TA. On the other hand, LIDCs have, on average, weaker TA and tend to collect lower tax revenues. There are significantly large variations in TA strength in EMs, where some countries have weaker TA similar to LIDCs while others have stronger TA similar to AEs.

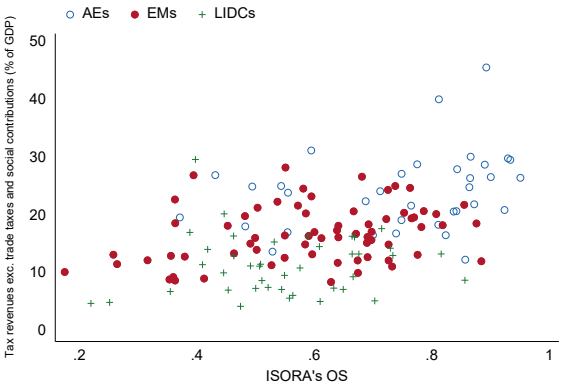
[15] The LTO indicator reflects only the presence of an LTO, not necessarily the performance or efficiency of the office/program. Another reason is that some more modern tax administrations in AEs are using taxpayer segmentation and specialization without resorting to an LTO.

Figure 2. Comparisons of TA Strength, Practices, and Characteristics Across Income Group Levels

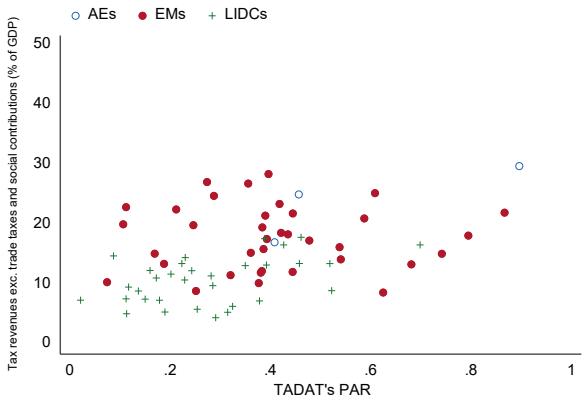
Sources: Authors' calculations using ISORA and TADAT.
Notes: This figure shows the comparisons of ISORA and TADAT indexes across country group income levels.

Figure 3. Tax Revenues and ISORA and TADAT Aggregate Indexes

Panel A. ISORA's OS



Panel B. TADAT's PAR

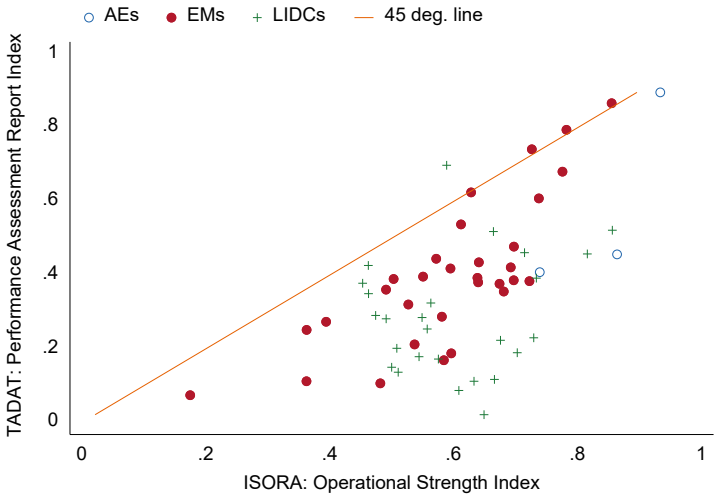


Sources: Authors' calculations using IMF's WEO, IMF's WoRLD, OECD's TRD, UNU-WIDER's GRD, ISORA and TADAT.

d) **The two aggregate indexes, ISORA's OS and TADAT's PAR, are correlated**. Figure 4 compares the two aggregate indexes used to measure overall TA strength. Though both indexes are strongly correlated (the coefficient of correlation is 0.61), the TADAT's PAR is generally lower than ISORA's OS for almost all countries, and the gap between the two indexes is higher for LIDCs. This may reflect the nature of ISORA data (based on self-reporting by the authorities), as opposed to TADAT assessments conducted by external experts. Further, these aggregate indexes are computed using different practices and characteristics of RA, with different aggregation methodologies for ISORA (which uses equally weighted additive components) and TADAT (which uses M1 and M2 methods as described above, and components assessed within the MD not considered as additive). So, the figure should be interpreted as a broad indication of a high correlation between the two (rather than a precise estimation of their relation).

e) **Sub-components of TA are also correlated.** The correlations among ISORA and TADAT sub-indexes are also high. Table AII.3 indicates strong correlations among the different ISORA, TADAT indexes, and between both ISORA and TADAT indexes. This suggests that; i) within tax administration, good practices or characteristics in one dimension are generally associated with good practices or characteristics in other dimensions, and ii) a reform aiming at improving TA strength should cover practices comprehensively and be implemented in an integrated manner. However, the high correlations across these sub-indexes make it difficult to gauge the relative importance of one practice or characteristic over others in raising revenues and analyze their interactions.

Figure 4. Scatter Plot Between ISORA and TADAT Aggregate Indexes



Sources: Authors' calculations using ISORA and TADAT.

C. Empirical Methodology

12. **TA practices and characteristics are slow-moving variables requiring a robust empirical approach**. Our measures of the strength of TA and its practices and characteristics do not change much over time in general. We, therefore, hypothesize that the values of the indexes measuring TA strength and the strength of its practices remain constant over the period of study from 2010 to 2019.[16] Practically, we input the value of the ISORA's OS and seven sub-indexes in the year 2017 to all years over the period of study. Similarly, we input the value of the PAR, POAs, and HLIs in the latest TADAT assessment, or the average values for these indexes when there are repeated assessments, to all years over the period of study.

13. **To estimate the revenue yields from the strength of TA, we employ a newly developed model by Chatelain and Ralf (2021)**. This model allows robust estimation of the effects of endogenous time-invariant variables that are correlated with individual effects. The model uses a pretest procedure to determine exogenous internal instruments among the time-varying variables systematically. Once the exogenous instrumental variables are determined through the pretest procedure, and if it is not under-identified, Hausman and Taylor's (1981) estimator (HT) can be applied (see for instance, some applications in Goh and Tham, 2013; Bouvatier, 2014).

- **The pretest procedure prevents the HT estimator from being potentially biased by wrongly assuming that all internal instruments are exogenous.** First, it consists of estimating a random effects estimator that includes all time-varying variables, all time-invariant variables, and most importantly, all averages over time of all time-varying variables. Second, it selects, as internal instruments, the subset of time-varying variables whose coefficients associated with their average over time are statistically non-

[16] In a robustness check, we also re-estimate our baseline models over a period of study 2015–19 for ISORA indexes only (we cannot do the same for TADAT due to insufficient data) assuming that the values of the indexes measuring TA strength and the strength of its practices remain constant over the period 2015–19. The findings of this robustness are similar to our baseline findings especially for the range estimates of revenue yields, except that the point estimates for revenue yields are lower than in the baseline for Models 2 and 3 (i.e., when tax staff and taxpayer base are added). The results are available upon request.

significant (i.e., not different from zero).[17] Third, the HT estimator can be estimated by specifying the exogenous instrumental variables, which allows estimating the effects of time-invariant variables. Chatelain and Ralf (2021) show through a Monte Carlo simulation that the HT estimator using the instrumental variables determined by the pretest procedure is a viable estimator that is overperforming other competing estimators of endogenous time-invariant variables.[18]

- **After applying the pretest procedure, the following four variables are internally selected as time-varying instrumental variables**: i) terms of trade, ii) the share of agriculture to GDP, iii) control of corruption, and iv) the proxies of tax policy.[19]

14. **Further, we use this method to account for the possible endogeneity of the TA strength and its practices and characteristics**. The state of tax administration could be potentially endogenous, as policy measures to improve it can be anticipated, correlated with past changes in economic activity and revenue performance, and motivated by concerns about fiscal sustainability or other developmental needs. All this implies a possible reserve causality between tax revenues and TA. For instance, Ebeke et al. (2016) show that the adoption of large taxpayer unit (LTUs) and Semi-Autonomous Revenue Agency (SARA) can be explained by various policy and structural variables. Specifically, they find that the probability of conducting TA reforms tends to be lower with a higher level of age dependency ratio, the size of the agricultural sector, and the non-resources tax-to-GDP ratio. Moreover, the presence of IMF-supported programs and past TA reforms would increase the likelihood of conducting new TA reforms. These findings confirm that TA practices and characteristics are potentially endogenous.

15. **Against this background and after identifying the exogenous time-varying instrumental variables, our model can be written as follows**:

$$Tax_{i,t} = \delta_1 Z_i + \beta_1 X_{1,it} + \beta_2 X_{2,it} + \tau_t + \mu_i + \varepsilon_{it} \quad (1)$$

where $Tax_{i,t}$ is the tax-to-GDP ratio excluding trade taxes and social security contributions, Z_i is a set of potentially endogenous time-invariant variables of interest, $X_{1,it}$ is a set of control endogenous time-varying variables, $X_{2,it}$ is a set of control exogenous time-varying variables identified by the pretest procedure, τ_t are time fixed-effects included to capture common shocks to tax revenues, μ_i are the unobserved country random effects, and ε_{it} is the error term. Our set of control variables X include variables as identified by the existing literature to be good determinants of tax revenues. They include i) macroeconomic variables such as per capita GDP and its square, inflation, trade openness, external debt to GDP, and terms of trade, ii) variables capturing the structure of the economy and quality of institutions such as oil exports to GDP, the share of agriculture to GDP, and control of corruption, and iii) proxies of tax revenue policy.

[17] Specifically, Mundlak (1978) proved that the estimator of the parameter related to the average over time of time-varying variables corresponds to the difference between the within estimator (which is not biased because of endogeneity) and the between estimator (which may be biased because of endogeneity). Hausman and Taylor (1981) demonstrated that testing the null hypothesis of the equality between the within estimator and the between estimator allows us to test the null hypothesis of the correlation between individual random effects and a given explanatory variable, and therefore identify internal instruments. We take a step forward and use an upward testing procedure for instrument selection that tests a sequence of joint null hypotheses of an increasing number of parameters of the average over time of time varying variables. We also test whether the instruments are strong by looking at the correlation between the instrumental and endogenous variables, which shows good correlations.

[18] As demonstrated by Pesaran and Zhou (2018), the “two-step approach” by Hsiao (2003) is unbiased and consistent for a large sample and finite period when the time-invariant variables are uncorrelated with the individual effects.

[19] For the simplicity of presentation, we do not present the results from the pretest strategy for all the specifications. The results can be obtained upon request.

16. **Following Chang et al. (2020), the effect of tax policy change is proxied in three ways.** Our preferred proxy for tax policy is the difference between the tax revenues projected in the budget for year (t) and the actual tax revenues of the preceding year (t-1), both relative to the respective GDP. The budgeted revenues reflect the expected effects of tax policy changes and macroeconomic fluctuations; normally, the latter move tax revenues broadly in proportion to GDP, so the ratio would be little affected. It is possible that the budgeted revenues also include some envisaged gains from tax administration reforms. However, since we only look one year ahead, such effects should be small, as usually tax administration reforms yield incremental gains over several years. Moreover (and perhaps more importantly), to the extent such gains are included in the budgeted revenues, the effect would be that our proxy would overstate the role of tax policy and, by implication, understate the revenue contribution of tax administration reforms in the estimation process. Therefore, our approach seems, if anything, conservative.[20] We also conducted sensitivity analysis with other proxies for tax policy: (i) the difference between the budgeted revenues in year (t) and the budgeted revenues in year (t-1), to account for possible projection bias; and (ii) the top tax rates for VAT, CIT, and PIT (see, Chang et al., 2020 for a detailed discussion of these proxies of tax policy).[21]

17. **We estimate the model in equation (1) using different variables of interest**. First, we estimate it for the ISORA's OS and TADAT's PAR aggregated indexes, capturing the strength of TA across all practices and characteristics. Second, we estimate the model by including all the sub-indexes in a multivariate setting to identify those that are predominantly more critical to raising more revenues. Our findings are presented in the next section.

