International Tax and Public Finance
https://doi.org/10.1007/s10797-024-09858-5

The effects of a risk-based approach to tax examinations: evidence from a tax pilot programme in Tanzania

Amina Ebrahim[1] · Elineema Kisanga[2] · Ezekiel Swema[3] · Vincent Leyaro[2] · Edwin P. Mhede[4] · Ephraim Mdee[3] · Heikki Palviainen[5] · Jukka Pirttilä[6]

Accepted: 31 May 2024

Abstract
Technical assistance and increased use of ICT in tax administrations hold promise for greater revenue collection. Yet, the evidence on how these activities work in the real-world circumstances of developing countries is scant. The paper attempts to fill this gap by evaluating an intervention undertaken jointly by the Finnish and Tanzanian revenue administrations. The programme introduced a new risk-based case selection method for enhancing the effectiveness of firm tax examinations in Tanzania complemented by supplementary staff training and developing management practices. Our results, stemming from a difference-in-differences analysis and administrative data from the Tanzanian Revenue Authority, indicate that the intervention increased the corrected amount of taxable income by approximately 20% during the first year of the new approach.

Keywords Risk-based approach · Firm tax audits · Tanzania

JEL Classification C93 · H26

1 Introduction

A key goal for many tax authorities in developing countries is to enhance compliance among taxpayers to increase tax revenue collection. Simultaneously, donor countries are strongly committed to assisting developing economies in reaching this objective. Indeed, tax and supporting the tax systems constitute a critical part of the 17th Sustainable Development Goal.[1] In 2015 already, the Addis Ababa Action

[1] See: https://sdgs.un.org/goals/goal17.

Elineema Kisanga passed away in July 2023. He was one of the authors of the original version and the first revision but did not participate in this second revision.

Extended author information available on the last page of the article

Published online: 03 September 2024

Agenda[2]—via, in particular, the Addis Tax Initiative[3]—brought development collaboration and technical assistance to the centre stage to enhance domestic revenue mobilization (DRM). The Addis Tax Initiative signatories committed to doubling their support for DRM and provided USD370 million in official development assistance for DRM in 2019 (Addis Tax Initiative, 2021).

Tax authorities use various enforcement strategies, including random audits and examinations. Still, risk-based approaches are common and, with the onset of machine learning, hold great promise for improving the detection probabilities of non-compliant taxpayers (see Khwaja et al., 2011 for a review of tax practices). However, limited systematic evidence on the impact of these risk-based strategies is available (at least outside revenue organizations). In our understanding, few earlier studies investigate how risk-based compliance interventions work in low-income developing countries like Tanzania.

In this paper, we attempt to fill the gap by examining the revenue impacts of a risk-based tax examination pilot implemented in Tanzania. The Tanzanian Revenue Authority (TRA) and the Finnish Tax Administration (VERO) planned the pilot jointly. The intervention is a new method for flagging taxpayers for tax examination based on a data-driven risk assessment. Tax examination refers to the process where revenue authority staff scrutinize the tax returns based on existing or newly acquired information without conducting traditional (field) audits. Hence, the examinations could also be classified as desk audits. The intervention was designed to improve the practice of choosing taxpayers to be examined, which relied on staff discretion. A crucial part of the intervention was teaching and management training related to the process and establishing systematic communication methods to inform examined taxpayers about the process.

We use a difference-in-differences design, comparing revenue developments of firms handled by Dar es Salaam tax offices (where the pilot took place) with corresponding firms in five major tax offices in Tanzania (Arusha, Dodoma, Mbeya, Morogoro, and Mwanza). The data stem from the TRA's administrative information system and are complemented with supplementary material used in the pilot.

Our paper contributes to the literature in four main ways. *First*, the results improve our understanding of risk-based tax examinations in a low-income county with a lower-than-usual tax administrative capacity. *Second*, this is one of the first studies to evaluate whether and how technical assistance financed by donor countries helps the receiving countries enhance their tax administrations and revenue collection. *Third*, while conventional wisdom is that tax avoidance and evasion are rampant in developing countries, more evidence is needed on how administrative interventions help lower the extent of these activities. Therefore, this paper is one step toward enhancing our understanding of this matter. *Fourth*, and more broadly, the paper contributes to the expanding literature that provides causal evidence about enforcement using administrative data from developing countries.

[2] See: https://sustainabledevelopment.un.org/index.php?page=view&type=400&nr=2051&menu=35.

[3] See: https://www.addistaxinitiative.net/.

Our context is Tanzania, a lower middle-income sub-Saharan African (SSA) developing country that strives to raise the tax-to-GDP ratios, like many other SSA countries. Besides substantial tax reforms since independence, studies have shown that tax evasion and avoidance remain rampant in Tanzania and are present in almost all sectors of the economy (Wadhawan & Gray, 1998). Tax evasion was 35–55% of the total revenue collected by 2010, including tax evasion from the informal sector (ESRF, 2010). Tax evasion is also rife in the trade sector through under-invoicing customs duties at the port of entry.[4]

Traditional deterrence measures, such as tax audit threats and penalties, remain the main measures to enhance tax compliance in Tanzania, as in many other developing countries. However, despite the deterrence methods employed, tax revenue collection performance in Tanzania is lower than the existing potential. Tax revenue has been approximately 12% of GDP in the last decade (TRA 2022), below the sub-Saharan African average of 18% of GDP.

The results from our analysis suggest that the intervention increased the adjusted income from the taxpayers by about 15% and the difference between adjusted and initially reported income by about 20%. In contrast, it did not lead to more examined cases. These results align with the reform's goals; the idea was to target examination efforts better and study the selected cases more thoroughly. The intervention was also highly cost-effective.

The rest of the paper proceeds as follows. A review of the earlier related literature is offered in Sect. 2. The Tanzanian context is explained in Sect. 3, whereas Sect. 4 describes the intervention we examine. Data and the empirical methods are covered in Sect. 5. Section 6 presents the results, and Sect. 7 concludes.

2 Literature review

Our study is related to the analysis of tax evasion pioneered by Allingham and Sandmo (1972).[5] Recent work, such as Kleven et al. (2011) and Pomeranz (2015), highlights the crucial role of third-party information in improving compliance. However, what is also key is the ability to act on the information (Carillo et al., 2017).

An expanding body of work investigates the deterrence impact of tax audits on future tax compliance of audited taxpayers with somewhat mixed results; see, e.g., Advani et al. (2023), DeBacker et al. (2015), DeBacker et al., (2018a, b), as well as Li et al. (2019). Kotsogiannis et al. (2022) offer an evaluation of audits in Rwanda. Their results suggest that audits lead to higher subsequent corporate income tax reporting in audited firms compared to a control group generated by matching techniques. In addition, a positive result emerges from field audits, while desk audits lead to counter-deterrent effects.

