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

Miguel Almunia
CUNEF & CEPR

Jonas Hjort
Columbia U. & CEPR & NBER

Justine Knebelmann
Paris School of Economics

Lin Tian
INSEAD & CEPR

August 2020

Abstract

Are firms sophisticated maximizers, or do they consistently make errors? Using transaction-level data from Ugandan value-added tax returns, we show that sellers and buyers report different amounts 79 percent of the time, despite invoices being easily cross-checked. We estimate that 27 percent of firms are disadvantageous misreporters—they misreport own sales and purchases such that their tax liability increases—while 73 percent are advantageous misreporters. Many firms—especially disadvantageous misreporters—fail to (or under-) report transactions they themselves reported at customs, increasing their VAT liability. Unilateral VAT misreporting cost Uganda about USD 383 million in foregone 2013-2016 tax revenue.

*miguel.almunia@cunef.edu, hjort@columbia.edu, justine.knebelmann@psemail.eu, lin.tian@insead.edu.

We thank David Henning and Claude Raisaro for outstanding research assistance, and Michael Best, Ray Fisman, Francois Gerard, Jim Hines, Wojciech Kopczuk, David Margolis, Joana Naritomi, Daniel Reck, Floris Zoutman and many seminar and conference participants for helpful comments and suggestions. We thank the Uganda Revenue Authority (URA) 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 the true costs and benefits to the firm. This assumption underlies the models that guide our understanding of how firms behave. Strategic decision-making by firms is by and large taken as self-evident.

There is, however, growing evidence that some firms deviate from optimal behavior (DellaVigna & Gentzkow, 2019; Kremer *et al.* , 2019; Hortacsu *et al.* , 2019). If a significant proportion consistently makes mistakes, the consequences for theory and policy design would be far-reaching. Consider how firms in low-income countries should be taxed—one of the most important questions for economic development (Besley & Persson, 2009; Kleven *et al.* , 2016). The value-added tax (VAT)—now in use in 166 countries around the world—is popular among economists in part because of its enforcement properties. The seller and buyer have asymmetric (mis)reporting incentives and invoices for the same transaction can easily be cross-checked in firm-to-firm transactions (Ebrill *et al.* , 2001; Kopczuk & Slemrod, 2006; Pomeranz, 2015). This is thought to make the VAT "self-enforcing," but the argument assumes that firms are sophisticated enough to infer the likelihood of cross-checks and to accurately keep track of their sales and purchases.[1]

In this paper, we study the sophistication of firms' decision-making in a low-income country context by analyzing their tax reporting behavior. We use 2013-2016 transaction-level VAT and customs records on all domestic and international trade involving the 22,388 VAT-registered firms in Uganda. In the first part of our analysis, we document that sellers and buyers report different transacted amounts in 79 percent of reported firm-pair×month VAT observations, using an approach akin to Fisman & Wei (2004)'s cross-checking of "mirror" customs data. In 60 percent of mismatch transactions we find a "seller shortfall," namely the seller reporting the lower value, and in the remaining 40 percent a "buyer shortfall." The latter cases are harder to rationalize since the buyer reporting less than the seller raises one or both firms' tax liability, other things equal.

In the second part of our analysis, we develop a fixed-effects methodology that estimates what fraction of each reporting discrepancy can be attributed to the seller vs. the buyer, controlling for other firms' time-invariant characteristics. Combining each firm's estimated reporting discrepancies as buyer and seller in turn allows us to categorize firms' reporting behavior. Some report accurately on average; some overreport total purchases and/or underreport total sales such that the firm's overall liability decreases—what we

[1] Another reason why economists recommend the use of the VAT is that the tax in theory does not distort production decisions. Note also that the self-enforcement property breaks down in sales to final consumers, because in that case the buyer doesn't have to file VAT invoices.

interpret as strategic behavior in a low-enforcement context and label *advantageous* misreporting; and some make *disadvantageous* reporting mistakes that increase the firm's overall liability.[2]

We find that 73 percent of VAT-registered Ugandan firms are advantageous misreporters and 27 percent are disadvantageous misreporters. Among advantageous misreporters, 11 percent "look small" by underreporting both sales and purchases and the firm's value-added (a form of behavior first identified by Carrillo *et al.* (2017) in Ecuador that may benefit the firm). Another 76 percent are "conspicuous" advantageous misreporters that underreport their sales and overreport their purchases. The remaining 13 percent "look big" by overreporting both sales and purchases. Seventy-two (63) percent of firms classified as advantageous (disadvantageous) in any given year remain in the same category the following year.