D. Results

1. Overall Strength of Tax Administration

18. **Our baseline results are presented in Table 2 and Table 3, using ISORA and TADAT data, respectively, and Figure 5 depicts them graphically (with also the estimated range).** The results are presented for a sample of all countries (columns 1–3) and EMs and LIDCs only (columns 4–5). They present the estimated impacts on tax revenues of the overall strength of TA, as well as those of the control variables.[22] Key findings from these results are as follows:

- **The overall strength of the TA is associated with an increase in tax revenues.** The coefficients associated with ISORA's OS and TADAT's PAR are positive and statistically significant in all specifications; they have a slightly higher magnitude in EMs and LIDCs. Moreover, the magnitudes are economically very

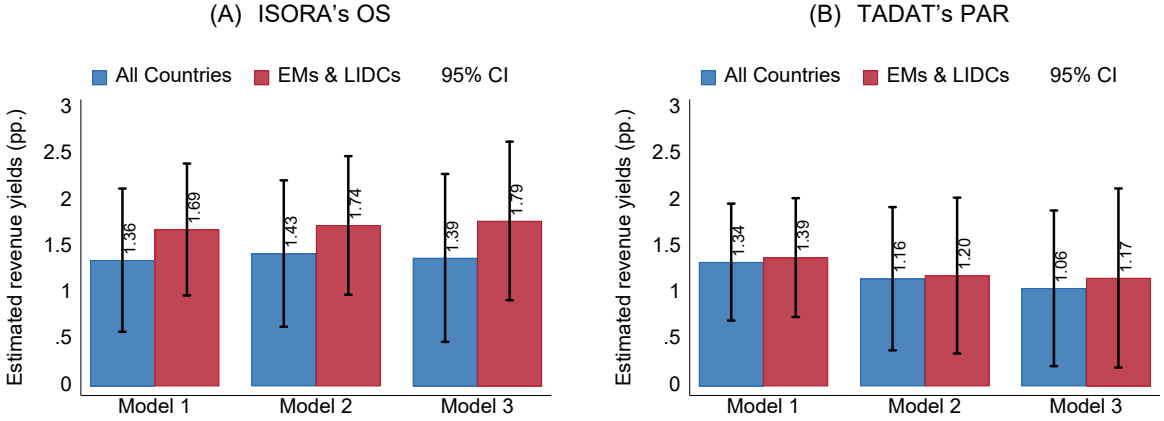
[20] As caveat for this exercise, it should be noted that our tax policy measure could be overstated or understated due to the following reasons: First, revenue fluctuations are not proportional to GDP (e.g., due to commodity price swings, weather-related or other supply shocks, or in recessions/booms where the tax elasticity to GDP could be below/above 1). Thus, it is likely that our tax policy measure would then be somewhat correlated with the macro indicators in the regression – multicollinearity – which, if anything, would increase the standard errors of the estimated coefficients and reduce their inferred significance. Second, it uses an estimate of the actual revenues in year (t-1) as the basis; and the factual actual revenues could significantly differ from that estimate. However, this could be a problem only to the extent that the estimation error in (t-1) is systematic, e.g., always has the same sign, which may not be the case a priori. Third, nowcasting tax revenues projected in the budget for year (t), for certain tax categories and certain countries, can be affected by bias (e.g., systematic pessimistic patterns and gaps in the implementation of fiscal policy) or serial correlation (i.e., once made, the error is likely to persist).

[21] The results of these robustness checks can be obtained upon request. We also find that our findings are not driven by outliers/influential data points. When dropping these influential points/countries, our findings remain similar for our variable of interest, although the coefficients are slightly lower.

[22] The goodness of fit measures (adjusted R-squared) indicate that our models could explain up to 56 percent of the variation of tax revenues over the period 2010–19.

meaningful (see below). These findings suggest that countries can mobilize substantially more revenues by conducting reforms to strengthen their TA.

Figure 5. Estimated Revenue Yields from an Increase of the Strength of Tax Administration



Sources: Authors' calculations using IMF's WEO, IMF's WoRLD, OECD's TRD, UNU-WIDER's GRD, ISORA, TADAT, and other sources (see Table AI.1).
Notes: Revenue yields are estimated using models presented in Table 2 (left chart) and Table 3 (right chart). Model 1 corresponds to columns 1 and 4, Model 2 to columns 2 and 5, and Model 3 to columns 3 and 6. The chart shows on y-axis the revenue yields resulting from an increase in the overall strength of TA by an amount equivalent to the change between the 40th percentile and the 60th percentile.

- **The point estimates show revenue yields of 1.3-1.8 percentage points of GDP for a feasible set of reforms as demonstrated by several countries (Table 1 and Figure 5).** Specifically, we find that an increase in the ISORA's OS by an amount equivalent to the change between the 40th percentile and the 60th percentile of the IMF member countries is associated with an increase in revenues by up to 1.8 pp. of GDP (with a 95 percent confidence interval of 0.5-2.6 pp. of GDP). For TADAT's PAR, these figures stand at up to 1.4 pp. of GDP (with a range of 0.2-2.1 pp. of GDP).[23] The cases of improving the TA strength with this magnitude (by an amount equivalent to the change between the 40th percentile and 60th percentile) are not uncommon. Based on the ISORA's OS, we find in our sample that 17 countries experienced such an improvement in their ISORA's OS index over the period 2014-17. Those with the highest increase are Grenada, St. Vincent and Grenadines, Uruguay, Greece, Ghana, Poland, and Myanmar.

Table 1. Estimated Revenue Yields from an Increase in TA Strength from the 40th to the 60th Percentile (Percentage points of GDP)

	All countries	EMs and LIDCs
Using ISORA OS	1.4 (0.5 – 2.2)	1.8 (1.0 – 2.6)
Using TADAT PAR	1.3 (0.2 – 2.0)	1.4 (0.2 – 2.1)

Sources: Authors' calculations using IMF's WEO, IMF's WoRLD, OECD's TRD, UNU-WIDER's GRD, ISORA, TADAT, and other sources (see Table AI.1).
Notes: The figures in the brackets indicate the range of estimates based on the 95 percent confidence intervals.

[23] Our baseline findings are robust to using an alternative sample excluding all commodity exporters (subject to volatility). The results are available upon request.

- **There is an inverted-U relationship between tax revenues and the staffing of tax administration for all countries, but this relationship does not hold in EMs and LIDCs.** This indicates that the staff level of TA is positively associated with tax revenues, and the returns are diminishing above a certain level of staff. At a lower level of staff, the increase in staffing would be associated with larger tax revenue collection. In contrast, the marginal return of increasing staffing would turn negative, passing the threshold of tax staff between 0.15 to 0.22 percent of the labor force (depending on the specifications).
- **Contrary to Chang et al. (2020), we find that the coefficients associated with the taxpayers' base as measured by the share of active taxpayers in the labor force are positive but non-significant in all our specifications.** However, when not controlling for the strength or staffing of TA, these coefficients turn significant. This indicates that, on average, the strength and staffing of TA may matter more to raise revenues than a broadening of the taxpayers' base.[24]
- **Some macroeconomic and structural variables significantly impact tax revenues.** Our findings show that trade openness is positively and statistically associated with higher revenues, consistent with the literature (Leuthold, 1991; Ghura, 1998; Chang et al., 2020). The effects of external debt on revenues are mixed, as they can reduce the incentives for domestic mobilization or create incentives from tax policy and TA reforms to mobilize more revenues needed to service public external debt (Tanzi, 1992). We find that increasing external debt to GDP leads to higher revenues in some of the specifications. The structure of the economy also matters for revenue collection. The coefficients associated with oil exports to GDP are positive and statistically significant, reflecting the correlation between oil export revenues and income revenues of oil companies. We also find that higher inflation is associated with lower revenues. Stronger control of corruption is also associated with higher revenues (in TADAT specifications). Not surprisingly, tax policy, as captured by its proxy, is positively and significantly associated with tax revenues. Our findings also indicate that several other variables do not significantly impact tax revenues, after controlling for others, including real GDP per capita and its squared, terms of trade, and share of agriculture in GDP.

19. **Overall, our results show that the strength of the TA—as measured by the ISORA's OS or the TADAT's PAR—is positively associated with higher revenues after controlling for tax policy and several macroeconomic and structural variables.** Moreover, staffing of TA also matters, but the returns are diminishing, passing a certain threshold (above 0.22 percent of the labor force). Tax policy measures and several macroeconomic and structural variables, including trade openness, external debt to GDP, and oil exports to GDP, are also positively associated with tax revenues. In contrast, inflation is negatively associated with tax revenues. Our findings are qualitatively and quantitatively robust to using other proxies of tax policy.

[24] These results where staffing of tax administration and taxpayers' base are included separately can be obtained upon request. Following Chang et al (2020), the size of active taxpayers is measured by the sum of taxpayers on register in PIT, CIT, tax withheld by employers from employees (e.g., PAYE), and VAT, without distinguishing the type of taxpayers. The observed active taxpayers may also not capture public servants who do not have a tax identification number, which is common in developing countries. Such measurement error may possibly create attenuation bias to our estimates.

Table 2. Baseline Results, Overall, Strength, ISORA

	(1) All	(2) All	(3) All	(4) EMs & LIDCs	(5) EMs & LIDCs	(6) EMs & LIDCs
Operational Strength, 2017	0.1333***	0.1403***	0.1357***	0.1658***	0.1701***	0.1748***
	(0.0386)	(0.0395)	(0.0453)	(0.0356)	(0.0374)	(0.0428)
#Tax Staff/LaborForce		0.0631**	0.0557*		0.0588	0.0453
		(0.0309)	(0.0321)		(0.0442)	(0.0489)
Sq(#Tax Staff/Labor Force)		-0.0142*	-0.0122*		-0.0124	-0.0080
		(0.0094)	(0.0093)		(0.0151)	(0.0175)
Active Taxpayer/Labor Force			0.0067			0.0094
			(0.0135)			(0.0190)
Log (PerCapitaGDP_PPP), lagged	0.0106	-0.0044	-0.0055	0.0009	0.0161	-0.0210
	(0.0567)	(0.0623)	(0.0598)	(0.0690)	(0.0750)	(0.0677)
Square log (PerCapitaGDP_PPP), lagged	-0.0003	0.0001	0.0003	0.0004	-0.0011	0.0012
	(0.0034)	(0.0037)	(0.0036)	(0.0041)	(0.0045)	(0.0041)
Trade openness (out of GDP), lagged	0.0285***	0.0192**	0.0203**	0.0346***	0.0254**	0.0232**
	(0.0089)	(0.0098)	(0.0095)	(0.0093)	(0.0101)	(0.0097)
External debt (out of GDP), lagged	0.0057	0.0037	0.0071*	0.0027	-0.0079	0.0021
	(0.0050)	(0.0051)	(0.0037)	(0.0084)	(0.0092)	(0.0085)
CPI, lagged	-0.0341*	-0.0372	-0.0261	-0.0313	-0.0259	-0.0206
	(0.0199)	(0.0231)	(0.0203)	(0.0197)	(0.0223)	(0.0201)
Terms of Trade (2000=1), lagged	-0.0126	-0.0040	0.0118	-0.0160*	-0.0075	0.0081
	(0.0092)	(0.0098)	(0.0075)	(0.0092)	(0.0100)	(0.0070)
Oil exports (as a % of GDP), lagged	0.4271**	0.3552**	0.3062*	0.4383***	0.3617**	0.3190*
	(0.1692)	(0.1540)	(0.1624)	(0.1700)	(0.1599)	(0.1689)
Log (Agri, %GDP), lagged	-0.0057	-0.0064	-0.0055	-0.0049	-0.0088	-0.0083
	(0.0057)	(0.0073)	(0.0075)	(0.0062)	(0.0081)	(0.0077)
Control Corruption, lagged	0.0075	0.0020	-0.0003	0.0095	0.0054	0.0042
	(0.0055)	(0.0063)	(0.0064)	(0.0060)	(0.0061)	(0.0058)
Budgeted_Actual, Nontrade_tax, Combined	0.1403***	0.1401***	0.1718***	0.1525***	0.1550***	0.1834***
	(0.0514)	(0.0519)	(0.0544)	(0.0520)	(0.0534)	(0.0546)
Constant	-0.0102	0.0446	0.0238	0.0018	-0.0481	0.0810
	(0.2450)	(0.2685)	(0.2527)	(0.2867)	(0.3111)	(0.2773)
Observations	768	551	482	682	470	416
Number of countries	92	75	65	80	63	55
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Adj. Overall R2	0.3884	0.4325	0.4553	0.3494	0.3695	0.4185

Sources: Authors' calculations using IMF's WEO, IMF's WoRLD, OECD's TRD, UNU-WIDER's GRD, ISORA, TADAT, and other sources (see Table AI.1).
Notes: Estimation of equation using Hausman-Taylor estimator. Robust standard errors in parentheses. *, **, *** denote statistical significance at the 1, 5, 10 percent levels, respectively. Tax policy changes is proxied by the difference between the budgeted revenue of year *t* as share of projected GDP and the actual revenue of year *t-1* as share of actual GDP.

20. **Based on the results, a new operational tool, IMF FAD Tax Administration - Assessment and Yield Tool (TA-AYT), is developed to systematically estimate the contribution of TA reforms in a country (**Box 1**) through strengthening the country's tax administration capacity.** The TA-AYT is the first user-friendly Excel-based tool developed by the IMF Fiscal Affairs Department that provides useful information and benchmarking of the strength of TA across IMF member countries and estimates the revenue yields from TA reforms. The tool fulfills the need for estimating revenue yields from undertaking reforms that can be mapped to the components of the TA's OS index.