Basri et al. (2020) focus on an additional enforcement angle, namely the role of dedicated tax offices to specific segments of taxpayers and the impacts of being

[4] See Epaphra (2015) for reasons for customs tax evasion in Tanzania.

[5] See Slemrod (2019) for an extensive review.

handled by a medium-sized taxpayer office (MTO) in Indonesia. The introduction of the MTO involved greater scrutiny of the taxpayers dealt with by the new bureau and resulted in a definite increase in taxable income and taxes paid.

Battaglini et al. (2023) examine to what extent machine learning (ML) can help to better target tax audits in the Italian context. Focusing on small businesses' tax audits, they note that ML methods can lead to substantial increases in the revenue implications of audits. In comparison, in the Finnish/TRA collaboration, the risk-based tool was based on more general knowledge about the perceived riskiness of different types of taxpayers.

In a closely related paper,[6] Bachas et al. (2021) compare discretionary and risk-based tax audits in Senegal. The research team designed a selection mechanism that relied on discrepancies and assessed the revenue raised by the new risk-based versus discretionary audits that the tax authority had used before. According to the results, the discretionary and risk-based audits uncovered similar amounts of hidden income. Still, because the selected taxpayers were, on average, larger in discretionary audits, these audits raised more revenue than risk-based audits. They also found that inspectors were less likely to initiate risk-based audits. One of their takeaways is that, especially when third-party information is scarce, discretionary methods may benefit from case-by-case details related to the likelihood of finding evasion.

In our understanding, there is no prior hard evidence on the impacts of the new risk-based tax examination process (the first inspection before any more thorough audits) on revenues in developing countries. In addition, there is very little research, in general, on how technical assistance works in tax administration despite the policy focus on support for tax revenue mobilization.

3 Overview of tax system and performance in Tanzania

Following a series of reforms, such as the formation of the Tanzania Revenue Authority (TRA) in 1995, the introduction of the Value-Added Tax (VAT) in 1998, and the initiation of the Large Taxpayers Department (LTD) within TRA in 2001, the Tanzanian tax system is now similar to tax systems in many other developing countries in structure and composition. Reforms have continued lately via strengthening the tax administrative systems and introducing advanced technology devices, such as electronic fiscal devices (EFD) and electronic tax stamps (ETS). As a result, tax collection relative to GDP has been approximately 12% in recent years (See Table A.1 in the online annex), whereas it remained below 10% before 2008.[7] Like other developing countries, indirect taxes form the largest share (7.1%) of the tax-to-GDP ratio (Table A.1). Domestic VAT and VAT on imports are the main contributing taxes to total revenue (Table A1). Corporate income tax (CIT) and Pay-As-You-Earn (PAYE) account for a rather large share of total revenue at 12.3% and 12.1% in 2021, respectively (Table A2). The CIT rate in Tanzania is 30%, in line with the CIT

[6] We learnt of this work after submitting our own to this journal.

[7] According to Government Revenue Dataset of UNU-WIDER.

average in most African countries. In addition, there is a presumptive tax regime for small businesses, which is based on an annual turnover below TZS100 million and does not require filing final returns with audited books of accounts. Those self-employed with no non-incorporated businesses should also pay personal income tax, which has graduated rates (with the top marginal tax rate of 30%). One of the main issues undermining income tax collection in Tanzania is the existence of tax exemptions and waivers offered to investors. The large informal sector contributes little tax, and the revenue authority has specific initiatives to target the sector.

4 Description of the pilot programme

Unrealistically, all tax returns were supposed to be examined in the old tax examination process. However, in practice, resource constraints—such as skills shortages among the officials and the lack of time—implied that only a subset of returns was examined. There were no unified criteria for selecting taxpayers for a more detailed examination, and there was no difference in this respect between the treated and the control areas before the intervention occurred. The earlier selection process, therefore, left considerable room for tax examiners to exercise discretion when choosing taxpayers for a more scrutinized examination.

The new examination method was set up for several reasons, including increasing revenue collected, improving the tax officers' skills, treating taxpayers equally, focusing examination efforts on risky taxpayers, and reducing the time spent examining taxpayers with limited risks. The new method involved using an Excel spreadsheet developed between TRA and VERO. The spreadsheet raises a flag when the information provided by the taxpayer on specific aspects of the tax calculation is found to be outside of reasonable bounds pre-determined by TRA and coded into the spreadsheet. These pertain to, for instance, the developments in firms' sales and expenses over time.[8] Tax officers involved in the process received training and instructions for using the Excel spreadsheet. In practice, the tax return is filed by the taxpayer and collected by the tax officer. Then, the data handler enters the return information into the TRA information system (called iTax) and Excel spreadsheet.

Further examination is then conducted where the spreadsheet indicates this is advised. If needed, adjustments are made and then recorded in the iTax system. The new tax examination process did not change the conduct of actual tax audits, which remains a heavier form of intervention. Originally, it was intended that the pilot would be implemented for income tax and value-added tax (VAT) payers. However, because the information system for the VAT was different, the pilot was only implemented for corporate income taxpayers (CIT) and, in principle, for personal income taxpayers (PIT). The selection criteria were, however, designed with the corporate income taxpayers in mind.

It is important to note that the intervention consisted of the new selection method and considerable training and capacity development of TRA tax officers

[8] Due to security reasons, the exact characteristics of the procedure cannot be disclosed.

to implement the new approach. The tax examiners were trained to comply with the new process and treat taxpayers based on similar procedures. The whole intervention package, therefore, also contained:

- Work manuals, i.e., work instructions that included information on return filing and examination process as well as performance evaluation (for example, Key Performance Indicators);
- Letter templates used in the communication between TRA and the taxpayers;
- Trained examination champions (around 20, 1–2 per office) who provided support in everyday examination processes;
- One main training event and hands-on support on-site for tax examiners;
- Process chart, which provided information on how the process for examination cases should flow;
- Data capturers who were trained to insert information into the Excel sheet and instructions developed to support the data capturers;
- Attitudes sensitization toward selecting the riskiest taxpayers into a careful examination rather than examining all taxpayers superficially.