In a series of robustness checks, we analyze several ways in which our estimates could under- or overestimate the prevalence of reporting mistakes. We re-estimate our model assuming extensive final sales underreporting, finding that the proportion of disadvantageous firms is still large. When we account for estimation uncertainty, the ratio of advantageous to disadvantageous misreporters remains stable around three, although in this case we cannot rule out neutral tax-reporting behavior for a substantial share of firms.

In the third part of our analysis we consider how sophisticated and less sophisticated firms behave in higher state capacity contexts. The case for the VAT assumes some degree of capacity to cross-check firms' tax reports. Our results suggest that low-income countries may not have such capacity. However, like models of firms' response to other public policies, the self-enforcing VAT hypothesis ultimately rests on a more fundamental assumption: that firms behave strategically. Confused firms may not respond as anticipated to enforcement incentives.

To investigate, we take advantage of goods being more closely monitored when moving through customs.[3] We compare an import transaction report at customs versus the *same firm's* report of the same transaction on the credit side of its domestic VAT records. While, as expected, double reports are more consistent when the same firm makes both reports and one of the two is at customs, we find discrepancies in a remarkable 48 percent of such cases. In particular, we again find evidence of firm mistakes. Firms reduce their tax liability by overreporting their imported inputs in VAT returns in 14 percent of

[2]We interpret *systematic* underreporting of a firm's liability as strategic behavior and systematic overreporting of a firm's liability as mistakes. By classifying any systematic, self-advantageous reporting errors as strategic behavior, we possibly underestimate the true extent of reporting inaccuracies.

[3]It is well documented that tariffs are more stringently enforced than domestic taxes, perhaps because goods have to physically clear customs (Riezman & Slemrod, 1987; Keen & Lighart, 2002; Emran & Stiglitz, 2005; Keen & Lighart, 2005; Baunsgaard & Keen, 2010; Cagé & Gadenne, 2018).

import transactions, while they increase their liability by underreporting in VAT returns in 34 percent of transactions. Importantly, the latter form of disadvantageous behavior is significantly more common among firms classified as disadvantageous misreporters in domestic VAT data.

Overall, our findings suggest that the majority of Ugandan firms are sophisticated enough to respond to weak tax enforcement by considerably underreporting their tax liability, as conventional models of firm behavior assume. However, a non-negligible proportion consistently make costly errors. We quantify the consequences of such errors and (self-) advantageous misreporting for tax collection. We estimate that the government revenue *gain* due to reporting errors by disadvantageous misreporters is large—around USD 138 million during 2013-2016. However, the revenue loss due to misreporting by advantageous misreporters is even larger, at around USD 522 million. On net, unilateral VAT misporting cost the Ugandan government around USD 383 million, or 4 percent of total tax revenue collected, during 2013-2016.

This paper contributes to three related but distinct strands of the literature on firm behavior and taxation. First, we provide what to our knowledge are the first direct estimates of the extent of mistakes in an economy-wide population of firms. The methodology we develop allows us to classify individual firms' behavior as self-advantageous or not, and we observe 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 Tourek, 2018; DellaVigna & Gentzkow, 2019; Hortacsu *et al.* , 2019; Kremer *et al.* , 2019; Hjort *et al.* , 2020).[4]

Second, we provide new evidence on how tax evasion 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 tariff rate. However, our focus is on variation in enforcement capacity, linking our analysis with existing work on the causes and consequences of state capacity (Besley & Persson, 2009, 2010; Acemoglu *et al.* , 2015; Best *et al.* , 2019; Page & Pande, 2018). We also build on existing studies of more-vs.-less attentive taxpayers' response to tax rates (Chetty *et al.* , 2009; Reck, 2016; Aghion *et al.* , 2018; Rees-Jones & Taubinsky, 2018).[5]

[4]Tourek (2018) documents another form of suboptimal taxpayer behavior—firms reporting identical amounts in their income tax year after year—in neighboring Rwanda.

