Strategic or Confused Firms?
Evidence from "Missing" Transactions in Uganda*

Miguel Almunia
CUNEF
& CEPR & CAGE

Jonas Hjort
Columbia U.
& CEPR & NBER

Justine Knebelmann
Paris School
of Economics

Lin Tian
INSEAD

May 9, 2019

Abstract

Are firms sophisticated maximizers, or do they consistently make errors? We study this question in Uganda. We show that sellers and buyers report different amounts in 79 percent of transactions subject to value-added tax (VAT), despite invoices being easily comparable. We estimate that 29 percent of firms misreport own sales and purchases such that their liability *increases*. However, 71 percent are self-advantageous misreporters. Only such firms misreport less when exposed to tighter enforcement (at customs, where exchange rate-variation-induced imports pass through). Despite the gain from firm errors, overall, unilateral VAT misreporting cost Uganda USD 446 million in revenue from 2013-2016.

*miguel.almunia@cunef.edu, hjort@columbia.edu, justine.knebelmann@psemail.eu, lin.tian@insead.edu. We thank David Henning for outstanding research assistance, and Raymond Fisman, Michael Best, Francois Gerard, Jim Hines, Wojciech Kopczuk, Joana Naritomi, Claude Raisaro, and seminar participants at several universities and conferences for helpful comments and suggestions. We thank the Uganda Revenue Authority for data sharing and excellent collaboration. We gratefully acknowledge funding from the International Growth Centre (IGC), the British Academy, and the Leverhulme Foundation. The views in this paper are those of the authors, and do not necessarily represent those of the Uganda Revenue Authority.

1 Introduction

In economics, firms are seen as sophisticated organizations—maximizers that make constrained but optimal decisions by carefully assessing their true costs and benefits to the firm. This assumption is the starting point of the models that guide our understanding of how firms respond to public policies, for example when deciding whether to evade taxes, the decision we consider in this paper. With some exceptions, strategic decision-making by firms is by and large taken as self-evident.[1]

There is growing evidence to suggest that firms often deviate from optimal behavior, however.[2] If a significant proportion consistently makes mistakes, the consequences for theory and policy design would be far-reaching. Consider one of the most important questions for economic development: how firms in low-income countries should be taxed (Besley & Persson, 2009; Kleven *et al.* , 2016). Economists favor the value-added tax (VAT) in part because firms are thought to generally comply with a tax that is reported by both buyers and sellers, whose invoices can be cross-checked.[3] However, this argument assumes that firms are sophisticated enough to infer that cross-checks can occur and to accurately keep track of their sales and purchases.

In contrast to firms' sophistication, *states'* sophistication is studied in-depth (see e.g. Besley & Persson, 2013). Limited state capacity is in fact at the core of the argument for use of the VAT (though capacity to cross-check some reports is assumed). However, the "self-enforcing" VAT hypothesis ultimately rests on the assumption that firms are sophisticated. Confused firms may not respond as anticipated to the incentives generated by the tax and the low-but-non-negligible level of enforcement thought to prevail in most of the 166 countries around the world that now use the VAT.

In this paper, we study the sophistication of firms' decisionmaking in a low-income country context by analyzing their tax reporting behavior. We use transaction-level data from VAT returns and customs records for all domestic and international transactions during 2013-2016 involving Ugandan VAT-registered firms. First, we document widespread VAT discrepancies at the firm-pair×month level, using an approach akin to Fisman & Wei (2004)'s cross-checking of "mirror data". Next, we develop a firm-as-buyer and firm-as-seller fixed-effects methodology that allows us to estimate what fraction of each discrepancy is due to each of the two firms. This in turn enables us to assess the extent to which misreporting firms overreport total purchases and/or underreport total sales such that the firm's overall liability decreases, as opposed to making liability-increasing reporting errors. Third, we evaluate how firms that engage in self-advantageous versus -disadvantageous misreporting change their tax behavior *when the tax authority's capacity is en-*

[1] There are of course good reasons to think that firms are more sophisticated decisionmakers than individuals, who often make mistakes (Bernheim *et al.* , 2019).

[2] See e.g. DellaVigna & Gentzkow (2017); Tourek (2018); Kremer *et al.* (2019) and references therein.

[3] The VAT is argued to be self-enforcing in firm-to-firm trade for two reasons (Ebrill *et al.* , 2001; Kopczuk & Slemrod, 2006; Pomeranz, 2015). First, transactions between VAT-registered firms generate a "double" paper trail, as both sides of the transaction must keep a copy of the invoice. Second, seller and buyer have asymmetric reporting incentives. Another reason why economists recommend the use of the VAT is that the tax in theory does not distort production decisions.

hanced. We do this by constructing an exchange rate variation-based instrument for whether input purchases pass through customs—i.e., are imported rather than acquired from domestic sellers—where a degree of monitoring happens automatically.

In the first step of our analysis, we show that, in data on all domestic trade between the 19,161 VAT-registered firms in Uganda, sellers and buyers report different amounts in 79 percent of reported firm-pair×month observations.[4] This finding suggests that Uganda's authorities have limited capacity to detect and discipline VAT misreporting. In 60 percent of mismatch transactions we find a "seller shortfall," namely the seller reporting a lower value than the buyer. In the remaining 40 percent we observe a "buyer shortfall."

At face value, while seller shortfall is consistent with a standard model of VAT evasion in which the tax authority has low cross-checking capacity, the extent of buyer shortfall we observe points towards firm reporting errors. This is because buyer (seller) shortfall implies an *increase* (decrease) in the firm pair's combined tax liability, other things equal. However, Carrillo *et al.* (2017) point out that a firm that underreports its value-added may strategically choose to underreport *both* its sales and purchases if e.g. the firm believes that the audit probability increases with firm size. Such "looking small" behavior can potentially explain buyer shortfall while simultaneously benefiting the responsible firm, as can "looking big" behavior (i.e., overreporting both sales and purchases).

To distinguish between strategic misreporting and reporting errors, we first quantify the contribution of seller- and buyer-specific factors towards each discrepancy, using a firm-as-buyer and firm-as-seller fixed effects model. Summing up a firm's two estimated fixed effects allows us to categorize the firm as an Advantageous or a Disadvantageous misreporter depending on whether its *net* misreporting position decreases or increases the firm's overall tax liability.

We find that 71 percent of VAT-registered Ugandan firms are *Advantageous* misreporters and 29 percent *Disadvantageous* misreporters. The proportion of firms we categorize as making errors is likely an underestimate.[5] Among Advantageous misreporters, only 2 percent "look small" by underreporting both sales and purchases (and the firm's value-added). Another 77 percent are "Conspicuous" Advantageous misreporters that underreport their sales and overreport their purchases. The remaining 21 percent "look big" by overreporting both sales and purchases. The year-to-year correlation of a firm being categorized as engaging in advantageous misreporting is 0.77, while that of the firm being categorized as engaging in disadvantageous misreporting is 0.62.

The estimated government revenue *gain* due to reporting errors by Disadvantageous misreporters is large—around USD 131 million during 2013-2016. However, the revenue loss due to misreporting by Advantageous misreporters is even larger, at around USD 577 million. On net, unilateral VAT misporting cost the Ugandan government around USD 446 million, or 4 percent of total tax revenue collected, during our data period. This occurs despite the characteristics of the

[4]To avoid false discrepancies, we allow for firms aggregating individual transactions in a given month; errors in the reported transaction month; and rounding errors.

[5]This is because we classify firms whose reporting errors happen to decrease their liabilities as Advantageous misreporters. We show in the Appendix that potential underreporting of sales to final consumers (which cannot be cross-checked) has little impact on the proportion of firms classified as Disavantageous misreporters.

VAT that were thought to make the tax self-enforcing.

In the third and final part of our analysis, we show that Advantageous misreporters respond to enhanced tax authority capacity by misreporting less, while Disadvantageous misreporters do not. To do so, we take advantage of the fact that imported goods pass through customs and hence are "automatically" observed by authorities, making tax evasion riskier (Riezman & Slemrod, 1987; Keen & Lighart, 2002; Emran & Stiglitz, 2005; Keen & Lighart, 2005; Baunsgaard & Keen, 2010; Cagé & Gadenne, 2018). To generate exogenous variation in how exposed transactions involving a given firm are to oversight, we interact real exchange rate shocks with baseline import shares at the firm-country of origin level, akin to Bastos *et al.* (2018). This allows us to estimate how month-to-month changes in the share of a firm's initial inputs that are imported (versus purchased domestically) affect the firm's misreporting in *domestic* transactions. We find that a one standard deviation increase in import share leads to a 16 percent decrease in the firm's seller shortfall. This response is entirely driven by firms classified as Advantageous misreporters.

Overall, our findings suggest that, in Uganda, the majority of firms are sophisticated enough to respond to weak tax enforcement by considerably underreporting their tax liability, as conventional models of firm behavior assume. Such firms also respond to higher monitoring "rationally"– that is, by evading less. However, a non-negligible proportion—almost one third—of Ugandan firms consistently make costly tax reporting errors. Such "confused" firms also do not change their behavior when tax authority oversight increases.

This paper contributes to three related but distinct strands of literature on firm behavior and taxation. First, we provide what to our knowledge are the first direct estimates of the extent of strategic behavior-vs.-errors among firms. We can do so because the methodology we develop allows us to classify individual firms' behavior as self-advantageous or not, and because we observe the behavior of the entire population of formal, non-micro firms in Uganda's economy. Our analysis builds on an emerging body of evidence of seemingly erroneous behavior among firms (see among others DellaVigna & Gentzkow, 2017; Tourek, 2018; Kremer *et al.* , 2019).

Second, we provide new evidence on how tax evasion in a low-income country responds to the state's enforcement capacity, and in particular how firms characterized by different degrees of sophistication respond. In this sense, our analysis builds most closely—methodologically and thematically—on Fisman & Wei (2004)'s mirror data approach to measuring how tariff evasion responds to the size of the tariff. We also build on existing studies of more-vs.-less attentive taxpayers' response to tax rates (Chetty & Looney, 2009; Akcigit *et al.* , 2018; Rees-Jones & Taubinsky, 2018).[6] However, our focus is on variation in enforcement capacity, linking this paper with existing work on the causes and consequences of state capacity (Besley & Persson, 2009, 2010; Acemoglu *et al.* , 2015; Best *et al.* , 2018; Page & Pande, 2018).

Finally, we show that the VAT is far from self-enforcing in low state capacity settings. Building

[6] Chetty & Looney (2009); Akcigit *et al.* (2018); Benzarti (2018); Gillitzer & Skov (2018); Rees-Jones & Taubinsky (2018) provide direct evidence of tax-reporting mistakes by *individuals*. Like this paper, Akcigit *et al.* (2018) show evidence that more sophisticated taxpayers tend to react as theory predicts to tax incentives, while less sophisticated taxpayers do not.

on a recent body of work studying how policy should be tailored to context (Laffont, 2005; Best *et al.* , 2015, 2018; Duflo *et al.* , 2018; Hansman *et al.* , 2019), our analysis—especially in combination with other evidence that third-party reporting may not in itself generate tax compliance (Carrillo *et al.* , 2017; Almunia & Lopez-Rodriguez, 2018; Waseem, 2018b)—qualifies the common argument that developing countries are especially likely to benefit from use of the VAT (see, e.g., Bird & Gendron, 2007).[7] The massive magnitude of the revenue loss from VAT evasion we document in Uganda—and the corresponding cross-country patterns documented by Cagé & Gadenne (2018)—suggest that the production efficiency benefits of VATs relative to tariffs are at least in part offset by capacity-constrained governments' ability to raise revenues on international versus domestic transactions.

2 Background

2.1 The Value-added Tax (VAT) in Uganda

Uganda's tax-to-GDP ratio, at 13 percent in 2016, is below the African and OECD averages of 18 and 34 percent (OECD, 2018), while the ratio of its tax administration costs to tax revenues (2.4 percent) is comparable to the average in other low-income countries (IMF, 2013; Lemgruber *et al.* , 2015).

The VAT was introduced in 1996 and in 2016 contributed 32 percent of Uganda's total tax revenue (excluding revenue from tariffs), similar to elsewhere in Africa (OECD, 2018). The Ugandan VAT follows a relatively standard design with a general rate of 18 percent, a credit-invoice system and standard exemptions (e.g., financial services) and zero-rating (e.g., exports). See Appendix A.1 for more details.

Since 2012 all Ugandan VAT firms must file their monthly VAT declarations electronically, within 15 days after the transaction month ended.[8] As a result, the Uganda Revenue Authority (URA) has detailed data in electronic format for all VAT firms in recent years. Additionally, VAT firms are required to submit detailed transaction-level records—spreadsheets listing each sale and purchase to/from other VAT-registered firms. This implies that, every month, the URA receives two reports for each transaction between any two VAT firms—one from the seller and one from the buyer.

[7] Tax evasion research has demonstrated the importance of third-party reporting in developed countries (Slemrod *et al.* , 2001; Kleven *et al.* , 2011; Kleven, 2014), but also its limitations (Slemrod *et al.* , 2017; Almunia & Lopez-Rodriguez, 2018). The existing literature shows that in middle-income countries whose enforcement capacity significantly exceeds Uganda's—Brazil, Chile, Ecuador, India, and Pakistan—authorities' ability to cross-check VAT records tends to reduce evasion (Ebrill *et al.* , 2001; Pomeranz, 2015; Carrillo *et al.* , 2017; Mittal & Mahajan, 2017; Waseem, 2018a).

[8] In the data, about 80 percent of VAT returns are reported within 15 days of the return month and another 9 percent within the next month.

2.2 Data

Our analysis exploits the complete administrative data from VAT-registered firms' monthly electronic declarations between fiscal years 2013 and 2016.[9] The monthly VAT data contains information at the firm level, including a scrambled firm Tax Identification Number (TIN), the declaration date, total sales (amount and the corresponding VAT charged), total purchases (amount and the corresponding VAT paid), and total VAT liabilities. The tax return also contains data from the spreadsheets detailing each transaction, including the transaction date, the seller and buyer TINs, transaction value, and the VAT charged or paid. Sales to final consumers and to non-VAT firms are recorded only as monthly aggregates. Importantly, the transaction-level information reported in the VAT schedules is consistent with the aggregate data.[10] This suggests that the transaction-level records constitute meaningful paper trails for firms' VAT declarations and liabilities.