Figure 1 provides the process flow chart of the pilot programme. The figure shows the sequential steps starting from a customer's income return from the perspectives of a consumer, data capturer, and examiner. At first, the data capturer captures the tax return information into iTax and Excel and adds the Excel sheet print in the tax return. If the tax return requires examination, the return is examined (desk audited) based on the instructions of the pilot programme. Next, the tax examiner discusses the examination results with the customer by using letter templates provided by the pilot programme. After obtaining the required information, the assessment is automatically created and sent to the customer. Our research setting allows us to identify the joint impact of the training, new procedures, and the new selection into tax examination.

In an interim assessment of the pilot intervention in September 2019, the TRA and VERO teams were satisfied with the experiences. They noted that the examination process had become smoother and more uniform, requiring less time from the officials. They also highlighted that examination staff resources were still limited at TRA and that some errors in entries to Excel sheets were found.

The pilot was implemented in tax regions such as Ilala, Temeke, Kinondoni, and Kariakoo, which are administratively located in the Dar es Salaam region. The Large Taxpayer Department (LTD) was also included in the pilot and considered a tax region in Dar es Salaam. In the LTD, the selection criteria had to be further developed because of the different characteristics of the taxpayers. We exploit the fact that the pilot was only implemented in the Dar region in our empirical strategy. According to VERO officials, the Dar region was selected for logistical reasons and to avoid time-consuming travel.

5 Empirical strategy and data

In this study, we use administrative firm-level data collected by the Tanzania Revenue Authority (TRA) covering all taxpayers in the examined tax regions. The data come from income tax returns from 1 July 2015 to 30 June 2020. Each income year contains more than 25,000 observations. Variables include a set of income variables, taxpayer types and industry, and the regional location of the firm. Because the pilot was introduced on 1 July 2019, we divided our sample into financial years (tax years). Our first examined year covers 1 July 2016–30 June 2017, and the last studied year covers 1 July 2019–30 June 2020, our treatment period. Examinations during each period pertain to information from tax years which have ended.

The treatment group includes the tax regions located in the Dar es Salaam region. The control areas represent the country's largest five other tax regions: Arusha, Dodoma, Mbeya, Morogoro, and Mwanza. As discussed above, Dar es Salaam is the largest region, so we include an extensive set of control variables related to these differences.

The empirical strategy follows the standard difference-in-differences (DD) approach:

$$Y_{i,r,t} = \alpha + \beta Dar_r + \gamma POST_t + \delta Dar_r \times POST_t + \theta X_{i,r,t} + \varepsilon_{i,r,t} \tag{1}$$

where $Y_{i,r,t}$ is the outcome variable for taxpayer i in a region r (tax office) and year t. The variable Dar_r is the treatment variable (equal to 1 if the tax office is in the Dar es Salaam region). We consider the following three outcomes: adjusted taxable income (henceforth, adjusted income), extra income defined as the difference between the adjusted (final) income the self-declared income, and an indicator variable for a firm that has adjusted its income. We use the inverse hyperbolic sine (IHS) transformation for the two continuous variables to retain zero and negative values and the log transformation, which discards those observations. The practical relevance of the not-strictly positive observations is limited since the examination process often finds some positive correction amount. Extra income is our preferred outcome variable. The extra income variable directly captures the potential effect of the intervention. Still, it may also be interesting to document the effect's relevance on final income, measured by the adjusted income variable.

Few adjustments may be initiated by the taxpayer, and we cannot identify this in our data. However, we do not expect this behaviour to change because of the pilot. $POST_t$ controls for time-variant conditions for everyone (equals to 1 for the post-intervention period). We observe the tax adjustment decision date; therefore, the $POST_t$ variable receives the value of 1 after 1 July 2019. The adjustment decisions are recorded until June 2020. Our policy parameter of interest is δ. This parameter shows how much the pilot increased adjusted income in the Dar es Salaam region tax offices compared to selected tax offices in large areas outside Dar es Salaam. Notice that our outcome variable captures the first-round effects of the examination process during the pilot period. We do not yet have the data for analysing a possible deterrence impact on subsequent self-reported incomes.[9]

We add control variables, $X_{i,r,t}$, to the basic model to control for any industry or other differences between firms unrelated to the treatment. *First*, a more parsimonious set of controls ('Controls 1' in what follows) is added, including industry dummies, tax type (CIT vs PIT), and ownership type. *Second,* in 'Controls 2', we add interactions between industry and period (the after dummy), tax type and period, and ownership type and period. Furthermore, indicator variables for the company's size (deciles of initially declared income), and aggregate VAT revenues from the tax region in a year are added. The purpose of this latter variable and the interactions with period dummies is to account for differences in firm characteristics between the treated and the control group and potential changes in the efficiency of examinations during the pandemic between the treated and the control areas.

Because companies are not typically examined, audited, or inspected in consecutive years, we use a cross-sectional difference in differences.

We compare firm characteristics in the control and treatment groups before and after pilot intervention in Table 1. Firms in the treatment area are larger and more profitable, but the IHS and the log transformation make these groups more comparable. The majority of firms in the retail and service sectors are liable for corporate income tax. Hence, the corporate income taxpayers' share is much larger than personal income taxpayers.

A DD approach assumes that differences between the treatment and control groups are stable over time. We investigate the parallel trends assumption for adjusted and extra income in Fig. 2. Apart from the dummy (adjusted or not), the trends appear very similar before the treatment took place, and even for the outcome 'Adjusted (0/1)', the year 2019 interaction with the treatment dummy is not statistically significant in an event study specification. The decline in incomes in 2020 is related to a changing distribution where the mean in unadjusted, level, variables increases but the median drops in treatment and control areas alike. This is reflected as a decline in logged outcomes, since the log transformation suppresses large values. A likely reason for this change is another intervention, conducted by the TRA, which brought additional cases for examinations in all tax regions, in both treatment and control regions between December 2019 and May 2020. These were cases which had not been examined within a five-year span. The Large Taxpayer Department was excluded from this campaign, and hence it increased the share of smaller taxpayers in the tax examination samples. While it is not ideal that the timing of this TRA-led intervention coincided with the VERO pilot, its impact on the DiD result is arguably limited because it should have affected treatment and control groups alike.

[9] A less discretionary selection of examined cases could also lead to fewer corruption opportunities for the tax officials. While this is a possible channel of influence, we are unable to examine it in the absence of internal audit information. We revert to this matter in Sect. 6.