Table 3. Baseline Results, Overall, Strength, TADAT

	(1) All	(2) All	(3) All	(4) EMs & LIDCs	(5) EMs & LIDCs	(6) EMs & LIDCs
TADAT: PAR: Performance assessment report	0.1328***	0.1150***	0.1047**	0.1376***	0.1183***	0.1158**
	(0.0319)	(0.0391)	(0.0425)	(0.0324)	(0.0425)	(0.0489)
#Tax Staff/LaborForce		0.1304***	0.1231***		0.1114*	0.1033*
		(0.0437)	(0.0431)		(0.0603)	(0.0582)
Sq(#Tax Staff/Labor Force)		-0.0429***	-0.0393**		-0.0333	-0.0284
		(0.0159)	(0.0155)		(0.0259)	(0.0245)
Active Taxpayer/Labor Force			0.0110			0.0025
			(0.0146)			(0.0229)
Log (PerCapitaGDP_PPP), lagged	0.0049	-0.0288	-0.0184	0.0027	-0.0199	-0.0192
	(0.0629)	(0.0820)	(0.0836)	(0.0649)	(0.0930)	(0.0932)
Square log (PerCapitaGDP_PPP), lagged	-0.0006	0.0012	0.0007	-0.0004	0.0006	0.0007
	(0.0037)	(0.0049)	(0.0050)	(0.0038)	(0.0056)	(0.0056)
Trade openness (out of GDP), lagged	0.0230**	0.0252**	0.0243**	0.0209**	0.0248**	0.0233*
	(0.0095)	(0.0115)	(0.0120)	(0.0095)	(0.0122)	(0.0125)
External debt (out of GDP), lagged	0.0170***	0.0029	0.0030	0.0178***	0.0001	0.0020
	(0.0063)	(0.0142)	(0.0140)	(0.0063)	(0.0158)	(0.0152)
CPI, lagged	-0.0589*	-0.0057	-0.0050	-0.0579*	0.0007	0.0003
	(0.0325)	(0.0248)	(0.0250)	(0.0330)	(0.0254)	(0.0257)
Terms of Trade (2000=1), lagged	-0.0079	-0.0056	0.0030	-0.0077	-0.0043	0.0047
	(0.0131)	(0.0107)	(0.0102)	(0.0133)	(0.0109)	(0.0103)
Oil exports (as a % of GDP), lagged	0.5636***	0.5534***	0.5677***	0.5682***	0.5459***	0.5606***
	(0.1299)	(0.1005)	(0.0797)	(0.1264)	(0.0980)	(0.0755)
Log (Agri, %GDP), lagged	-0.0058	-0.0049	-0.0015	-0.0064	-0.0082	-0.0047
	(0.0059)	(0.0069)	(0.0074)	(0.0058)	(0.0070)	(0.0074)
Control Corruption, lagged	0.0097*	0.0030	-0.0004	0.0105*	0.0038	0.0008
	(0.0058)	(0.0074)	(0.0071)	(0.0061)	(0.0078)	(0.0076)
Budgeted_Actual, Nontrade_tax, Combined	0.0957*	0.1347*	0.1857***	0.0928	0.1342*	0.1855***
	(0.0564)	(0.0692)	(0.0698)	(0.0580)	(0.0705)	(0.0710)
Constant	0.0857	0.1858	0.1151	0.0897	0.1666	0.1325
	(0.2715)	(0.3451)	(0.3472)	(0.2793)	(0.3879)	(0.3860)
Observations	536	315	299	518	297	281
Number of countries	60	41	39	57	38	36
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Adj. Overall R2	0.4928	0.5422	0.5664	0.4706	0.4954	0.5158

Sources: Authors' calculations using IMF's WEO, IMF's WoRLD, OECD's TRD, UNU-WIDER's GRD, ISORA, TADAT, and other sources (see Table AI.1).
Notes: Estimation of equation using Hausman-Taylor estimator. Robust standard errors in parentheses. *, **, *** denote statistical significance at the 1, 5, 10 percent levels, respectively. Tax policy changes is proxied by the difference between the budgeted revenue of year *t* as share of projected GDP and the actual revenue of year *t-1* as share of actual GDP.

21. **For the practical use of this tool, extra caution is needed, with due consideration for an overall reform design and country-specific circumstances.** Indeed, TA reforms work best when they are introduced in a well-designed and properly sequenced package, accounting for the interactions and synergies between their components while fully taking into account the country's circumstances (e.g., capacity level, resource constraint). Small and isolated improvements in some practices and procedures may not automatically translate into higher yields and could even pose a fiscal risk (if the expected revenue yield is not achieved). Moreover, the tool only covers some not all possible tax administration reforms that might be needed to raise revenues. Still, it may help tax administration experts identify bottlenecks and weaknesses hindering tax administration operations.

Box 1. The New IMF FAD Tax Administration - Assessment and Yield Tool (TA-AYT)

The New IMF FAD TA-AYT is structured in three modules and designed to support policymakers, CD experts, and IMF country teams in estimating the revenue yields from tax administration reforms. It uses the ISORA dataset (ISORA 2016 and ISORA 2018 vintages) to measure the strength of TA and its practices, based on the operational strength and seven sub-indexes we constructed. Other sources for data are also used, as shown in the "Index" tab (see also Table AI.1).

TA-AYT helps benchmark the performance of TA in a particular country against comparators over time (Module 1). It provides an overview of the TA strength—overall and across the different practices or characteristics of TA, while allowing for comparisons (see "Select" tab) with two individual countries, one income group (AEs, EMDEs, EMs, LICDs), and one regional group (according to IMF geographical department, AFR, APD, EUR, MCD, and WHD; or geographical classifications, East Asia and Pacific, Europe and Central Asia, Latin American and Caribbean, Middle East and North Africa, North America, South Asia, and Sub-Saharan Africa). It helps identify policy gaps needed to strengthen the country tax administration. Moreover, Module 1 provides useful information on the evolution of tax revenues, excluding trade taxes and social contributions over time, and the correlations between tax revenues and the indexes measuring the strength of TA and its practices.

Using the regression results presented in this paper, TA-AYT provides the goodness of fit of the model (Module 2). Given that we estimate average effects, Module 2 also presents two charts allowing the user to assess the goodness of fit of the model for the country of interest and its comparators and identify if the model helps predict accurately the tax revenues for the country of interest.

Interestingly, TA-AYT can be used to estimate revenue yields from past and future TA reforms (Module 3). To do so, it allows the user to provide information on the potential outcomes and duration of country tax administration reforms by answering questions related to tax administration practices (used to compute the indexes). Then, it constructs new values for the indexes measuring the strength of tax administration and its practices, in line with the TA reforms. It also provides comparison between the strength of tax administration and its practices between the latest available (from ISORA) and new inputs by the user. Finally, it uses the regression results and the new values of the indexes computed based on the information about TA reforms imputed by the user to estimate a range of revenue yields for the country of interest, as well as the cumulative yields over time.

While TA-AYT provides useful cross-country perspectives and allows estimating revenue yields, the macro-level analysis may mask important specificity of individual countries. In this regard, it should be used with a complementary exercise, including an in-depth and specific review of the strength of tax administration. The revenue yields should be carefully validated with granular and country-specific characteristics. The applicability of the findings coming out from the tool should also be carefully assessed as there is no one-size-fits-all solution in well-designing and sequencing tax administration reforms.

Thus, in practice, using the yield estimation would call for extra caution. Taking into account country-specific circumstances (which may not be fully captured in the model analysis above), it would be prudent to use the lower econometric bound of the estimated yields (with 95 percent confidence), in case the results are used for the purpose of fiscal projections. This is because an overestimation of the yields would be more difficult to address in the course of the budget than an underestimation. Moreover, the lags between reform implementation and results could be country specific. So, for a practical purpose, the use of the estimated results at the lower bound would be encouraged, while preparing for a possible upside risk.

2. Specific Components of Tax Administration: Multivariate Models

23. **Now, we shift our focus to each specific component of TA.** The analyses above focus on the overall strength of TA, without looking into specific components (what aspects of TA improved and contributed to revenue yields?). While we attempt to explore this issue, it is a challenging task, because i) each component

of TA is correlated with each other (thus, we may not clearly distinguish the impact of one component from others); and ii) singling out one component for the analysis (without considering other components) could overestimate its impact (omitted variable problem). With these caveats, we extend the analyses with multivariate models.[25]

24. **For this purpose, we analyze the relative importance of specific practices or characteristics of TA, by including all the sub-indexes in our models** (the results are presented in Table AIII.1 and Table AIII.2 for ISORA and TADAT, respectively). [26] While acknowledging the caveats for this exercise (mentioned above), all the sub-indexes are included, as a selective use of the sub-indexes could lead to biased results (due to a possible omitted variable problem). Looking at the adjusted R-squared, the multivariate models have a higher explanatory power of the variation of tax revenues than the models with aggregate indexes. This suggests that the multivariate models add information to the analysis, while we acknowledge that the high correlations across the practices of TA make it difficult to precisely gauge the relative importance of one practice or characteristic over others in raising revenues.

- **For ISORA, enhancing compliance risk management (both for all countries and EMs and LIDCs), strengthening public accountability, and establishing LTO (for EMs and LIDCs only) can significantly raise revenues.** These are more critical to raise revenues. These findings confirm the previous results presented in Chang et al. (2020).

- **For TADAT, we find that improving accountability and transparency and timely filing of tax declarations could yield higher revenues.** However, our results show that almost all coefficients associated with POAs are non-significant when controlling for the staffing of TA and taxpayers' base (columns 2–3 and 5–6).

III. Estimating the Impacts Based on Reform Episodes

A. Data and Stylized Facts

25. **This section presents the impact of TA reforms in the period after the 2007-08 GFC based on country experiences**. Exploiting a novel database by Akitoby et al. (2020), a sample of LIDCs and EMs with sustained revenue mobilization efforts are identified and used to create a treatment group of countries with observed TA reforms (as seen in Table 4).[27] Many countries in the sample pursued tax administration and tax

[25] In light of these caveats, the results of the multivariate model are not used for the TA-AYT as this model could suffer from multicollinearity that makes the statistical significance and inference less reliable: multicollinearity can cause wider confidence intervals that can lower the statistical significance of regression coefficients. Separately, we also explore the impact of each practice and characteristic of the TA on tax revenues in single variate models (results are available upon request). The results of the single variate models show that almost all the coefficients associated with the different sub-indexes are positive and statistically significant, while we may not distinguish the impact of one component from others (as these sub-indexes are correlated with each other). In summary, our emphasis lies in the comprehensive development of TA reforms, where various TA functions mutually support one another, rather than focusing solely on individual practices or measures separately.

[26] Robustness checks using alternative proxies of tax policy can be obtained upon request. Our findings are qualitatively and quantitatively unchanged using these alternative proxies.

[27] Please follow this link to download the dataset: https://www.imf.org/-/media/Files/Publications/WP/2018/datasets/wp18234.ashx

policy reforms in parallel and all saw some increase in revenue ratio in the year after the TA reforms (see Figure AIV.2).

Table 4. TA Reforms in EMs and LIDCs After the Global Financial Crisis (GFC)

Country (Years of TA Reforms)	Management, governance & HR	Large taxpayers' office & segmentation	IT system	Registration & filing	Audit & verification program	Management of payment obligations	Improving compliance
Afghanistan (09-10)							
Algeria (08-11)							
Armenia (13-14)							
Bahamas (15)							
Belize (13-14)							
Burkina Faso (09-13)							
Burundi (09-11)							
Cambodia (12-15)							
CAR (08-10)							
Congo (11-15)							
Gambia (11-15)							
Guinea-Bissau (09-15)							
Jamaica (12-15)							
Liberia (09-12)							
Malawi (12-14)							
Maldives (11-15)							
Mauritania (10-13)							
Nepal (07-10)							
Philippines (10-15)							
Rwanda (10-14)							
Senegal (10-15)							
Sierra Leone (10-11)							
Tuvalu (13)							
Uganda (13-15)							

Legend:

TA reforms COVERED this practice

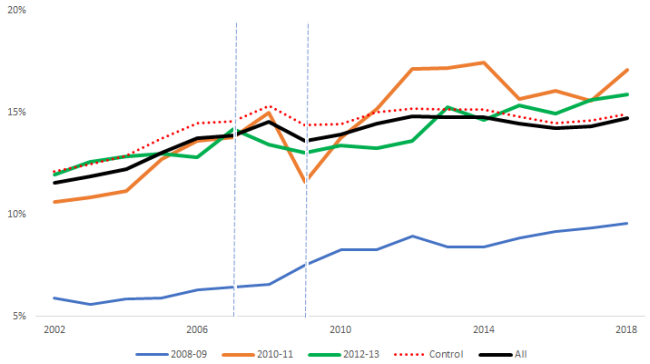
TA reforms DID NOT COVER this practice

Sources: Akitoby et al. (2020).

26. **Not surprisingly, countries made faster progress in tax collection by adopting TA reforms** (Figure 7). For comparison, we present the tax performance of countries adopting TA reforms and other EMs and LIDCs not adopting the reforms (control group). For the former, we further disaggregate the countries by

the timing of TA reforms, as the developments of tax performance differ by the adoption year of TA Reforms.[28] Notably, the country group adopting TA measures immediately after the GFC in 2008-09 had considerably lower average revenue ratio relative to other groups but also saw some significant uptick since the GFC and in the years after 2014. In contrast, the country group adopting TA measures in 2010-11 appears to have had the highest growth in tax-to-GDP ratio since the GFC and the highest average in recent years. Similarly, the country group adopting TA measures in 2012-13 saw a steady increase in tax-to-GDP ratio in the years after adoption. The control group, on the other hand, appears to have had a near-flat growth rate in tax-to-GDP ratio since the GFC. While the differences between the treated and control groups (in tax performance) are obvious, they may reflect other factors, and thus, causality cannot be inferred without adequate control for confounding factors (e.g., macroeconomic and structural factors) and unobserved heterogeneity across countries (e.g., tax policy changes).

