BUILDING FISCAL CAPACITY IN DEVELOPING COUNTRIES: EVIDENCE ON THE ROLE OF INFORMATION TECHNOLOGY

Merima Ali, Abdulaziz B. Shifa, Abebe Shimeles, and Firew Woldeyes

Weak fiscal capacity is a major challenge in low-income countries. Recently, governments have adopted information technology to modernize tax collection; however, there is little evidence on the impact of such reforms. We narrow this gap using unique administrative firm-level panel data covering all business taxpayers in Ethiopia. We find a robust increase in value-added tax collections and reported sales following the adoption of electronic sales register machines (ESRMs), without decreasing formal employment. These effects are larger among downstream firms. ESRM adoption is also associated with a decrease in entry into downstream sectors.

Keywords: developing economy, fiscal capacity, information technology, taxation

JEL Codes: H26, O10, O55

I. INTRODUCTION

Economic development requires a state capable of mobilizing fiscal resources to finance the provision of essential public goods — a capacity that developing countries often lack. Weak fiscal capacity of states has thus received increased attention in the political economy of development (e.g., Bird, 1989; Tanzi and Zee, 2000; Acemoglu, 2005; Besley and Persson, 2010, 2011; Baskaran and Bigsten,

Merima Ali: CHR Michelsen Institute, Norway and Syracuse University, Syracuse, NY, USA (merima.ali@cmi.no); Abdulaziz B. Shifa: Maxwell School, Syracuse University, Syracuse, NY, USA (abshifa@maxwell.syr.edu); Abebe Shimeles: African Economic Research Consortium & University of Cape Town, School of Economics, Nairobi, Kenya (a.shimeles@afdb.org); Firew Woldeyes: Ethiopian Development Research Institute, Addis Ababa, Ethiopia (w._rew@edri-eth.org)

Electronically published September 3, 2021

National Tax Journal, volume 74, number 3, September 2021.
 https://doi.org/10.1086/715511

2013). Governments with the bare minimum of tax administrative infrastructure, as is typical of developing countries, struggle to enforce tax compliance, partly due to lack of reliable records on earnings by taxpayers. Thus, the potential of information technology (IT) to gather and analyze large amounts of data on taxpayers at relatively minimal cost has attracted the attention of tax authorities in developing countries. Tax reform efforts to enhance monitoring of earnings and to improve tax collection "in developing countries have generally centered on information technology" (Bird and Zolt, 2008, p. 794). Nevertheless, there has been little systematic empirical evidence on the impact of these reforms. In this study, using administrative firm-level panel data on a large number of business taxpayers, we examine the impact of electronic sales register machines (henceforth, ESRMs) on tax revenues in the context of a developing country.

The focus of our study is a reform to expand the adoption of ESRMs in Ethiopia — a sub-Saharan African country with one of the lowest per capita incomes in the world and minimal fiscal capacity. Starting in 2008, the Ethiopian Revenue and Customs Authority (ERCA) required businesses to adopt ESRMs. The program was rolled out over many rounds, and in 2014, more than 70,000 firms adopted the new system to conduct business transactions. The machines register sales and print receipts. The transactions are then reported via a network to an ERCA server. Once a firm starts using ESRMs, ERCA receives daily data on the firm's revenue. With the traditional, paper-based receipts, this would have been prohibitively expensive and nearly impossible to accomplish.

Even though ESRMs have the potential to provide more accurate data on transactions to help minimize tax evasion, it is not obvious whether developing countries can effectively harness the technology to generate higher tax revenues. It depends on the effect of ESRMs on compliance among existing registered firms (i.e., the intensive margin), as well as their effect on the tax base (i.e., the extensive margin).

With regard to using ESRMs to increase compliance among existing taxpayers, developing countries may face technical and administrative challenges. Operation of the machines requires a reliable supply of electricity and network infrastructure, as well as availability of technical expertise that can administer the network. In addition, processing of the massive data collected to aid decision-making would not be an easy fit, given the possible lack of requisite technical expertise, coordination failures among public agencies, and weak enforcement of tax laws. Thus, in settings where institutions are weak, as is the case in developing countries, the impact of gaining extra information on earnings may be minimal.

The effect on the extensive margin is also important because, even though ESRM adoption may lead to increased compliance among registered taxpayers, whether the compliance increases the overall tax revenue depends on changes in the tax base. On the one hand, increasing taxes on final sales may lower demand, decrease production, and encourage exit out of (or discourage entry into) the formal sector. This effect implies that ESRMs can lead to erosion of the tax base. On the other hand, ESRM adoption may affect the decision of firms in a way that incentivizes them to operate at larger scales, hence expanding the tax base. For example, in the absence of ESRM adoption, to evade taxes, firms may operate on a smaller scale. This would be the case if, as firms become larger, they tend to leave detectable transactions (due, e.g., to increased use of legal contracts) that make tax evasion harder and undermine their incentives to grow. However, if the adoption of ESRMs leads to more accurate revenue records and a uniform level of enforcement across both small and large businesses, firms may no longer have the incentives to operate on smaller scales for the purpose of evading taxes.

Hence, to provide a fuller picture of the effect of ESRMs on overall fiscal capacity, we empirically examine the impact of ESRMs on both value-added tax (VAT) collections by registered taxpayers (i.e., the intensive margin) and proxies of the tax base (i.e., the extensive margin). We examine the effect of ESRMs on tax compliance by estimating changes in reported sales and VAT following ESRM adoption. As a proxy for the effect of ESRM adoption on the tax base, we examine the effects on formal employment and rates of entry/exit. Data on formal employment are sourced from official payroll records and hence represent only formal activities of the firm. Rates of entry/exit are defined as the numbers of newly entered/exited firms as a share of the total number of firms in the preceding period.

Baseline results from difference-in-difference (DID) estimates show that firms report higher sales and pay more VAT following ESRM adoption. We find that employment does not appear to decrease after the adoption of ESRMs. To mitigate the possible bias that may arise due to selection into ESRM adoption, we exploit potential variations in the intensity of treatment with respect to tax oversight induced by ESRMs. We do this by comparing changes in the outcome variables between upstream and downstream firms because the latter are more likely to underreport their revenues before the adoption of ESRMs and hence to face a higher intensity of changes in reporting incentives from increases in enforcement (Kleven et al., 2011; Naritomi, 2019). Results from the comparison between upstream and downstream firms show that the effects are larger among the latter.

We also undertake a matching difference-in-difference (MDID) analysis, which has increasingly been used in the evaluation literature to address endogeneity concerns in studies using nonexperimental data (Girma and Görg, 2007; Becerril and Abdulai, 2010). We use detailed information about firm characteristics so as to limit the comparison between adopters and nonadopters to a set of firms that are otherwise similar across several features. Because comparison units in the matching analysis are constructed based on observable characteristics, one cannot directly test the assumption that the comparison units constitute valid counterfactuals for causal identification. However, results from a placebo analysis do not appear to suggest the violation of the identification assumption in the matched samples. The estimated results from the MDID analysis deliver similar patterns — reported sales and VAT collections increase without decreases in employment.

On average, rates of ESRM adoption across sectors or locations do not show significant relationships with rates of net entry into the formal sector. However, for downstream sectors, higher adoption rate is associated with a reduction in entry into those sectors.

This paper contributes to the growing literature that assesses the impact of various policy experiments on tax collection, partly enabled by the greater availability of administrative tax records from developing countries (Kleven and Waseem, 2013; Best et al., 2015; Gupta et al., 2017; Pomeranz and Vila-Belda, 2019). One of the crucial challenges for tax authorities in developing countries is the lack of accurate information on earnings (e.g., Boadway and Sato, 2009; Gordon and Li, 2009; Olken and Singhal, 2011). Although we focus on the effect of ESRMs to address this challenge, several recent studies examine the impact of reporting by third parties to the tax authority.[1] The evidence on effectiveness of these reforms appears to be mixed. Carrillo, Pomeranz, and Singhal (2017) examine the impact of a policy experiment in Ecuador in which the tax authority attempted to improve compliance of corporate taxpayers using third-party information. Carrillo, Pomeranz, and Singhal (2017) find that, due to weak enforcement capacity of the tax authority, the intervention was ineffective in lowering tax evasion. On the other hand, Pomeranz (2015) and Mittal and Mahajan (2017) find that paper trails facilitated by VAT improve compliance. Using data from Brazil, Naritomi (2019) finds that incentivizing consumers to request receipts and inform on noncompliant firms improved compliance. Using data from Costa Rica, Brockmeyer and Hernandez (2016) find improvements in compliance due to third-party reporting through withholding on sales taxes.

The paper also complements a few recent studies that have examined the effects of digitization of tax data in developing countries. Using data from Tajikistan, Okunogbe and Pouliquen (2018) show that electronic tax filing (instead of in-person filing) lowers tax evasion, decreases compliance cost, and lowers bribe payments. Eissa et al. (2014) report that the adoption of electronic billing machines by firms in Rwanda increased VAT collections. Fan et al. (2018) also find that digitization of invoices increased VAT collection in China.

The rest of the paper is structured as follows. Section II describes the institutional context. In Section III, we describe the data and present descriptive evidence. Section IV presents results from the comparison of upstream and downstream firms. In Section V, we report results from matching analysis. Section VI presents results on entry/exit. Concluding remarks follow in Section VII.