Table 1 Descriptive statistics before and after the treatment period

	Before	After						
	Mean, control	Mean, treatment	Difference	t-test	Mean, control	Mean, treatment	Difference	t-test
Outcomes								
Initial reported Income	33.19	346.95	−313.76	−6.41***	49.02	517.37	−468.36	−6.79***
Adjusted income	123.33	582.47	−459.14	−6.96***	197.22	1,006.08	−808.87	−6.36***
Extra income	90.14	235.51	−145.38	−5.68***	148.20	488.71	−340.51	−4.27***
IHS adjusted income	4.33	4.56	−0.23	−8.59***	4.24	4.60	−0.36	−8.57***
Log adjusted income (+)	3.63	3.86	−0.23	−8.44***	3.53	3.90	−0.37	−8.46***
IHS extra income	3.68	3.82	−0.14	−4.88***	3.54	3.87	−0.34	−7.14***
Log extra income (+)	2.94	3.09	−0.15	−4.76***	2.75	3.11	−0.35	−6.66***
Ownership type								
Limited Company	0.57	0.72	−0.15	−17.51***	0.65	0.70	−0.05	−3.74***
Sole proprietor	0.42	0.27	0.15	17.08***	0.33	0.30	0.04	3.05**
Tax type								
Corporate Tax	0.58	0.72	−0.15	−17.06***	0.67	0.70	−0.04	−3.04**
Personal Income Tax	0.42	0.28	0.15	17.06***	0.33	0.30	0.04	3.04**
Industry								
Agriculture	0.01	0.00	0.01	4.25***	0.01	0.01	0.01	2.50*
Mining	0.02	0.01	0.01	5.32***	0.02	0.01	0.01	2.97**
FIRE	0.05	0.06	−0.02	−4.65***	0.04	0.08	−0.04	−8.43***
Construction	0.06	0.06	0.00	0.90	0.05	0.05	−0.00	−0.70
Wholesale and retail	0.40	0.32	0.07	8.61***	0.30	0.34	−0.04	−3.05**
Manufacturing	0.09	0.10	−0.01	−2.41*	0.09	0.09	−0.00	−0.40
Information and communication	0.01	0.03	−0.02	−8.42***	0.02	0.04	−0.02	−3.95***
Transportation and storage	0.06	0.11	−0.05	−10.22***	0.05	0.09	−0.05	−7.91***

Table 1 (continued)

	Before	After						
	Mean, control	Mean, treatment	Difference	t-test	Mean, control	Mean, treatment	Difference	t-test
Accommodation, arts, health, admin, professional	0.24	0.24	0.00	0.18	0.36	0.24	0.12	10.08***
Observations	15,572				8527			

Notes: "Before" refers to the period from 1 June 2016 to 30 June 2019, and "After" from 1 July 2019 to 20 June 2020. FIRE means finance, insurance and real estate

Source: authors' estimates based on TRA iTax data

Table 2 Estimated treatment effects

	(1) Adjusted 0/1	(2) IHS adjusted income	(3) IHS Extra income	(4) Log adjusted income	(5) Log extra income
Basic DD	0.0187	0.1814**	0.2199**	0.1293	0.1994**
	(0.0201)	(0.0908)	(0.0975)	(0.0873)	(0.0908)
DD with controls 1	0.0008	0.1340	0.1988*	0.1670**	0.2592***
	(0.0189)	(0.1086)	(0.1083)	(0.0824)	(0.0845)
DD with controls 2	0.0194	0.2077***	0.2206**	0.1534***	0.1919***
	(0.0161)	(0.0791)	(0.0992)	(0.0566)	(0.0653)
Observations	28,464	28,464	28,464	24,947	24,099

Notes (1) Controls 1 include industry, tax, and ownership types. Controls 2 includes interactions between industry and period, tax type and period, ownership type and period, firm size and aggregate VAT revenues from the tax region in a year. (2) Standard errors are clustered at the tax region industry level. (3) Significance levels: ***1%, **5% and *10%

Source: authors' estimates based on TRA iTax data

6 Results and discussion

6.1 Main results

We now present the estimates of the DD regression for all our outcomes of interest. Table 2 shows three sets of DD results: basic DD with no controls and DD with two different sets of control variables, as described in the empirical strategy section. Controls 1 includes a basic set of controls. Controls 2 includes interaction terms to control for possible differences between the treated and the control areas during the pandemic.

Our preferred specification is the final one, with the most extensive set of controls, since adding these helps also account for time-varying differences between industries and regions (as captured by VAT revenues).

Column (1) estimates a linear probability model for an indicator variable of adjustment taking place or not. The second and the third columns use the IHS of the adjusted final income or extra income as outcomes, whereas Columns (4) and (5) use the log transformation, discarding observations with zero or negative values.

The first column (1) shows no impact on the number of adjustments. However, in all other outcomes, we document increased income from the tax examinations in Dar es Salaam tax regions during the pilot. Interpreting the magnitude of the result is more straightforward for the case of log incomes, and the increase is 15 log points for adjusted income and 19 log points for the extra income (row DD with controls 2). The impact on extra income is more considerable because of a smaller base, and the estimates are statistically significant. The coefficient 0.1534 for log adjusted income means that the pilot increased adjusted income by approximately 16.5% ($\exp(0.1534) \approx 1.165$), while the effect on extra income is found to be bigger (21%). Recently, Chen and Roth (2023) have criticized percentage interpretations with log-like transformations because the percentage effect is not well defined on the extensive margin. This may be less problematic in our setting because we have a relatively low number of zero observations (11.9% in the pre-treatment period). As shown in Table 2, the results are very similar to the logged extra income (positive values only) and IHS extra income outcomes. In addition, we estimate a Poisson regression that is not sensitive to rescaling as a robustness check.

The results from the adjusted indicator suggest the effect arises from having more income from firms with income adjustments rather than many more adjustments in the treatment area after the pilot. This effect is in line with the rationale of the pilot, which was to increase the efficiency of examinations rather than conducting more investigations. The pilot appears to have successfully reduced the number of less efficient desk audits and replaced them with desk audits of firms with more detectable evasion.

We also present event study graphs for all outcomes in Fig. 3, where the coefficients of interactions between year dummies and the treatment groups are depicted. The event study regressions include the treatment dummy, the set of year dummies, and their interactions. In addition, the extensive set of control variables (controls 2) is used. This analysis demonstrates no statistically significant differences between the treated and the control groups before the pilot took place for the outcome variables we used. The difference in the logged and IHS outcome variables is seen clearly during the 2020 pilot year, indicating that the adjusted and extra income increased relatively more in the treated area.

6.1.1 Subgroup analysis

Table 3 reports the extra income outcome results for three subgroups: CIT payers, PIT payers, and those firms not handled by the Large Taxpayer Department (LTD). Most taxpayers are corporations, so the positive impact arises from the corporate income side. When the firms managed by the LTD are dropped, the treatment impact declines, but its statistical significance remains in the model with controls 2.