Figure 6. Tax-to-GDP Ratio by Country Group



Sources: Authors' calculations using IMF's WEO, IMF's WoRLD, OECD's TRD, UNU-WIDER's GRD, and Akitoby et al. (2020).
Note 1: Country group 2008-09: Afghanistan, Algeria, Burkina Faso, Burundi, Central African Republic, Guinea-Bissau, Liberia, Nepal. Country group 2010-11: Congo Central Republic, Mauritania, Philippines, Rwanda, Senegal, Sierra Leone, Gambia, Maldives. Country group 2012-13: Malawi, Jamaica, Tuvalu, Uganda, Belize, Armenia. Control group: Most other EMs and LIDCs.
Note 2: Excluding trade taxes and social contributions.

B. Empirical Methodology

27. **While these TA reform episodes are identified, there are still notable challenges to estimating the impacts of TA reforms.** The actual impacts of TA reforms are difficult to isolate and quantify. For example, causal conclusions cannot be drawn by observing simple before-and-after changes in the tax-to-GDP ratio of a country, as factors other than TA reform measures may influence the ratio over time, as noted above (tax policy changes, macroeconomic and structural factors, etc.). It would also be erroneous to simply compare observed tax-to-GDP ratios of countries that have undertaken TA reforms with those without such reforms due to potential biases stemming from differences in unobservable characteristics between countries. In practice, this means one can only observe a world where TA reforms have occurred in the countries in the sample, and not a world in which they did not. Further, TA reforms are complex and come in many shapes and forms and could hardly be compressed into one "TA reform event" that provides a uniform impact on tax-to-GDP ratios. In this light, this section estimates the impact of TA reforms, using a quasi-experimental approach. It first

[28] The difference in the groups could be due to various factors including compositional aspects, variation in RA and tax policy reforms and/or macroeconomic conditions.

hypothetically constructs a counterfactual to compare the two states (before and after TA reforms) and then compares the performance of the sample of countries with observed TA reforms (treated units) relative to the performance of the control group of countries with no TA reforms.

28. **Difference-in-differences (DID) is an increasingly popular strategy to estimate causal relationships.** In its canonical format, it involves the identification of a treatment group, a control group, a pre-treatment period, and a post-treatment period. Estimating the average treatment effect on the treated (ATT) requires differencing the change in the outcome variable for the treated pre- and post-treatment from the change in outcome variable for units in the control group pre- and post-treatment. The crucial identifying assumption is that the treated group has similar trends to the control group in the absence of treatment (the so-called parallel trends assumption). In so doing, the variation in the outcome variable can be explained across time and groups.

29. **Applying a DID estimation strategy to TA reforms gives, however, rise to four complications**: i) countries often adopt TA reforms during different time periods from one another, i.e., there is often variation in treatment timing; ii) crucially, the identification assumption of parallel trends in the absence of TA reforms might not hold in practice, i.e., in the absence of TA reforms, the average tax-to-GDP ratios for treated and comparison groups might not have followed parallel paths over time; iii) countries' TA reforms may differ both in terms of scope and impact, i.e., variation in treatment effect; and iv) episodes of revenue growth in EMs and LIDCs are often associated with tax policy reforms or exogenous shocks. That is, it is difficult to isolate TA reforms gains from those stemming from tax policy changes such as the reduction of exemptions, and/or rate increases in indirect taxes, or exogenous effects, such as commodity price hikes.

30. **In this study, these identification challenges are addressed by exploiting three recent extensions to the canonical DID, namely:**

1) **Difference-in-differences in a staggered-adoption design** (Callaway and Sant'Anna, 2021). The analysis first estimates the group-time average treatment effect, i.e., the ATT of TA reforms for country group i at time t, where a "country group" is defined as a set of countries introducing TA reforms in a specific two-year period. It then uses an aggregation of parameters (across all country groups) to form summary measures of the average causal effects of TA reforms. In so doing, this extension of the canonical DID allows for staggered-adoption designs, where treated countries can adopt TA reforms at different time periods, while control countries never adopt. In addition, this enables estimation of impact in terms of duration (i.e., the x-axis shows time elapsed after TA reforms).

- **A natural way to generalize the ATT from the canonical two periods and two groups case to the multiple periods case is to estimate the group-time average treatment effects through the following identification equation by Callaway and Sant'Anna (2021).**

$$ATT(g,t)=E[Y_t-Y_{t-1}|G=g]-E[Y_t-Y_{t-1}|C=1] \qquad (2)$$

  This is the average effect of TA reforms for countries in group *g* at period *t* and where *Y* is the tax-to-GDP ratio and *C* is an indicator variable for never-treated control group. That is, the pre- and post- TA reforms tax-to-GDP ratios for the group initiating TA reforms in the period *t* is differenced from the pre- and post-tax-to-GDP ratios for the countries in the control group across the same time periods.

- **Next, it would be necessary to aggregate the resulting group-time average treatment effects to estimate the overall ATT.** This can be done in two ways: i) first, the ATT for each group (across all time periods) can be computed and then averaged across groups to summarize the overall effect of TA reforms;

ii) ATTs of countries that have been exposed to the treatment (TA reforms) for exactly *x* time periods can be averaged to estimate how effects of TA reforms vary with elapsed time. This method has the added benefit of circumventing the issue of changing the composition of groups (a common shortcoming of the first method).

2) **Synthetic Difference-in-Differences (SDID) with common support** (Arkhangelsky et al., 2021). To explain variation in tax-to-GDP ratios across time and groups with greater reliability, SDID estimates ATT using DID between treated units and synthetic control units, where synthetic control units are chosen as an optimally weighted function of countries without TA reforms and pre- TA reforms time periods. In so doing, the resulting SDID estimator relies on weighing control observations to achieve parallel trends with treated observations as opposed to just assuming the parallel common support implicit in the canonical DID. The SDID reweighs the unexposed control countries to make their time trend parallel to countries in the treatment before the TA reforms, and then applies a staggered DID analysis to this reweighed panel. In practice, the estimation procedure follows three steps: i) a least square estimation of country weights defining a synthetic control group using pre-treatment data; ii) a least square estimation of time weights defining a synthetic pre-treatment period using control data; iii) a DID application in a staggered design (iterating on subsets of the data) to the resulting synthetic panels.

3) **Controlled Interrupted Time-series Analysis (CITSA) (Linden, 2015).** To estimate the ATT of different TA reform packages, CITSA can be applied to understand how the tax-to-GDP ratio for a specific country changed after TA reforms relative to a control group of countries. As an extension of the canonical DID, it allows the assumption of parallel trends to be verified, and for differences in trend between two groups to be adjusted for, while controlling for time-varying confounders, including contemporaneous events, that affect both groups.[29] To do so, the CITS design implicitly controls for differences in the baseline mean and trends (i.e., it does not require the common trend assumption) between the treatment and comparison group (in contrast to the DID and the SDID methodologies that control for the baseline mean only and require the common support assumption).

31. **The CITSA regression model can be specified as follows:**

$$Y_t = \beta_0 + \beta_1 T_t + \beta_2 X_t + \beta_3 X_t T_t + \beta_4 Z + \beta_5 Z T_t + \beta_6 Z X_t + \beta_7 Z X_t T_t + \in t \quad (3)$$

Z is a dummy variable to denote the cohort assignment (=1 for countries with TA reforms), the coefficients β_0 to β_3 represent the control group, and the coefficients β_4 to β_7 represent tax-to-GDP ratio of the country that adopted TA reforms. More specifically, β_0 represents the intercept or the starting level of tax-to-GDP ratio for the control group. β_1 is the slope or trajectory pre- TA reform period (i.e., before treated units introduced TA reforms) for the control group. β_2 represents the change in the level that occurs in the period immediately following the intervention (compared with the counterfactual) for the control group. β_3 represents the difference between pre-intervention and post intervention slopes in tax-to-GDP for the control group. β_4 represents the difference in the level (intercept) of the tax-to-GDP ratio between control group and treated country prior to the TA reforms, β_5 represents the difference in the slope (trend) of tax-to-GDP ratio prior to TA reforms adoption, β_6 indicates the difference between the treated and control groups in the level of the tax-to-GDP ratio immediately following introduction of TA reforms, and β_7 represents the difference between the treated and control groups in

[29] In particular, co-interventions and other events concurrent with the intervention.

the slope (trend) of tax-to-GDP ratio after initiation of TA reforms compared with pre- TA reform period (akin to a DID of slopes).

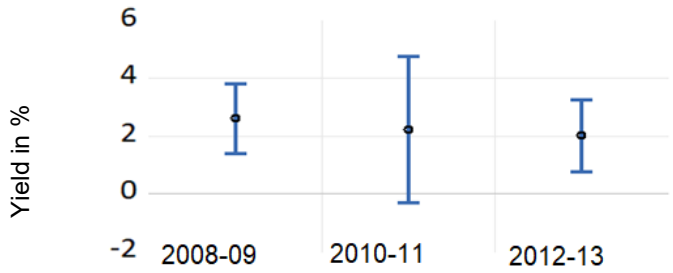
32. **The DID methodologies above likely partially control the impact of tax policy changes**. As an additional robustness test, the effect of tax policy changes is controlled for by proxying it by the difference between the budgeted revenue and the actual revenue in the DID regressions (as done in Chang et al., 2020 and earlier in this paper) in which the budgeted revenue reflects planned tax policies when the budget is prepared. Despite this, the inclusion of this control variable did not yield a statistically significant coefficient nor change the magnitudes or signs of key coefficients of interest. While imperfect, we can infer that the above DID specifications are able to absorb most of the systematic impact of tax policy changes, ideally leaving unrelated idiosyncratic noises in the error term or in some instances inflating the impact of TA reforms somewhat (provided tax policy changes on average increase tax-to-GDP ratio).

C. Results

1. Difference-in-Differences in a Staggered-Adoption Design

33. **The results from the first stage of the staggered DID point to an average increase in tax revenues by about 2 pp. of GDP for the three country groups adopting TA reforms (Table 7).** The 2008-09 country group had the highest mean average ATT (2.6 pp. of GDP), while the 2012-13 had somewhat lower impact of TA reforms (2 pp. of GDP). The results for the 2010-11 country group had a large variance and a statistically less significant coefficient, but the point estimate shows an increase in tax-to-GDP ratio by 2.2 pp of GDP. While it appears that the 2010-11 cohort experienced the largest increase in tax revenues after TA reforms (steepest slope in Figure 6), the 2008-09 cohort likely had the highest ATT because these countries had the longest exposure to TA reforms but also started from the lowest base given the GFC.

Figure 7. Group Time ATT (Yield in %)

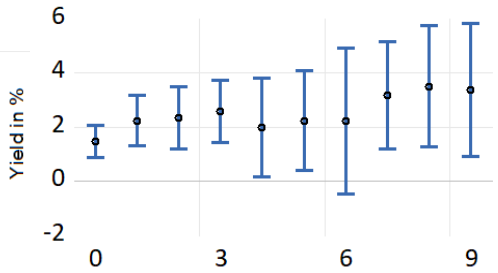


Sources: Authors' calculations using IMF's WEO, IMF's WoRLD, OECD's TRD, UNU-WIDER's GRD, and Akitoby et al. (2020).
Notes: Country group 2008-09: Afghanistan, Algeria, Burkina Faso, Burundi, Central African Republic, Guinea-Bissau, Liberia, Nepal. Country group 2010-11: Congo Central Republic, Mauritania, Philippines, Rwanda, Senegal, Sierra Leone, Gambia, Maldives. Country group 2012-13: Malawi, Jamaica, Tuvalu, Uganda, Belize, Armenia.

34. **Further analyses (by aggregating the resulting group-time average treatment effects) reveal more granular results, with yield increasing over time from about 1.5 pp. of GDP in the 1st year following the reforms to more than 3 pp. of GDP after the 6th year.** Since the group time ATT does not impose any restriction on treatment effect heterogeneity across groups or across time, an overall average effect of TA reforms can be formed through aggregation (the second stage of the staggered DID). Here, to circumvent the issues of the changing composition of groups, an aggregation scheme based on exposure is

used. Figure 8 shows how the effect of TA reforms varies with the length of exposure. The ATT increases in the first three years from about a yield of about 1.5 pp. of GDP in the 1st year after TA reforms adoption to about a yield of 2.6 pp. of GDP in the 3rd year. It then goes down a bit and stabilizes around 2.2 pp. of GDP before increasing to just above 3 pp. of GDP after the 6th year.[30]

Figure 8. Duration ATT (Yield in %)



Sources: Authors' calculations using IMF's WEO, IMF's WoRLD, OECD's TRD, UNU-WIDER's GRD, and Akitoby et al. (2020).

Notes: Years after TA reforms. ATT effect heterogeneity with respect to length of exposure to TA reforms. 2011 and 2013 country groups are dropped from the aggregate ATT after year 7 and year 5, respectively.