Our dataset contains 22,388 unique VAT-registered firms submitting at least one monthly VAT return between 2013 and 2016, and the transactions data cover 15,569 sellers and 19,421 buyers, leading to 3,373,183 seller-buyer-month observations.[11]

The data on imports comes from customs declarations submitted to the URA between 2012 and 2016. These declarations are transaction-specific and submitted electronically. The data includes information on the number and type of items imported, the value of the goods, and the date of import. The TIN of the importer allows us to directly match the customs data to the domestic transactions. 8,672 VAT-registered firms import at least once. Further data description can be found in Appendix A.1.

3 Discrepancies in VAT Declarations

3.1 Conceptual background

For a transaction between a seller s and a buyer b on date j, let y^S_{sbj} and y^B_{sbj} denote the output VAT charged (as reported by the seller s) and the input VAT paid (as reported by the buyer b). In the analysis, we aggregate transactions at the monthly level, so we define $Y^S_{sbt} \equiv \sum_{j \in J_t} y^S_{sbj}$ and $Y^B_{sbt} \equiv \sum_{j \in J_t} y^B_{sbj}$, where t denotes the transaction month. We define "seller shortfall" as a situation in which the total VAT charged is *lower* than the total VAT paid, i.e., $Y^S_{sbt} < Y^B_{sbt}$, and "buyer shortfall" as $Y^S_{sbt} > Y^B_{sbt}$.

Seller shortfall may be due to the seller underreporting output VAT or the buyer overreporting input VAT (or both). In either case, it implies a potential revenue loss for the government, as the reported tax liability is lower than the true liability. Symmetrically, buyer shortfall may be due to the seller overreporting output VAT or the buyer underreporting input VAT (or both), which implies a potential revenue *gain* for the government. Other things equal, buyer shortfall thus

[9] We refer to fiscal year 2013/14 as 2013.

[10] The aggregate firm-level VAT data match the sum of individual transaction data for 97 percent of the cases.

[11] Out of 22,388 firms, 19,161 have non-missing firm-as-buyer and/or firm-as-seller fixed effect estimated as described in Section 4 and therefore make up our main sample of analysis.

points towards errors by firms.

However, it might be rational for buyers to understate their purchases if they are simultaneously understating their sales, e.g. because this allows them to report a less suspicious (say, non-negative) VAT liability. Carrillo *et al.* (2017) provide evidence of such "looking small" behavior in Ecuador. In theory, buyer shortfall cases could also be due to sellers engaging in liability-reducing "looking big" behavior by overstating both their purchases and sales while underreporting their value added. These observations underscore that transaction-pair level discrepancies do not in themselves allow us to distinguish between sophisticated, self-advantageous tax evasion and systematic reporting errors.

3.2 Discrepancies

In this subsection, we quantify the VAT reporting discrepancies in Uganda at the seller-buyer-month level for the 2013-2016 period. The average monthly reported VAT liability is slightly negative, and the median is zero, as is common in developing countries (Lemgruber *et al.* , 2015; Pomeranz, 2015). 56 percent of monthly VAT returns report a nil or negative tax liability.

Table 1 displays the average proportion of *firms* that in a full fiscal year report respectively a positive, zero, and negative (i) value-added and (ii) VAT liability. The reported value-added is negative or zero for around 15 percent of firms. However, the reported VAT liability is zero or negative for 45 percent of firms. This number varies from 44 percent among smaller VAT-registered firms to 50 percent among the largest firms.[12]

Figure 1 provides a graphical illustration of the discrepancies between seller and buyer-declared amounts. The vertical axis measures the inverse hyperbolic sine transformation of the total monthly amounts declared by sellers, and the horizontal axis that of the total monthly amounts declared by buyers. Each dot represents a seller-buyer-month observation and the solid straight line is the 45-degree line. Sellers and buyers report the same amount in 21 percent of the observations. We observe seller shortfall in 47 percent of the cases and buyer shortfall in the remaining 32 percent.

We observe these widespread discrepancies despite taking a number of steps to avoid detecting false discrepancies. First, we use transaction dates (the month in which a transaction took place) rather than filing dates to account for the fact that sellers and buyers may not file a given transaction in the same month. Second, we minimize mismatched transactions by (i) identifying discrepancies using firms' aggregate monthly records rather than individual transactions and (ii) not labelling as discrepancies cases where the seller and buyer declare the same amount with only a one or two-month lag. Finally, we allow for a rounding of 1,000 Ugandan Shillings (about USD 0.30).

The dashed curve in Figure 1 is a binned scatter plot showing the average amounts reported by sellers for different values of the buyer-reported amounts. The curve lies systematically below the

[12] Zero or negative reported VAT liabilities occur in combination with positive reported value-added because the VAT liability includes offsets carried over from the previous month. Since it is difficult to claim refunds, many firms choose to carry-over offsets instead.

45-degree line, implying that seller shortfall is quantitatively more important than buyer shortfall in aggregate terms. Additionally, the distance to the 45-degree line appears to be increasing with the transaction amount, suggesting that the fraction of the transaction amount being unreported gets relatively larger as the amount of transaction increases.

4 Classifying Firms' Reporting Behavior

In this section we evaluate the extent to which Ugandan firms underreport their value-added—sales minus purchases—such that their liability falls, as opposed to making reporting errors that increase the firm's liability. To do so we first estimate what share of each discrepancy is due to each of the two firms.

4.1 Assigning the blame: two-way fixed-effects analysis

We allocate a share of the responsibility for each discrepancy to the seller and the buyer based on the aggregate reporting accuracy of each firm in all their transactions, i.e., across all periods and with all trading partners. The starting point is a two-way fixed-effects model inspired by Abowd *et al.* (1999, 2002). We define the discrepancy between buyer f, and seller f' in month t as $d_{ff't} \equiv Y^B_{ff't} - Y^S_{ff't}$, such that $d_{ff't} > 0$ implies seller shortfall and $d_{ff't} < 0$ implies buyer shortfall. Then, we estimate the following regression:

$$d_{ff't} = \delta^b_f + \delta^s_{f'} + \delta_t + r_{ff't}, \tag{1}$$

where δ^b_f and $\delta^s_{f'}$ denote buyer and seller fixed effects (defined at the firm level), respectively; δ_t is a month fixed effect; and $r_{ff't}$ is an error term. Since $d_{ff't}$ is the nominal value of the discrepancy, δ^b_f (respectively $\delta^s_{f'}$) can be interpreted as a firm's average contribution to discrepancies as a buyer (seller), in monetary terms.[13]

As shown in Abowd *et al.* (1999, 2002), the two-dimensional fixed effects are separately identified only within a "connected set" of firms, which in our context refers to firm-pairs that are linked by transaction and all of such firms' trade-partners. The largest connected set observed during our 2013-2016 data period covers over 99 percent of all observations, 90 percent of sellers, and 94 percent of buyers. Following the existing literature, we thus restrict our analysis to this largest connected set of firms. Appendix A.2 provides technical details on the two-way fixed-effects estimation and the firm classification procedure that follows in the next subsection.

[13] In Appendix A.2 we show results from running (1) with various controls included. These are generally very similar.

4.2 Firm-level reporting behavior

We next construct a firm-level discrepancy measure, Q_f, adding up the two estimated fixed effects for f weighted by the relative values of its sales (Y_f^s) and purchases (Y_f^b) over the study period:

$$Q_f \equiv \left(\frac{Y_f^s}{Y_f^s + Y_f^b}\right)\hat{\delta}_f^s + \left(\frac{Y_f^b}{Y_f^s + Y_f^b}\right)\hat{\delta}_f^b. \tag{2}$$

With the Q_f measure in hand, we can formalize our classification of a firm's aggregate reporting behavior. A firm engages in *Advantageous* misreporting behavior if $Q_f > 0$, meaning that it reports, at the aggregate level, in a way that *reduces* its VAT liability. Symmetrically, a firm engages in *Disadvantageous* misreporting behavior if $Q_f < 0$, which implies that it reports in a way that *increases* its overall VAT liability.

We further classify Advantageous misreporters into three subcategories. First, a firm engaging in *Conspicuous* advantageous behavior is one for which $\hat{\delta}_f^s \geq 0$ and $\hat{\delta}_f^b \geq 0$. This implies that the firm both underreports its sales and overreports its purchases, and hence appears not to be concerned with hiding its tax evasion from the tax authorities. Second, a firm engaging in *Looking-small* advantageous behavior is one for which $\hat{\delta}_f^s \geq 0$ and $\hat{\delta}_f^b < 0$. This implies that the firm underreports its sales and underreports its purchases. Finally, a firm engaging in *Looking-big* advantageous behavior is one in which $\hat{\delta}_f^s < 0$ and $\hat{\delta}_f^b \geq 0$. This implies that the firm overreports its sales and overreports its purchases.

The results are shown in the first row of Table 2. We find that only 85 out of 19,161 Ugandan VAT-eligible firms report consistently on average, while 13,528 or 71 percent are *Advantageous* misreporters. These estimates suggest that when the VAT is implemented in a low-state capacity context, where systematic cross-checks appear not to occur, the majority of firms misreport so as to lower their VAT liability.

However, we also find that 5,548 or 29 percent of firms misreport in a *Disadvantageous* way. A substantial share of firms thus simply make reporting errors. This result underscores the importance of accounting for heterogeneity in firm sophistication in theory and policy design. It also foreshadows our results in Section 5, where we study how firms respond to increased oversight by the tax authority.

Of the 13,528 Ugandan firms that misreport in an advantageous way, 77 percent are Conspicuous advantageous misreporters, only 2 percent are Looking-small advantageous misreporters, and the remaining 21 percent are Looking-big advantageous misreporters. The high proportion of Conspicuous advantageous misreporters suggests that the majority of Ugandan firms believe that the tax authority is unlikely to detect evasion by monitoring firms' reported value-added.[14]

As seen in Figure 2, the average Q_f measure is similarly distributed across most of the distribution of firm size, suggesting that advantageous and disadvantageous misreporting occurs with

[14]The more surprising set of firms engaging in Looking-big behavior may for example be due to such firms believing that the tax authority pays more attention to small than big firms (see e.g. Amodio *et al.* , 2019), or misclassifying sales to final consumers as sales to other VAT-eligible firms, while simultaneously overreporting their purchases.

comparable frequency among smaller, medium-sized, and somewhat larger VAT-registered firms. However, Figure 2 also shows that the average Q_f measure markedly increases among the largest firms, suggesting that the largest firms are more sophisticated tax (mis)reporters than other firms. A more detailed comparison of the observable characteristics of the two types of firms is in Appendix A.2.

4.3 Robustness analysis

Our methodology, starting with the two-way fixed effects regression (1) and thereafter classifying firm types using the resulting Q_f measure, is one particular way to characterize Ugandan firms based on "missing" transactions. While allowing us to shed new light on the sophistication of firms' decisionmaking, our method has potential limitations.

First, we cannot distinguish truthful reporting from collusive evasion in which both seller and buyer misreport the transaction by the same amount. This implies that we may underestimate the total extent of VAT evasion in Uganda.

Second, our definition of disadvantageous behavior may be too conservative, as some errors may lower the firm's VAT liability, leading us to underestimate the true extent of firm errors. One way to shed light on the frequency of such "idiosyncratic but advantageous" errors relative to systematic flaws in a firm's tax reporting is to investigate the consistency of firm behavior over time. Doing so also allows us to probe the prediction accuracy of our classification method. We thus re-do the estimation of (1) and the classification of firms via (2) separately for each year in our sample. As shown in Appendix A.2, we find that 77 (resp., 62) percent of firms classified as Advantageous (Disadvantageous) misreporters in year t stay within that classification also in the subsequent year. This suggests that we primarily capture *systematic* components of firms' reporting behavior. It also suggests that disadvantageous behavior not surprisingly is somewhat less consistent over time than advantageous behavior.

A third limitation of our methodology is that we do not observe the level of misreporting of sales to final consumers. If firms engaging in buyer shortfall underreport a large enough share of sales to final consumers, their total misreporting may in principle be advantageous.[15] To address this concern, we repeat the original estimation assuming that all firms underreport a fixed proportion of their sales to final consumers. As seen in Appendix Table A.7, the proportion of Advantageous firms increases to 75 percent when we assume that all firms underreport final sales by 10 percent. Even assuming an extreme and arguably implausible degree of misreporting of sales to final consumers—40 percent—the share of Disadvantageous firms is about 16 percent.[16]

[15] If firms are underreporting their sales to final consumers, it might be rational to also underreport their input purchases to follow a Looking-small strategy.

[16] Assuming that the entire VAT compliance gap estimated for Uganda is due to evasion on sales to final consumers—which this paper shows is far from the case—would imply that firms misreport sales to final consumers by 50 percent (IMF, 2014).

4.4 Revenue consequences

We documented in Sub-section 3.2 that Ugandan sellers and buyers report different values in 79 percent of VAT transactions, and that 40 percent of such mismatch transactions involve a "buyer shortfall" and 60 percent a "seller shortfall". This suggests that, in aggregate, the revenue consequences of VAT misreporting for the Ugandan government—which gets about 32 percent of its total tax revenue from the VAT—are likely adverse and potentially large, but also that there may be significant positive revenue consequences from the observed disadvantageous misreporting.

An increased (or decreased) liability attributed to one firm may have different revenue consequences from one attributed to the other firm involved in a given transaction (see Appendix A.3). To proceed, we thus divide up the "blame" for a given reporting discrepancy $d_{ff't}$ using the relevant seller and buyer fixed effect estimated in Sub-section 4.1. If the two fixed effects have the same sign, we assign shares of the discrepancy proportionally, based on the relative size of each firm's fixed effect. If, instead, the two fixed effects have opposite signs, we assign the entire discrepancy to the firm whose fixed effect matches the sign of the discrepancy. Details are in Appendix A.3.

The results are reported in the bottom rows of Table 2. Our estimates imply that the Ugandan government would have lost USD 131 million in tax revenues during 2013-2016 if (only) disadvantageous misreporting were eliminated. If (only) advantageous misreporting were eliminated, our estimates imply a revenue *gain* of about USD 577 million. If both forms of misreporting were eliminated, our estimates imply a revenue gain of USD 446 million, or about 33 percent of the total VAT collected.[17] These estimated revenue consequences are very similar if we use an alternative way to apportion discrepancies based on the estimated fixed effects, and also if we naively assume that all instances of seller shortfall are entirely due to sellers and all instances of buyer shortfall due to buyers, as shown in Appendix A.3.