Table 4, in turn, reports the DD results by industries. Most companies operate in the service sector, and that is where the effect stems from. The treatment impact is not precisely estimated for the other sectors due to fewer firms in these groups. When all the other sectors are combined and services are dropped, the treatment impact remains non-significant (not reported). The coefficient is positive for other industries apart from agriculture and mining when a more extensive set of controls is used.

6.2 Robustness

As the first robustness check, we employ coarsened exact matching (CEM) to our main outcome variable, the IHS of extra income. Coarsened exact matching is motivated by the differences between the larger taxpayers in the pilot and control areas. CEM coarsens the selected variables into strata and performs exact matching on the coarsened data (Iacus et al., 2012).

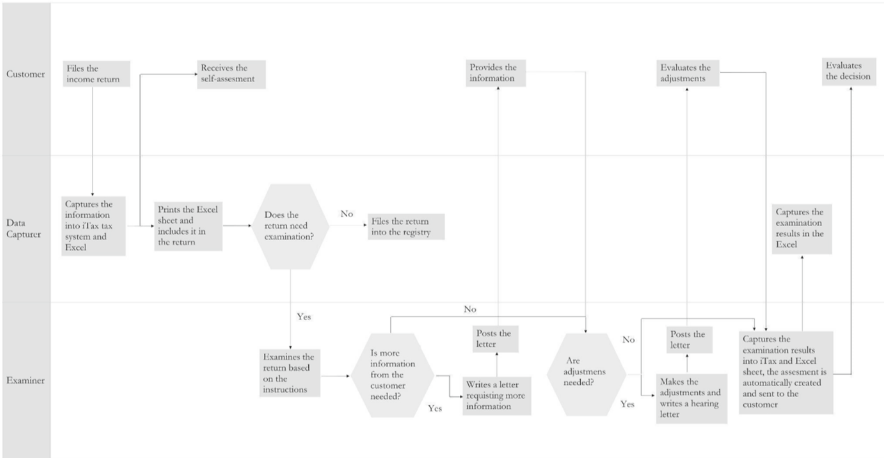

Fig. 1 Process flow chart of the pilot programme. *Note*: Process flow chart presenting the steps from the perspectives of a consumer, data capturer, and examiner. *Source*: Authors' simplified depiction from the VERO pilot process flow chart

As matched variables, we use tax type, taxpayer category, industry, and firm size, which is coarsened using equally sized bins based on quantiles. Table 5 shows the estimation results. The estimated treatment coefficient is now 0.2552, which is somewhat larger than our previous estimations and supports the positive effect of the intervention. Figure 4 plots the treatment effects on the pre-treatment years and our treatment year. The figure shows that the treatment effects do not differ statistically significantly from zero in the pre-treatment years, but we find a positive effect for the treatment year.

In CEM matching, there is a trade-off between internal and external validity. The more bins one uses, the more precise the results are, but the results may not be externally valid. After matching, we lose 461 firms from the treatment group out of 21,023 firms. Since we cannot find a match to 2.2% of the firms in the treatment group, the results are not fully externally valid. However, the share of unmatched firms is small, and the CEM matching shows that the intervention positively affected adjusted taxable income and tax revenues.

As the second robustness analysis, we further study the role of zero observations in the outcome variables. In the main analysis, we use the inverse hyperbolic sine to include zero observations in adjusted income or extra income. Here, we employ the PPML model of Santos Silva and Tenreyro (2006). The benefit of the PPLM model is that it allows for including zero observations in our analyses. We trim our sample with observations above 99.5th, 99th, 97.5th, and 95th percentiles dropped. There is a lot of dispersion in our outcomes and large standard errors without trimming. Trimming is also motivated by making the right tail of the treated and comparison groups more similar. Figure A1 illustrates how, before treatment, the right tail of the treated group is much longer than that of the comparison group, but there is a better balance after the trimming. The results from the PPML analysis, shown in Table 6 and Fig. 5, confirm the positive impact on extra income if a smaller share of

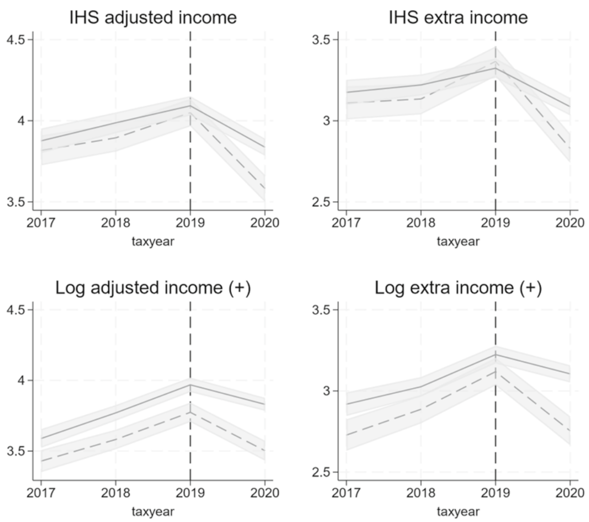

Fig. 2 Mean adjusted taxable and extra incomes for the treatment and control groups. *Note*: The log adjusted income (+) and log extra income (+) outcomes do not include zero or negative values. *Source*: Authors' calculations from TRA iTax data

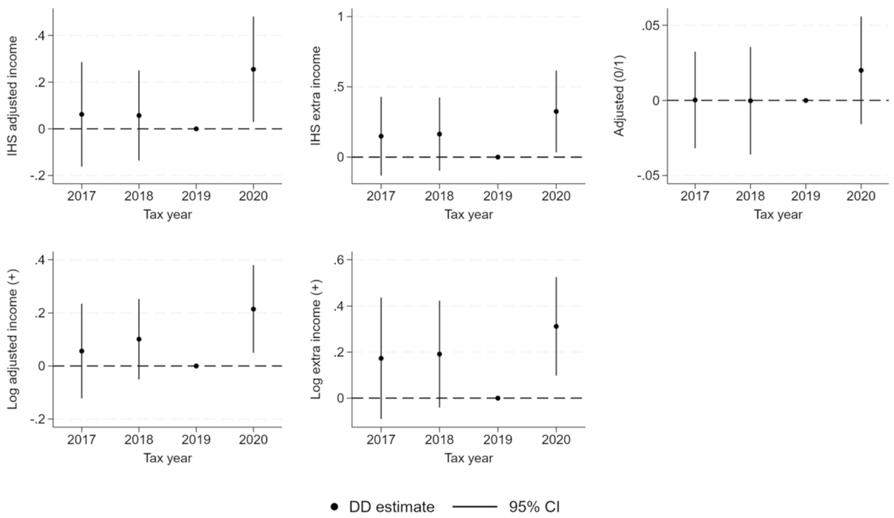

Fig. 3 Event study graphs. *Notes*: The figure shows the results of the treated dummy interaction with the year indicator. Standard errors are clustered at the tax region industry level. *Source*: authors' estimates based on iTax data

observations is dropped (first two columns in Table 6), whereas if a greater share is discarded, the significance disappears.