II. BACKGROUND TO ESRM REFORM IN ETHIOPIA

Ethiopia is a low-income country with limited fiscal capacity. In 2015, Ethiopia's gross domestic product (GDP) per capita was US$1,632 in current purchasing

[1] In contrast, a number of studies assess the effect of increasing the amount of information available directly to taxpayers (as opposed to the tax authority), either in the form of threats against tax evasion or a moral appeal to comply with tax laws (e.g., Fellner, Sausgruber, and Traxler, 2013; Castro and Scartascini, 2015; Shimeles, Gurara, and Woldeyes, 2017; Brockmeyer et al., 2019). For a detailed review of this literature, see Slemrod (2019).

power parity, about half of the average in sub-Saharan Africa.[2] The country has registered strong growth in recent years, with an annual average rate of 10.8 percent from 2003 to 2016. However, the revenue collection remained low. During the period 2003–2016, the tax revenue in Ethiopia contributed to only 12 percent of GDP, which is much smaller than the sub-Saharan average of about 18 percent (GRD, 2018). The country also relies heavily on taxes on international trade — a kind of tax that is relatively easy to enforce but tends to be more distortionary to the economy. More than 40 percent of Ethiopia's tax revenue came from taxes on international trade, which is a very high ratio, even by the standards of developing countries (GRD, 2018).

To improve the efficiency of tax collection and increase tax revenue, Ethiopia initiated the Tax Systems Reform Program (TSRP) in 2003 (IMF, 2010; World Bank, 2013).[3] The TSRP included institutional reforms to improve tax collection, tax policies to broaden the tax base, and investment in modern IT to increase the audit and enforcement capacity. As part of the reform to broaden the tax base, Ethiopia introduced the VAT in 2003. Since its introduction, the VAT rate has been set at 15 percent. The VAT has now become a significant source of government revenue, contributing nearly one-fifth of the domestic total tax revenue and half of indirect tax revenue. The solid line in Figure 1 plots the number of VAT-registered taxpayers, which started with about 6,000 firms in 2003 and gradually expanded, reaching more than 140,000 firms by the end of 2014.

All VAT-registered businesses are required by law to issue receipts.[4] The receipts form the basis for auditing the volumes and types of transactions when businesses file their VAT returns at the end of every month. However, the lack of accurate records of sales and purchases make it difficult for the tax authority to verify all the earnings in the receipts. Also, the auditing requires a substantial amount of an accountant's time to manually review all receipts. To help address this problem, ERCA pressed ahead with another major reform, namely, the introduction of ESRMs.

Following a regulation issued by the Council of Ministers in 2007, ESRM adoption started in 2008.[5] Because ESRMs report transactions automatically to ERCA servers upon registration of transactions, their adoption could help verify the consistency between the amount of transactions filed in tax returns and the amount registered by ESRMs.

[2] The 2015 per capita GDP for sub-Saharan Africa was US$3,755 (World Development Indicators online data bank, accessed on April 8, 2019).

[3] The TSRP was supported by the World Bank as part of the Public Sector Capacity Building Program Support Project that was undertaken during the period 2003–2010 (World Bank, 2013).

[4] Failure to issue receipts may lead to legal punishments — both financial and prison sentences. The legislation for VAT registration exempted smaller firms. For a detailed theoretical discussion on the optimal VAT threshold, see Keen and Mintz (2004).

[5] See Schreiber (2018) for a more detailed discussion on Ethiopia's tax reforms.

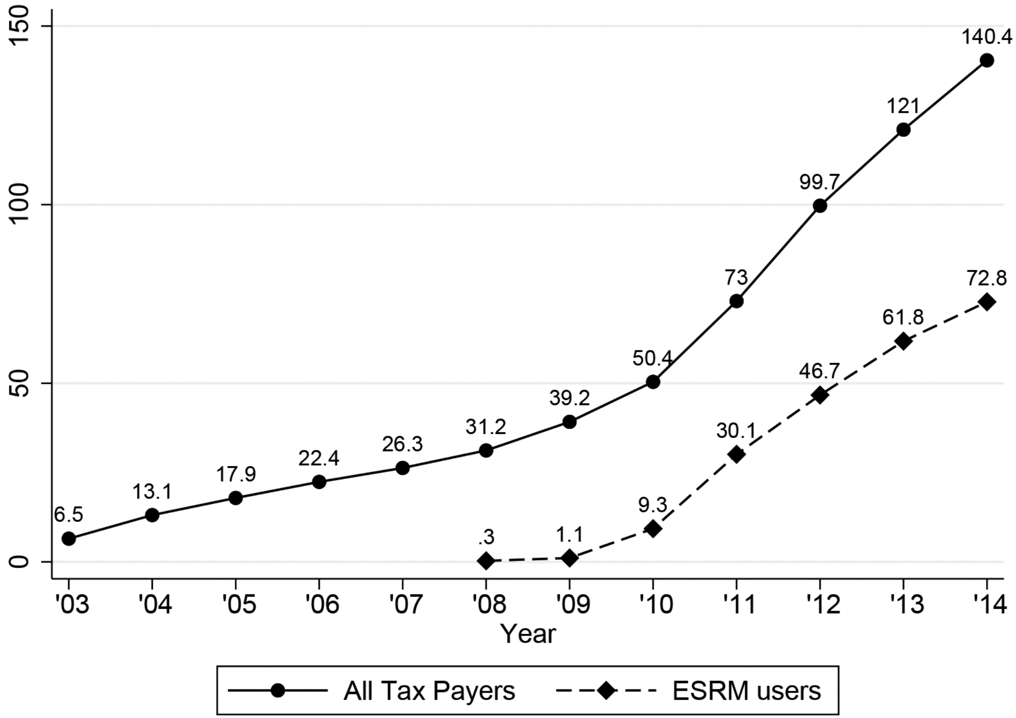

Figure 1. Number of VAT payers and ESRM adopters ('000).

Firms were required to buy the machines at their own cost from government-approved suppliers. The machines typically cost up to US$650 (based on the current exchange rate) — a significant sum for many businesses in Ethiopia. The implementation also requires ERCA's capacity to handle the information system and monitor compliance. Constrained by these factors, the adoption took place over an extended period of time.[6] Early in the implementation, the rollout was meant to occur through a series of ad hoc directives issued by ERCA. These directives set out broad guidelines to identify types of firms that should adopt ESRMs.[7] Managers of regional ERCA offices would then identify specific firms that should adopt ESRMs. However, according to our conversation with ERCA officials, the actual implementation was driven mainly by the discretion of regional managers and their logistical considerations. Because the machines are quite expensive, larger firms tended to be prioritized for the adoption of ESRMs, as they were considered to be less burdened by the cost of the machines. Even though there have been no directives issued since December 2010, adoption has continued mostly under the discretion of regional managers, who enforced adoption with a threat to suspend business licenses if firms refused to adopt. Figure 1 shows that the adoption of ESRMs began with a few hundred firms in 2008. By 2014, about 73,000 taxpayers (of more than 140,000 VAT-registered taxpayers) had adopted ESRMs.

The enforcement of ESRM use by adopted firms was instituted with legal and surveillance mechanisms. Harsh legal punishments, including long prison sentences, were enacted against VAT-registered businesses that evaded VAT.[8] The surveillance took place in the form of random visits by ERCA detectives, in which the detectives pose as ordinary customers and check whether sellers issue receipts. Such random visits have thus far led to several prosecutions and prison sentences.[9] In addition to the data sent through ESRMs, all VAT-registered businesses are required to file VAT returns every month (using a standard form). Businesses need to maintain their records for audit purposes. Due to the readily available data collected by ESRMs, ERCA would have an additional source of information on ESRM adopters during audits to investigate the accuracy of VAT filings. Thus, for ESRM adopters, the ease with which ERCA can assess the discrepancies between VAT filings and ESRM records means that ESRM adopters could face a higher risk of detection for underreporting their sales.

[6] Once ERCA managers identify a firm that should adopt ESRMs, the machines are installed at the firm's sales outlets/stores. This is done in the presence of IT technicians from ERCA, who assess whether the installations satisfy the technical requirements and standards set by ERCA.

[7] For example, the first directive, which was issued in 2008, required those firms operating in the hospitality industry (such as hotels, bars, and restaurants) to adopt ESRMs. The directive continued until March 2010 and used factors like size, business sectors, and location of business activities to determine which firms should adopt ESRMs.

[8] These penalties were included in Proclamation No. 609/2008, which was enacted in 2008.

[9] Local media reports extensively on such arrests. See, e.g., https://www.ethiopianreporter.com/index.php/article/13981 and http://ethionetblog.blogspot.com/2010/12/6500.html?m=1.

III. DATA AND DESCRIPTIVE RESULTS

A. Data and Econometric Framework

Our data set contains administrative tax records on the universe of VAT-registered firms in Ethiopia, covering the period from January 2003 — the year of VAT introduction in Ethiopia — to the end of 2014.[10] The administrative records provide information on several firm characteristics, such as sales, employment, VAT collections, location, types of business activity (sector), ownership structure, age, and date of ESRM adoption. Because all VAT-registered businesses must file their taxes every month, data on our key outcome variables — reported sales, VAT collections, and employment — are available on a monthly basis. Since the administrative data are from the firm's actual tax records, they provide accurate information on reported sales and VAT. The employment data also provide accurate information on formal employment, that is, employees who are officially reported in payroll records for tax filings. Using these data, we estimate the effect of ESRMs based on the DID equation:

$$y_{i,t} = \theta_1 \, (Post_t{}^{*} Treated_i) + \theta_2 Post_t + \theta_3 Treated_i + \theta_4 + \epsilon_{i,t}, \tag{1}$$

where the subscripts i and t identify the firm and the period. $Post_t$ is a dummy for the posttreatment period (i.e., after the period of ESRM adoption). $Treated_i$ is an indicator for whether the firm belongs to the treated group (i.e., whether the firm adopted ESRMs in the adoption period). θ_4 and $\theta_3 + \theta_4$ are group-specific means for the comparison and treated groups, respectively. The coefficient of interest is θ_1. It captures the change in level differences between the treatment and comparison groups following the adoption of ESRMs. The dependent variable $y_{i,t}$ represents any of the outcome variables (sales, VAT, and employment).