[5]Chetty *et al.* (2009); Aghion *et al.* (2018); Benzarti (forthcoming); Gillitzer & Skov (2018); Rees-Jones & Taubinsky (2018) provide direct evidence of tax-reporting mistakes by *individuals*. Like this paper, Aghion *et al.* (2018) show evidence that more sophisticated taxpayers tend to react as theory predicts to tax incen-

Finally, we show that the VAT is far from self-enforcing in low state capacity settings. Building on work studying how policy should be tailored to context (Laffont, 2005; Best *et al.* , 2015, 2019; 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).[6] 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 revenue on domestic vs. international 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 similar to 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 design of the Ugandan VAT is relatively standard, with a general rate of 18 percent, a credit-invoice system and standard exemptions (e.g., financial services) and zero-rating (e.g., exports). Appendix A.1 provides details.

Since 2012 all Ugandan VAT-firms must file their monthly VAT declarations electronically, within 15 days of the transaction month ending.[7] These must include detailed transaction-level records—spreadsheets listing each sale to and purchase from other VAT-registered firms. This implies that, every month, the Uganda Revenue Authority (URA) receives two reports for each transaction between any two VAT-firms.

tives, while less sophisticated taxpayers do so to a lesser extent.

[6] 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, China, 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; Naritomi, 2019; Fan *et al.* , 2019).

[7] 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' declarations between 2013 and 2016.[8] The monthly VAT data contain information at the firm level, including a scrambled firm Tax Identification Number (TIN), the declaration date, total sales/purchases (amount and VAT charged/paid), and total VAT liabilities. The tax return also contains data from the spreadsheets, called VAT "schedules", detailing each transaction. These include the transaction date, the seller and buyer TINs, the transaction value, and the VAT charged or paid. Schedule 1 (VS1) contains all sales transactions to other VAT-registered firms. Sales to final consumers or non-VAT firms are recorded only as a monthly aggregate. Schedules 2, 3, and 4 contain domestic input purchases, imports, and administrative expenses. Importantly, the transaction-level information reported in the VAT schedules is consistent with the firm-level reports in 97 percent of cases, suggesting that the transaction-level records constitute meaningful paper trails.

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.[9]

The data on imports comes from customs declarations submitted to the URA between 2012 and 2016. These are transaction-specific, submitted electronically, and include the value of the goods imported, the type and number of items, and the date of import. The TIN of the importer allows us to match the customs data to the domestic VAT data. 9,998 VAT-registered firms import at least once. More information is in Appendix A.1.

3 Discrepancies in VAT Declarations

In this section, we document massive VAT reporting discrepancies in Uganda at the seller-buyer-month level.

3.1 Conceptual background

For a date j transaction, 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). We aggregate transactions at the monthly level and 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

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

[9] Out of 22,388 firms, 19,137 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.

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 financial gain for one or both firms, 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 financial *loss* for one or both firms.

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

3.2 Discrepancies

Ugandan firms' average monthly reported VAT liability for the 2013-2016 period is slightly negative, and the median is zero, as is common in developing countries (Lemgruber *et al.* , 2015; Pomeranz, 2015). While only 15 percent of firms report negative or zero value-added in a full fiscal year, the reported VAT *liability* is zero or negative for 52 percent of firms (see Table A.1). This proportion is quite similar across VAT-registered firms of different sizes. Many firms can report positive value-added but zero or negative VAT liability because the latter includes offsets carried over from previous months. Since refund claims are restricted, offsets are typically carried-over.

We observe seller shortfall in 47 percent and buyer shortfall in 32 percent of seller-buyer-month observations. Figure 1 provides a graphical illustration of these discrepancies. The vertical axis measures the (inverse hyperbolic sine of the) total monthly amounts declared by sellers, and the horizontal axis that of the total monthly amounts declared by buyers. The data are grouped into a grid where the color of each square represents the number of observations, going from 1 (lightest gray) to more than 50,000 (black). Squares on the 45-degree line correspond to observations where seller and buyer-reported amounts match. Observations above (below) that line correspond to cases of buyer (seller) shortfall. Sellers and buyers report the same amount in only 21 percent of the observa-

tions.