Third, instead of point estimates, we examine bounds for the estimates. For this, we use the honest DD approach by Rambachan and Roth (2023), which

Table 3 Estimated treatment effects for IHS of extra income by subgroups

	(1) All	(2) CIT	(3) PIT	(4) No LTD
Basic DD	0.2199**	0.3295**	0.0461	0.1746*
	(0.0975)	(0.1281)	(0.1051)	(0.0969)
DD with controls 1	0.1993*	0.2599*	0.0269	0.1601
	(0.1080)	(0.1443)	(0.1070)	(0.1044)
DD with controls 2	0.2207**	0.2679**	0.0574	0.1886*
	(0.0987)	(0.1316)	(0.1053)	(0.0958)
Observations	28,464	20,372	8,092	27,412

Notes (1) Controls 1 include tax, and ownership types. Controls 2 includes interactions between tax type and period, ownership type and period, firm size and aggregate VAT revenues from the tax region in a year. (2) Standard errors are clustered at the tax region industry level. (3) Significance levels: ***1%, **5% and *10%

Source authors' estimates based on TRA iTax data

Table 4 Estimated treatment effects for IHS of extra income by sector

	(1) Agric. and mining	(3) Manuf	(4) Services	(5) Other
Basic DD	−0.5658	−0.1059	0.3124***	−0.0868
	(0.4706)	(0.2422)	(0.1131)	(0.1580)
DD with controls 1	−0.5558	−0.0842	0.3416***	−0.0355
	(0.4563)	(0.2346)	(0.1152)	(0.1550)
DD with controls 2	−0.1772	−0.0127	0.3274***	0.1370
	(0.5142)	(0.2575)	(0.1002)	(0.1827)
Observations	626	4,489	21,864	1,485

Notes (1) Controls 1 include tax and ownership types. Controls 2 includes interactions between tax type and period, ownership type and period, firm size and aggregate VAT revenues from the tax region in a year.) Standard errors are clustered at the tax region industry level. (3) Significance levels: ***1%, **5% and *10%

Source authors' estimates based on TRA iTax data

Table 5 Estimated treatment effects after CEM matching

	IHS extra income
Basic DD	0.2552**
	(0.1216)
Observations	27,907

Notes Standard errors clustered at the tax region industry level. Significance levels: ***1%, **5% and *10%

Source authors' estimates based on ITAX data

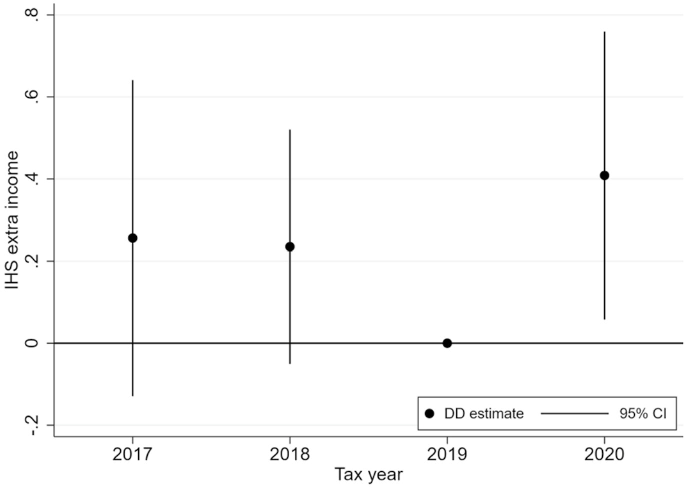

Fig. 4 Event study graphs after CEM matching. *Notes*: Standard errors are clustered at the tax region industry level. *Source*: authors' estimates based on TRA iTax data

Table 6 PPML with trimming changes

	Extra income (99.5th Pi)	Extra income (99th Pi)	Extra income (97.5th Pi)	Extra income (95th Pi)
Basic DD	0.4561***	0.2559**	0.0919	0.0911
	(0.1468)	(0.1126)	(0.0968)	(0.0809)
DD with Controls 1	0.4808***	0.2749***	0.1195	0.1145
	(0.1387)	(0.0986)	(0.0860)	(0.0702)
DD with Controls 2	0.4727***	0.2469**	0.1177	0.1085
	(0.1523)	(0.1011)	(0.0940)	(0.0800)
	27,847	27,704	27,278	26,566

Note Standard errors are clustered at the tax region industry level. Significance levels: ***1%, **5% and *10%

Source Authors' calculations from TRA iTax data

allows for the possibility that the parallel trends assumption does not hold well. However, in our case, the method's power may be limited by the relatively small number of periods. Using the honest DD approach, Figure A2 plots the confidence intervals if the pre-trends are extrapolated for the main outcomes. The figure shows robust confidence intervals with different $\bar{M}$ values. The breakdown value for our significant outcomes is around 1. This means that our treatment effects are robust to the post-treatment violations of parallel trends that are no larger than the maximal pre-treatment violation of parallel trends. Figure A2 shows that the results are robust to divergences in parallel trends, but the lower

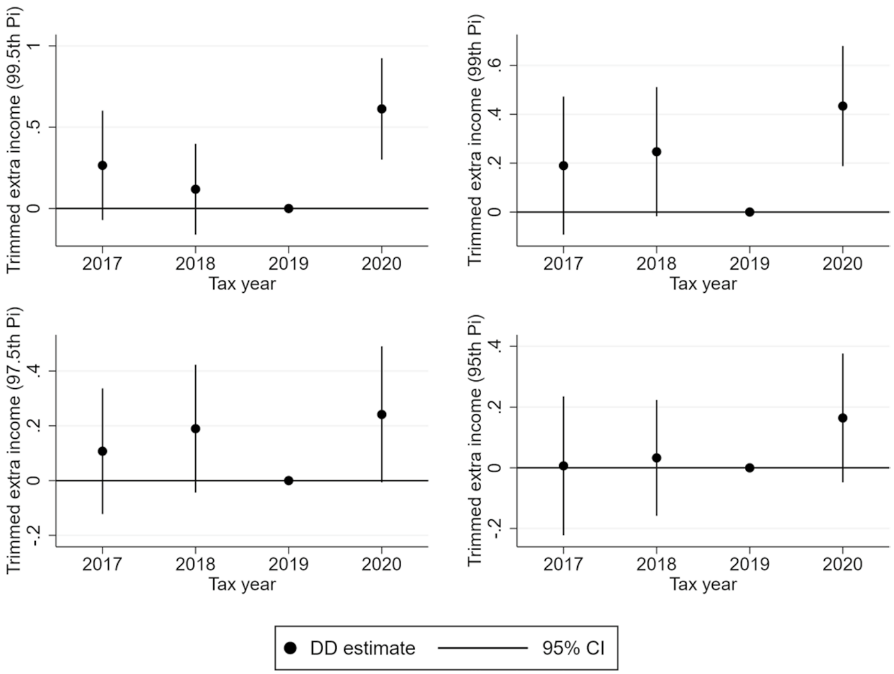

Fig. 5 PPML event graphs. *Note*: Standard errors are clustered at the tax region industry level. *Source*: Authors' calculations from TRA iTax data