2. Synthetic Difference-in-Differences with Common Support

35. **The results from the Synthetic Difference-in-Differences with Common Support, in which both time and unit weights are used to create a reliable counterfactual, show an estimated yield of around 2.2 pp. of GDP, which is very similar to the results from the staggered DID (Table 5).** The ATT for countries with TA reforms is obtained from the application of the SDID estimator repeatedly, in which optimal unit and time-specific weight vector is calculated for each adoption period. The estimator is simply a weighted average of the three country-group ATTs, where the weight applied to any country group ATT is equal to the proportion of countries with TA reforms that originate in a specific country group. Supporting calculations are presented in Figure 9. The left-hand chart in Panel A illustrates trends in tax-to-GDP ratio over time for countries that adopted TA reforms in 2008-09 and the weighted average of control countries, with the weights used to average pre-TA reforms time periods at the bottom of the graphs (in green). The right-hand chart shows the country-by-country tax-to-GDP ratio difference, with the weights indicated by dot size and the estimated effect (weighted average of the differences)—indicated by a horizontal line. Panels B and C present the results for the 2010-11 and 2012-13 country groups, respectively.

Table 5. Aggregated ATE

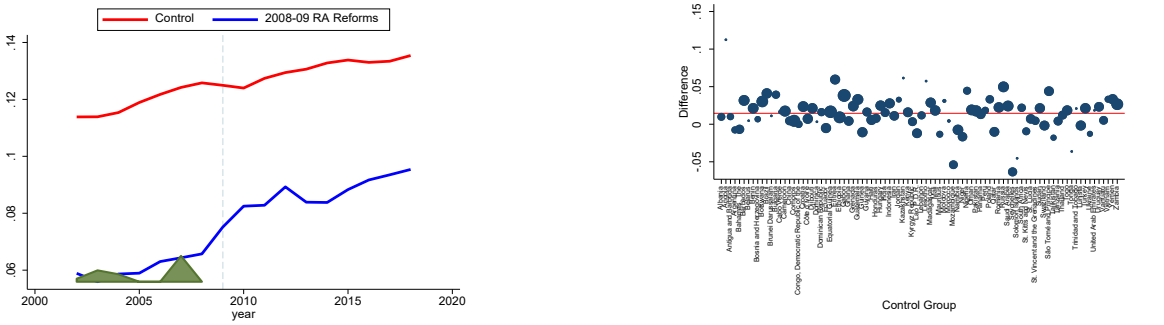
	Yield in %	Std. Err.	t	P>t
Average ATE	2.168	0.00609	3.56	0.000

Sources: Authors' calculations using IMF's WEO, IMF's WoRLD, OECD's TRD, UNU-WIDER's GRD, and Akitoby et al. (2020).

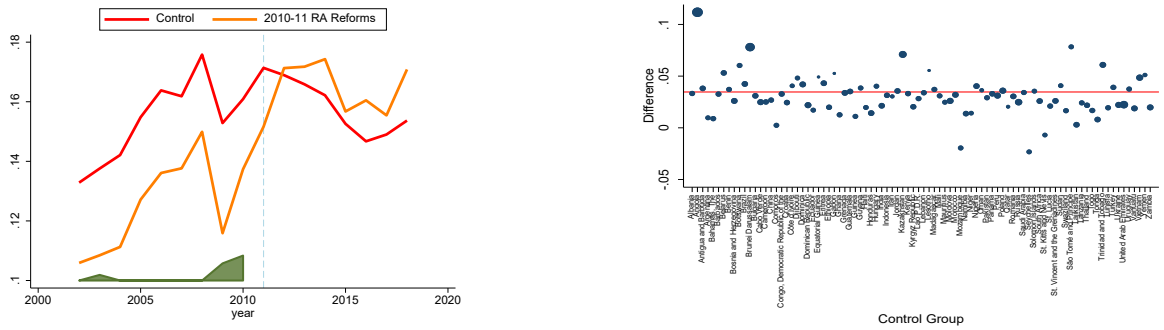
[30] One caveat, however, is that the sample runs until 2018 and that since countries in the 2011 and 2013 country groups are not observed over a 9-year period, they are dropped from the aggregation of ATTs after the 7th and 5th years, respectively.

Figure 9. Country Group vs. Control Group

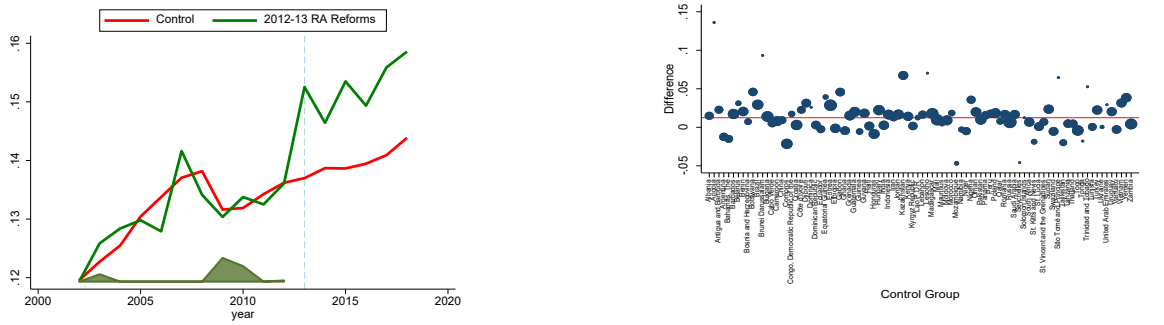
Panel A. 2008-09



Panel B: 2010-11



Panel C: 2012-13



Sources: Authors' calculations using IMF's WEO, IMF's WoRLD, OECD's TRD, UNU-WIDER's GRD, and Akitoby et al. (2020).

Notes: The chart on the left-hand side shows trends in revenue ratio over time for country groups in the sample and the relevant weighted average of control countries are shown, with the weights used to average pretreatment time periods at the bottom of the graphs. Here, the SDID allows for an intercept in the estimation weights (i.e., unlike traditional Synthetic Control Method where the goal of the weight is to perfectly match the treatment group, the SDID only seek to mimic the trend of the treatment group even if the two groups have a very different intercept as shown in the left-hand charts of Panel A and B). The estimated effect is indicated by applying a DID analysis to this reweighted panel. The charts on the right-hand side, show the country-country adjusted revenue ratio difference, with the weights indicated by dot size and the weighted average of these differences: the estimated effect—indicated by a horizontal line.

3. Controlled Interrupted Time-series Analysis Case Studies

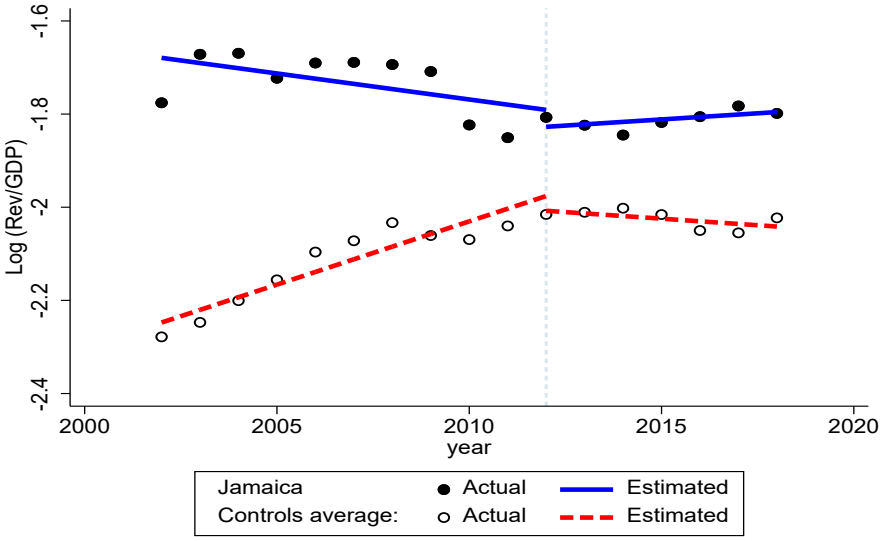
36. **The results below present the estimation of the effect of TA reforms in three country case studies relative to control countries, when tax-to-GDP ratios are ordered as a time series.** The yield estimates are broadly in line with the preceding results based on country groups. While the DID method applied can partially control for tax policy measures, we expect to see larger yields from these regressions, if tax policy measures in the treated unit (country undertaking TA reforms) increase tax-to-GDP ratio over time relative to the control group (i.e., the estimated yields could be upward bias). Importantly, TA policies are often implemented as a package, reinforcing each other, and they work by elevating the whole performance of the TA, which in turn leads to the improvement of voluntary compliance. In line with this, the results from the case studies below indicate that the establishment of LTO and its segmentation and the improvement in compliance risk management are essential parts of TA reforms. The result highlights that the yield gains of a comprehensive TA reform package would be reduced without those measures.

a. Jamaica – Comprehensive TA Reforms

37. **The impact of TA reforms in Jamaica is of particular interest since the authorities used a series of measures to bolster tax collection between 2012 and 2015.** The revenue administrative measures adopted include improving compliance, mandatory electronic filing and payment, modernizing the tax administration system, facilitating electronic filing and payment, creating a semi-autonomous revenue agency, expanding the staff at the LTO, and strengthening the LTO with increased audit coverage. Legislative changes were also adopted to increase the powers of the tax administration (Akitoby et al., 2020). Figure 10 illustrates the trends of logarithm of tax-to-GDP ratio over time for Jamaica relative to a control group of countries with no TA reforms. The table at the bottom presents the key coefficients of interest from the CITSA regression for Jamaica. Specifically, β4 shows that Jamaica had a much larger tax-to-GDP ratio relative to the countries in the control, despite there being no statistically significant difference in trend between the two (β5).

38. **Jamaica saw a statistically significant sustained difference of 4.9 pp. of GDP in tax-to-GDP ratio than the control group.** This can be extrapolated to a yield difference of about 0.8 pp. of GDP on annual basis. Interestingly, there was no difference in the level of tax-to-GDP ratio immediately following the introduction of TA reforms, suggesting that TA reforms took time to yield gains.

Figure 10. CITSA Jamaica vs. Controls Average



Dependent variable: Log (REV/GDP)	Coefficient	std. err.	t	P>t
β4 - difference in the level of revenue ratio before RA Reforms	0.568	0.0644	8.82	0.000
β5 - difference in the trend in revenue ratio before RA Reforms	-0.038	0.0121	-3.17	0.002
β6 difference in the level of revenue ratio immediately following introduction of RA Reforms	-0.0048	0.0892	-0.05	0.958
β7 difference in the trend of revenue ratio after initiation of RA reforms compared with preinte	0.049	0.0191	2.57	0.010

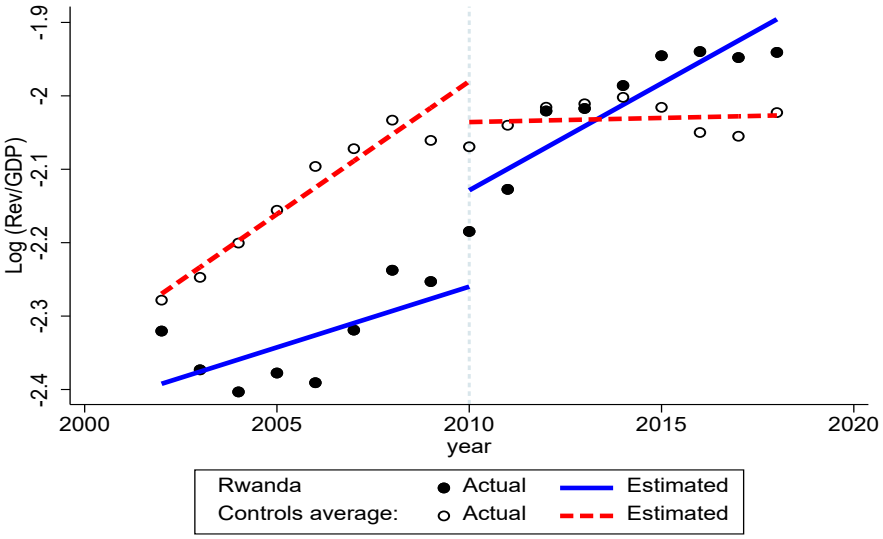
Sources: Authors' calculations using IMF's WEO, IMF's WoRLD, OECD's TRD, UNU-WIDER's GRD, and Akitoby et al. (2020).

b. Rwanda – Comprehensive TA Reforms (Less LTO and Segmentation)

39. **Rwanda adopted similar measures as Jamaica, over the period 2010-14**. Specifically, the authorities introduced electronic filing and payment systems and implemented electronic registration in addition to improvements in tax compliance and collection of arrears. There were, however, no observed reforms related to LTO or segmentation. The results for Rwanda (Figure 11) show that prior to the TA reforms in 2010, the country had a lower tax-to-GDP ratio than the control groups, though statically significant differences in level and trend could not be established due to the large variances in the coefficients, possibly driven by the GFC.

40. **In the period after the TA reforms, there was a considerable uptick in the tax-to-GDP ratio of Rwanda in levels relative to the control, with a sustained revenue yield of 4.7 pp. of GDP.** The large immediate uptick of revenue could potentially be due to tax policy measures, specifically the increases in the tax rate for imported construction materials and excise duty on airtime of mobile phones in addition to removal of some VAT exemptions (Akitoby et al., 2020).