5 Enhanced Tax Authority Capacity and VAT Evasion by Strategic and Confused Firms

In this section we study how Ugandan firms change their reporting behavior when the tax authority's capacity is enhanced. To do so we leverage the fact that imported goods are subject to greater oversight at customs, making tax evasion riskier than in domestic transactions (Riezman & Slemrod, 1987; Keen & Lighart, 2002, 2005; Emran & Stiglitz, 2005; Baunsgaard & Keen, 2010; Cagé & Gadenne, 2018). We use exchange rate variation to construct an instrument that shifts firms into or out of importing. We then compare the reporting response to higher tax authority oversight of firms that are sophisticated enough to systematically underreport their tax liability under mini-

[17] The fact that many Ugandan firms have positive outstanding balances with the URA helps explain why the revenue consequences of eliminating disadvantageous misreporting are smaller (in absolute value) relative to those of eliminating advantageous misreporting than the estimated relative size of the magnitude of the two forms of misreporting themselves. This, in combination with the correlation between individual firms' buyer and seller shortfalls (see Sub-section 4.2), also helps explain why the revenue gain from eliminating all VAT misreporting is smaller than the sum of the revenue gain from eliminating respectively disadvantageous and advantageous misreporting.

mal tax enforcement in the domestic economy and that of firms whose misreporting in domestic transactions appears to be due to systematic errors.

5.1 Exchange rate fluctuations and imports

To account for endogeneity in firms' importing decisions, we instrument for the proportion of a firm's inputs that are imported by interacting the firm's baseline import shares from different countries and fluctuations in monthly real exchange rates (RER), following Bastos *et al.* (2018). To maximize power in the first stage, we restrict the list of countries of origin to the 10 countries from which Ugandan firms as a whole import the most in 2012 (baseline year). Our first stage specification is:

$$ImportShare_{it} = \sum_{c=1}^{10} \beta_c log(RER)_{ct} * S_{ic} + \beta_{11} sales_{it} + \beta_{12} inputs_{it} + \gamma_i + \gamma_t + \epsilon_{it}, \tag{3}$$

where $ImportShare_{it}$ is the share of inputs that firm i imports from any country of origin in month t, $log(RER)_{ct}$ is the log of the RER between the Ugandan shilling and the currency of country c in month t, and S_{ic} is the share of inputs that i imports from country c in 2012. γ_i and γ_t are firm and year-month fixed effects, and $sales_{it}$ and $inputs_{it}$ dummies that control for respectively the firm's sales and inputs decile.

The first stage results are reported in Table A.10 in Appendix A.4. In both the full sample and the two subsamples consisting of Advantageous and Disadvantageous misreporters, the first stage analysis generates large Kleibergen-Paap rk Wald F statistics and estimates of β_1 to β_{10} are all negative, as expected, and significant for the most part.

5.2 Results

In the second stage we regress the domestic VAT misreporting that our estimates from Section 4 indicate that (potentially) importing firm i is responsible for in month t and that is costly to the tax authority—$SellerShortfall_{it}$—on the instrumented share of its initial inputs that are imported and hence exposed to enhanced oversight $\widehat{ImportShare}_{it}$.[18] We run:

$$SellerShortfall_{it} = \delta_1 \widehat{ImportShare}_{it} + \delta_2 sales_{it} + \delta_3 inputs_{it} + \gamma_i + \gamma_t + \epsilon_{it}. \tag{4}$$

The results are shown in Table 3. In the IV specifications shown in columns 1-3, we see that a one standard deviation increase in the share of imported inputs leads to a decrease in seller short-

[18] The shortfall amounts are computed using only domestic transactions so that, with controls for the firm's total level of sales and inputs included, we avoid any mechanical or circular relationship between the share of imports and shortfall amounts. Specifically, we add up the shares that firm i is estimated to be responsible for of all discrepancies the firm is involved in during the relevant month that display a seller shortfall, the form of VAT misreporting that in general reduces the firm's tax liability and revenues collected by the URA. To do so we use the firm's two estimated fixed effects from Section 4 and those of its domestic trade partners. $SellerShortfall_{it}$ is the inverse hyperbolic sine transformation of total seller shortfall amounts for firm i in month t.

fall of 16.2 percent in the full sample and 21.3 percent in the sample of Advantageous misreporters. The effect is close to zero in magnitude and not statistically significant for Disadvantageous misreporters. In Appendix A.4, we show that these results are robust to various alternative approaches.

In sum, our results in this section suggest that sophisticated firms—those that tend to respond to minimal general tax enforcement in domestic trade by underreporting their tax liability—also strategically respond to enhanced monitoring by misreporting less. On the other hand, "confused" firms—firms that consistently make reporting errors—also appear to respond less to variation in the state's enforcement capacity.

6 Conclusion

In this paper we study the sophistication of firms' decisionmaking, using tax reporting behavior in a low enforcement setting—Uganda—as a lab for analyzing the extent to which firms make decisions that appear to benefit themselves. Exploiting transaction-level data from VAT returns, we document widespread discrepancies between seller and buyer reports, with dramatic consequences for tax revenue collected. By comparing a given firm's misreporting of sales and purchases over time, we show that, while a majority of firms misreport in a way that reduces their tax liability, a non-negligible fraction—29 percent—misreports such that the firm's tax liability increases. We also show that when exchange-rate variation induces firms to import a higher share of their imports, implying stricter oversight, firms classified as self-advantageous misreporters choose to misreport less in onward trade, while firms whose misreporting appears to be due to errors do not. We interpret our findings as indicating that (i) the proportion of firms that appear not to engage in sophisticated optimization as usually assumed is high—which has important implications for theory and policy—but (ii) the majority of firms nevertheless respond to low state capacity by evading taxes, as traditional economic theory predicts.

References

ABOWD, JOHN M., KRAMARZ, FRANCIS, & MARGOLIS, DAVID N. 1999. High Wage Workers and High Wage Firms. *Econometrica*, **67**(2), 251–333.

ABOWD, JOHN M, CREECY, R, & KRAMARZ, FRANCIS. 2002. *Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data*. Tech. rept. US Census Bureau, Washington DC.

ACEMOGLU, DARON, GARCIA-JIMENO, CAMILO, & ROBINSON, JAMES A. 2015. State Capacity and Economic Development: A Network Approach. *American Economic Review*.

AKCIGIT, UFUK, AGHION, PHILIPPE, LEQUIEN, MATHIEU, & STANTCHEVA, STEFANIE. 2018. *Tax Simplicity and Heterogeneous Learning*. NBER Working Paper 24049.

ALMUNIA, MIGUEL, & LOPEZ-RODRIGUEZ, DAVID. 2018. Under the Radar: The Effects of Monitoring Firms on Tax Compliance. *American Economic Journal: Economic Policy*, **10**(1), 1–38.

ALMUNIA, MIGUEL, GERARD, FRANCOIS, HJORT, JONAS, KNEBELMANN, JUSTINE, NAKYAMBADDE, DOROTHY, RAISARO, CLAUDE, & TIAN, LIN. 2017. *An Analysis of Discrepancies in Tax Declarations submitted Under Value-Added Tax in Uganda*. International Growth Centre Project Report.

AMODIO, FRANCESCO, CHOI, JIEUN, DE GIORGI, GIACOMO, & RAHMAN, AMINUR. 2019. *Bribes vs. Taxes: Market Structure and Incentives*. Mimeo McGill.

BASTOS, PAULO, SILVA, JOANA, & VERHOOGEN, ERIC. 2018. Export Destinations and Input Prices. *American Economic Review*, **108**(2), 353–92.

BAUNSGAARD, THOMAS, & KEEN, MICHAEL. 2010. Tax revenue and (or?) trade liberalization. *Journal of Public Economics*, **94**(9-10), 563–577.

BENZARTI, YOUSSEF. 2018. How Taxing Is Tax Filing? Using Revealed Preferences to Estimate Compliance Costs. *National Tax Journal*.

BERNHEIM, DOUGLAS, DELLAVIGNA, STEFANO, & LAIBSON, DAVID. 2019. *Handbook of Behavioral Economics, Vol. 2*. Elsevier.

BESLEY, T., & PERSSON, T. 2010. State capacity, conflict and development. *Econometrica*, **78**, 1–34.

BESLEY, TIMOTHY, & PERSSON, TORSTEN. 2009. The origins of state capacity: Property rights, taxation, and politics. *American Economic Review*, **99**, 1218–44.

BESLEY, TIMOTHY, & PERSSON, TORSTEN. 2013. Taxation and Development. *Pages 51 – 110 of:* ALAN J. AUERBACH, RAJ CHETTY, MARTIN FELDSTEIN, & SAEZ, EMMANUEL (eds), *handbook of public economics, vol. 5*. Handbook of Public Economics, vol. 5. Elsevier.

BEST, MICHAEL, BROCKMEYER, ANNE, KLEVEN, HENRIK, SPINNEWIJN, JOHANNES, & WASEEM, MAZHAR. 2015. Production versus Revenue Efficiency with Limited Tax Capacity: Theory and Evidence from Pakistan. *Journal of Political Economy*, **123**(6), 1311–1355.

BEST, MICHAEL CARLOS, HJORT, JONAS, & SZAKONYI, DAVID. 2018. *Individuals and Organizations as Sources of State Effectiveness*. NBER Working Paper 23350.

BIRD, RICHARD, & GENDRON, PIERRE-PASCAL. 2007. *The VAT in Developing and Transitional Countries*. Cambridge and New York: Cambridge University Press.

CAGÉ, JULIA, & GADENNE, LUCIE. 2018. Tax Revenues, Development, and the Fiscal Cost of Trade Liberalization, 1792-2006. *Explorations in Economic History*, **70**(C), 1–24.

CARD, DAVID, HEINING, JORG, & KLINE, PATRICK. 2013. Workplace Heterogeneity and the Rise of West German Wage Inequality. *The Quarterly Journal of Economics*, **128**(3), 967–1015.

CARRILLO, PAUL, POMERANZ, DINA, & SINGHAL, MONICA. 2017. Dodging the taxman: Firm

misreporting and limits to tax enforcement. *American Economic Journal: Applied Economics*, **9**(2), 144–164.

CHETTY, RAJ KORY KROFT, & LOONEY, ADAM. 2009. Salience and Taxation: Theory and Evidence. *American Economic Review*.

DELLAVIGNA, STEFANO, & GENTZKOW, MATTHEW. 2017. *Uniform Pricing in US Retail Chains*. NBER Working Paper No. 23996.

DUFLO, ESTHER, GREENSTONE, MICHAEL, PANDE, ROHINI, & RYAN, NICHOLAS. 2018. The Value of Discretion in the Enforcement of Regulation: Experimental Evidence and Structural Estimates from Environmental Inspections in India. *Econometrica*.

EBRILL, LIAM, BODIN, JEAN-PAUL, KEEN, MICHAEL, & SUMMERS, VICTORIA. 2001. *The Modern VAT*. Washington: International Monetary Fund.

ELECTORAL COMMISION. 2016. *The 2016 General Elections Statistics*. (link) Accessed on 23 April 2019.

EMRAN, M. SHAHE, & STIGLITZ, JOSEPH E. 2005. On Selective Indirect Tax Reform in Developing Countries. *Journal of Public Economics*, **89**(4), 599–623.

FISMAN, RAYMOND, & WEI, SHANG-JIN. 2004. Tax Rates and Tax Evasion: Evidence from "Missing Imports" in China. *Journal of Political Economy*, **112**(2), 471–496.

GILLITZER, CHRISTIAN, & SKOV, PEER EBBESEN. 2018. The use of third-party information reporting for tax deductions: evidence and implications from charitable deductions in Denmark. *Oxford Economic Papers*, **70**(3), 892–916.

HANSMAN, CHRISTOPHER, HJORT, JONAS, & LEÓN, GIANMARCO. 2019. Interlinked Firms and the Consequences of Piecemeal Regulation. *Journal of the European Economic Association*.

IMF. 2013. *Uganda - Staff Report for 2013 Article IV Consultation*. International Monetary Fund Country Report, volume 13/215.

IMF. 2014. Revenue Administration Gap Analysis Program-The Value-Added Tax Gap Uganda.

KEEN, MICHAEL, & LIGHART, JENNY. 2002. Coordinating Tariff Reduction and Domestic Tax Reform. *Journal of International Economics*, **56**(2), 489–507.

KEEN, MICHAEL, & LIGHART, JENNY. 2005. Coordinating Tariff Reduction and Domestic Tax Reform under Imperfect Competition. *Review of International Economics*, **13**(2), 385–390.

KLEVEN, HENRI. 2014. How can Scandinavians tax so much? *Journal of Economic Perspectives*, **28**(4), 77–98.

KLEVEN, HENRIK, KNUDSEN, MARTIN, KREINER, CLAUS, PEDERSEN, SOREN, & SAEZ, EMMANUEL. 2011. Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark. *Econometrica*, **79**(3), 651–692.

KLEVEN, HENRIK, KREINER, CLAUS, & SAEZ, EMMANUEL. 2016. Why Can Modern Governments Tax So Much? An Agency Model of Firms as Fiscal Intermediaries. *Economica*, **83**, 219–246.

KOPCZUK, WOJCIECH, & SLEMROD, JOEL. 2006. Putting Firms into Optimal Tax Theory. *American Economic Review, Papers and Proceedings*, **96**, 130–134.

KREMER, MICHAEL, RAO, GAUTAM, & SCHILBACH, FRANK. 2019. Behavioral Development Economics. *In:* BERNHEIM, DOUGLAS, DELLAVIGNA, STEFANO, & LAIBSON, DAVID (eds), *Handbook of Behavioral Economics*, vol. 2. Elsevier.

LAFFONT, JEAN-JAQUES. 2005. *Regulation and Development*. Cambridge University Press.

LEMGRUBER, ANDREA, MASTERS, ANDREW, & CLEARY, DUNCAN. 2015. *Understanding Revenue Administration: An Initial Data Analysis Using the Revenue Administration Fiscal Information Tool*. International Monetary Fund Fiscal Affairs Department Paper Series.

MITTAL, SHEKHAR, & MAHAJAN, APRAJIT. 2017. *VAT in Emerging Economies: Does Third Party Verification Matter*. Mimeo UCLA.

OECD. 2018. *Revenue Statistics in Africa 2018*. OECD Centre for Tax Policy and Administration, African Union Commission (AUC) and African Tax Administration Forum (ATAF).

PAGE, LUCY, & PANDE, ROHINI. 2018. Ending Global Poverty: Why Money Isn't Enough. *Journal of Economic Perspectives*, **32**, 173–200.

POMERANZ, DINA. 2015. No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax. *American Economic Review*, **105**(8), 2539–69.