We construct four sets of estimation samples, depending on the period in which the treatment firms adopted ESRMs. We then run the DID Equation (1) on each of these samples. In constructing the estimation samples, we focus on firms that first adopted ESRMs in 2011 and 2012, because these periods offer large numbers of firms adopting ESRMs (Figure 1) but the economy still had many nonadopting firms to serve as control firms in the analysis.[11]

In the regression analysis, we aggregate the series into semester-level (half-yearly) series. The reason for choosing semester-level frequency is twofold. First, the semester-level series is an intermediate option in the trade-off between minimizing noise from using a lower frequency (e.g., a year) and capturing the dynamics by using a higher frequency (e.g., a month). Second, the number of firms that adopt ESRMs in a given month (or quarter) tends to be too few to construct samples with reasonable numbers

[10] The administrative data set was acquired by making an official request to ERCA.

[11] However, we also ran the regressions for firms that adopted in 2010 and in the first half of 2013, and we found similar results.

Table 1

Number of Firms in the Treatment and Comparison Groups

		Period of ESRM Adoption			
		2011:s1	2011:s2	2012:s1	2012:s2
		[1]	[2]	[3]	[4]
[I]	Window period	2010:s1–12:s1	2010:s2–12:s2	2011:s1–13:s1	2011:s2–13:s2
[II]	All new adopters	13,001	7,747	8,126	8,611
[III]	Treated firms	8,157	2,613	3,309	2,123
[IV]	Comparison firms	10,270	10,253	11,306	14,485

Note: Row II shows the number of firms that first adopted ESRMs during the corresponding period of ESRM adoption (given in the top panel). Row III shows the number of adopters that were active during the window period and hence are included in the treatment group. Row IV shows the number of firms that had not adopted ESRMs by the end of the window period and hence are included in the comparison group.

of treated firms. This is the case particularly in the matching analysis, because we can use only the observations that satisfy the balancing criteria.[12]

Columns 1–4 of Table 1 report the number of observations in each of these four samples. Column 1 describes the observations that we use to estimate the impact on firms that adopted ESRMs in the first semester of 2011 (2011:s1), that is, the treatment group is composed of firms that first adopted ESRMs in 2011:s1. In comparing the trends between treatment and control firms, we consider a window of two periods before and after ESRM adoption, so that in Equation (1), t = 2010:s1, 2010:s2, 2011:s1, 2011:s2, 2012:s1. These periods represent the preadoption time (2010:s1, 2010:s2), the adoption period (2011:s1), and the postadoption period (2011:s2, 2012:s1). The control group consists of firms that had not adopted ESRMs by the end of the comparison period (2012:s1).

In the data set, there are a total of 13,001 firms that first adopted ESRMs in 2011: s1 (Row II of Table 1). Of these, 8,157 (63 percent) were active during the window period 2010:s1–2012:s1 (Row III) and hence constitute the firms in the treatment group. We consider a firm to be active if it reported positive activity (i.e., taxes or sales) both before and after the window period. The rest of the new adopters (37 percent) are mostly newly registered VAT payers that were not active during the two preadoption periods (2010:s1 and 2010:s2). There were 10,270 firms that were active during the window period but had not adopted ESRMs by the end of the window period and hence comprise the sample of firms in the comparison group (Row IV).

In the second sample (Column 2), the treatment group consists of firms that adopted ESRMs in 2011:s2. Correspondingly, we examine the trends during the periods t =

[12] However, findings from the more disaggregated quarterly series deliver similar results. We report them in the Appendix. Tables A6 and A7 show these results from the quarterly series, using unmatched and matched samples, respectively.

2010:s2, 2011:s1, 2011:s2, 2012:s1, and 2012:s2. The set of control firms in this sample is selected from firms that had not adopted ESRMs by 2012:s2. In the remaining two samples, the treated firms consist of firms that adopted ESRMs in 2012:s1 and 2012:s2 (Columns 3 and 4, respectively). The control groups are selected from firms that had not adopted ESRMs by the end of the corresponding comparison periods.

It is worth noting that the average size of firms included in our regressions is likely to be larger than those in the population. As shown in Table 1, not all of the new adopters are included in the treated groups (Row II vs. Row III). The excluded sample mostly includes firms that are newly registered for VAT and that were not active during some of the preadoption times in the window period. Because new entrants tend to be smaller than incumbent firms, the exclusion of the former means that the treated group in our sample consists of relatively larger firms (than the population of new adopters). As shall be shown in Section IV, the impact of ESRMs tends to be larger among smaller firms. Hence, the estimated impacts for the treatment group in our regression samples are likely to represent lower bounds for the excluded ones.

B. Baseline Results

Table 2 reports the DID estimates (Equation (1)). Because distributions of sales, VAT, and employment are not normal, due to both the presence of observations with zero values and some outliers on the right tail, we use the log transformations (adding 1) as dependent variables.[13] This transformation helps minimize the problem of outliers and enables us to use all observations.[14] The outcome variables are listed in the first column. Robust standard errors clustered at the firm level are in parentheses.

In Table 2, we report the results for each of the four samples separately. We have also run the estimations in which we pooled all the observations in a single regression with firm and time fixed effects. Estimates from these pooled regressions deliver similar results.

The first row in Table 2 presents the estimated impact of ESRMs on reported sales. Reported sales show a statistically significant increase following ESRM adoption. The second row presents the estimated impact of ESRMs on VAT. The coefficient shows that VAT collections also show a statistically significant increase following ESRM adoption.

As briefly noted in Section I, ESRMs may affect not only tax reporting but also output. On the one hand, if ESRMs increase effective taxes on final sales through stronger enforcement, firms could face decreases in demand for their output due to

[13] As a robustness check, we also estimated using Poisson quasi-maximum likelihood (MLE) regressions (Olken and Singhal, 2011; Kondo et al., 2015), and we found similar results. Alternatively, we consider as the dependent variable a dummy indicating whether the firm reported positive values. We also checked by adding smaller numbers, such as 0.1 and 0.01. These transformations deliver similar results.

[14] Moreover, the effect of ESRMs is likely to depend on some base values due to factors such as inflation and firm size, making log-transformed dependent variables more appropriate.

Table 2

Baseline Results from DID Estimates

	Period of ESRM Adoption			
	2011:s1	2011:s2	2012:s1	2012:s2
	[1]	[2]	[3]	[4]
Dependent variable				
Sales	0.83	1.12	1.46	0.86
	(0.03)	(0.05)	(0.04)	(0.05)
VAT	0.41	0.63	1.01	0.64
	(0.02)	(0.04)	(0.03)	(0.04)
Employment	0.40	0.40	0.11	0.10
	(0.01)	(0.02)	(0.01)	(0.02)
$\mathbb{1}_{Sales > 0}$	0.11	0.15	0.24	0.12
	(0.01)	(0.01)	(0.01)	(0.01)
$\mathbb{1}_{VAT > 0}$	0.07	0.11	0.24	0.12
	(0.01)	(0.01)	(0.01)	(0.01)
$\mathbb{1}_{Employment > 0}$	0.19	0.21	0.08	0.05
	(0.00)	(0.01)	(0.01)	(0.01)
Observations	92,135	64,330	73,075	83,040
Firms	18,427	12,866	14,615	16,608
Treated	8,157	2,613	3,309	2,123
Comparison	10,270	10,253	11,306	14,485

Note: The table reports the estimated impact of ESRM adoption. We report the coefficient θ_1 from the DID regression equation: $y_{i,t} = \theta_1(Post_t{}^*Treated_i) + \theta_2 Post_t + \Gamma X_{i,t} + \epsilon_{i,t}$. The leftmost column lists the dependent variables. The estimates are provided for four samples (corresponding to each of Columns 1–4). In Column 1, the treated group consists of firms that adopted ESRMs during the first semester of 2011 (2011:s1). The comparison group consists of firms that either never adopted ESRMs or adopted after 2012:s2. Similarly, in Columns 2–4, the treated group adopted ESRMs during 2011:s2, 2012:s1, and 2012:s2, respectively. The comparison group in each sample consists of either firms that never adopted ESRMs or firms that adopted ESRMs a year after the treated firms in the respective sample (see Table 1). Robust standard errors, clustered by firms, are in parentheses.

the increase in the after-tax prices. On the other hand, ESRMs could improve business records and management oversight, thereby helping to expand output (Akcigit, Alp, and Peters, 2016).

The estimated changes in reported sales and VAT do not distinguish the changes between reported and actual activity. Therefore, they are not satisfactorily informative about the effect of ESRMs on output. In fact, one cannot rule out the possibility that output may decrease while reported sales and VAT collections increase. This could happen if firms pay more taxes (due to improved compliance), even if their output decreases. Output is not directly observed in our data set. However, we have data on employment, and therefore we use employment as an alternative dependent variable to examine the effect of ESRMs on output and firm size. Because the employment data are drawn from the administrative tax records, they only include workers who are reported in the official payrolls, that is, they exclude workers who are employed informally. Thus, we consider the employment numbers as proxies for formal output (instead of all output).

Because our estimates capture the effect on only formal employment, part of the observed change in formal employment could happen due to firms shifting some of their informal employees to formal status, instead of due to hiring of new workers. However, because the tax base is likely to depend on the formal activities, using formal employment data can have the benefit of providing a more accurate picture of how ESRMs affect the tax base. The third row in Table 2 shows that the coefficient on employment is positive and significant, suggesting that firms did not decrease their formal production in response to ESRM adoption.