Figure 11. CITSA Rwanda vs. Controls Average



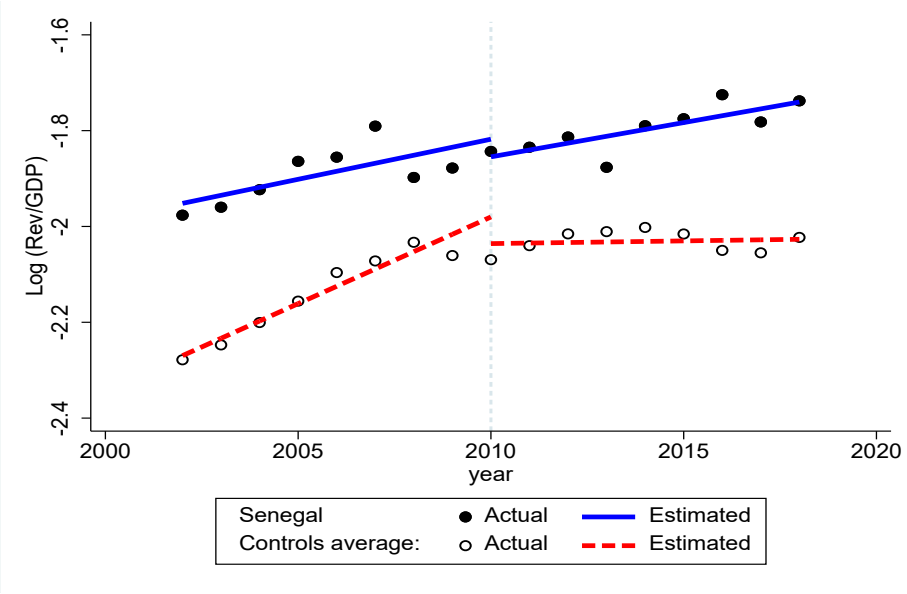
Dependent variable: Log (REV/GDP)	Coefficient	std. err.	t	P>t
β4 - difference in the level of revenue ratio before RA Reforms	-0.1226	0.068	-1.8	0.072
β5 - difference in the trend in revenue ratio before RA Reforms	-0.0196	0.0151	-1.3	0.193
β6 difference in the level of revenue ratio immediately following introduction of RA Reforms	0.1868	0.0917	2.04	0.042
β7 difference in the trend of revenue ratio after initiation of RA reforms compared with preinte	0.0476	0.0194	2.45	0.014

Sources: Authors' calculations using IMF's WEO, IMF's WoRLD, OECD's TRD, UNU-WIDER's GRD, and Akitoby et al. (2020).

c. Senegal – Comprehensive TA Reforms (Less Improving Compliance)

41. **Senegal implemented various TA reforms between 2010 and 2015, leading to an estimated revenue yield of 3.3 pp. of GDP**. These included improvements in taxpayer segmentation, and e-filing and e-payment of taxes, which helped to strengthen efficiency. However, the authorities had no observed reforms related to improving compliance. The results of the CITSA regression (Figure 12) show that prior to the TA reforms in 2010, the country had a higher tax-to-GDP ratio in level than the control group, though no statically significant difference in trend could be established. In the period after the TA reforms, Senegal saw a sustained difference of 3.3 pp. of GDP (β7) relative to the control group, but no immediate impact following introduction of TA reforms.

Figure 12. CITSA Senegal vs. Controls Average



Dependent variable: Log (REV/GDP)	Coefficient	std. err.	t	P>t
β4 difference in the level of revenue ratio before RA Reforms	0.3178	0.0479	6.64	0.000
β5 difference in the trend in revenue ratio before RA Reforms	-0.019	0.0118	-1.65	0.099
β6 difference in the level of revenue ratio immediately following introduction of RA Reforms	0.0187	0.074	0.25	0.799
β7 difference in the trend of revenue ratio after initiation of RA reforms compared with preinte	0.0326	0.014	2.27	0.023

Sources: Authors' calculations using IMF's WEO, IMF's WoRLD, OECD's TRD, UNU-WIDER's GRD, and Akitoby et al. (2020).

IV. Concluding Remarks

42. **In light of the challenges to quantify the impact of TA strength and reforms on tax revenues and the scarcity of analyses on this topic in the literature, this paper explores more robust empirical approaches to estimate potential yields from TA reforms.** As existing TA indexes are slow-moving and a country's TA status is endogenously determined, a robust empirical framework is needed to estimate the revenue yields of TA reforms with cross-country data. A detailed examination of specific TA reform episodes is also needed to account for the specificity of each reform episode, which better identifies the heterogeneity and dynamics of the impact of TA reforms over time. Specifically, this paper employed innovative panel data methods (a Hausman-Taylor cross-country panel regression and several recently developed event-study models) to estimate the revenue yields from various TA features, as measured by comprehensive ISORA and TADAT indicators.

43. **Our key findings are the following:**

- **The strength of TA generally changes only moderately over time, while there are some exceptional cases.** Looking at the developments of the ISORA index between 2014 and 2017, we find that, in most countries, the changes in the TA strength were modest, likely reflecting the difficulties in building TA capacity and sustaining capacity-building efforts. Nevertheless, 17 countries strengthened their TA during

the period (moved from the 40th percentile to the 60th percentile of the OS index). This suggests that there is scope for countries to strengthen TA, despite such difficulties.

- **TA strength and reforms are associated with a significant increase in tax revenues.** The panel regression results show that an increase in the TA strength by an amount equivalent to the change between the 40th percentile and the 60th percentile of the IMF member countries is associated with an increase in tax revenues by up to 1.8 pp. of GDP. We also provide a range estimate of revenue yields, based on 95 percent confidence intervals, ranging from 0.5 to 2.6 pp. of GDP using ISORA and 0.2 to 2.1 pp. of GDP using TADAT for such an increase in the indexes. The impact assessment shows similar results, with sustained TA reforms after the GFC leading to an increase in tax revenues between 2 to 3 pp. of GDP, with the treatment effect increasing over six years, in line with developments in three country cases (Jamaica, Rwanda, and Senegal).
- **While significant TA reforms may lead to some initial gains, it could take time for the full results to be realized.** The event-study analysis reveals that the revenue yields are increasing over time to more than 3 pp. of GDP after the 6th year following a far-reaching and comprehensive reform across several practices of tax administration.
- **To achieve revenue gains of this magnitude, a set of TA reforms should be properly implemented.** TA reforms can take different forms, and multiple TA measures were implemented in successful reform cases (e.g., Jamaica, Rwanda, and Senegal). Among the TA measures, we found specific TA reform areas with higher yields. Those measures include: (i) strengthening compliance risks management, (ii) enhancing public accountability, (iii) establishing LTO, (iv) strengthening accountability and transparency, and (v) enhancing timely filing of tax declarations. That said, attribution is not an exact science, and we would emphasize the need to develop TA integrally, as different TA functions support and reinforce each other.

44. **This paper offers a comprehensive database in benchmarking a country's state of TA against comparators across multiple dimensions and estimating revenue yields that would be useful for the IMF country operations and capacity development of its member countries.** An operational tool is developed in parallel to systematically estimate the contribution of past or future TA reforms for IMF member countries (Box 1). From a policy perspective, the finding of this paper usefully identified specific reforms of higher yields that help in reform prioritization. The operational tool gathers information on countries' TA practices. Then, using the regression results from our analyses, it constructs new values for the indexes measuring the strength of TA and its practices, in line with the TA reforms, to estimate a range of revenue yields from country TA reforms.

45. **While this paper provides useful cross-country perspectives, the macro-level analysis masks important specificity of individual countries.** In this regard, the findings of this paper should be carefully validated with granular country-level case studies. The applicability of the findings coming out from the analysis to each country should also be carefully assessed as there is no one-size-fits-all solution in well-designing and sequencing tax administration reforms. Taking into account country-specific circumstances (which may not be fully captured in the model analyses), it would be prudent to use the yield estimates at the lower bound, in case the results are used for fiscal projections, especially for countries with capacity constraints, weaker ownership, and institutions. Indeed, TA reforms work best when they are introduced in a well-designed and properly sequenced package, accounting for the interactions and synergies between their components, while fully taking into account the country's circumstances. Small and isolated improvements in some practices and procedures may not automatically translate into higher yields and could even pose a fiscal risk (if the expected revenue yield is not achieved).

Annex I. Data Descriptions and Sources

Table AI.1. Data Descriptions and Sources

	Variables	Descriptions	Sources
ISORA	Operational Strength	It measures the overall strength of the revenue administration across practices and characteristics.	ISORA dataset
	Compliance Risk Management	It measures the formal approach for identifying, assessing, and prioritizing key compliance risks (automated risk profiling, electronic audit, …)	ISORA dataset
	Third Party Data	It measures the use of computer-based information systems for processing various forms of third-party data and use of third-party in prefilling returns.	ISORA dataset
	Digitalization	It measures the provision of e-services, including e-filing, e-payment, and web-based information and communication services	ISORA dataset
	Service Orientation	It measures the service offerings facilitating compliance and the provisions to enhance taxpayer satisfaction with services offered.	ISORA dataset
	Public Accountability	It measures if key documents such as plans and achievements against plans are published and the use of external auditor	ISORA dataset
	Autonomy	It measures the institutional form, the degree of autonomy in managing expenditure and human resources, and ability to set performance standards, and provision of tax policy advice.	ISORA dataset
	Est. LTO	It takes the value one if there is an office or program dedicated to large taxpayers in operation	ISORA dataset
TADAT	PAR: Performance Assessment Report	It measures the overall strength of the revenue administration across practices and characteristics.	TADAT dataset
	POA1: Integrity of the Registered Taxpayer Base	It measures if all businesses, individuals, and other entities that are required to register are included in a taxpayer registration database. Information held in the database is complete, accurate, and up-to-date.	TADAT dataset
	POA2: Effective Risk Management	It measures if the risks to revenue and tax administration operations are identified and managed effectively.	TADAT dataset
	POA3: Supporting Voluntary Compliance	It measures if taxpayers have the necessary information and support to voluntarily comply at a reasonable cost to them.	TADAT dataset
	POA4: Timely Filing of Tax Declarations	It measures if taxpayers file tax declarations on time.	TADAT dataset
	POA5: Timely Payment of Taxes	It measures if taxpayers pay their taxes in full on time.	TADAT dataset
	POA6: Accurate Reporting in Declarations	It measures if taxpayers report complete and accurate information in their tax declarations.	TADAT dataset
	POA7: Effective Tax Dispute Resolution	It measures if the tax dispute resolution process is fair and independent, accessible to taxpayers, and effective in resolving disputed matters in a timely manner.	TADAT dataset
	POA8: Efficient Revenue Management	It measures if tax revenue collections are fully accounted for, monitored against expectations, and analyzed to inform government revenue forecasting. Legitimate tax refunds are paid promptly.	TADAT dataset
	POA9: Accountability and Transparency	It measures if the tax administration is transparent in the conduct of its activities and accountable to the government and community.	TADAT dataset
Dependent and Control Variables	Tax revenues excluding trade taxes and social security contributions (% of GDP)	Total tax revenues minus trade taxes and social security contributions as a percent of GDP	Authors' calculations using the IMF's WEO, the IMF's WoRLD, the OECD's TRD, and the UNU-WIDER's GRD.
	Log (PerCapitaGDP_PPP), lagged	One-year lagged value of the log of GDP per capita in purchasing power parity	Authors' calculations using the IMF's WEO dataset
	Square log (PerCapitaGDP_PPP), lagged	One-year lagged value of the squared of the log of GDP per capita in purchasing power parity	Authors' calculations using the IMF's WEO dataset
	Trade openness (out of GDP), lagged	One-year lagged value of trade openness (sum of exports and imports over GDP)	Authors' calculations using the IMF's WEO dataset
	External debt (out of GDP), lagged	External debt as percent of GDP	Authors' calculations using the IMF's WEO and World Bank's WDI datasets
	CPI, lagged	One-year lagged value of inflation, percent change of consumer price index, period average	Authors' calculations using the IMF's WEO and World Bank's WDI datasets
	Terms of Trade (2000=1), lagged	One-year lagged value of terms of trade (2000=1)	Authors' calculations using the IMF's WEO and World Bank's WDI datasets
	Oil exports (as a % of GDP), lagged	One-year lagged value of oil exports as percent of GDP	Authors' calculations using the IMF's WEO dataset
	Log (Agri, %GDP), lagged	One-year lagged value of the log of the share of agriculture in GDP	Authors' calculations using the World Bank's WDI dataset
	Control Corruption, lagged	One-year lagged value of the log of control of corruption index	Authors' calculations using the World Bank WGI dataset
	Budgeted_Actual, Nontrade_tax, Combined	Proxy of tax policy measures: the difference between the budgeted revenue of this year as share of projected GDP and the actual revenue of last year as share of actual GDP	Authors' calculations using the IMF's WEO, the IMF's WoRLD, the OECD's TRD, and the UNU-WIDER's GRD.
	Top Combined CIT Rate (%), lagged	One-year lagged value of top CIT rate (%)	IMF's Tax Rate Database (DART) and IMF's GFS dataset
	Top Combined PIT Rate (%), lagged	One-year lagged value of top PIT rate (%)	IMF's Tax Rate Database (DART) and IMF's GFS dataset
	Standard VAT Rate (%), lagged	One-year lagged value of VAT rate (%)	IMF's Tax Rate Database (DART) and IMF's GFS dataset
	Budgeted_L.budgeted, Nontrade_tax, Combined	Proxy of tax policy measures: the difference between the budgeted revenue of this year as share of projected GDP and the budgeted revenue in last year as share of projected GDP	Authors' calculations using the IMF's WEO, the IMF's WoRLD, the OECD's TRD, and the UNU-WIDER's GRD.
	VAT compliance gap	The VAT compliance gap is the difference between the actual VAT revenues, determined on the accrued collections basis, and potential VAT revenues, estimated using the current policy framework.	Authors' calculations using the IMF's RA-GAP and other sources
	Average daily number of structural conditionalities (SCs) on revenue administration reforms	The number of SCs on RA reforms divided by the program duration in days	IMF's MONA Dataset
	Average daily number of SCs on revenue administration reforms met	The number of SCs on RA reforms met divided by the program duration in days	IMF's MONA Dataset
	The share of SCs on RA reforms met	The number of SCs on RA reforms met divided by the total number of SCs	IMF's MONA Dataset
	Data on Revenue Administration Reforms	Identified countries and years with sustained period of RA reforms	Akitoby et al. (2020)