REES-JONES, ALEX, & TAUBINSKY, DMITRY. 2018. Attention Variation and Welfare: Theory and Evidence from a Tax Salience Experiment. *Review of Economic Studies*.

RIEZMAN, RAYMOND, & SLEMROD, JOEL. 1987. Tariffs and collection costs. *Review of World Economics*, **123**(3), 545–549.

SLEMROD, JOEL, BLUMENTHAL, MARSHA, & CHRISTIAN, CHARLES. 2001. Taxpayer Response to an Increased Probability of Audit: Evidence from a Controlled Experiment in Minnesota. *Journal of Public Economics*, **79**, 455–483.

SLEMROD, JOEL, COLLINS, BRETT, HOOPES, JEFFREY, RECK, DANIEL, & SEBASTIANI, MICHAEL. 2017. Does Credit-card Information Reporting Improve Small-business Tax Compliance? *Journal of Public Economics*, **149**, 1–19.

TOUREK, GABRIEL. 2018. *Targeting in Tax Behavior: Evidence from Rwandan Firms*. Mimeo Harvard.

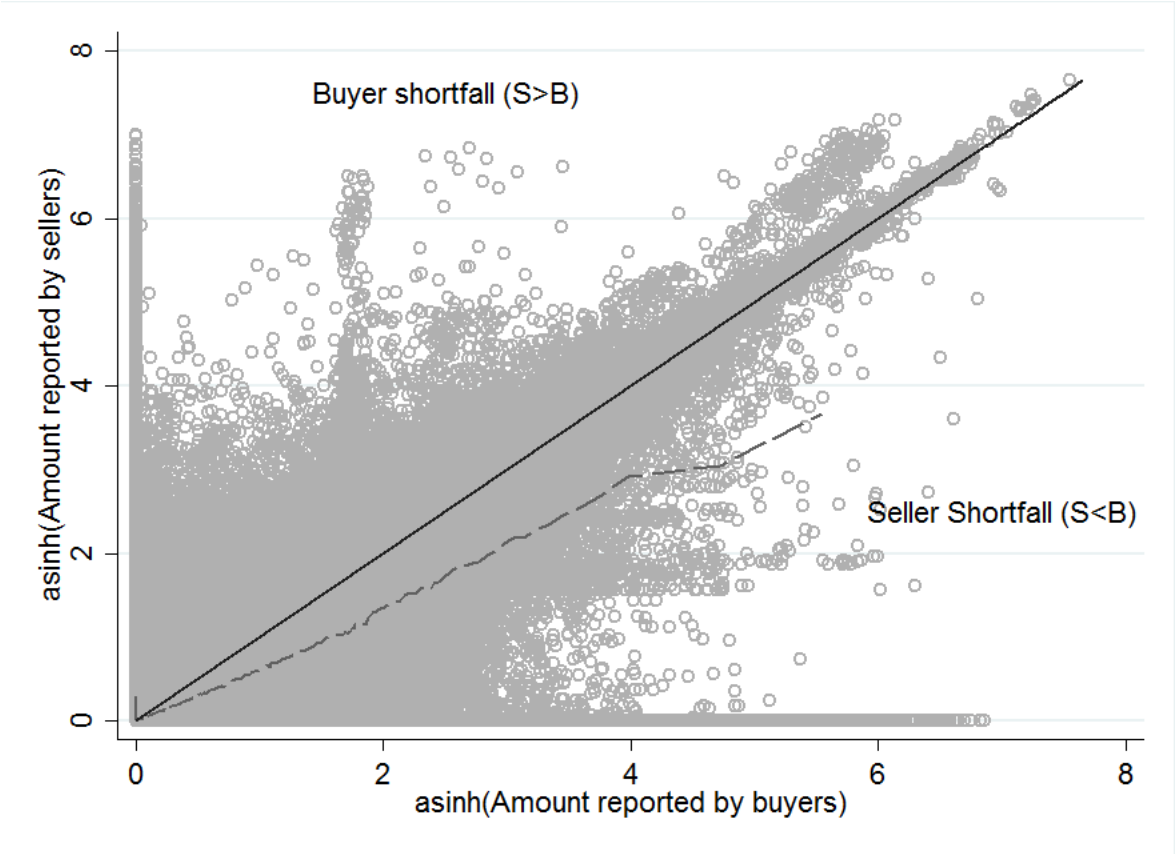
UGANDA REVENUE AUTHORITY. 2016. *Domestic Tax Laws of Uganda: Updated and Tracked Compendium*.

WASEEM, MAZHAR. 2018a. *Information, Asymmetric Incentives, Or Withholding? Understanding the Self-Enforcement of Value-Added*. Mimeo University of Manchester.

WASEEM, MAZHAR. 2018b. Taxes, informality and income shifting: Evidence from a recent Pakistani tax reform. *Journal of Public Economics*, **157**, 41 – 77.

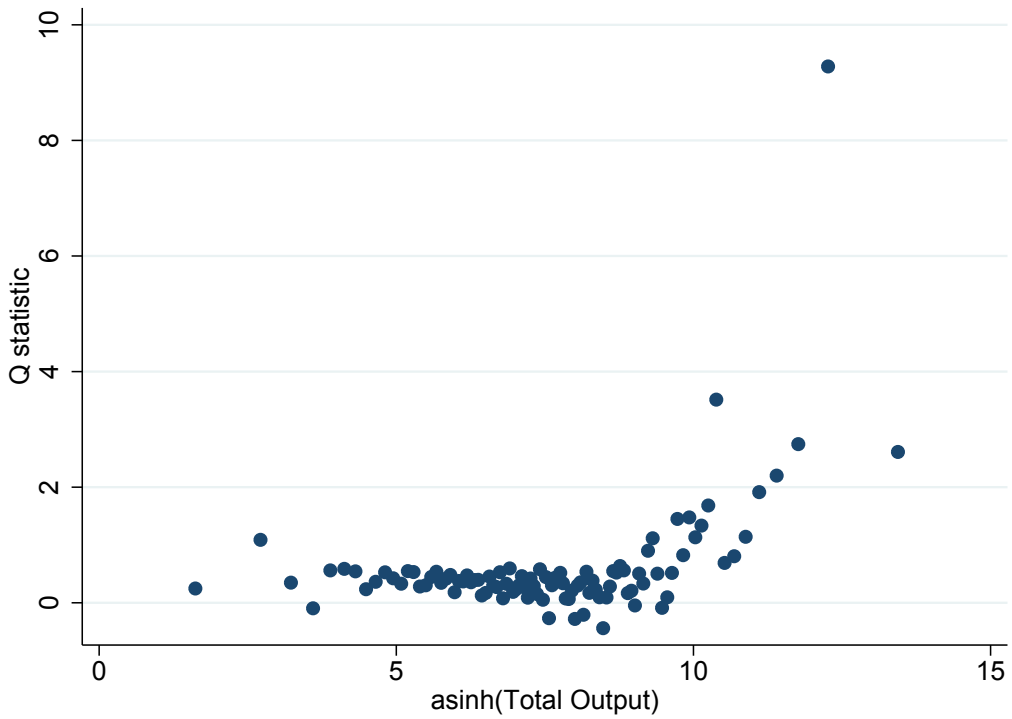
Figures

FIGURE 1
DOMESTIC VAT AMOUNTS DECLARED BY SELLERS VS BUYERS



Notes: In this figure, we plot, in gray circles, the inverse hyperbolic sine transformation of amounts reported by sellers over that by buyers for all transactions in fiscal years 2013-2016. The solid black line is the identity line, on which all observations would be if there were no reporting discrepancies. Points above the solid black line are cases of buyer shortfall; points below are cases of seller shortfall. The gray dashed line is a binned scatter plot of the inverse hyperbolic sine transformation of the amounts reported by sellers over that reported by buyers. Data source: VAT Schedules data for fiscal years 2013-2016.

FIGURE 2
Q STATISTIC OVER SIZE.



Notes: In this Figure, we plot a firms estimated Q statistic (Q^f in Equation (2)) over the inverse hyperbolic sine transformation of a firms total output in the estimation period. Data source: VAT Schedules and Monthly Summary data for fiscal years 2013-2016.

Tables

TABLE 1
DISTRIBUTION OF VALUE-ADDED AND VAT LIABILITY BY FIRM SIZE

		(1) Value added	(2) Output-Input VAT	(3) VAT liability
	Share > 0	84.33%	77.36%	48.26%
All VAT firms	Share $= 0$	5.12%	7.43%	6.47%
(N = 22,388)	Share < 0	10.55%	15.21%	45.27%
	Share > 0	93.08%	77.75%	48.64%
LTO firms	Share $= 0$	0.81%	0.77%	1.28%
(N = 738)	Share < 0	6.11%	21.49%	50.07%
	Share > 0	91.85%	79.94%	50.69%
MTO firms	Share $= 0$	0.71%	1.39%	1.41%
(N = 1,635)	Share < 0	7.43%	18.66%	47.91%
	Share > 0	82.82%	77.00%	47.92%
Other VAT firms	Share $= 0$	5.95%	8.62%	7.44%
(N = 20,015)	Share < 0	11.22%	14.39%	44.63%

Notes: Data source: VAT Monthly Summary data for fiscal years 2013-2016. Column (1) shows total value added over the fiscal year, including goods that are VAT-exempt. Column (2) shows the difference between total output VAT and total input VAT. Column (3) shows total tax liability over the fiscal year, taking into account VAT credits carried over from previous fiscal year (2012). Firms can display a positive Output-Input VAT, but a nil or negative VAT liability once offsets are subtracted. LTOs are firms with an annual turnover above 15 billion Ugandan Shillings (USD 4.1 million) and/or belonging to specific sectors such as oil and mining, banking, insurance, and government departments. MTOs are firms with a turnover above 2 billion Ugandan Shillings (USD 550,260, threshold increased to 5 billion Ugandan Shillings/USD 1.3 million in 2015). Other VAT firms refer to VAT-paying firms with an annual turnover lower than the MTO threshold.

TABLE 2
SUMMARY STATISTICS AND REVENUE CONSEQUENCES BY FIRM TYPE

	All	(1) Consist	(2) Disadv.	(3) Adv.	(3a) Conspic.	(3b) Looking Small	(3c) Looking Big
No. of distinct firms	19,161	85	5,548	13,528	10,371	345	2,812
Percentage of all firms	(100%)	(0%)	(29%)	(71%)	(54%)	(2%)	(15%)
Total net VAT due	1,554,101	531	864,525	689,045	439,360	49,896	199,789
Seller shortfall							
Number of distinct firms with seller shortfall	17,255	29	4,902	12,324	9,185	343	2,796
Total net VAT due from firms with seller shortfall	1,275,946	11	760,049	515,886	345,909	36,425	133,552
Total VAT subject to seller shortfall	900,099	57	101,680	798,362	455,863	175,719	166,779
Buyer shortfall							
Number of distinct firms with buyer shortfall	18,000	67	5,287	12,646	9,507	341	2,798
Total net VAT due from firms with buyer shortfall	1,316,829	236	798,553	518,039	345,640	38,634	133,765
Total VAT subject to buyer shortfall	727,373	649	528,417	198,307	65,996	48,720	83,591
Correcting seller shortfall and buyer shortfall							
Impact on total net VAT due	446,224	26	−130,753	576,950	359,323	131,119	86,508
Percentage of total VAT collected	32.8%	0.0%	−9.6%	42.4%	26.4%	9.6%	6.4%

Notes: Data Source: VAT Schedules and Monthly Summary data for fiscal years 2013-2016. We include firms with at least one non-missing fixed effect estimates. Revenue consequences are calculated by correcting the VAT liability in the last month of the year for the total VAT under seller shortfall and under buyer shortfall. *Definitions:* (1) Consistent: $Q(f) = 0$; (2) Disadvantageous: $Q(f) < 0$; and (3) Advantageous: $Q(f) > 0$. Under Advantages, firms are further categorized into: (3a) Conspicuous Advantageous: $\hat{\delta}_s(f) \geq 0$ and $\hat{\delta}_b(f) \geq 0$; (3b) Looking small Advantageous: $\hat{\delta}_s(f) \geq 0$ and $\hat{\delta}_b(f) < 0$; and (3c) Looking big Advantageous: $\hat{\delta}_s(f) < 0$ and $\hat{\delta}_b(f) \geq 0$. $Q(f)$ is calculated as a weighted average of the estimated firm-as-buyer fixed effect and firm-as-seller fixed effect, i.e., : $Q(f) = \hat{\delta}_b \times \frac{p}{p+s} + \hat{\delta}_s \times \frac{s}{p+s}$ where s denotes total sales and p stands for total purchases. All values are in thousands of USD.

TABLE 3
EFFECT OF ENHANCED TAX AUTHORITY OVERSIGHT ON VAT COMPLIANCE BY SOPHISTICATED AND CONFUSED FIRMS

	2SLS		
Dependent variable	$asinh(Seller\ shortfall\ amounts)$		
	(1)	(2)	(3)
Sample	Full	Advantageous	Disadvantageous
ImportShare	-0.558***	-0.772***	-0.081
	(0.150)	(0.175)	(0.272)
Sales decile	Yes	Yes	Yes
Inputs decile	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes
Month-Year FE	Yes	Yes	Yes
N	442519	314669	127850
Mean of dep.	0.90	1.03	0.57
Kleibergen-Paap LM stat.	344.261	286.455	74.104
Kleibergen-Paap Wald F stat.	53.101	46.381	11.443

Notes: Data Source: Customs, VAT Schedules and Monthly Summary data for fiscal year 2013-2016. This regression analyzes whether having a larger share of imported inputs has an effect on seller shortfall amounts. Observations are at the firm-month level. The dependent variable is the inverse hyperbolic sine transformation of the amount of seller shortfall a given firm has for all its transactions in a given month. Seller shortfall amounts are assigned using the estimated firm fixed effects. Firms are classified into Advantageous and Disadvantageous based on the value of $Q(f)$, as explained in Section 4.4. We instrument *ImportShare*—the share of a firm's inputs which are imported—using a set of interactions between firm-level baseline import shares and real exchange rate at the country of origin-month level, for Uganda's top 10 trading partners, as described in Appendix Section A.4. First stage results are reported in Table A.10 in the Appendix. In all columns, we include dummies that control for the deciles of firm sales and inputs. Standard errors, clustered at the firm level, are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

A Appendix

A.1 Background on the VAT in Uganda

A.1.1 Institutional background

The Ugandan VAT – introduced in 1996 – follows a relatively standard design. A general rate of 18 percent applies to all sales, with the usual exemptions for necessities and some services.[19] Firms with an annual turnover above 50 million Ugandan Shillings (USD 13,700)—a threshold raised to 150 million Ugandan Shillings (USD 41,100) in fiscal year 2015-16—are required to be registered for the VAT, while smaller firms can choose to pay a simplified turnover tax.[20] As in other countries, exports are zero-rated, but the VAT applies to imports. The VAT on imports is directly paid at customs, and can be credited as input in the VAT declarations.[21] VAT firms are required to submit monthly VAT declarations to the Uganda Revenue Authority (URA). Payments of positive tax liabilities are due within 30 days of the declaration. Refunds in the case of negative VAT liabilities are restricted. Negative liabilities of less than 5 million Ugandan Shillings (USD 1,370) can only be carried over as offset against future VAT liabilities (indefinitely). If the stock of negative liabilities is above this threshold, firms may request a refund but this triggers a desk audit by the URA. The strict regulation of VAT refunds is common practice in low-income countries (Lemgruber *et al.* , 2015).