Turning to the dummy indicators for whether we observe positive values for the outcome variables, the results point to similar patterns. The likelihood that one observes positive values for reported sales, VAT collections, and employment increases significantly following the adoption of ESRMs.

IV. COMPARISON OF UPSTREAM VERSUS DOWNSTREAM FIRMS

With regard to minimizing underreporting of revenues, one could expect ESRMs to have a stronger effect among firms that evade prior to the adoption of ESRMs. For firms that comply prior to the adoption of ESRMs, reported sales are already accurate, and hence the adoption of ESRMs should have little effect on the accuracy of their reports (Kleven et al., 2011; Naritomi, 2019). We thus examine how the effect of ESRMs varies across firms that are likely to differ with respect to the incentives they have for VAT compliance.

An important aspect of VAT is the paper trail that it leaves behind across the value chain. Unlike in the case of a retail sales tax, for which firms collect taxes only from final consumers, VAT is also paid on input purchases by all firms across the value chain. Moreover, firms can claim credit for the VAT that they pay on input purchases, that is, instead of paying 15 percent of their total sales, firms pay 15 percent of the total value added (sales minus input purchases). This means that input buyers act as third-party reporters because to claim credit for taxes paid on input purchases, they must report their purchases to the tax authority (Pomeranz, 2015). In contrast, final consumers receive no such credits on their tax payments and hence lack similar incentives to report their purchases.

Due to this difference in the incentive to report purchases between input buyers and final goods buyers, downstream firms (e.g., retailers) could have a weaker incentive for VAT compliance than their upstream counterparts (e.g., wholesalers) (Naritomi, 2019). Because further enforcement tools are unlikely to affect the behavior of already compliant firms (Kleven et al., 2011), one could expect the impact of ESRM adoption to be relatively stronger for downstream firms than for upstream ones.

Figure 2. Trends for firms in downstream versus upstream sectors. The figure plots the means of reported sales and employment around the semester of adoption (the six months between the vertical lines). The means are plotted for downstream firms in the treated groups (solid thick lines), upstream firms in the treated group (solid thin lines), and firms in the comparison groups (broken lines). The means are normalized by each group's means of the respective variable in the pre-adoption period.

We do not have data from input-output tables across firms. Instead, we use sector names provided in the administrative data to categorize firms into downstream and upstream ones. The former group consists of firms in sectors that are likely to engage mostly in retail sales to final consumers, such as grocery retailers, beauty salons, and restaurants. The latter group includes wholesalers, whose buyers are likely to be other firms (rather than final consumers).[15]

Figure 2 displays how the impact of ESRMs varies between firms in downstream versus upstream sectors. We plot the monthly sales and employment trends for three groups of firms: nonadopters, downstream adopters, and upstream adopters. The plots for VAT track reported sales very closely and hence are not reported. The trends are plotted for 12 months around the semesters of ESRM adoption. The months in the semesters of ESRM adoption are represented by the middle periods indicated by the two vertical lines. The plots to the left (right) of this middle area represent preadoption (postadoption) mean values aggregated across all of the four samples. For ease of comparison, the plotted values are normalized by each group's means of the respective outcome variable in the preadoption period.

[15] Some of the firms are dropped from Table 3 because we are not able to determine whether the firm is in a downstream sector.

The focus on the comparison between downstream and upstream firms, as compared with adoption status, has two main identification advantages. First, because both upstream and downstream firms have adopted ESRMs, this comparison avoids potential bias that could result from endogenous selection into ESRM adoption. Second, one could be concerned that anticipated adoption may affect the reaction of firms in the periods prior to the adoption of ESRMs. Such anticipatory effect on future adopters could create a systematic difference between adopters and nonadopters prior to the adoption of ESRMs, violating the parallel trend assumption. The comparison between upstream and downstream firms avoids such bias because the period of adoption is held the same for both groups of firms.

The pattern in Figure 2 confirms that the effect of ESRMs is indeed larger on downstream firms than their upstream counterparts. For the months prior to the semester of adoption, the upstream and downstream firms track each other very closely, rendering a strong credibility to the identifying assumption. The two groups diverge considerably after the adoption of ESRMs, with downstream firms showing a larger upward shift.

Table 3 reports the estimated impact of ESRMs on downstream versus upstream firms. In addition to allowing for group-specific means for the two sets of firms, we include, in the regression equation, an interaction term $Post \times Treated \times DS$, in which DS (for "downstream") is a dummy for whether the firm is in a downstream sector:

$$\begin{aligned} y_{i,t} = {} & \kappa_1\,(Post_t \times Treated_i \times DS) + \kappa_2\,(Post_t \times Treated_i) \\ & + \kappa_3\,Post_t + \kappa_4\,Treated_i + \kappa_5 + \epsilon_{i,t}. \end{aligned} \tag{2}$$

As a robustness check, we also consider the sector's size (as measured by the mean of sales in the sector) as an alternative proxy for the sector's position in the value chain. Because firms in upstream sectors (e.g., wholesalers) tend to be larger than those in downstream ones, volumes of sales are likely to be larger in the former. Thus, we expect the impacts of ESRMs to be larger in sectors dominated by smaller firms. We report these results in Table 4. The interaction term, $Post \times Size$, indicates how the effects of ESRMs vary depending on the sector's size. The variable *Size* represents the lag of the sector's mean of sales per firm (log scale) during the year prior to the period of ESRM adoption.

For sales and VAT, the coefficients on $Post \times DS$ are positive, implying a larger effect of ESRMs on reported sales and VAT collections on downstream firms than their upstream counterparts. Similarly, Table 4 also shows that the impact of ESRMs on reported sales and VAT collections decreases with an increase in the sector's size (i.e., the coefficients on $Post \times Size$ are negative). In both tables, the effect on employment is either significantly larger for downstream firms or shows no significant difference. Thus, the general picture from both Tables 3 and 4 is that the impact of ESRMs on reported sales and VAT tends to be larger for downstream firms than for upstream ones, and this extra effect on sales and VAT for downstream firms appears to have occurred without decreasing employment.

Table 3

Effect on Upstream versus Downstream Firms

		Period of ESRM Adoption			
		2011:s1	2011:s2	2012:s1	2012:s2
Dependent var	Right-hand-side var	[1]	[2]	[3]	[4]
Sales	*Post* × *Treated*	0.44	0.64	0.80	0.89
		(0.04)	(0.09)	(0.17)	(0.15)
	Post × *Treated* × *DS*	0.37	0.37	1.09	0.24
		(0.06)	(0.11)	(0.17)	(0.16)
Value-added tax	*Post* × *Treated*	0.34	0.52	0.62	0.49
		(0.03)	(0.07)	(0.12)	(0.11)
	Post × *Treated* × *DS*	0.29	0.23	0.68	0.30
		(0.04)	(0.08)	(0.12)	(0.12)
Employment	*Post* × *Treated*	0.25	0.36	0.18	0.14
		(0.01)	(0.03)	(0.03)	(0.03)
	Post × *Treated* × *DS*	0.40	0.29	0.07	0.03
		(0.02)	(0.04)	(0.03)	(0.04)
$\mathbb{1}_{Sales > 0}$	*Post* × *Treated*	0.03	0.06	0.10	0.12
		(0.01)	(0.01)	(0.02)	(0.02)
	Post × *Treated* × *DS*	0.05	0.05	0.19	0.04
		(0.01)	(0.02)	(0.02)	(0.02)
$\mathbb{1}_{VAT > 0}$	*Post* × *Treated*	0.03	0.07	0.11	0.09
		(0.01)	(0.01)	(0.02)	(0.02)
	Post × *Treated* × *DS*	0.05	0.05	0.18	0.06
		(0.01)	(0.02)	(0.02)	(0.02)
$\mathbb{1}_{Employment > 0}$	*Post* × *Treated*	0.14	0.21	0.10	0.08
		(0.00)	(0.01)	(0.01)	(0.01)
	Post × *Treated* × *DS*	0.14	0.10	0.05	0.00
		(0.01)	(0.02)	(0.02)	(0.02)
Observations		55,705	32,345	39,215	41,620
Firms		11,141	6,469	7,843	8,324
Treated		6,724	2,111	2,702	1,640
Comparison		4,417	4,358	5,141	6,684

Note: The table reports the estimated impact of ESRM adoption. We report the coeficients θ_1 and θ_2 from the DID regression equation: $y_{i,t} = \theta_1(Post_t{*}Treated_i) + \theta_2(Post_t{*}Treated_t{*}DS_i) + \Gamma X_{i,t} + \epsilon_{i,t}$. The variable *DS* is an indicator for whether the firm is in a downstream sector. The leftmost column lists the dependent variables. The estimates are provided for four samples (corresponding to each of Columns 1–4). In Column 1, the treated group consists of firms that adopted ESRMs during the first semester of 2011 (2011:s1). The comparison group consists of firms that either never adopted ESRMs or adopted after 2012:s2. Similarly, in Columns 2–4, the treated group adopted ESRMs during 2011:s2, 2012:s1, and 2012:s2, respectively. The comparison group in each sample consists of either firms that never adopted ESRMs or firms that adopted ESRMs a year after the treated firms in the respective sample. Robust standard errors clustered by firms are in parentheses.