bounds of confidence intervals are close to zero for the IHS adjusted income and IHS extra income outcomes for lower M bar values (<1).

7 Discussion

Our estimations show that the pilot increased the corrected amount of taxable income by approximately 20%. While the setup has specific challenges, such as a large difference in the mean size of the firm in the treatment versus control areas, the main results are still reasonably robust, since a similar picture emerges from the matched sample.

The confounding effect of COVID-19 is limited because the tax returns are from the time before COVID-19. Still, the pandemic may have impacted tax revenue generation differently in the treated and control areas. We attempt to control this possibility by including time-varying VAT revenues by region and industry indicator variables interacting with the reform dummy. A more extended follow-up period would also be ideal. The COVID-19 pandemic implied that the connections between TRA and VERO had to be moved to take place virtually as of March 2020, which may have limited the full potential of the intervention. There was also a simultaneous TRA-led campaign which targeted cases that had not been examined in the past five years, bringing additional, predominantly smaller cases for examinations. This campaign affected both treatment and control group in principle in the same way, however.

The positive result in our study contrasts the null finding by Bachas et al. (2021). There can be various reasons for this difference. One plausible factor is the country's varying quality level of the conventional examination and audit process. Another is that the Tanzanian intervention was relatively wide-ranging in that it also contained extensive training, communication, and project management support, most likely contributing to its overall positive impact.

There is evidence that corruption is prevalent in developing countries (Olken & Barron, 2009; Olken & Olken, 2007), and Tanzania may not be that different. Okunogbe and Pouliquen (2022) find that a move to an e-filing system in taxation led to a decline in corruption because of reduced opportunities for extortion. In principle, a more efficient selection could also facilitate greater corruption opportunities if the selection algorithm finds cash-rich companies. On the other hand, a move to a less discretionary case selection limits staff freedom to pick suitable taxpayers for examination. The selection criteria were also primarily determined by the technical assistance team, and this may further limit the worry that corruption could be behind the revenue increase.

Tax examinations reveal detected evasion, which can be turned into an estimate of a tax gap, defined as the additional income uncovered in the examination divided by the sum of the originally reported income and the additional income. In our data, this amounts to approximately 22%. This may be compared to recent estimates of the tax gap in Pakistan, which exceeds 40% (Best et al., 2021), and Zambia, where the CIT and VAT gap stood at approximately 50% (Adu-Ababio et al., 2023). The Tanzanian figure is smaller, but one needs to remember it is not based on full audits, which is different to the Pakistani and Zambian case.

Finally, we discuss some of the project's cost–benefit properties. The Finnish development assistance component costs approximately €60,000. One year before the reform, the additional income from the treatment groups was €575 million. When multiplied by an estimated treatment effect (0.22) and the corporate income tax rate (0.3), the estimated tax revenue increases to €38 million. In other words, the intervention cost for the donor per additional revenue is tiny. The costs to TRA are unknown, but no additional revenue authority staff were hired to implement the pilot. The benefits of the projects in terms of additional tax revenue dominate any additional intervention costs.

8 Conclusions

This paper examined the revenue impacts of a new risk-based method of tax examinations (desk inspection) in Tanzania. The approach was jointly developed with the Finnish tax administration and represented a capacity development programme to increase the efficiency of tax processes. Our paper, therefore, also serves as an evaluation of a technical assistance project in the tax area. Based on a difference-in-differences estimation strategy and administrative tax data from TRA, the results indicate that the pilot increased the upward correction in taxable income by approximately 20%. The impacts were concentrated in the service sector and among corporate income taxpayers.

One factor that most likely contributed to the positive result is that the previous tax examination process was untargeted. In principle, all cases had to be examined, which left little scope for in-depth analysis. The reform focused the examination efforts on more suspicious instances, leaving more time for investigating the tax returns of selected cases. Another key ingredient was that the intervention included extensive training, dedicated '*examination champions*' in tax offices, structured communication protocols for examined taxpayers, and project management support for the whole process. These features of the intervention and the characteristics of the baseline situation should be taken into account when considering whether similar projects would lead to positive outcomes elsewhere.

In a context with little consistency in selecting taxpayers for examination, the pilot process has provided additional structure. The pilot standardized the examination process and provided detailed inspection instructions. This has translated into reduced examination process times using a systematic approach and clearly defined selection criteria. However, there are still possible efficiency gains via scaling up the new policy. In addition, TRA is integrating the new examination process into a thoroughly modernized overall information system. Therefore, the longer-term impacts of the process are still uncertain. Finally, an interesting topic for a follow-up study would be the potential deterrence impact of tax examination on subsequent initially declared incomes.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10797-024-09858-5.

Acknowledgements We thank seminar audiences at TRA, WIDER, and the IIPF congress for their feedback. Useful comments by two anonymous referees are gratefully acknowledged. We are especially thankful to the Editor, Nadine Riedel, for her guidance in the revision process.

Funding Open Access funding provided by University of Helsinki (including Helsinki University Central Hospital).

References

Addis Tax Initiative. (2021). *2019 ATI monitoring report*. Bonn: International Tax Compact.

Adu-Ababio, K., Koivisto, A., Lungu, E., Mwale, E., Msoni, J., & Musole, K. (2023). Estimating tax gaps in Zambia: A bottom-up approach based on audit assessments. WIDER working paper 2023/25. Helsinki: UNU-WIDER. https://doi.org/10.35188/UNU-WIDER/2023/333-8.