While the rules regarding VAT declaration and payment are similar across all VAT firms,[22] the URA categorizes firms into three groups for monitoring and enforcement purposes: large taxpayers are handled by a specific Large Taxpayer Office (LTO); medium-size taxpayers are handled by the Medium Taxpayer Office (MTO); and smaller firms are handled by the local URA offices spread out across the country.[23] For further institutional details and descriptive statistics on the VAT system Uganda, see Almunia *et al.* (2017).

A.1.2 Computation of revenue consequences

To compute revenue consequences of misreporting in the VAT, we rely on firms' monthly VAT declarations, and then aggregate the revenue implications at the yearly level.[24] When discrepancies are detected at the monthly level within a firm pair, we need estimates of how much of the reporting gap is due to the buyer and the seller, so that we can calculate the overall firm-level reporting

[19] For instance, unprocessed agricultural products and medical, educational and financial services are exempted from VAT. Another set of goods and services are zero-rated. A firm producing zero-rated goods may claim input tax credits, while VAT paid on inputs used in the production of exempted goods cannot be recovered (Uganda Revenue Authority, 2016).

[20] This turnover tax replaces both the VAT and the CIT. Firms below the registration threshold may choose to enter the VAT system on a voluntary basis. After the threshold was increased, the majority of firms between the new and the old threshold remained in the VAT system.

[21] Total VAT revenues are divided almost equally between the contributions from the domestic VAT and the VAT on imports.

[22] With the exception that firms with an annual turnover below 200 million Ugandan Shillings (USD 55,026) may apply for their VAT to be calculated using cash basis accounting.

[23] LTOs are firms with an annual turnover above 15 billion Ugandan Shillings (USD 4.1 million) and/or belonging to specific sectors such as oil and mining, banking, insurance, and government departments. MTOs are firms with a turnover above 2 billion Ugandan Shillings (USD 550,260, threshold increased to 5 billion Ugandan Shillings/USD 1.3 million in 2015). STOs are firms with an annual turnover lower than the MTO threshold, but above 50 million Ugandan Shillings (13,700 USD, threshold increased to 150 million Ugandan Shillings/USD 41,100 in 2015). Below this threshold, which is the same as the mandatory VAT registration threshold, firms are classified as Micro Taxpayers.

[24] The fiscal year in Uganda runs from July to June.

discrepancies and the corresponding VAT liability. This is necessary because an increased (or decreased) liability attributed to one firm involved in a given transaction may have different revenue consequences from one attributed to the other firm involved in the transaction. For example, if a firm reports a negative VAT liability in a given month, "correcting" one case of seller shortfall may still leave it with a negative liability vis-a-vis the tax authority. Our main results aggregate the revenue consequences over the 2013-2016 period. Yearly results are show in Figure A.2. The foregone revenue as a share of total VAT collected varies between 24 and 33 percent.

A.2 Two-way fixed effect analysis

In this section, we present further details for the two-way fixed effect analysis and results from the robustness checks .

A.2.1 Comparison of advantageous and disadvantageous firms

After classifying firms into Advantageous and Disadvantgeous type as described in Section 4, we compare the observable characteristics of each firm-type. Results are shown in Table A.6. We regress a dummy variable for being an Advantageous firm, on a set of firm characteristics. To facilitate comparison, all variables are standardized and have unit standard deviation. We display results for the OLS regression (Columns 1 and 2), and for a LASSO regression (Column 3). The LASSO results show that the characteristics which are significantly different across firm types are the following: Advantageous firms are less likely to belong to the Medium or Large Taxpayers Office (MTO or LTO). This seems consistent with the idea that MTO and LTO firms are under higher scrutiny. Advantageous firms have a higher ratio of sales to final consumers, and are more dowstream. This seems consistent with the idea that VAT compliance is stronger higher up in the production chain. Advantageous firms are more likely to be in the manufacturing and wholesale and retail, sectors, and less likely to be in the mining, transportation/accomodation, financial, real estate and public administration and sectors.

A.2.2 Panel estimation

Exploiting the panel dimension of the data, we investigate if firms that have self-advantageous reporting behaviors in one year tend to be the same ones that have them in the next year. This allows us to verify whether our classification is consistent over time. There are several ways of doing this, here we present three alternative versions.

In our baseline version, we compute the transition matrix by comparing a firm's classifications for different years. That is, we run Equation (1) separately for each year in the sample:

$$d_{ff't} = \delta^b_{fy} + \delta^s_{f'y} + \delta_t + r_{ff't}, \tag{A.1}$$

where y = Fiscal Year 2013 to Fiscal Year 2016.

Since the buyer and seller fixed effects are only identified within a "connected" set (Abowd *et al.* , 1999), we follow Card *et al.* (2013) and restrict the analysis to the largest connected set of buyer-seller network for each year. Table A.8 shows the results as a transition matrix laying out firms' classification in year $t+1$ conditional on their year t classification. As shown in Panel A, we find that 71 percent of firms stay within their classification in the following year.

While the approach above allows for the most number of firm-pairs to be included in the two-way fixed effect analysis, the sample of buyers and sellers included in the analysis vary across

different time periods, as the set of connected firms changes. To overcome this problem, we consider two alternative methods that allow us to conduct the analysis with a fixed set of firm-pairs across all years. In the first method, we first identify the largest connected set for each year, just as we did in the first version. We then find the common set of firm-pairs that appear in all years. In doing so, 5,835 firms remain in the analysis.

In the second version, we pool data over the years together and include a year dummy interacted with the fixed effects:

$$d_{ff't} = \delta_f^b \times \text{Year}_t + \delta_{f'}^s \times \text{Year}_t + \delta_t + r_{ff't}, \tag{A.2}$$

In both of these alternative specifications, the percentage of firms that stay within their classification from one year to the next remains approximately the same as before, at 77 and 71 percent respectively (see table A.8 Panel B and C for details).

A.2.3 Robustness

We also re-run the two-way fixed effect regression by including controls that affect the propensity of two firms to trade with each other. The objective is that by controlling for these, the likelihood for a seller to trade with a particular buyer is as good as randomly assigned. Specifically, we include two variables, one accounting for geographical proximity, and one accounting for sectoral complementarity. The first one is a dummy variable for whether two firms are located in the same sub-county.[25] The second one is the share of products from the seller's sector that are sold to the buyer's sector. To compute this, we use the official aggregate sector-level Input-Output tables calculated by the Ugandan Bureau of Statistics for financial year 2009. Introducing the controls decreases the sample of firms from 19,161 to 18,651.

The results are show in Table A.5. They are similar to what we obtained when running the regression without controls: 70 percent of firms are classified as Advantageous and 29 percent are classified as Disadvantageous. Among the Advantageous firms there are slightly more Conspicuous and Looking-small firms than in the baseline approach (80 compared to 77 and 3 compared to 2 percent, respectively), and slightly less Looking-Big firms (17 compared to 21 percent).

A.3 Firm type classification and revenue consequence computation

A.3.1 Firm type classification

For the purpose of classification, we assign a value of 0 if a firm has a missing seller or buyer fixed effect. A firm has a missing fixed-effect if the firm, either as a seller or as a buyer, is not in the largest connected set. Firms that never appear in the largest connected set (and therefore are missing both seller and buyer fixed effects) are dropped from the sample of analysis. We also perform robustness checks with: 1) dropping cases where the seller fixed effects or buyer fixed effects are missing; 2) removing weights in constructing Q_f , i.e., $Q_f \equiv \hat{\delta}_f^s + \hat{\delta}_f^b$; and 3) dropping cases where the seller fixed effects or buyer fixed effects are missing, and removing weights. Results for (1) are shown in Table A.4, remaining results are available upon request.

A.3.2 Details on revenue consequences

In the baseline approach, we divide the "blame" for each reporting discrepancy using the estimated fixed effects. The idea is to assign shares of the discrepancy proportionally based on the

[25] Uganda is divided up into a total of 1,403 sub-counties (Electoral Commision, 2016).

relative sizes of each firm's fixed effect. We present our methodology formally here. Let $s_{it} \in [0,1]$ be the share of the discrepancy assigned to buyer 1 and seller 2. Then:

$$s_{1t} = \begin{cases} \frac{\hat{\delta}_1^b}{\hat{\delta}_1^b + \hat{\delta}_2^s} & \text{if } \hat{\delta}_1^b \cdot \hat{\delta}_2^s > 0 \\ 0.5 & \text{if } \hat{\delta}_1^b = \hat{\delta}_1^s = 0 \\ 1 & \text{if } \hat{\delta}_1^b \cdot \hat{\delta}_2^s < 0 \text{ and } \hat{\delta}_1^b \cdot d_{12t} > 0 \end{cases}$$

For example, suppose $\hat{\delta}_1^b = 30$ and $\hat{\delta}_2^s = 10$. For seller shortfall cases ($d_{12t} > 0$), we assign $s_{1t} = 0.75$ and $s_{2t} = 0.25$. In the case of buyer shortfall ($d_{12t} < 0$), we assign $s_{1t} = 0.25$ and $s_{2t} = 0.75$. If the two relevant fixed effects have opposite signs, e.g. $\hat{\delta}_1^b = 30$ and $\hat{\delta}_2^s = -10$, we assign $s_{1t} = 1$ and $s_{2t} = 0$ in case of seller shortfall, and $s_{1t} = 0$ and $s_{2t} = 1$ in case of buyer shortfall.

A.3.3 Alternative revenue consequences

In the revenue consequence evaluation, we also consider an alternative method to assign a given discrepancy observed for a firm pair to the buyer and the seller involved. This approach uses information on the relative contributions of the two firms, revealed by their estimated fixed effects.

For a given discrepancy $d_{ff't}$ in a given month t between the two firms involved (say, a buyer $f = 1$ and a seller $f' = 2$), we first calculate the difference in the two estimated fixed effects for the two firms involved, i.e., $\hat{\delta}_1^b - \hat{\delta}_2^s$. If the absolute value of d_{12t} is greater than the absolute value of the difference, we allocate the discrepancy between the firm pair such that the assigned discrepancies reflect the difference in the estimated fixed effects.[26] If the absolute value of d_{12t} is less than the absolute value of the difference, we assign all the discrepancy to the more offending firm in the direction of the discrepancy. This means for a seller shortfall case, the entire discrepancy is assigned to the firm with a higher value of the fixed effects; whereas for a buyer shortfall case, the entire discrepancy is assigned to the firm with a lower value of the fixed effects. More formally, we assign the reporting discrepancies, for a given firm $f = 1$ in month t, according to the following equation:

$$d_{1t} = \begin{cases} \frac{d_{12t} + (\hat{\delta}_1^b - \hat{\delta}_2^s)}{2}, & \text{if } \mid d_{12t} \mid > \mid \hat{\delta}_1^b - \hat{\delta}_2^s \mid . \\ d_{12t} \frac{\max(\hat{\delta}_1^b - \hat{\delta}_2^s, 0)}{\hat{\delta}_1^b - \hat{\delta}_2^s}, & \text{if } \mid d_{12t} \mid \leq \mid \hat{\delta}_1^b - \hat{\delta}_2^s \mid \text{ and } d_{12t} > 0. \\ d_{12t} \frac{\min(\hat{\delta}_1^b - \hat{\delta}_2^s, 0)}{\hat{\delta}_1^b - \hat{\delta}_2^s}, & \text{if } \mid d_{12t} \mid \leq \mid \hat{\delta}_1^b - \hat{\delta}_2^s \mid \text{ and } d_{12t} < 0. \end{cases} \tag{A.3}$$

In Column 2 of Table A.2, we report the revenue consequence calculations using the approach described above. The revenue loss due to misreporting remains similar to our baseline approach: the adjusted revenue implications amount to 27 percent of VAT revenue over the whole time period.

As a further robustness check we also calculate the revenue consequences when using fixed effects from a regression that includes controls that affect the propensity of firm trade. The revenue loss again remains similar, the adjusted revenue implication amount to 26 percent of VAT revenue.

[26] For example, if d_{12t} is 60, $\hat{\delta}_1^b$ is 30, and $\hat{\delta}_2^s$ is 20, the assigned discrepancies for the buyer $f = 1$ and the seller $f = 2$ are 35 and 25, respectively. Note that the difference in $\hat{\delta}_1^b$ and $\hat{\delta}_2^s$ of 10 is preserved in the assignment. If d_{12t} is 60, $\hat{\delta}_b$ is 30, and $\hat{\delta}_2^s$ is 30, the assigned discrepancies for the buyer $f = 1$ and the seller $f = 2$ are 30 and 20, respectively. Again, the difference in $\hat{\delta}_1^b$ and $\hat{\delta}_2^s$ of 0 is preserved in the assignment.

A.4 Exchange rate fluctuations and imports : Instrumentation strategy

A.4.1 Instrumentation strategy

We instrument the share of a firm's inputs that are imported (versus purchased domestically) using exogenous variations in exchange rates, similar to Bastos *et al.* (2018).[27] In establishing the relevance of our instrument, we follow Bastos *et al.* (2018) and first assess the extent to which imports respond to RER fluctuations in the full set of firms and countries of origin by running:

$$ImportShare_{ict} = \alpha_1 log(RER)_{ct} + \alpha_2 log(RER)_{ct} * S_{ic} + \gamma_i + \gamma_t + \epsilon_{ict}, \tag{A.4}$$

where $ImportShare_{ict}$ is the share of firm i's inputs in year t that are imported from country c, $log(RER)_{ct}$ is the RER between the Ugandan shilling and the currency of country c in year t.[28] S_{ic} is either a dummy that is equal to one if firm i imports from country c in the baseline year 2012, or alternatively the share of inputs that i imported from country c in 2012. And finally, γ_i and γ_t are the firm and year fixed effects.