Table 4

Effect of ESRMs by Sector Size

		Period of ESRM adoption			
		2011:s1	2011:s2	2012:s1	2012:s2
Dependent var	Right-hand-side var	[1]	[2]	[3]	[4]
Sales	*Post* × *Treated*	2.31	3.11	3.84	4.05
		(0.17)	(0.32)	(0.37)	(0.43)
	Post × *Treated* × *Size*	−0.10	−0.13	−0.17	−0.21
		(0.01)	(0.02)	(0.03)	(0.03)
VAT	*Post* × *Treated*	1.57	1.85	2.32	2.88
		(0.12)	(0.23)	(0.26)	(0.30)
	Post × *Treated* × *Size*	−0.08	−0.08	−0.09	−0.15
		(0.01)	(0.02)	(0.02)	(0.02)
Employment	*Post* × *Treated*	1.25	0.75	0.16	0.08
		(0.06)	(0.14)	(0.10)	(0.15)
	Post × *Treated* × *Size*	−0.06	−0.02	−0.00	0.00
		(0.00)	(0.01)	(0.01)	(0.01)
$\mathbb{1}_{Sales > 0}$	*Post* × *Treated*	0.28	0.46	0.72	0.63
		(0.03)	(0.05)	(0.06)	(0.08)
	Post × *Treated* × *Size*	−0.01	−0.02	−0.03	−0.03
		(0.00)	(0.00)	(0.00)	(0.00)
$\mathbb{1}_{VAT > 0}$	*Post* × *Treated*	0.24	0.41	0.72	0.67
		(0.03)	(0.05)	(0.06)	(0.08)
	Post × *Treated* × *Size*	−0.01	−0.02	−0.03	−0.04
		(0.00)	(0.00)	(0.00)	(0.00)
$\mathbb{1}_{Employment > 0}$	*Post* × *Treated*	0.56	0.39	0.19	0.12
		(0.02)	(0.06)	(0.05)	(0.06)
	Post × *Treated* × *Size*	−0.02	−0.01	−0.01	(0.06)
		(0.00)	(0.00)	(0.00)	(0.00)
Observations		92,135	64,330	73,075	83,040
Firms		18,427	12,866	14,615	16,608
Treated		8,157	2,613	3,309	2,123
Comparison		10,270	10,253	11,306	14,485

Note: The table reports the estimated impact of ESRM adoption. We report the coefficients θ_1 and θ_2 from the DID regression equation: $y_{i,t} = \theta_1 (Post_t{}^*Treated_i) + \theta_2 (Post_t{}^*Treated_i{}^*Size_i) + \Gamma X_{i,t} + \epsilon_{i,t}$. The variable *Size* is the sector's average firm size. The leftmost column lists the dependent variables. The estimates are provided for four samples (corresponding to each of Columns 1–4). In Column 1, the treated group consists of firms that adopted ESRMs during the first semester of 2011 (2011:s1). The comparison group consists of firms that either never adopted ESRMs or adopted after 2012:s2. Similarly, in Columns 2–4, the treated group adopted ESRMs during 2011:s2, 2012:s1, and 2012:s2, respectively. The comparison group in each sample consists of either firms that never adopted ESRMs or firms that adopted ESRMs a year after the treated firms in the respective sample. Robust standard errors clustered by firms are in parentheses.

V. MATCHING ESTIMATES

Compared with the approach in which all of the nonadopters are included in the control group, matching is desirable, because it minimizes the likelihood of bias if the nonadopters are considerably different from the adopters (Blundell and Dias, 2000). Once observations from treatment and control groups are matched (based on pretreatment characteristics), standard matching methods use observations in the matched sample to estimate the difference in (weighted) mean outcome levels between treatment and control groups. However, because we have longitudinal data, we estimate the difference in mean differences (instead of the difference in mean levels) by using the DID Equation (1).

This procedure of combining matching and DID methods helps to exploit the advantages of both methods. Whereas the standard matching estimator (of differences in levels) would require the strong assumption that, in the absence of the treatment, levels of outcome variables should be the same across treatment and control groups in a matched sample, causal interpretation in the MDID requires a relatively weaker assumption by allowing for unobserved time-invariant differences between the two groups (Smith and Todd, 2005). Hence, the combination of matching and DID "has the potential to improve the quality of non-experimental evaluation results significantly" (Blundell and Dias, 2000, p. 438).

A. Estimation Samples

Using detailed information provided by the administrative tax records in the data set, we construct our estimation samples by matching the treatment and control groups across several characteristics (18 covariates). We use two broad sets of matching covariates in the data set. The first set of variables includes a number of time-invariant firm characteristics: a location indicator for whether the firm is in Addis Ababa; two indicators for each sector — one for retail and another for other services; an indicator for whether the firm is a sole proprietorship (as opposed to limited liability); and two dummies indicating the firm's size category, as registered by ERCA (small, medium, and large). Treatment and control groups may exhibit differential trends if, for example, firm growth opportunities could vary across locations, sectors, and/or legal status, and these factors could in turn be correlated with the likelihood of ESRM adoption. Thus, matching on these covariates helps minimize bias that could result from their confounding effects.

The next set of matching covariates include time-varying variables, which are meant to match the control and treatment groups with respect to characteristics that could change over time. Hence they are meant to address the concern that the estimated effects may be confounded by differential time trends between the treatment and comparison groups due to, for example, temporary shocks that affect the two groups differently. For example, past growth might be correlated with future growth due to persistence in productivity shocks. We thus match treated and control firms with respect to pretreatment dynamics in sales, tax payments, and employment by matching on the values of sales, VAT, and employment during the two pretreatment periods (i.e., the two periods preceding the period of adoption). To account for shocks at sector and local levels, we match the pretreatment values of average sales at district and sector levels.[16]

Standardized differences are commonly used to examine the similarity of distribution of matching variables between treatment and comparison groups (Rosenbaum and Rubin, 1985). In Table A1 (Appendix is available online), we report mean standard bias (MSB) and other balance statistics.[17] There is no universally accepted cutoff value for MSB to decide whether the matching is satisfactory. However, some authors suggest that an MSB of 10 or 5 percent could roughly be considered as an indicator for negligible imbalance (Rosenbaum and Rubin, 1985; Caliendo and Kopeinig, 2008). Thus, in constructing the matched sample, we prune observations until — rather conservatively — the MSBs are below 5 percent for each of the matching covariates.

We provide a more detailed comparison of the treatment and control groups in Tables A2–A5, both before and after matching. In the unmatched samples, adopters tend be larger, as indicated by the mean values of sales and employment. They are also more likely to be registered as large firms (instead of small or medium). Adopters have a higher probability of being sole proprietorships. They are more likely to be located Addis Ababa (the capital), and in districts and sectors with larger firms (as indicated by mean sector sales and district sales). Adopters tend to be younger and are more likely to operate in retail and other services. Our matching estimates minimize biases if these differences between treated and control firms affect the timing of ESRM adoptions in a way that correlates with trends in the outcome variables.

Table 5 shows the number of firms in the matched sample. In Column 1, the treated group consists of firms that adopted ESRMs in 2011:s1. There are 7,909 and 4,785 firms in the treated and comparison groups, respectively. This represents 97 percent of the

[16] We use the caliper-matching algorithm developed by Lechner, Miquel, and Wunsch (2011), which involves four major steps. First, probability of treatment (propensity score) is estimated using a standard probit model. Second, for each treatment unit, Mahalanobis distances between the treatment unit and comparison units are computed (over a subset of the matching covariates and the estimated propensity scores). Third, each treatment unit is matched with a set of comparison units that lie within a given distance (or radius) from the treatment unit. Finally, control units are weighted, based on their similarity (as defined by the Mahalanobis distance) with the treatment units with which they are matched.

[17] Mean standard bias (in percent) is calculated as the difference in sample mean of the outcome variable between treated and control groups ($\bar{X}_T - \bar{X}_C$), divided by the square root of the average of the sample variance of the outcome variable for the treated and control groups (Rosenbaum and Rubin, 1985):

$$MSB = 100 \times \frac{\bar{X}_T - \bar{X}_C}{\sqrt{\frac{V_T(X)+V_C(X)}{2}}}.$$

Table 5

Number of Firms after Matching (as Percentage of Firms before Matching)

		Period of ESRM Adoption			
		2011:s1	2011:s2	2012:s1	2012:s2
		[1]	[2]	[3]	[4]
[I]	Window period	2010:s1−12:s1	2010:s2–12:s2	2011:s1–13:s1	2011:s2–13:s2
[II]	Treated firms	7,909 (97%)	2,537 (97%)	3,192 (97%)	1,977 (93%)
[III]	Comparison	4,785 (47%)	4,518 (44%)	5,704 (51%)	6,320 (44%)

Note: The table shows the number of treated firms (Row II) and comparison firms (Row III) in the matched samples. The figures in parentheses show the number of firms in the matched sample as a percentage of the number of firms in the unmatched sample (Table 1). The period of adoption for each treated group is given in the top panel.

treated and 47 percent of the comparison firms in the unmatched sample (Table 1). So balance is achieved by trimming observations mainly in the comparison groups. This is also true for all the other columns.

B. Empirical Results

Table 6 reports results from the MDID estimates using the matched samples. Consistent with the results in the unmatched sample (Table 2), the estimated coefficients are significant for all of the outcome variables. Thus, although the number of firms in the unmatched samples substantially exceeds that of the matched ones, the results from the two samples are generally similar.

Figure 3 plots trend differences (in reported sales and employment) between treated and comparison firms. The visual displays help provide a relatively transparent look at the data. We present the plots for the sample in which the period of adoption is 2012:s2, the most recent period of adoption in our analysis. The other three samples also display similar trends (see Figure A1). Dotted lines represent the 95 percent confidence intervals. These plots confirm the patterns reported in Table 6 — the treatment groups display significantly larger shifts for each of the outcome variables.