Annex II. Descriptive Statistics and Stylized Facts

Table AII.1. Summary Statistics

		Obs.	Mean	Sd	CV	Median	Min	Max
ISORA	Operational Strength	146	0.617	0.173	0.280	0.630	0.175	0.951
	Compliance Risk Management	146	0.562	0.305	0.543	0.667	0.000	1.000
	Third Party Data	146	0.332	0.258	0.777	0.273	0.000	0.955
	Digitalization	146	0.532	0.327	0.615	0.583	0.000	1.000
	Service Orientation	146	0.719	0.203	0.282	0.769	0.091	1.000
	Public Accountability	146	0.573	0.248	0.433	0.583	0.000	1.000
	Autonomy	146	0.713	0.222	0.311	0.760	0.077	1.000
	LTO	146	0.890	0.313	0.352	1.000	0.000	1.000
	#Rev.Staff/LaborForce	123	1.118	0.845	0.756	0.857	0.060	3.523
	#ActiveTaxpayer/LaborForce	113	0.654	0.731	1.118	0.338	0.000	2.922
TADAT	PAR: Performance assessment report	77	0.356	0.188	0.528	0.370	0.021	0.896
	POA1: Integrity of the Registered Taxpayer Bas	77	0.351	0.250	0.712	0.333	0.000	1.000
	POA2: Effective Risk Management	77	0.252	0.216	0.857	0.200	0.000	1.000
	POA3: Supporting Voluntary Compliance	77	0.529	0.268	0.507	0.524	0.000	1.000
	POA4: Timely Filing of Tax Declarations	77	0.275	0.288	1.047	0.167	0.000	1.000
	POA5: Timely Payment of Taxes	77	0.414	0.232	0.560	0.381	0.000	0.952
	POA6: Accurate Reporting in Declarations	76	0.253	0.196	0.775	0.244	0.000	1.000
	POA7: Effective Tax Dispute Resolution	77	0.479	0.224	0.468	0.533	0.000	0.933
	POA8: Efficient Revenue Management	76	0.271	0.261	0.963	0.167	0.000	1.000
	POA9: Accountability and Transparency	77	0.369	0.238	0.645	0.333	0.000	0.905

Notes: This table shows the summary statistics for ISORA (only using the year 2017, latest available data), and TADAT. Sd stands for standard deviation, and CV for coefficient of variation.

Table AII.2. Correlations Between ISORA and TADAT Indexes Across Years

	Variables	Correlation	Obs.
ISORA	Operational Strength	0.92	121
	Compliance Risk Management	0.85	121
	Third Party Data	0.88	121
	Digitalization	0.75	121
	Service Orientation	0.83	121
	Public Accountability	0.81	121
	Autonomy	0.83	121
	LTO	0.61	121
	#Rev.Staff/LaborForce	0.95	90
	#ActiveTaxpayer/LaborForce	0.93	85
TADAT	PAR: Performance assessment report	0.75	11
	POA1: Integrity of the Registered Taxpayer Bas	0.40	11
	POA2: Effective Risk Management	0.41	11
	POA3: Supporting Voluntary Compliance	0.81	11
	POA4: Timely Filing of Tax Declarations	0.63	11
	POA5: Timely Payment of Taxes	0.35	11
	POA6: Accurate Reporting in Declarations	0.15	11
	POA7: Effective Tax Dispute Resolution	0.27	11
	POA8: Efficient Revenue Management	0.54	10
	POA9: Accountability and Transparency	0.70	11

Notes: This table shows the correlations between the i) ISORA's indexes in 2014 (first year available) and 2017 (last year available), and ii) the TADAT's indexes from first and last assessments (for repeated assessments only). Note that given the low number of observations, the correlations between the TADAT's indexes should be cautiously considered.

Table AII.3. Correlations among ISORA and TADAT Indexes, and between ISORA and TADAT Indexes

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
(1)		Operational Strength	1																		
(2)		Compliance Risk Management	0.694***	1																	
(3)		Third Party Data	0.631***	0.511***	1																
(4)		Digitalization	0.749***	0.404***	0.393***	1															
(5)	ISORA	Service Orientation	0.776***	0.521***	0.465***	0.527***	1														
(6)	ISORA	Public Accountability	0.686***	0.326***	0.395***	0.415***	0.494***	1													
(7)	ISORA	Autonomy	0.687***	0.369***	0.337***	0.445***	0.593***	0.391***	1												
(8)	ISORA	LTO	0.354***	0.011	-0.144***	0.149***	0.101***	0.188***	0.151***	1											
(9)	ISORA	#Rev.Staff/LaborForce	0.219***	0.262***	0.284***	0.185***	0.031	0.148***	0.041	-0.029	1										
(10)	ISORA	#ActiveTaxpayer/LaborForce	0.532***	0.469***	0.644***	0.366***	0.348***	0.336***	0.239***	-0.061*	0.538***	1									
(11)		PAR: Performance assessment report	0.590***	0.315***	0.394***	0.435***	0.423***	0.350***	0.302***	0.253***	0.237***	0.410***	1								
(12)	TADAT	POA1: Integrity of the Registered Taxpayer Bas	0.420***	0.349***	0.224***	0.251***	0.289***	0.256***	0.139***	0.248***	0.252***	0.228***	0.812***	1							
(13)	TADAT	POA2: Effective Risk Management	0.568***	0.382***	0.364***	0.406***	0.471***	0.249***	0.299***	0.215***	0.049	0.322***	0.810***	0.648***	1							
(14)	TADAT	POA3: Supporting Voluntary Compliance	0.525***	0.076*	0.261***	0.460***	0.345***	0.391***	0.336***	0.310***	0.127**	0.216***	0.812***	0.609***	0.584***	1						
(15)	TADAT	POA4: Timely Filing of Tax Declarations	0.505***	0.284***	0.319***	0.413***	0.325***	0.365***	0.202***	0.169***	0.292***	0.507***	0.826***	0.617***	0.604***	0.560***	1					
(16)	TADAT	POA5: Timely Payment of Taxes	0.333***	0.086**	0.302***	0.222***	0.217***	0.239***	0.234***	0.127***	0.362***	0.269***	0.732***	0.493***	0.463***	0.516***	0.677***	1				
(17)	TADAT	POA6: Accurate Reporting in Declarations	0.466***	0.339***	0.420***	0.388***	0.292***	0.187***	0.153***	0.149***	0.149***	0.451***	0.822***	0.604***	0.762***	0.636***	0.679***	0.518***	1			
(18)	TADAT	POA7: Effective Tax Dispute Resolution	0.395***	0.163***	0.329***	0.238***	0.337***	0.208***	0.193***	0.244***	0.126**	0.362***	0.672***	0.386***	0.485***	0.621***	0.496***	0.498***	0.483***	1		
(19)	TADAT	POA8: Efficient Revenue Management	0.378***	0.281***	0.382***	0.191***	0.345***	0.211***	0.223***	0.016	0.215***	0.339***	0.778***	0.671***	0.550***	0.533***	0.586***	0.559***	0.598***	0.386***	1	
(20)	TADAT	POA9: Accountability and Transparency	0.672***	0.357***	0.281***	0.547***	0.462***	0.378***	0.390***	0.346***	0.115**	0.239***	0.782***	0.669***	0.703***	0.654***	0.534***	0.417***	0.576***	0.428***	0.540***	1

Notes: This table shows the correlations among the i) ISORA's indexes (upper-left quadrant), ii) the TADAT's indexes (lower-right quadrant), and between ISORA and TADAT indexes (lower-left quadrant).

Annex III. Results for Multivariate Models

Table AIII.1. Multivariate Models, ISORA

	(1) All	(2) All	(3) All	(4) EMs & LIDCs	(5) EMs & LIDCs	(6) EMs & LIDCs
Compliance Risk Management	0.0555***	0.0626***	0.0714***	0.0596***	0.0633**	0.0664**
	(0.0212)	(0.0237)	(0.0256)	(0.0213)	(0.0255)	(0.0270)
Third Party Data	0.0218	0.0372	0.0265	0.0331	0.0390	0.0504
	(0.0293)	(0.0356)	(0.0435)	(0.0272)	(0.0386)	(0.0453)
Digitalization	0.0160	0.0173	0.0181	0.0058	0.0076	-0.0008
	(0.0176)	(0.0196)	(0.0205)	(0.0165)	(0.0195)	(0.0199)
Service Orientation	-0.0231	-0.0150	-0.0389	-0.0303	-0.0138	-0.0254
	(0.0307)	(0.0304)	(0.0360)	(0.0348)	(0.0349)	(0.0401)
Public Accountability	0.0177	0.0103	0.0206	0.0437*	0.0407	0.0593*
	(0.0259)	(0.0285)	(0.0331)	(0.0262)	(0.0281)	(0.0315)
Autonomy	0.0117	0.0062	0.0063	0.0149	0.0097	0.0045
	(0.0278)	(0.0318)	(0.0321)	(0.0277)	(0.0338)	(0.0332)
Est. LTO	0.0220	0.0077	0.0235	0.0688***	0.0465*	0.0687**
	(0.0330)	(0.0345)	(0.0381)	(0.0192)	(0.0275)	(0.0300)
#Tax Staff/LaborForce		0.0690**	0.0670**		0.0591	0.0491
		(0.0293)	(0.0307)		(0.0438)	(0.0507)
Sq(#Tax Staff/Labor Force)		-0.0161*	-0.0161*		-0.0131	-0.0096
		(0.0086)	(0.0083)		(0.0149)	(0.0183)
Active Taxpayer/Labor Force			0.0038			0.0033
			(0.0135)			(0.0206)
Log (PerCapitaGDP_PPP), lagged	0.0131	0.0030	-0.0016	0.0031	0.0182	-0.0169
	(0.0578)	(0.0652)	(0.0628)	(0.0692)	(0.0758)	(0.0682)
Square log (PerCapitaGDP_PPP), lagged	-0.0005	-0.0004	-0.0001	0.0003	-0.0011	0.0010
	(0.0034)	(0.0039)	(0.0038)	(0.0042)	(0.0046)	(0.0042)
Trade openness (out of GDP), lagged	0.0285***	0.0190*	0.0199**	0.0351***	0.0258**	0.0232**
	(0.0089)	(0.0099)	(0.0095)	(0.0094)	(0.0102)	(0.0097)
External debt (out of GDP), lagged	0.0056	0.0028	0.0064	0.0029	-0.0077	0.0022
	(0.0051)	(0.0053)	(0.0040)	(0.0083)	(0.0092)	(0.0084)
CPI, lagged	-0.0344*	-0.0372	-0.0264	-0.0311	-0.0259	-0.0201
	(0.0201)	(0.0234)	(0.0207)	(0.0198)	(0.0223)	(0.0202)
Terms of Trade (2000=1), lagged	-0.0128	-0.0038	0.0120	-0.0161*	-0.0074	0.0087
	(0.0092)	(0.0099)	(0.0075)	(0.0092)	(0.0101)	(0.0070)
Oil exports (as a % of GDP), lagged	0.4289**	0.3588**	0.3099*	0.4368**	0.3635**	0.3198*
	(0.1694)	(0.1543)	(0.1627)	(0.1704)	(0.1611)	(0.1703)
Log (Agri, %GDP), lagged	-0.0053	-0.0052	-0.0043	-0.0054	-0.0079	-0.0078
	(0.0057)	(0.0075)	(0.0078)	(0.0061)	(0.0083)	(0.0077)
Control Corruption, lagged	0.0070	0.0011	-0.0008	0.0091	0.0048	0.0038
	(0.0055)	(0.0062)	(0.0063)	(0.0060)	(0.0061)	(0.0058)
Budgeted_Actual, Nontrade_tax, Combined	0.1402***	0.1407***	0.1723***	0.1523***	0.1539***	0.1828***
	(0.0519)	(0.0533)	(0.0560)	(0.0521)	(0.0538)	(0.0555)
Constant	-0.0043	0.0446	0.0262	-0.0307	-0.0705	0.0324
	(0.2483)	(0.2797)	(0.2643)	(0.2845)	(0.3101)	(0.2757)
Observations	768	551	482	682	470	416
Number of countries	92	75	65	80	63	55
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Adj. Overall R2	0.4107	0.4526	0.4924	0.4060	0.3982	0.4747

Notes: Estimation of equation using Hausman-Taylor estimator. Multivariate Models. Robust standard errors in parentheses. *, **, *** denote statistical significance at the 1, 5, 10 percent levels, respectively. Tax policy changes is proxied by the difference between the budgeted revenue of year *t* as share of projected GDP and the actual revenue of year *t-1* as share of actual GDP.