The results are displayed in Table A.9. The coefficient of interest is always significant, and negative as expected. The estimates imply that a 10 percent increase in the RER reduces the share of imports from a country from which the firm imports at baseline by 1.2 percentage points.[29]

To increase statistical power in the first stage of our 2SLS estimation (equation 3), we restrict to a subset of main countries of origin. Using the 10 top countries of origin yields the highest first stage F-statistic. These countries are: China, France, India, Indonesia, Japan, Kenya, South Africa, United Arab Emirates, United Kingdom, United States. Results are displayed in Table A.10.

A.4.2 Sample and Definitions

We include VAT firms from our main sample of analysis, meaning firms for which we estimate fixed effects as seller and as buyer in the two-way fixed effect regression. As in the rest of the paper, if one of the two fixed effects is missing, it is set to zero. The sample includes firms that never import.[30]

The endogenous regressor $ImportShare_{it}$ is computed as total monthly imports divided by the sum of imports and domestic purchases. The mean and standard deviation of the import share for each sample are displayed in Table A.13: they are, respectively: 0.136 and 0.316 (full sample), 0.133 and 0.314 (advantageous misreporters), 0.142 and 0.32 (disadvantageous misreporters). The figures for the restricted sample used for robustness checks, where we keep only firms for which both the fixed effect as seller and as buyer were obtained, are displayed in the second row of Table A.13.

A.4.3 Robustness

Our main 2SLS results show that a one standard deviation increase in the share of imported inputs leads to a decrease in seller shortfall of 16.2 percent in the full sample and 21.3 percent in

[27] In Bastos *et al.* (2018), the instrumented variable is the firm's $exports$.

[28] $RER = \frac{EP_c}{P}$ where E is the nominal exchange rate, P_c the price index in country c and P the domestic price index. An increase means that more Ugandan shillings are needed to purchase a given basket of goods from the country of origin.

[29] When relying on initial $shares$ of imports from different countries, we find that an increase in RER reduces imports in cases where the initial share is above 0.05 percent. See Table A.9.

[30] We also exclude the 85 consistent firms from the analysis. Inclusion of these consistent firms does not change the results significantly. The results are available upon request.

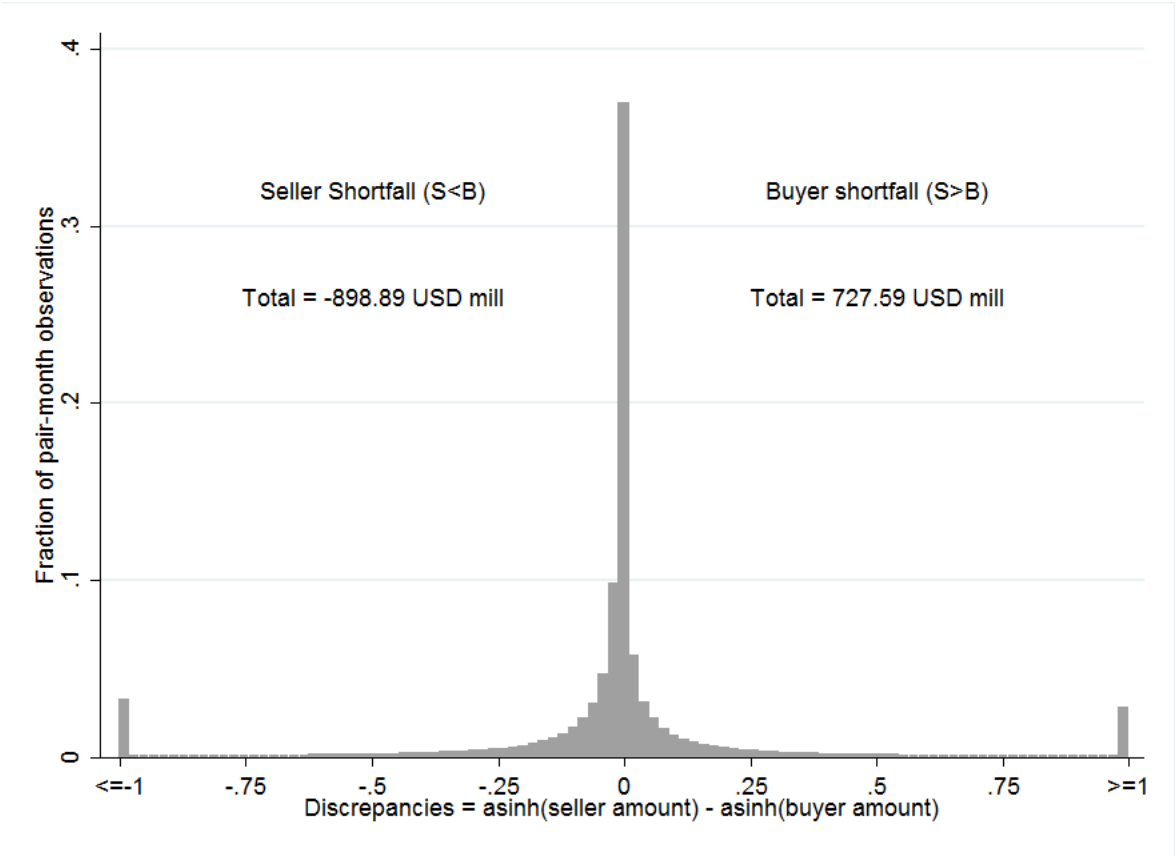
the sample of Advantageous misreporters, while the effect is close to zero in magnitude and not statistically significant for Disadvantageous misreporters(Table 3). The corresponding OLS estimates are shown in Table A.11, and are, respectively 7.8 percent; 8.4 percent; and 6 percent. The larger IV estimates are likely due to the LATE these estimates capture—the effect on compliance for firms that are shifted into importing because of favorable terms of trade—whereas the correlational OLS estimates show the average relationship between imports and compliance for all firms. Weak instrument problems are unlikely, as seen in the high first stage F-statistics.

Our results are robust, and quantitatively similar, when seller shortfall is measured using our alternative methodology to assign discrepancies (described in Appendix section A.3). Columns 5 and 6 in Table A.12 show that a one standard deviation increase in the share of imports decreases seller shortfall amounts by 21.3 percent for advantageous firms (the same impact as the one measured in our main specification), while the effect is not significant for disadvantageous firms.

We also conduct the same analysis on the restricted sample, including only firms for which none of the estimated fixed effects (as seller, and as buyer) is missing. We find the same results: a significant impact of the share of imports on reporting behavior for advantageous firms, and a non-significant one for disadvantageous firms. The effect for advantageous firms is slightly stronger than in the unrestricted sample: a one standard deviation increase in the share of imports leads to a 22.7% decrease in seller shortfall amounts (results available upon request).

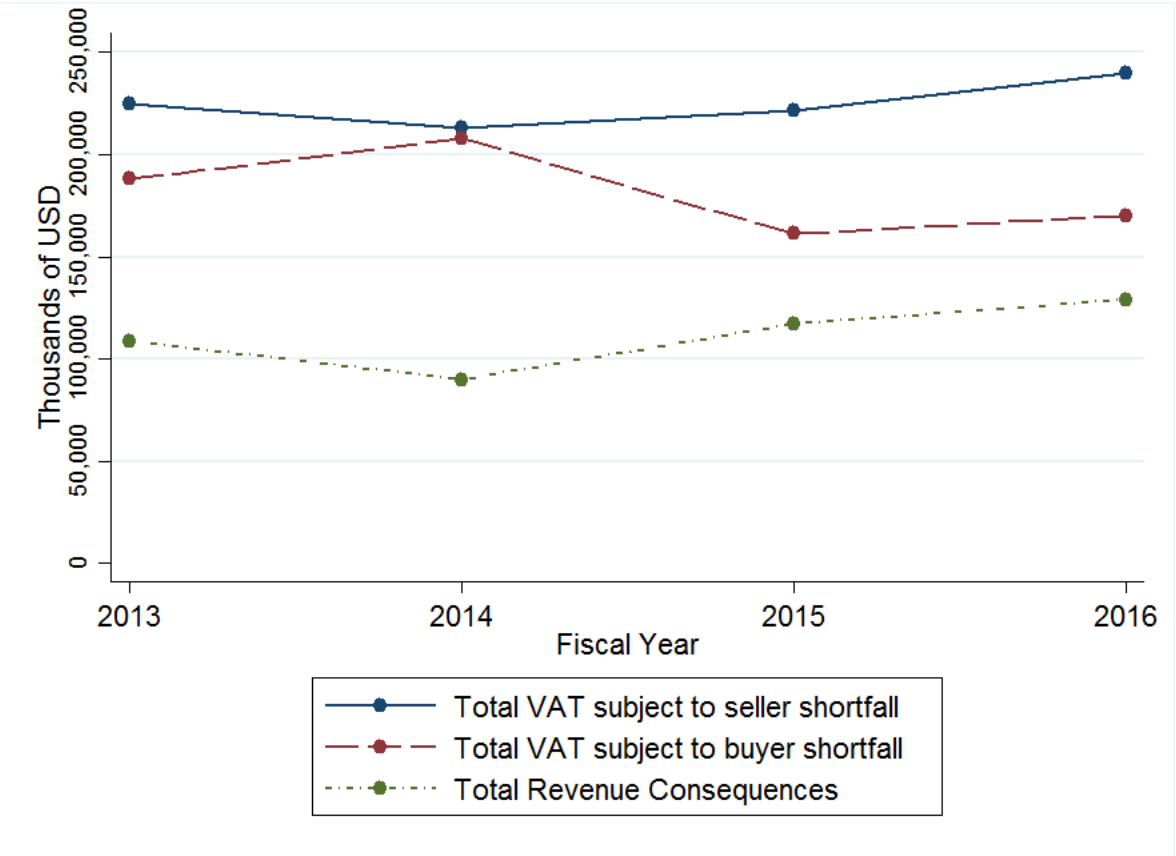
Figures

FIGURE A.1
DISTRIBUTION OF REPORTING DISCREPANCIES IN THE DOMESTIC VAT



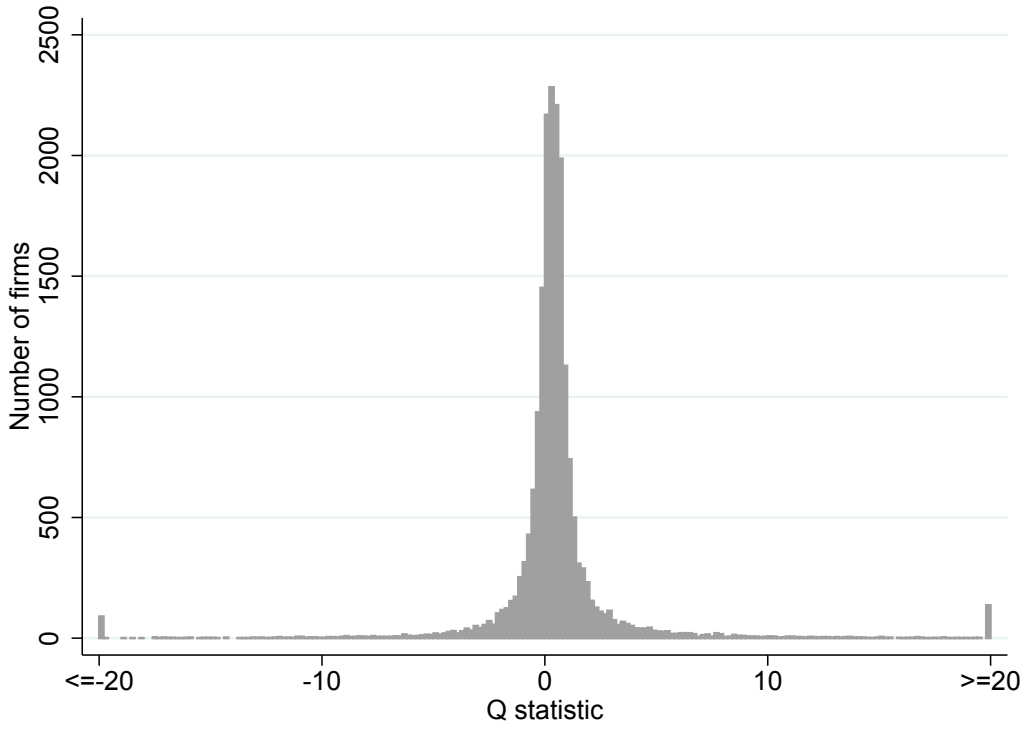
Notes: In this Figure, we show the distribution of discrepancies in the reporting of transactions by sellers and buyers for fiscal years 2013-2016. Data source: VAT Schedules data. Calculated by taking the difference between VAT charged in VS1 and VAT paid in VS24. We use the inverse hyperbolic sine transformation of VS1 and VS24. Share ≥ 1: 0.028; Share ≤ -1: 0.031.

FIGURE A.2
EVOLUTION OF REPORTING DISCREPANCIES IN THE DOMESTIC VAT OVER TIME



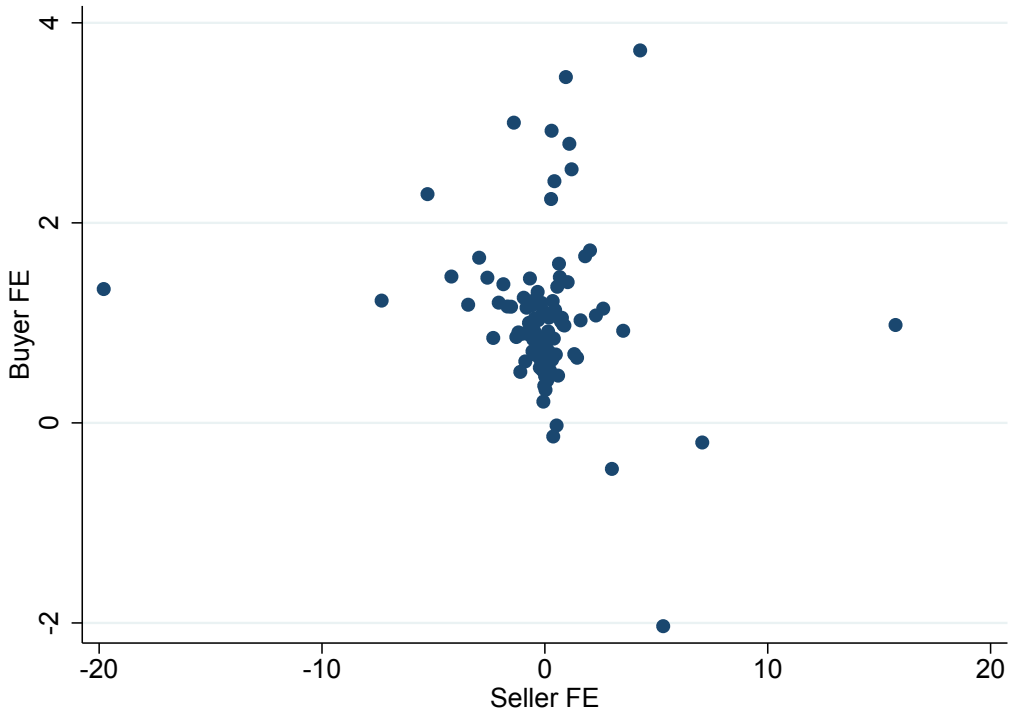
Notes: Data sources: VAT Schedules and Monthly Summary data for fiscal years 2013-2016. In this Figure, we show the evolution over time of total VAT under seller shortfall, total VAT under buyer shortfall, and the resulting revenue consequences. We include firms where either the firm-as-buyer and/or firm-as-seller fixed effects are not missing. Discrepancies are assigned to firms based on each firm's estimated fixed-effects, as described in 4.4. Revenue consequences are calculated by taking the difference between VAT charged in VS1 and VAT paid in VS24, and correcting the VAT liability in the last month of the year for the total VAT under seller excess and under buyer excess. All values are in thousands of USD.