A crucial identification assumption in the matching analysis is that, in the absence of ESRM adoption, treatment and comparison groups within the matched samples would exhibit similar trends with respect to the outcome variables. Even though the similarity between treated and comparison units across detailed matching covariates helps to reinforce the plausibility of this assumption, one cannot directly test its empirical validity. Often a "placebo" comparison of treatment and comparison groups provides the best option in assessing the validity of the assumption that treatment and comparison groups would have had similar trends in the absence of the treatment.

Figure 4 provides such a comparison during periods that are presumably unaffected by ESRM adoption. To this end, we consider the placebo window period to be the five semesters preceding the window periods in the main regressions. We plot trend

Table 6

Results from Matching DID Estimates

	Period of ESRM Adoption			
	2011:s1	2011:s2	2012:s1	2012:s2
	[1]	[2]	[3]	[4]
Dependent variable				
Sales	0.99	1.01	1.20	1.01
	(0.10)	(0.11)	(0.11)	(0.08)
VAT	0.65	0.67	0.81	0.69
	(0.08)	(0.08)	(0.07)	(0.05)
Employment	0.16	0.14	0.08	0.13
	(0.03)	(0.05)	(0.02)	(0.02)
$\mathbb{1}_{Sales\,>\,0}$	0.17	0.19	0.22	0.16
	(0.02)	(0.02)	(0.02)	(0.01)
$\mathbb{1}_{VAT\,>\,0}$	0.17	0.19	0.22	0.16
	(0.02)	(0.02)	(0.02)	(0.01)
$\mathbb{1}_{Employment\,>\,0}$	0.07	0.08	0.06	0.06
	(0.02)	(0.02)	(0.01)	(0.01)
Observations	63,470	35,275	44,480	41,485
Firms	12,694	7,055	8,896	8,297
Treated	7,909	2,537	3,192	1,977
Comparison	4,785	4,518	5,704	6,320

Note: The table reports the estimated impact of ESRM adoption. We report the coefficient θ_1 from the DID regression equation: $y_{i,t} = \theta_1 \, (Post_t{}^{*}Treated_i) + \theta_2 Post_t + \Gamma X_{i,t} + \epsilon_{i,t}$. The leftmost column lists the dependent variables. The estimates are provided for four matched samples (corresponding to each of Columns 1–4). In Column 1, the treated group consists of firms that adopted ESRMs during the first semester of 2011 (2011:s1). The comparison group consists of firms that either never adopted ESRMs or adopted after a year (after 2012:s2). Similarly, in Columns 2–4, the treated group adopted ESRMs during 2011:s2, 2012:s1, and 2012:s2, respectively. The comparison group in each matched sample consists of either firms that never adopted ESRMs or firms that adopted ESRMs a year after the treated firms in the respective matched sample. Robust standard errors clustered by firms are in parentheses.

differences (in reported sales and employment) between treated and comparison groups during the placebo window periods. In Figure 4, we focus on the sample in which the period of adoption is 2012:s2. Results for the other samples are found to be similar (see Figure A2).

The placebo period in Figure 4 covers 2009:s1–2011:s1 and represents the placebo comparison corresponding to the most recent matched sample in the main regression (i.e., the five semesters preceding the window period 2011:s2–2013:s2). We first match firms that adopted ESRMs in 2012:s2 (i.e., the treatment period in

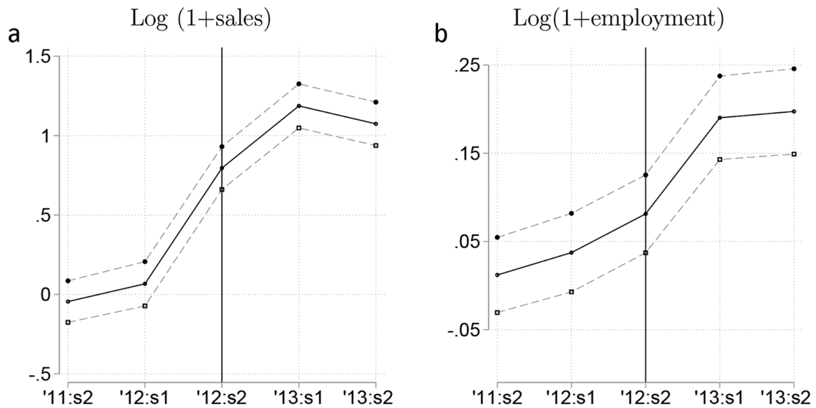

Figure 3. Trend differences (with 95 percent confidence intervals) between treated and comparison units, by period of adoption. The figure presents trend differences (in sales and employment) between treated and comparison for the matched sample with the most recent period of adoption, that is, the second semester of 2012 (2012:s2). The dotted lines represent the 95 percent confidence intervals. The vertical lines show the period of ESRM adoption for the treated group.

the main regression) with those that had not adopted until then, based on their characteristics in the first two semesters (2009:s1 and 2009:s2), and we next compare the differences post–2010:s1. Thus, the trend breaks, as we cross the vertical line in Figure 4, represent the placebo effects of ESRMs. The motivation for this analysis is

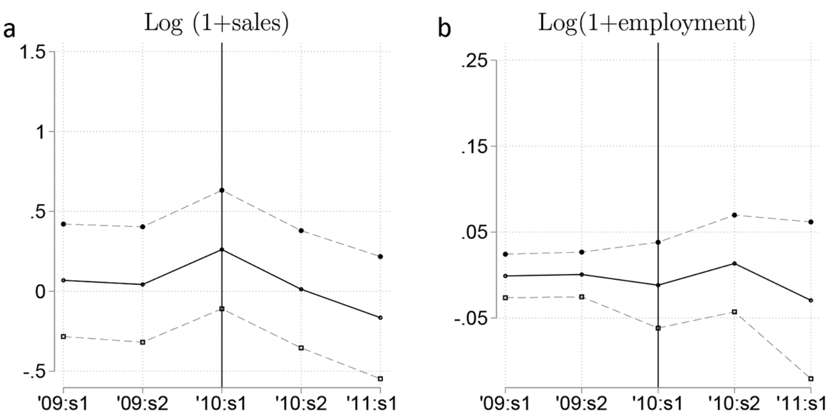

Figure 4. Placebo impact of ESRM adoption — trend differences between treated and comparison groups during the placebo window periods. This figure presents a placebo comparison (of reported sales and employment) between treated and comparison groups for the sample in which the treated firms adopted ESRMs in the second semester of 2012 (2012:s2). The dotted lines represent the 95 percent confidence intervals. The mean differences are plotted for the placebo window period (i.e., the five semesters prior to the matching periods). The vertical lines show the placebo period of ESRM adoption for the treated group.

that if one assumes that the treated and comparison groups would have similar trends in the absence of ESRM adoption, then one would expect no significant breaks beyond the vertical line. According to Figure 4, this appears to be the case. The trend differences between the two groups are relatively small and insignificant, providing no indication that the treated and comparison groups would diverge in the absence of ESRM adoption.[18]

C. Robustness

We undertake a number of robustness checks. First, we assess robustness of the results to the choice of alternative matching algorithms. Instead of using caliper matching, which uses all of the observations within a predefined distance, we implement nearest-neighbor matching, in which one uses only the nearest observation(s). We report these results in Table A8.[19] We also checked the sensitivity of the results to the exclusion of some of the matching covariates. Among others, we undertook the analysis after dropping all time-invariant variables from the list of matching covariates. We also checked dropping the pretreatment values of the outcome variables from the set of matching covariates. The results remain unchanged.

The next robustness checks relate to the spillover effects, which are particularly important in assessing the effect of ESRMs on overall outcomes (such as total tax revenue). Spillovers may arise, for example, if the adoption of ESRMs by some firms affects outcomes for nonadopters. If ESRM adopters become more competitive (vis-à-vis nonadopters), nonadopters may lose some of their market share. In this case, the regression coefficients may overestimate the effect of ESRMs. To the contrary, the coefficients could underestimate the effect if ESRM adopters become less competitive and lose some of their market share to nonadopters.

To address this concern, we reran the analysis, after excluding (from the comparison group) observations located in districts with relatively high levels of treatment rates. We first ranked districts according to the share of firms that adopted ESRMs in each district during the period in which the treatment firms adopted ESRMs. We then dropped all untreated observations located in districts where the treatment rate is above the median district and reran the analysis. Table A9 reports the results. The matched sample size is now smaller, as we have fewer firms to match. However, the results remain the same.

One could be concerned that the spillover effects could work through sectors rather than locations. Thus, we also ranked sectors according to rate of ESRM adoption (i.e., the fraction of firms in the sector that adopted ESRMs). We then

[18] Scales of the vertical axes are chosen so that magnitudes of the displayed effects in Figures 3 and 4 are visually comparable with each other.

[19] In the nearest-neighbor matching, we measure distance between observations defined on the basis of propensity scores (Rosenbaum and Rubin, 1983, 1985; Dehejia and Wahba, 2002). We have also implemented Mahalanobis distance matching and found similar results.

reran the analysis after excluding observations (from the comparison group) in which the treatment rate for the firm's sector is above that of the median sector. We found similar results.

We also checked the robustness of the results to allowing time fixed effects to vary by sector or location (by including sector- or district-specific fixed effects). Note that the control group consists of firms that either adopted late or those that never adopted until the most recent period in our data (2014). As the latter group (i.e., nonadopters) could be too dissimilar from the adopters, we reran the analysis after dropping them. Because the majority of firms in our data are sole proprietorships (Tables A2–A5), we also estimated separately for sole proprietorships. The results remain the same from all of these estimations.

D. Contribution to Overall VAT Revenue

The estimated effects, so far, show that VAT collection increased significantly following ESRM adoptions. However, because the dependent variable in the VAT regressions is $\log(1 + VAT)$, instead of $\log(VAT)$, the coefficient cannot be directly interpreted as the elasticity of VAT collections to ESRM adoption.