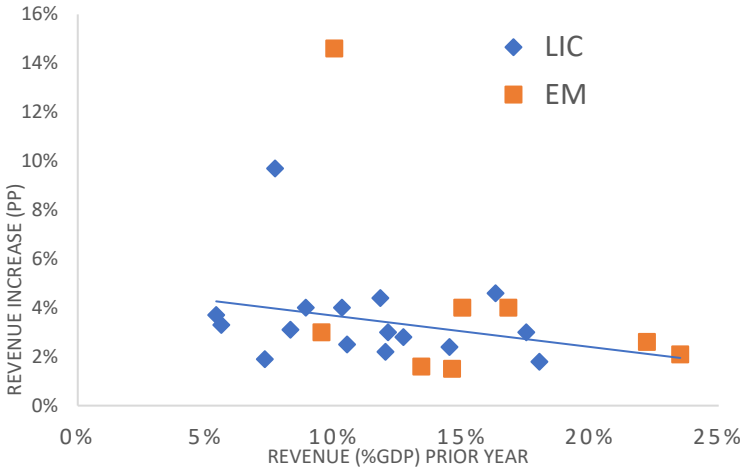
Table AIII.2. Multivariate Models, TADAT

	(1) All	(2) All	(3) All	(4) EMs & LIDCs	(5) EMs & LIDCs	(6) EMs & LIDCs
TADAT: POA1: Integrity of the Registered Taxpayer Base	-0.0402	-0.0327	-0.0435	-0.0360	-0.0276	-0.0440
	(0.0367)	(0.0394)	(0.0397)	(0.0373)	(0.0400)	(0.0362)
TADAT: POA2: Effective Risk Management	-0.0156	0.0106	0.0456	-0.0205	0.0178	0.0523
	(0.0343)	(0.0551)	(0.0561)	(0.0347)	(0.0547)	(0.0551)
TADAT: POA3: Supporting Voluntary Compliance	0.0296	0.0170	0.0156	0.0284	0.0278	0.0225
	(0.0372)	(0.0433)	(0.0444)	(0.0363)	(0.0441)	(0.0443)
TADAT: POA4: Timely Filing of Tax Declarations	0.0635*	0.0786*	0.0555	0.0581*	0.0773*	0.0591
	(0.0328)	(0.0417)	(0.0407)	(0.0344)	(0.0399)	(0.0398)
TADAT: POA5: Timely Payment of Taxes	0.0262	-0.0097	-0.0005	0.0337	0.0033	0.0110
	(0.0349)	(0.0379)	(0.0382)	(0.0347)	(0.0389)	(0.0376)
TADAT: POA6: Accurate Reporting in Declarations	-0.0137	-0.0002	0.0231	-0.0046	-0.0101	0.0266
	(0.0447)	(0.0773)	(0.0813)	(0.0436)	(0.0815)	(0.0827)
TADAT: POA7: Effective Tax Dispute Resolution	-0.0209	-0.0168	-0.0219	-0.0289	-0.0288	-0.0323
	(0.0280)	(0.0332)	(0.0359)	(0.0278)	(0.0329)	(0.0336)
TADAT: POA8: Efficient Revenue Management	-0.0244	-0.0310	-0.0524	-0.0242	-0.0318	-0.0538
	(0.0333)	(0.0530)	(0.0530)	(0.0332)	(0.0545)	(0.0545)
TADAT: POA9: Accountability and Transparency	0.1081***	0.0740	0.0544	0.1098***	0.0592	0.0387
	(0.0388)	(0.0659)	(0.0686)	(0.0406)	(0.0662)	(0.0676)
#Tax Staff/LaborForce		0.1181***	0.1093**		0.0892	0.0745
		(0.0429)	(0.0446)		(0.0585)	(0.0604)
Sq(#Tax Staff/Labor Force)		-0.0390**	-0.0331**		-0.0249	-0.0151
		(0.0166)	(0.0169)		(0.0260)	(0.0259)
Active Taxpayer/Labor Force			0.0065			0.0009
			(0.0153)			(0.0238)
Log (PerCapitaGDP_PPP), lagged	0.0053	-0.0376	-0.0276	0.0070	-0.0257	-0.0242
	(0.0626)	(0.0833)	(0.0843)	(0.0646)	(0.0939)	(0.0936)
Square log (PerCapitaGDP_PPP), lagged	-0.0005	0.0017	0.0013	-0.0005	0.0011	0.0012
	(0.0037)	(0.0050)	(0.0051)	(0.0038)	(0.0056)	(0.0056)
Trade openness (out of GDP), lagged	0.0234**	0.0263**	0.0251**	0.0218**	0.0250**	0.0234*
	(0.0095)	(0.0115)	(0.0120)	(0.0098)	(0.0120)	(0.0124)
External debt (out of GDP), lagged	0.0176***	0.0037	0.0044	0.0181***	0.0016	0.0036
	(0.0061)	(0.0140)	(0.0140)	(0.0062)	(0.0159)	(0.0151)
CPI, lagged	-0.0589*	-0.0057	-0.0047	-0.0578*	0.0006	0.0007
	(0.0326)	(0.0246)	(0.0246)	(0.0332)	(0.0255)	(0.0255)
Terms of Trade (2000=1), lagged	-0.0081	-0.0050	0.0037	-0.0079	-0.0038	0.0054
	(0.0129)	(0.0103)	(0.0100)	(0.0131)	(0.0107)	(0.0103)
Oil exports (as a % of GDP), lagged	0.5603***	0.5385***	0.5532***	0.5638***	0.5337***	0.5481***
	(0.1303)	(0.1009)	(0.0782)	(0.1262)	(0.0990)	(0.0752)
Log (Agri, %GDP), lagged	-0.0064	-0.0066	-0.0030	-0.0070	-0.0098	-0.0063
	(0.0058)	(0.0070)	(0.0075)	(0.0058)	(0.0073)	(0.0075)
Control Corruption, lagged	0.0108*	0.0040	0.0004	0.0114*	0.0047	0.0013
	(0.0061)	(0.0079)	(0.0076)	(0.0063)	(0.0084)	(0.0081)
Budgeted_Actual, Nontrade_tax, Combined	0.0955*	0.1357**	0.1847***	0.0928	0.1343*	0.1842***
	(0.0562)	(0.0673)	(0.0685)	(0.0577)	(0.0686)	(0.0699)
Constant	0.0787	0.2409	0.1698	0.0711	0.2061	0.1731
	(0.2667)	(0.3486)	(0.3495)	(0.2751)	(0.3923)	(0.3895)
Observations	536	315	299	518	297	281
Number of countries	60	41	39	57	38	36
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Adj. Overall R2	0.5749	0.6028	0.6192	0.5535	0.5510	0.5740

Notes: Estimation of equation using Hausman-Taylor estimator. Multivariate Models. Robust standard errors in parentheses. *, **, *** denote statistical significance at the 1, 5, 10 percent levels, respectively. Tax policy changes are proxied by the difference between the budgeted revenue of year *t* as share of projected GDP and the actual revenue of year *t-1* as share of actual GDP.

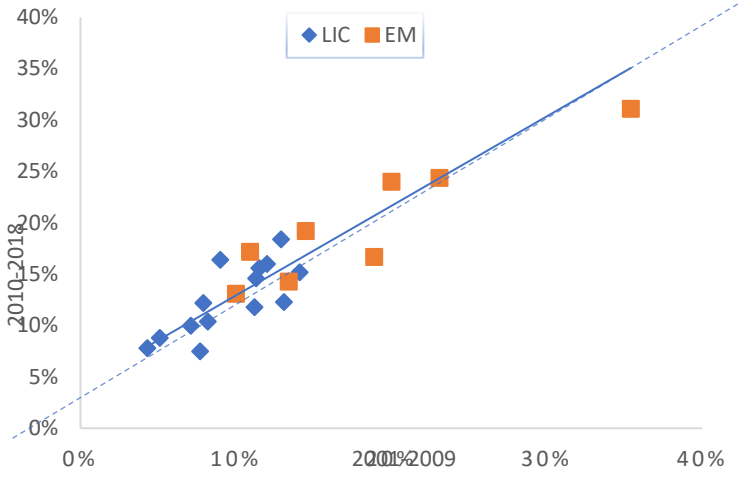
Annex IV: The Impact of Tax Administration Reforms in the Post-GFC

Figure AIV.1. Revenue Ratio Increase vs. Revenue Ratio prior Year



Sources: Akitoby et al. (2020).

Figure AIV.2. Pre- vs. Post- GFC Revenue/GDP Ratio (% of GDP)



Sources: Akitoby et al. (2020).

References

Akitoby, B., Baum, A., Hackney, C., Harrison, O., Primus, K., & Salins, V. (2020). Tax revenue mobilization episodes in developing countries. *Policy Design and Practice*, *3*(1), 1-29.

Arkhangelsky, D., Athey, S., Hirshberg, D.A., Imbens, G.W. & Wager, S. (2021). Synthetic difference-in-differences. *American Economic Review*, 111(12), pp.4088-4118.

Basri, M. C., M. Felix, R. Hanna & B. A. Olken (2021). Tax Administration versus Tax Rates: Evidence from Corporate Taxation in Indonesia. *American Economic Review* 111(12): 3827-3871.

Bouvatier, V. (2014). Heterogeneous bank regulatory standards and the cross-border supply of financial services. *Economic modelling*, *40*, 342-354.

Callaway, B. and Sant'Anna, P.H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2), pp.200-230.

Carrillo, P., D. Pomeranz and M. Singhal (2017). Dodging the taxman: Firm misreporting and limits to tax enforcement. *American Economic Journal: Applied Economics* 9(2): 144-164.

Chang, E. S., Gueorguiev, N., Gavin, E., & Honda, J. (2020). Raising Tax Revenue: How to Get More from Tax Administrations? IMF Working Paper No. 20/142

Chatelain, J. B., & Ralf, K. (2021). Inference on time-invariant variables using panel data: a pretest estimator. *Economic Modelling*, *97*, 157-166.

Crandall, W., Gavin, E., & Masters, A. (2021). Understanding Revenue Administration. International Monetary Fund, Fiscal Affairs Department, October 2021, Departmental paper series.

Crivelli, E. (2019). A basic tool to assess tax administration strength in emerging Europe. *Economics of Transition and institutional Change*, *27*(2), 425-446.

Dabla-Norris, E., F. Misch, M. D. Cleary and M. Khwaja (2020). The quality of tax administration and firm performance: evidence from developing countries. *International Tax and Public Finance*, 27, 514-551.

Ebeke, C., Mansour, M., & Rota-Graziosi, G. (2016). The Power to Tax in Sub-Saharan Africa: LTUs, VATs, and SARAs. FERDI Working Paper.

Goh, S. K., & Tham, S. Y. (2013). Trade linkages of inward and outward FDI: Evidence from Malaysia. *Economic Modelling*, *35*, 224-230.

Ghura, D. (1998). Tax Revenue in Sub-Saharan Africa: Effects of Economic Policies. *International Monetary Fund*.

Gupta, A. S. (2007). Determinants of Tax Revenue Efforts in Developing Countries"

Hausman, J. A., & Taylor, W. E. (1981). Panel data and unobservable individual effects. *Econometrica: Journal of the Econometric society*, 1377-1398.

Hsiao, Cheng. Analysis of Panel Data, Cambridge University Press, United Kingdom, 2003.

International Monetary Fund (2018). Sub Saharan Africa Regional Economic Outlook. Domestic Revenue Mobilization and Private Investment. International Monetary Fund, Washington D.C.

International Monetary Fund (2019). Fiscal Monitor. Curbing Corruption. International Monetary Fund, Washington D.C.

Leuthold, J. H. (1991). Tax shares in developing economies a panel study. *Journal of Development Economics*, *35*(1), 173-185.

Linden, A., 2015. Conducting interrupted time-series analysis for single-and multiple-group comparisons. The Stata Journal, 15(2), pp.480-500.

Morrow, P., M. Smart and A. Swistak (2022). VAT compliance, trade, and institutions. *Journal of Public Economics* 208: 104634.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica: journal of the Econometric Society*, 69-85.

Ngoma, M. M., & Krsic, N. (2017). Assessing the impact of tax administration reforms in Sub-Saharan Africa. *MPRA Paper*, (89275).

Pesaran, M. H., & Zhou, Q. (2018). Estimation of time-invariant effects in static panel data models. Econometric Reviews, 37(10), 1137-1171.

Pomeranz, D. (2015). No taxation without information: Deterrence and self-enforcement in the value added tax. *American Economic Review* 105(8): 2539-2569.

Tanzi, V. (1992). Fiscal policy and economic reconstruction in Latin America. *World Development*, *20*(5), 641-657.

PUBLICATIONS

Quantifying the Revenue Yields from Tax Administration Reforms

Working Paper No. WP/2023/231