FIGURE A.3
DISTRIBUTION OF Q STATISTIC.



Notes: In this Figure, we plot the distribution of firms estimated Q statistic (Q^f in Equation (2)). Data source: VAT Schedules data for fiscal years 2013-2016.

FIGURE A.4
CORRELATION BETWEEN BUYER AND SELLER FIXED EFFECTS



Notes: In this Figure, we plot a firms estimated buyer fixed effect over its estimated seller fe. Data source: VAT Schedules data for fiscal years 2013-2016.

Tables

TABLE A.1
AGGREGATE DOMESTIC VAT STATISTICS

	(1) Output VAT - Input VAT	(2) VAT offsets from previous year	(3) VAT liability (1) - (2)	(4) VAT due
All VAT Firms (N = 22,388)	1,830,374	67,500	1,762,874	1,361,909
LTO firms (N = 738)	1,466,848	29,646	1,437,203	979,532
MTO firms (N = 1,635)	222,911	14,055	208,855	214,868
Other VAT firms (N = 20,015)	140,615	23,799	116,816	167,509

Notes: Data source: VAT Monthly Summary data for fiscal years 2013-2016. All amounts are in thousand of USD. Column (1) shows total output VAT minus total input VAT. Column (2) shows the aggregate amount of VAT credits carried over from the previous fiscal year as offsets for current VAT dues. These are cases where the firms VAT liability in the previous year were negative. Column (3) is the aggregate VAT liability computed as (1) minus (2). Column (4) shows the total VAT amounts to be remitted to the URA, i.e., the VAT due.

TABLE A.2
SELLER SHORTFALL AND BUYER SHORTFALL IN THE DOMESTIC VAT ADJUSTING FOR FIRM-SPECIFIC CONTRIBUTION TO DISCREPANCIES

	(1) Main	(2) Alt.	(3) Naive
No. of distinct firms	19,161	19,161	19,161
Total net VAT due	1,554,101	1,554,101	1,554,101
Seller shortfall			
Number of distinct firms with seller shortfall	17,255	17,255	13,451
Total net VAT due from firms with seller shortfall	1,275,946	1,275,946	1,133,483
Total VAT subject to seller shortfall	900,099	900,099	900,099
Buyer shortfall			
Number of distinct firms with buyer shortfall	18,000	18,000	17,202
Total net VAT due from firms with buyer shortfall	1,316,829	1,316,829	1,262,514
Total VAT subject to buyer shortfall	727,373	727,373	727,373
Correcting seller shortfall and buyer shortfall			
Impact on total net VAT due	446,224	371,363	493,471
Percentage of total VAT collected	32.8%	27.3%	36.2%

Notes: Data source: VAT Schedules and Monthly Summary data for fiscal years 2013-2016. In this table we display the revenue consequence analysis using various methods to assign discrepancies to firms. Revenue consequences are calculated by taking the difference between VAT charged in VS1 and VAT paid in VS24, and correcting the VAT liability in the last month of the year for the total VAT under seller shortfall and under buyer shortfall. In column (1) (main approach), discrepancies are assigned to firms based on each firm's estimated fixed-effects, as described in 4.4. In column (2) (alternative approach) discrepancies are assigned to firms based on each firm's estimated fixed-effects, as described in A.3. In column (3) (naive approach), we assign all seller shortfall to the seller, and all buyer shortfall to the buyer. All values are in thousands of USD.

TABLE A.3
SELLER SHORTFALL AND BUYER SHORTFALL IN THE DOMESTIC VAT

Sample	(1) All firms	(2) Reporting any sales	(3) Positive sales to VAT firms	(4) Reciprocal reporting
		Panel A: Full sample		
No. of distinct firms	22,388	22,388	19,902	19,435
Percentage of all firms	(100%)	(100%)	(89%)	(87%)
Total net VAT due	1,555,848	1,555,848	1,541,647	1,532,692
Seller shortfall				
Number of distinct firms with seller shortfall	13,670	13,223	8,319	7,519
Total net VAT due from firms with seller shortfall	1,133,508	1,135,177	1,057,149	980,458
Total VAT subject to seller shortfall	900,353	610,508	499,301	424,915
Buyer shortfall				
Number of distinct firms with buyer shortfall	17,794	16,589	9,051	7,171
Total net VAT due from firms with buyer shortfall	1,262,552	1,267,812	1,197,015	1,047,951
Total VAT subject to buyer shortfall	727,664	578,811	375,591	293,240
Correcting seller shortfall and buyer shortfall				
Impact on total net VAT due	494,533	271,361	214,643	185,480
Percentage of total VAT collected	36.3%	19.9%	15.8%	13.6%
		Panel B: Study sample		
No. of distinct firms	19,161	19,161	19,043	18,787
Percentage of all firms	(100%)	(100%)	(99%)	(98%)
Total net VAT due	1,554,101	1,554,101	1,541,501	1,532,597
Seller shortfall				
Number of distinct firms with seller shortfall	13,451	13,074	8,314	7,515
Total net VAT due from firms with seller shortfall	1,133,483	1,135,152	1,057,128	980,441
Total VAT subject to seller shortfall	900,099	610,324	499,244	424,863
Buyer shortfall				
Number of distinct firms with buyer shortfall	17,202	16,287	9,019	7,171
Total net VAT due from firms with buyer shortfall	1,262,514	1,267,773	1,196,999	1,047,951
Total VAT subject to buyer shortfall	727,373	578,598	375,540	293,215
Correcting seller shortfall and buyer shortfall				
Impact on total net VAT due	493,471	272,180	216,142	185,289
Percentage of total VAT collected	36.2%	20.0%	15.9%	13.6%

Notes: Data source: VAT Schedules and Monthly Summary data for fiscal years 2013-2016. In this table we display the revenue consequence analysis for various categories of firms using a "naive" methodology where we assign all seller shortfall to sellers and all buyer shortfall to buyers. Revenue consequences are calculated by correcting the VAT liability in the last month of the year for the total VAT under seller shortfall and under buyer shortfall. Panel A considers the full sample of all active VAT-registered firms in our estimation period. Panel B considers the study sample we use in our two-way fixed effect analysis. *Definitions:* (1) Are all firms that are VAT-registered in the estimation period. (2) Are the VAT-registered firms that report sales in the estimation period. (3) Are the VAT-registered firms that report positive sales to other VAT-registered firms. This is different from (2) because the firm only reports sales to final consumers and/or because the firm reports negative sales to VAT-registered firms. (4) Are the firms from (3) where the buyer also reported a purchase from the seller for at least one month in the estimation period. All values are in thousands of USD.

TABLE A.4
SUMMARY STATISTICS AND REVENUE CONSEQUENCES BY FIRM TYPE. ROBUSTNESS ANALYSIS: RESTRICTED SAMPLE.

	All	(1) Consist	(2) Disadv.	(3) Adv.	(3a) Conspic.	(3b) Looking Small	(3c) Looking Big
No. of distinct firms	13,248	0	4,108	9,140	5,983	345	2,812
Percentage of all firms	(100%)	(0%)	(31%)	(69%)	(45%)	(3%)	(21%)
Total net VAT due	1,527,174	0	855,449	671,726	422,040	49,896	199,789
Seller shortfall							
Number of distinct firms with seller shortfall	13,010	0	3,952	9,058	5,921	343	2,794
Total net VAT due from firms with seller shortfall	1,279,477	0	758,153	521,323	351,734	36,390	133,200
Total VAT subject to seller shortfall	840,641	0	99,674	740,968	411,592	163,474	165,901
Buyer shortfall							
Number of distinct firms with buyer shortfall	12,982	0	4,062	8,920	5,787	341	2,792
Total net VAT due from firms with buyer shortfall	1,316,418	0	792,261	524,157	352,066	38,618	133,473
Total VAT subject to buyer shortfall	678,314	0	487,601	190,713	63,498	48,310	78,904
Correcting seller shortfall and buyer shortfall							
Impact on total net VAT due	405,737	0	−123,678	529,415	321,449	119,569	88,397
Percentage of total VAT collected	29.8%	0%	−9.1%	38.9%	23.6%	8.8%	6.5%

Notes: Data Source: VAT Schedules and Monthly Summary data for fiscal years 2013-2016. We include firms where both the firm-as-seller and firm-as-buyer fixed effects are not missing. Revenue consequences are calculated by correcting the VAT liability in the last month of the year for the total VAT under seller shortfall and under buyer shortfall. *Definitions:* (1) Consistent: $Q(f) = 0$. (2) Disadvantageous: $Q(f) < 0$. (3) Advantageous: $Q(f) > 0$. (3A) Conspicuous Advantageous: $\hat{\delta}_s(f) \geq 0$ and $\hat{\delta}_b(f) \geq 0$. (3B) Looking small Advantageous: $\hat{\delta}_s(f) \geq 0$ and $\hat{\delta}_b(f) < 0$. (3C) Looking big Advantageous: $\hat{\delta}_s(f) < 0$ and $\hat{\delta}_b(f) \geq 0$. $Q(f)$ is calculated as a weighted average of the estimated firm-as-buyer fixed effect and firm-as-seller fixed effect, i.e., : $Q(f) = \hat{\delta}_b \times \frac{p}{p+s} + \hat{\delta}_s \times \frac{s}{p+s}$ where s stands for total sales and p stands for total purchases. All values are in thousands of USD.

TABLE A.5
SUMMARY STATISTICS AND REVENUE CONSEQUENCES BY FIRM TYPE. ROBUSTNESS ANALYSIS: TWO-WAY FIXED EFFECTS WITH CONTROLS.

	All	(1) Consist	(2) Disadv.	(3) Adv.	(3a) Conspic.	(3b) Looking Small	(3c) Looking Big
No. of distinct firms	18,651	90	5,426	13,135	10,466	408	2,261
Percentage of all firms	(100%)	(0%)	(29%)	(70%)	(56%)	(2%)	(12%)
Total net VAT due	1,527,903	541	869,019	658,343	391,989	52,032	214,321
Seller shortfall							
Number of distinct firms with seller shortfall	16,804	31	4,816	11,957	9,306	406	2,245
Total net VAT due from firms with seller shortfall	1,255,178	11	751,930	503,237	289,800	43,717	169,720
Total VAT subject to seller shortfall	866,432	60	112,066	754,305	445,953	167,143	141,209
Buyer shortfall							
Number of distinct firms with buyer shortfall	17,516	70	5,170	12,276	9,623	404	2,249
Total net VAT due from firms with buyer shortfall	1,296,154	237	788,412	507,505	290,105	46,763	170,636
Total VAT subject to buyer shortfall	694,862	292	504,852	189,717	70,708	45,492	73,518
Correcting seller shortfall and buyer shortfall							
Impact on total net VAT due	422,512	38	−116,096	538,570	341,640	124,006	72,924
Percentage of total VAT collected	31.0%	0.0%	−8.5%	39.5%	25.1%	9.1%	5.4%

Notes: Data Source: VAT Schedules and Monthly Summary data for fiscal years 2013-2016. We include controls described in Section A.2 in the two-way fixed-effects model estimating firms' fixed effect as a seller and as a buyer. Revenue consequences are calculated by correcting the VAT liability in the last month of the year for the total VAT under seller shortfall and under buyer shortfall. *Definitions:* (1) Consistent: $Q(f) = 0$. (2) Disadvantageous: $Q(f) < 0$. (3) Advantageous: $Q(f) > 0$. (3A) Conspicuous Advantageous: $\hat{\delta}_s(f) \geq 0$ and $\hat{\delta}_b(f) \geq 0$. (3B) Looking small Advantageous: $\hat{\delta}_s(f) \geq 0$ and $\hat{\delta}_b(f) < 0$. (3C) Looking big Advantageous: $\hat{\delta}_s(f) < 0$ and $\hat{\delta}_b(f) \geq 0$. $Q(f)$ is calculated as a weighted average of the estimated firm-as-buyer fixed effect and firm-as-seller fixed effect, i.e., : $Q(f) = \hat{\delta}_b \times \frac{p}{p+s} + \hat{\delta}_s \times \frac{s}{p+s}$ where s stands for total sales and p stands for total purchases. All values are in thousands of USD.