ESRMs could affect VAT collection by (1) affecting the probability that a firm pays a positive amount, and/or (2) changing the amount of VAT, conditional on the firm paying positive VAT. In assessing the impact of ESRMs for overall revenue collection, one needs to consider both effects. Even though the semester-level series could be desirable to observe the dynamics (e.g., the trend shifts after ESRM adoption), interpretation of the combined impact of these two effects tends to be easier with a more aggregated series (e.g., annual instead of monthly). This happens because zero values become less frequent with an increase in the aggregation level. Thus, we estimate Equation (1) using an aggregated series and calculate some ballpark numbers on the magnitude of the contribution of ESRMs to government revenue. Table 7 presents these results for each of the four matched samples. As the outcome variable, we consider the *total* value of VAT in the pretreatment (posttreatment) periods if $Post = 0$ ($Post = 1$). That is, we estimate Equation (1) in which, for each firm, we have only one aggregated pretreatment and one aggregated posttreatment observation.

In Column 1, the dependent variable is a dummy for whether the aggregated VAT is positive, that is, whether the firm paid positive VAT during any of the pretreatment (posttreatment) periods if $Post = 0$ ($Post = 1$). ESRM adoption is associated with an increase of 13–18 percentage points in the likelihood of paying positive VAT. As compared with the share of comparison firms that paid any positive VAT in the posttreatment period, these estimates imply that ESRM adoption is associated with an 18–27 percent increase in the probability of reporting positive VAT (Column 2).

Column 3 reports the effect of ESRMs on log VAT, conditional on the firm paying positive VAT. That is, we estimate Equation (1) using only observations for which VAT is positive. According to Column 3, the estimated impact ranges from no

Table 7

Magnitude of Contribution of ESRMs to VAT Revenue

Period of Adoption	Prob(VAT > 0) Percentage Points [1]	Percent [2]	Log(*VAT*) given *VAT* > 0 [3]
2011:s1	0.17***	23.95	−0.002
	(0.01)		(0.076)
2011:s2	0.18***	26.90	−0.06
	(0.02)		(0.10)
2012:s1	0.14***	17.72	0.32***
	(0.02)		0.08)
2012:s2	0.13***	20.31	0.20***
	(0.01)		(0.01)

Note: The table reports the estimated impact of ESRMs using the aggregated VAT series. The figures are reported for each of the four matched samples. Columns 1 and 2 report the impact on the probability of paying positive VAT in percentage points and percent, respectively. Column 3 reports the impact on the mean value of log VAT among firms that paid positive VAT. Robust standard errors are in parentheses. Asterisks denote significance at the 1% (***) level.

significant effect to a significant increase of 0.32 log points. Because neither of the two effects (i.e., Columns 2 and 3) show a significant negative effect, the largest of these coefficients in each row constitute a lower bound on the combined effect of ESRMs. That is, if ESRMs increase the likelihood of positive collections without lowering the amount of payments among those who collect positive VAT, then the former effect could be considered as a lower bound, and vice versa (Mihaylova et al., 2011).[20] Then, considering the maximums of Columns 2 and 3, this suggests that log VAT collections increased by (at least) 18–32 percent. Because VAT contributes nearly 20 percent of the total domestic revenue collection in Ethiopia, this increase corresponds to a 3.6–6.4 percent rise in the overall revenue.

We wish to highlight some important precautions regarding the assumptions and interpretation of these ballpark estimates. First, the coefficients do not take spillover (general equilibrium) effects into account. Second, the coefficients are estimated using only a subset of firms (i.e., matched observations) in the data set. So the validity of these coefficients to observations in the rest of the sample could be limited, depending on how well the latter are represented by the estimation sample. Importantly, the impact of ESRMs appears to decrease with firm size (Table 4). That is, the estimated average effect might be driven up by smaller firms, whose weight in the overall revenue

[20] Let α and β, respectively, denote the effect on the probability of reporting positive values (Column 3) and the effect on log VAT conditional on the firm paying positive VAT. Then, the combined effect on the expected log VAT equals $\alpha + \beta + \alpha\beta$.

may not be so large. Thus, the aggregate effect of ESRMs could be lower than what is implied by the ballpark numbers.

VI. ESRMS AND FIRM ENTRY/EXIT

As discussed in Section I, the potential contribution of ESRMs to build fiscal capacity may be undermined if firms lower their output in response to increases in the cost of production (due to extra tax liability), thereby leading to erosion of the tax base. Although the positive effect of ESRMs on employment suggests that this may not be the case, ESRMs may still affect the tax base if they affect the decision of firms to enter into (or exit from) the formal sector. For example, if ESRMs make it more difficult for firms to evade taxes, existing firms may respond by leaving the formal sector and operating informally, where they can shun the adoption of ESRMs. Similarly, the threat of ESRM adoption may discourage potential new entrants into the formal sector.

Admittedly, identifying the causal effect of ESRM adoption on entry/exit is difficult. We nonetheless present the available preliminary evidence on the association between ESRM adoption and entry/exit. We consider the following regression:

$$NetEntry_{s,t} = \psi \times AdoptionRate_{s,t-1} + \gamma_t + \omega_s + \epsilon_{s,t}. \quad (3)$$

The outcome variable is the rate of net entry in sector s, period t. The rate of net entry is defined as the number of newly entered firms less those that exited (as a share of the total number of firms in period $t-1$). The variable $AdoptionRate_{s,t-1}$ is the share of firms that adopted ESRMs in sector s, as of period $t-1$. The variables γ_t and ω_s are time and sector fixed effects. The variable ψ is the coefficient of interest. A negative value of ψ would suggest that increases in a sector's rate of ESRM adoption are associated with lower rates of net entry.

The first column of Table 8 reports the estimate for ψ. The coefficient does not show a statistically significant relationship between rate of ESRM adoption and net entry rate. In Columns 2 and 3, we consider the rate of entry and exit separately. Entry (exit) in period t is defined as the number of firms that newly entered (exited) as a share of the total number of firms in period $t-1$. These results also do not show a statistically significant relationship between entry/exit and adoption rates.[21]

As noted in Section IV, firms in downstream sectors exhibited a larger increase in reported sales and VAT collections after ESRM adoption, suggesting that the enforcement effect of ESRMs is stronger in downstream sectors. Thus, one could expect the effect on firms' entry/exit choices to be larger on downstream sectors. In Columns 4–6, we include the interaction term $DS \times AdoptionRate_{s,t-1}$, for which DS is a dummy indicating whether the sector is a downstream one. In Column 4,

[21] We have also looked at the correlation between the rate of net entry into a district ($NetEntry_{d,t}$) and the lags of ESRM adoption rate by firms within that district ($AdoptionRated_{d,t-1}$). These results also show no significant association between the rate of ESRM adoption and net entry across districts.

Table 8

Net Entry and the Rate of ESRM Adoption

	Dependent Variable Is Rate of					
	Net Entry	Entry	Exit	Net Entry	Entry	Exit
	[1]	[2]	[3]	[4]	[5]	[6]
$AdoptationRate_{s,t-1}$	−0.07	−0.07	−0.01	−0.01	−0.01	−0.01
	(0.06)	(0.06)	(0.00)	(0.15)	(0.15)	(0.01)
$DS \times AdoptationRate_{s,t-1}$				−0.23**	−0.23**	0.00
				(0.09)	(0.10)	(0.00)
Observations = 2,249						
Sectors = 197						

Note: The table shows the correlations between the current rate of net entry, entry, and exit into a sector and the lags of ESRM adoption rate in the sector ($AdoptationRate_{s,t-1}$). The variable *DS* is a dummy indicating whether the sector is a downstream one. Robust standard errors clustered by district are in parentheses. Asterisks denote significance at the 5% (**) level.

the estimated coefficient on this interaction term is significantly negative, suggesting that a higher rate of ESRM adoption could make firms shy away from downstream sectors. Estimating the effect separately for entry and exit (Columns 5 and 6), the effect is significant only for entry, suggesting that the impact operates primarily through discouraging the entry of new firms.

VII. CONCLUSION

Limited fiscal capacity of states has received increased attention as an important constraint to economic development. Having an effective tax system requires a vast administrative infrastructure with the capacity for gathering, analyzing, and monitoring earnings information on a large number of taxpayers — a capacity that many developing countries lack. Thus, the advent of electronic systems has attracted governments in many developing countries, because the technology could provide a relatively cheap alternative for monitoring earnings information and improving fiscal capacity. In this study, using administrative data from Ethiopia, we document empirical evidence on one such policy experiment.

We find that VAT collections by firms increase following ESRM adoption. We also find that for registered taxpayers, the increased VAT collections do not seem to have decreased output (as indicated by employment levels). However, we find some evidence of a negative relationship between firm entry and the adoption of ESRMs, suggesting a possible erosion of the tax base on the extensive margin.

The policy of adopting ESRMs in Ethiopia was implemented with a relatively strong political commitment by the government to reform its tax collection (Schreiber, 2018). Whether other developing countries can successfully harness the potential of

ESRMs may depend on their governments' will to enforce their tax laws. Thus, the many reforms to adopt ESRMs across developing countries warrant further research to assess their impact and get a fuller picture.

It should also be noted that despite such impactful reforms, Ethiopia's tax/GDP ratio remains one of the lowest, even by the standards of the African continent; hence, there may exist significant room to further enhance the potential of ESRMs in revenue collection. ESRMs, even when fully effective, could only monitor the revenue aspect of a firm's business transactions. Input purchases, however, remain less observable to the revenue authorities, and they mostly rely on self-reporting by businesses. As the application of ESRM expands to small businesses and input suppliers, there is a possibility of triangulating expense reports by third parties to self-reported expenses. This requires, among other things, a competent data computation facility, reliable network, and auditing capacity. The installation and operation costs of ESRMs may also become a constraining factor as the adoption expands to smaller firms. In light of these challenges, the question of whether developing countries can effectively utilize the third-party information generated by ESRMs begs more research.