Advani, A., Elming, W., & Shaw, J. (2023). The dynamic effects of tax audits. *Review of Economics and Statistics, 105*(3), 545–561. https://doi.org/10.1162/rest_a_01101

Allingham, M. G., & Sandmo, A. (1972). Income tax evasion: A theoretical analysis. *Journal of Public Economics, 1*(3–4), 323–338. https://doi.org/10.1016/0047-2727(72)90010-2

Bachas, P., Brockmeyer, A., Ferreira, A., & Sarr, B. (2021). *How to target enforcement at scale? Evidence from tax audits in Senegal*. EDI working paper series 2022.

Basri, M., Felix, M., Hanna, R., & Olken, B. (2020). *Tax administration vs. tax rates: Evidence from corporate taxation in Indonesia, Mimeo*. Cambridge, MIT: Massachusetts Institute of Technology.

Battaglini, M., Guiso, L., Lacava, C., Miller, D., & Patacchini, E. (2023). Refining public policies with machine learning: The case of tax auditing. CEPR discussion paper no. 17796. Paris: CEPR Press. https://cepr.org/publications/dp17796

Best, M., Shah, J., & Waseem, M. (2021). *Detection without deterrence: Long-run effects of tax audit on firm behavior*. Manchester: University of Manchester.

Carrillo, P., Pomeranz, D., & Singhal, M. (2017). Dodging the Taxman: Firm misreporting and limits to tax enforcement. *American Economic Journal: Applied Economics, 9*(2), 144–164. https://doi.org/10.1257/app.20140495

Chen, J., & Roth, J. (2023). Logs with zeros? Some problems and solutions. *The Quarterly Journal of Economics*. https://doi.org/10.1093/qje/qjad054

DeBacker, J., Heim, B., Tran, A., & Yuskavage, A. (2015). Legal enforcement and corporate behavior: An analysis of tax aggressiveness after an audit. *Journal of Law and Economics, 58*, 291–324.

DeBacker, J., Heim, B., Tran, A., & Yuskavage, A. (2018a). Once bitten, twice shy? The lasting impact of enforcement on tax compliance. *Journal of Law and Economics, 61*, 1–35.

DeBacker, J., Heim, B., Tran, A., & Yuskavage, A. (2018b). The effects of IRS audits on EITC claimants. *National Tax Journal, 71*, 451–484.

Epaphra, M. (2015). Tax rates and tax evasion: Evidence from missing imports in Tanzania. *International Journal of Economics and Finance, 7*(2), 122–137. https://doi.org/10.5539/ijef.v7n2p122

ESRF. (2010). Informal sector taxation in Tanzania. TAKNET policy brief series number 012-2010. Dar es Salaam: Economic and Social Research Foundation. Available at: http://esrf.or.tz/docs/INFORMATIONSECTORTAXATIONINTANZANIA_12.PDF. Accessed 8 Feb 2021.

Iacus, S., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political Analysis, 20*, 1–24. https://doi.org/10.1093/pan/mpr013

Khwaja, M. S., Awasthi, R., & Loeprick, J. (Eds.). (2011). *Risk-based tax audits. Approaches and country experiences*. Washington, DC: World Bank. https://doi.org/10.1596/978-0-8213-8754-2

Kleven, H. J., Knudsen, M. B., Kreiner, C. T., Pedersen, S., & Saez, E. (2011). Unwilling or unable to cheat? Evidence from a Tax Audit Experiment in Denmark. *Econometrica, 79*(3), 651–692. https://doi.org/10.3982/ECTA9113

Kotsogiannis, C., Salvadori, L., Karangwa, J., Mukamana T. (2022). Do Tax Audits have a dynamic impact? Evidence from Corporate Income Tax Administrative Data. TARC discussion paper 035-21. Exeter: TARC.

Li, W., Pittman, J., & Wang, Z.-T. (2019). The determinants and consequences of tax audits. Some evidence from China. *Journal of the American Taxation Association, 41*, 91–122.

Okunogbe, O., & Pouliquen, V. (2022). Technology, taxation, and corruption: Evidence from the introduction of electronic tax filing. *American Economic Journal: Economic Policy, 14*, 341–372.

Olken, B. A., & Barron, P. (2009). The simple economics of extortion: Evidence from trucking in Aceh. *Journal of Political Economy, 117*, 417–452. https://doi.org/10.1086/599707

Olken, B., & Olken, B. (2007). Monitoring corruption: Evidence from a field experiment in Indonesia. *Journal of Political Economy, 115*, 200–249.

Pomeranz, D. (2015). No taxation without information: Deterrence and self-enforcement in the value added tax. *American Economic Review, 105*(8), 2539–2569. https://doi.org/10.1257/aer.20130393

Rambachan, A., & Roth, J. (2023). A more credible approach to parallel trends. *Review of Economic Studies, 90*, 2555–2591.

Santos Silva, J. M. C., & Tenreyro, S. (2006). The log of gravity. *The Review of Economics and Statistics, 88*(4), 641–658.

Slemrod, J. (2019). Tax compliance and enforcement. *Journal of Economic Literature, 57*(4), 904–954. https://doi.org/10.1257/jel.20181437

Tanzania Revenue Authority. (2022). Internal Monthly Revenue Reports (flash report). Dar es Salaam: TRA. Unpublished.

Wadhawan S. C., & Gray, C. (1998). Enhancing tax transparency in tax administration: A survey. African Economic Policy Discussion Paper 3. Arlington: EAGER Publications/BHM.

Authors and Affiliations

Amina Ebrahim[1] · Elineema Kisanga[2] · Ezekiel Swema[3] · Vincent Leyaro[2] · Edwin P. Mhede[4] · Ephraim Mdee[3] · Heikki Palviainen[5] · Jukka Pirttilä[6]

✉ Jukka Pirttilä
jukka.pirttila@helsinki.fi

Amina Ebrahim
amina@wider.unu.edu

Ezekiel Swema
eswema@tra.go.tz

Vincent Leyaro
vinleyaro@gmail.com

Edwin P. Mhede
edwin.mhede@dart.go.tz

Ephraim Mdee
efmdee@tra.go.tz

Heikki Palviainen
heikki.palviainen@gmail.com

1 United Nations World Institute for Development Economics Research, Katajanokanlaituri 6 B, 00160 Helsinki, Finland

2 School of Economics, University of Dar Es Salaam, P.O. Box 35045, Dar Es Salaam, Tanzania

3 Tanzania Revenue Authority, Sokoine Drive, Illala, Dar Es Salaam, Tanzania

4 Dar Rapid Transit Agency, Ubungo Maji, Morogoro Road, P.O. Box 724, Dar Es Salaam, Tanzania

5 Tampere University, Kalevantie 4, 33100 Tampere, Finland

6 University of Helsinki, Arkadiankatu 7, 00100 Helsinki, Finland