TABLE A.6
COMPARISON OF ADVANTAGEOUS AND DISADVANTAGEOUS FIRMS

Dep. Variable: Probability of Being Advantageous	*Panel A*		*Panel B*
	Coefficient	P-value	Coefficient
in Kampala	0.00	0.92	0.00
Distance to URA office	−0.04	0.00***	0.00
MTO/LTO	−0.06	0.00***	−0.06
VAT Payable	0.05	0.02**	0.00
VAT Due	−0.03	0.06*	0.00
Total input	0.05	0.04**	0.00
Total output	−0.05	0.05**	0.00
Ratio of sales to FC	0.10	0.00***	0.09
Number of clients	−0.01	0.29	0.00
Number of suppliers	0.01	0.20	0.00
Upstreamness	−0.02	0.00***	−0.02
Distinct outputs (all good codes)	−0.02	0.60	0.00
Distinct outputs (relevant good codes)	0.03	0.43	0.00
Distinct inputs (all good codes)	−0.04	0.41	0.00
Distinct inputs (relevant good codes)	0.02	0.60	0.00
Sectors:			
Agriculture, forestry, fishing	−0.01	0.45	0.00
Mining, Quarrying	−0.03	0.00***	−0.03
Manufacturing	0.01	0.30	0.02
Water, Electricity services	−0.01	0.31	0.00
Construction	−0.01	0.41	0.00
Wholesale and retail	0.00	0.00	0.02
Transportation, accomodation services	−0.03	0.00***	−0.01
Information, communication	−0.01	0.28	0.00
Financial services	−0.02	0.00***	−0.01
Real estate	−0.04	0.00***	−0.03
Professional, Admin, Other Services	−0.02	0.01***	0.00
Public Administration	−0.03	0.05**	−0.02
Education	−0.01	0.16	0.00
Health and social work	0.00	0.83	0.00
Arts and Entertainment	0.00	0.65	0.00

Notes: Data source: VAT Schedules and Monthly Summary data for fiscal years 2013-2016. This table shows the results of the regression of a firm-type dummy variable – equal to one if the firm is categorized as Advantageous and zero otherwise – on a set of firm characteristics. Panel A displays the results from a multivariate regression including all variables listed. Panel B display the results from a LASSO regression. All variables are standardized to have unit standard deviation. *in Kampala* is a dummy equal to one if the firm is in Kampala. *Distance* is calculated by assigning each firm to a Sub-county and calculating the distance from the center of the Sub-county to the closest URA office. *MTO/LTO* is a dummy variable equal to one if the firm is registered in the Medium or Large Taxpayers' Office (as of June 2017). *Vat Payable, Vat Due, Total inputs and Total Output* are totals over years 2013-2016. *Ratio of sales to FC*, is the ratio of total sales to final consumers over total sales. *Number of clients and Number of suppliers* are the totals over years 2013-2016. *Upstreamness* indicates the firms' distance to final consumption – larger values indicate that the firm is higher up in the production chain. It is computed by creating an input-output matrix, based on firm-to-firm good code transactions. *Distinct outputs and Distinct inputs* are the number of unique good codes within the firm's sales/purchases over the 2013-2016 period. Good codes are based on the universe of transactions from year 2014 and are obtained by applying a machine learning text algorithm to the text descriptions included in the VAT Schedules. Sector is the firm's sector as listed in the tax registry. We drop Consistent firms from the regressions.

TABLE A.7
SUMMARY STATISTICS FOR FIRM-TYPE, ASSUMING VARIOUS PERCENTAGES OF SALES TO FINAL CONSUMERS IS SUBJECT TO SELLER SHORTFALL

	10% of sales to FC		20% of sales to FC		30% of sales to FC		40% of sales to FC	
	No. of Firms	Share of firms	No. of Firms	Share of firms	No. of Firms	Share of firms	No. of Firms	Share of firms
Consistent	170	0.01	170	0.01	170	0.01	170	0.01
Disadvantageous	4555	0.24	3898	0.20	3486	0.18	3143	0.16
Advantageous	14436	0.75	15093	0.79	15505	0.81	15848	0.83
Conspicuous	11864	0.62	12818	0.67	13405	0.70	13846	0.72
Looking small	818	0.04	954	0.05	1042	0.05	1118	0.06
Looking big	1754	0.09	1321	0.07	1058	0.06	884	0.05

Notes: Data source: VAT Schedules and Monthly Summary data for fiscal years 2013-2016. This table presents summary statistics for firm-types, assuming various percentages of sales to final consumers are subject to seller shortfall. The sample is restricted to our study sample. *Definitions:* Consistent: $Q(f) = 0$. Disadvantageous: $Q(f) < 0$. Advantageous: $Q(f) > 0$. Conspicuous Advantageous: $\hat{\delta}_s(f) \geq 0$ and $\hat{\delta}_b(f) \geq 0$. Looking–small Advantageous: $\hat{\delta}_s(f) \geq 0$ and $\hat{\delta}_b(f) < 0$. Looking–big Advantageous: $\hat{\delta}_s(f) < 0$ and $\hat{\delta}_b(f) \geq 0$.$Q(f)$ is calculated as a weighted average of the estimated firm-as-buyer fixed effect and firm-as-seller fixed effect, i.e., : $Q(f) = \hat{\delta}_b \times \frac{p}{p+s} + \hat{\delta}_s \times \frac{s}{p+s}$ where s stands for total sales and p stands for total purchases. All values are in thousand USD.

TABLE A.8
FIRM-TYPE TRANSITION MATRIX

	Panel A: All firms			
	Advantageous (t)	Disadvantageous (t)	Consistent (t)	
Advantageous (t+1)	0.50	0.14	0.00	0.64
Disadvantageous (t+1)	0.15	0.21	0.00	0.36
Consistent (t+1)	0.00	0.00	0.00	0.00
	0.65	0.34	0.00	1.00
	Panel B: Common firm-pairs			
	Advantageous (t)	Disadvantageous (t)	Consistent (t)	
Advantageous (t+1)	0.55	0.12	0.00	0.67
Disadvantageous (t+1)	0.10	0.21	0.00	0.32
Consistent (t+1)	0.00	0.00	0.01	0.02
	0.65	0.33	0.02	1.00
	Panel C: Time dummies			
	Advantageous (t)	Disadvantageous (t)	Consistent (t)	
Advantageous (t+1)	0.49	0.14	0.00	0.64
Disadvantageous (t+1)	0.15	0.21	0.00	0.36
Consistent (t+1)	0.00	0.00	0.00	0.00
	0.64	0.35	0.00	1.00

Notes: Data source: VAT Schedules and Monthly Summary data for fiscal years 2013-2016. This table presents the transition matrix for firm classifications, using several versions. *Panel A* calculates the transition matrix matrix for all firms when we calculate the fixed effects separately for every fiscal year. *Panel B* presents the transition matrix restricted to firm-pairs that appear in every financial year. *Panel C* presents the transition matrix calculated when the yearly dummies are included in the regressions.

TABLE A.9
IMPORT RESPONSE TO REAL-EXCHANGE-RATE FLUCTUATIONS

	Dependent variable: Share of purchases imported from a given country			
log(RER)	-0.023524	-0.063563**	0.028678	0.043828*
	(0.011)	(0.014)	(0.015)	(0.018)
log(RER).1(any imports 2012)			-12.884944**	
			(2.791)	
log(RER).(imports share 2012)				-0.864913**
				(0.212)
R-squared	0.074	0.721	0.721	0.724
N	4226547	4042505	4042505	4042505
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	No	No	No
Origin FE	Yes	No	No	No
Firm*Origin FE	No	Yes	Yes	Yes

Notes: Data source: Customs, VAT Schedules and Monthly Summary data for fiscal years 2013-2016. This regression verifies that firms are less likely to import from a given country in a given year when the real exchange rate is less favorable. Observations are at the firm-country of origin level, for years 2013-2016. We include all VAT firms. The dependent variable is the share of total inputs which are imported from a given country by the firm. $Log(RER)$ is the real exchange rate between the Ugandan Shilling and the currency of the country of origin ($RER = \frac{EP_c}{P}$ where E is the nominal exchange rate, P_c the price index in country c and P the domestic price index, an increase means more units of UGX are needed to purchase a given basket of goods from the country of origin). *1(any imports 2012)* is a dummy equal to one if the firm is importing from a given country in the baseline year, 2012. *(imports share 2012)* is the share of a firm's purchases which are imported from a given country of origin in the baseline year, 2012. The dependent variable, *1(any imports 2012)* and*(imports share 2012)* are set to zero for any firm-country combination for which no import is recorded in a given year. Therefore in each year, all firm-country combinations exist. Standard errors are clustered at the firm-year level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

TABLE A.10
FIRST STAGE FOR THE 2SLS

Sample Dep. Variable	(1) Full ImportShare	(2) Advantageous ImportShare	(3) Disadvantageous ImportShare
United Arab Emirates	-0.006***	-0.006***	-0.005**
	(0.001)	(0.001)	(0.002)
China	-0.005***	-0.005***	-0.005***
	(0.001)	(0.001)	(0.001)
France	-0.006	-0.002	-0.017
	(0.008)	(0.007)	(0.015)
United Kingdom	-0.001	-0.001	-0.000
	(0.001)	(0.001)	(0.002)
Indonesia	-0.013***	-0.013***	-0.029**
	(0.003)	(0.003)	(0.014)
India	-0.007***	-0.008***	-0.005***
	(0.001)	(0.001)	(0.002)
Japan	-0.009***	-0.010***	-0.007***
	(0.001)	(0.001)	(0.002)
Kenya	-0.016***	-0.017***	-0.013***
	(0.001)	(0.002)	(0.002)
United States	-0.006***	-0.007***	-0.004
	(0.002)	(0.002)	(0.003)
South Africa	-0.006***	-0.005**	-0.008***
	(0.002)	(0.002)	(0.003)
Sales decile	Yes	Yes	Yes
Inputs decile	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes
Month-Year FE	Yes	Yes	Yes
N	442519	314669	127850

Notes: Data source: Customs, VAT Schedules and Monthly Summary data for fiscal years 2013-2016. This table displays the first stage results for Table 3. Observations are at the firm-month level, for years 2013 to 2016. The dependent variable is the share of a firms' inputs which are imported, computed as imports over total inputs (local purchases plus imports). The 10 instruments are computed as a firm's share of imports from country c in baseline (year 2012) interacted with the real exchange rate between country c's currency and the Ugandan shilling in a given month, for Uganda's top 10 trading partners (based on 2012 volumes of trade). In column (1) we include all firms (except the 85 Consistent firms), while column (2) and (3) show results when the sample is split between advantageous firms and disadvantageous firms. Firm type is determined based on estimated Q statistic $Q(f)$: advantageous when $Q(f) > 0$, disadvantageous when $Q(f) < 0$. We control for sales decile and inputs decile. Standard errors are clustered at the firm level (in parantheses). *** p < 0.01, ** p < 0.05, * p < 0.10.

TABLE A.11
EFFECT OF ENHANCED TAX AUTHORITY OVERSIGHT ON VAT COMPLIANCE BY SOPHISTICATED AND CONFUSED FIRMS - OLS

	OLS		
Dependent variable	$asinh(Seller\ shortfall\ amounts)$		
	(1)	(2)	(3)
Sample	Full	Advantageous	Disadvantageous
ImportShare	-0.256***	-0.280***	-0.194***
	(0.008)	(0.010)	(0.012)
Sales decile	Yes	Yes	Yes
Inputs decile	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes
Month-Year FE	Yes	Yes	Yes
N	442519	314669	127850
Mean of dep.	0.90	1.03	0.57

Notes: Data source: Customs, VAT Schedules and Monthly Summary data for fiscal years 2013-2016. This regression analyzes whether having a larger share of imported inputs has an effect on seller shortfall amounts. Observations are at the firm-month level, for years 2013 to 2016. The dependent variable is the inverse hyperbolic sine transformation of the amount of seller shortfall a given firm has for all its transactions in a given month. Seller shortfall amounts are assigned using the estimated firm fixed-effects. Firms are classified into Advantageous and Disadvantageous based on the value of $Q(f)$, as explained in Section 4.4. *ImportShare* is the share of a firm's inputs which are imported. In all columns, we include dummies that control for the deciles of firm sales and inputs. Standard errors, which are clustered at the firm level, are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

TABLE A.12
EFFECTS OF ENHANCED TAX AUTHORITY OVERSIGHT ON VAT COMPLIANCE - ROBUSTNESS CHECK: ALTERNATIVE ASSIGNATION OF DISCREPANCIES

Dependent variable	$asinh(Seller\ shortfall\ amounts)$					
	OLS			**IV**		
Sample	Full	Adv.	Disadv.	Full	Adv.	Disadv.
	(1)	(2)	(3)	(4)	(5)	(6)
ImportShare	-0.242***	-0.261***	-0.194***	-0.604***	-0.764***	-0.243
	(0.008)	(0.010)	(0.013)	(0.161)	(0.172)	(0.372)
Sales decile	Yes	Yes	Yes	Yes	Yes	Yes
Inputs decile	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Month-Year FE	Yes	Yes	Yes	Yes	Yes	Yes
N	442519	314669	127850	442519	314669	127850
Mean of dep.	0.94	1.03	0.72	0.94	1.03	0.72
Kleibergen-Paap LM stat.				344.261	286.455	74.104
Kleibergen-Paap Wald F stat.				53.101	46.381	11.443

Notes: Data source: Customs, VAT Schedules and Monthly Summary data for fiscal years 2013-2016. This regression analyzes whether having a larger share of imported inputs has an effect on seller shortfall amounts. Observations are at the firm-month level, for years 2013 to 2016. The dependent variable is the inverse hyperbolic sine transformation of the amount of seller shortfall a given firm has for all its transactions in a given month. Seller shortfall amounts are assigned using estimated firm fixed-effects for reporting behavior, using the alternative assignation method described in Appendix Section A.3. Firms are classified into Advantageous and Disadvantageous based on the value of $Q(f)$, the firm-specific quantity (see Section 4.4 for detailed description). Columns (1) to (3) report the OLS estimation. In 2SLS estimation, Columns (4) to (6), we instrument *ImportShare*—the share of a firm's inputs which are imported—using a set of interactions between firm-level baseline import shares and real exchange rate at the country of origin-month level, for Uganda's top 10 trading partners. First stage results are displayed in Table A.10 in the Appendix. In all columns, we include dummies that control for the deciles of firm sales and inputs. Standard errors are clustered at the firm level (in parentheses). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

TABLE A.13
IMPORT SHARE DESCRIPTIVE STATISTICS

	Import Share					
	Full		**Adv.**		**Disadv.**	
	Mean	SD	Mean	SD	Mean	SD
Main Sample	0.136	0.316	0.133	0.314	0.142	0.320
Restricted Sample	0.131	0.307	0.123	0.299	0.149	0.326

Notes: Data source: Customs, VAT Schedules and Monthly Summary data for fiscal years 2013-2016. This table displays descriptive statistics for the variable $Import\ Share$, computed as total monthly imports divided by the sum of imports and domestic purchases, for the firm-month observations included in the 2SLS sepecifications. The first row shows the statistics for the main sample used in the 2SLS analysis, while the second row shows statistics for the restricted sample, where we keep only firms for which both the fixed effect as seller and as buyer were obtained. Columns (1) and (2) correspond to all firms, Columns (3) and (4) to firms classified as Advantageous misreporters, and Columns (5) and (6) to firms classified as Disadvantageous misreporters.