ACKNOWLEDGMENTS

We thank editor William Gentry, coeditor Stacy Dickert-Conlin, and three anonymous referees for helpful comments. We are thankful to officials at the ERCA for access to the data and other relevant documents. We thank the International Center for Taxation and Development and the African Development Bank, both for financial support and useful feedback. We thank seminar participants at Cornell University, Working Group in African Political Economy workshops at New York University Abu Dhabi and Penn State University, and the 2018 Annual National Tax Association conference.

DISCLOSURES

The authors have no financial arrangements that might give rise to conflicts of interest with respect to the research reported in this paper.

REFERENCES

Acemoglu, Daron, 2005. "Politics and Economics in Weak and Strong States." *Journal of Monetary Economics* 52 (7), 1199–1226.

Akcigit, Ufuk, Harun Alp, and Michael Peters, 2016. "Lack of Selection and Limits to Delegation: Firm Dynamics in Developing Countries." NBER Working Paper No. 21905. National Bureau of Economic Research, Cambridge, MA.

Baskaran, Thushyanthan, and Arne Bigsten, 2013. "Fiscal Capacity and the Quality of Government in Sub-Saharan Africa." *World Development* 45 (C), 92–107.

Becerril, Javier, and Awudu Abdulai, 2010. "The Impact of Improved Maize Varieties on Poverty in Mexico: A Propensity Score-Matching Approach." *World Development* 38 (7), 1024–1035.

Besley, Timothy, and Torsten Persson, 2010. "State Capacity, Conflict, and Development." *Econometrica* 78 (1), 1–34.

Besley, Timothy, and Torsten Persson, 2011. *Pillars of Prosperity: The Political Economics of Development Clusters*. Princeton University Press, Princeton, NJ.

Best, Michael, Anne Brockmeyer, Henrik Kleven, and Johannes Spinnewijn, 2015. "Production vs Revenue Efficiency With Limited Tax Capacity: Theory and Evidence From Pakistan." *Journal of Political Economy* 123 (6), 1311–1355.

Bird, R. M., 1989. "The Administrative Dimension of Tax Reform in Developing Countries." In Gillis, Malcolm (ed.), *Lessons from Tax Reform in Developing Countries*, 315–346. Duke University Press, Durham, NC.

Bird, Richard M., and Eric M. Zolt, 2008. "Technology and Taxation in Developing Countries: From Hand to Mouse." *National Tax Journal* 61 (4), 791–821.

Blundell, Richard, and Monica Costa Dias, 2000. "Evaluation Methods for Non-Experimental Data." *Fiscal Studies* 21 (4), 427–468.

Boadway, Robin, and Motohiro Sato, 2009. "Optimal Tax Design and Enforcement with an Informal Sector." *American Economic Journal: Economic Policy* 1 (1), 1–27.

Brockmeyer, Anne, and Marco Hernandez, 2016. "Taxation, Information, and Withholding: Evidence from Costa Rica." Policy Research Working Paper 7600. World Bank, Washington, DC.

Brockmeyer, Anne, Spencer Smith, Marco Hernandez, and Stewart Kettle, 2019. "Casting a Wider Tax Net: Experimental Evidence from Costa Rica." *American Economic Journal: Economic Policy* 11 (3), 55–87.

Caliendo, Marco, and Sabine Kopeinig, 2008. "Some Practical Guidance for the Implementation of Propensity Score Matching." *Journal of Economic Surveys* 22 (1), 31–72.

Carrillo, Paul, Dina Pomeranz, and Monica Singhal, 2017. "Dodging the Taxman: Firm Misreporting and Limits to Tax Enforcement." *American Economic Journal: Applied Economics* 9 (2), 144–164.

Castro, Lucio, and Carlos Scartascini, 2015. "Tax Compliance and Enforcement in the Pampas: Evidence from a Field Experiment." *Journal of Economic Behavior and Organization* 116, 65–82.

Dehejia, Rajeev H., and Sadek Wahba, 2002. "Propensity Score-Matching Methods for Nonexperimental Causal Studies." *Review of Economics and Statistics* 84 (1), 151–161.

Eissa, Nada, Andrew Zeitlin, Saahil Karpe, and Sally Murray, 2014. "Incidence and Impact of Electronic Billing Machines for VAT in Rwanda." Technical Report, International Growth Center, London.

Fan, Haichao, Yu Liu, Nancy Qian, and Jaya Wen, 2018. "The Dynamic Effects of Computerized VAT Invoices on Chinese Manufacturing Firms." NBER Working Paper No. 24414. National Bureau of Economic Research, Cambridge, MA.

Fellner, Gerlinde, Rupert Sausgruber, and Christian Traxler, 2013. "Testing Enforcement Strategies in the Field: Threat, Moral Appeal and Social Information." *Journal of the European Economic Association* 11 (3), 634–660.

Girma, Sourafel, and Holger Görg, 2007. "Evaluating the Foreign Ownership Wage Premium Using a Difference-in-Differences Matching Approach." *Journal of International Economics* 72 (1), 97–112.

Gordon, Roger, and Wei Li, 2009. "Tax Structures in Developing Countries: Many Puzzles and a Possible Explanation." *Journal of Public Economics* 93 (7–8), 855–866.

GRD (Government Revenue Dataset), 2018, https://www.wider.unu.edu/project/government-revenue-dataset.

Gupta, Sanjeev, Michael Keen, Alpa Shah, and Genevieve Verdier, 2017. *Digital Revolutions in Public Finance*. International Monetary Fund, Washington, DC.

IMF (International Monetary Fund), 2010. *Ethiopia: 2010 Article IV Consultation and First Review of the Arrangement under the Exogenous Shocks Facility: Staff Report*. International Monetary Fund, Washington, DC.

Keen, Michael, and Jack Mintz, 2004. "The Optimal Threshold for a Value-Added Tax." *Journal of Public Economics* 88 (3–4), 559–576.

Kleven, Henrik J., and Mazhar Waseem, 2013. "Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan." *Quarterly Journal of Economics* 128 (2), 669–723.

Kleven, Henrik Jacobsen, Martin B. Knudsen, Claus Thustrup Kreiner, Søren Pedersen, and Emmanuel Saez, 2011. "Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark." *Econometrica* 79 (3), 651–692.

Kondo, Michelle C., Danya Keene, Bernadette C. Hohl, John M. MacDonald, and Charles C. Branas, 2015. "A Difference-In-Differences Study of the Effects of a New Abandoned Building Remediation Strategy on Safety." *PLoS ONE* 10 (7), e0129582.

Lechner, Michael, Ruth Miquel, and Conny Wunsch, 2011. "Long-Run Effects of Public Sector Sponsored Training in West Germany." *Journal of the European Economic Association* 9 (4), 742–784.

Mihaylova, Borislava, Andrew Briggs, Anthony O'Hagan, and Simon G. Thompson, 2011. "Review of Statistical Methods for Analysing Healthcare Resources and Costs." *Health Economics* 20 (8), 897–916.

Mittal, Shekhar, and Aprajit Mahajan, 2017. "VAT in Emerging Economies: Does Third Party Verification Matter?" Technical Report, SSRN, https://dx.doi.org/10.2139/ssrn.3029963.

Naritomi, Joana, 2019. "Consumers as Tax Auditors." *American Economic Review* 109 (9), 3031–3072.

Okunogbe, Oyebola Motunrayo, and Victor Maurice Joseph Pouliquen, 2018. "Technology, Taxation, and Corruption: Evidence from the Introduction of Electronic Tax Filing." Policy Research Working Paper WPS8452. World Bank, Washington, DC.

Olken, Benjamin A., and Monica Singhal, 2011. "Informal Taxation." *American Economic Journal: Applied Economics* 3 (4), 1–28.

Pomeranz, Dina, 2015. "No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax." *American Economic Review* 105 (8), 2539–2569.

Pomeranz, Dina, and José Vila-Belda, 2019. "Taking State-Capacity Research to the Field: Insights from Collaborations with Tax Authorities." *Annual Review of Economics* 11, 755–781.

Rosenbaum, Paul R., and Donald B. Rubin, 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70 (1), 41–55.

Rosenbaum, Paul R., and Donald B. Rubin, 1985. "Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score." *American Statistician* 39 (1), 33–38.

Schreiber, Leon, 2018. *Funding Development: Ethiopia Tries to Strengthen Its Tax System, 2007–2018, Innovations for Successful Societies*. Princeton University, Princeton, NJ.

Shimeles, Abebe, Daniel Zerfu Gurara, and Firew Woldeyes, 2017. "Taxman's Dilemma: Coercion or Persuasion? Evidence from a Randomized Field Experiment in Ethiopia." *American Economic Review* 107 (5), 420–424.

Slemrod, Joel, 2019. "Tax Compliance and Enforcement." *Journal of Economic Literature* 57 (4), 904–954.

Smith, Jeffrey A., and Petra E. Todd, 2005. "Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators?" *Journal of Econometrics* 125 (1–2), 305–353.

Tanzi, Vito, and Howell H. Zee, 2000. "Tax Policy for Emerging Markets: Developing Countries." *National Tax Journal* 53 (2), 299–322.

World Bank, 2013. *Ethiopia — Public Sector Capacity Building Program Support Project (PSCAP) (English)*. World Bank Group, Washington, DC.