We observe these widespread discrepancies despite taking a number of steps to avoid detecting false discrepancies. First, we use transaction dates rather than seller/buyer filing dates. Second, we minimize mismatched transactions by (i) 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, only with a one or two-month lag. Finally, we allow for rounding errors of 1,000 Ugandan Shillings (about USD 0.30).

The dashed curve in Figure 1 shows the average amount reported by sellers for different values of the buyer-reported amounts. The curve lies systematically below the 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 increases with the transaction amount, suggesting that the fraction of the transaction amount unreported is higher for larger transactions.

4 Classifying Firms' Reporting Behavior

In this section we show that most Ugandan firms engage in strategic tax reporting behavior, taking into account the country's low-enforcement environment, as economic theory predicts. We also show that, in contrast, a sizeable minority makes costly reporting mistakes. To do this we evaluate whether firms underreport their value-added—sales minus purchases—such that their liability falls, or erroneously overreport their value-added.

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:

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

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; δ_c is a constant and $r_{ff't}$ is a residual error term. Since $d_{ff't}$ is the nominal value of the discrepancy, δ^s_f can be interpreted as a firm's average discrepancy as a seller, in monetary terms, controlling for all time-invariant characteristics of

its clients, such as average size and reporting reliability. Similarly, $\delta^b_{f'}$ can be interpreted as a firm's average contribution to discrepancies as a buyer, controlling for all time-invariant characteristics of its sellers.[10]

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. We thus restrict our analysis to this largest connected set of firms.

4.2 Firm-level reporting behavior

We now formalize our classification of a firm's reporting behavior. We construct a firm-level discrepancy measure, Q_f, adding up the two estimated fixed effects for firm f:

$$Q_f \equiv w_s \cdot \hat{\delta}^s_f + w_b \cdot \hat{\delta}^b_f, \tag{2}$$

where the weights w_s and w_b represent the number of firm-trading partner monthly observations as a seller or buyer, respectively.[11] A firm engages in *advantageous* misreporting behavior if $Q_f > 0$, meaning that it reports in a way that reduces its aggregate 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 misreporting is one for which $w_s \cdot \hat{\delta}^s_f \geq 0$ and $w_b \cdot \hat{\delta}^b_f \geq 0$. 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 misreporting is one for which $w_s \cdot \hat{\delta}^s_f \geq 0$ and $w_b \cdot \hat{\delta}^b_f < 0$. This implies that the firm underreports its sales and underreports its purchases. Finally, a firm engaging in *looking-big* advantageous misreporting is one for which $w_s \cdot \hat{\delta}^s_f < 0$ and $w_b \cdot \hat{\delta}^b_f \geq 0$, thus overreporting its sales and its purchases.

The top panel of Table 1 shows the resulting classification of firms. We find that none of the 19,137 Ugandan VAT-eligible firms report consistently on average, while 14,026 firms

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

[11] More precisely, $\hat{\delta}^s_f = \hat{\delta}^{s'}_f + \delta_c$ and $\hat{\delta}^b_f = \hat{\delta}^{b'}_f + \delta_c$ where $\hat{\delta}^{s'}_f$ and $\hat{\delta}^{b'}_f$ are the fixed effects estimated in (1). By adding the mean discrepancy (δ_c) to the deviations from the mean ($\hat{\delta}^{s'}_f$ and $\hat{\delta}^{b'}_f$), $\hat{\delta}^s_f$ and $\hat{\delta}^b_f$ give us each firm's reporting discrepancies as a seller (respectively, a buyer) controlling for trading partners' effect and time variations.

(73 percent) are *advantageous* misreporters. These results suggest that when the VAT is implemented in a low-state capacity context without systematic cross-checks, the majority of firms misreport in order to lower their VAT liability. Of the firms that misreport in an advantageous way, 76 percent are conspicuous advantageous misreporters, only 11 percent are looking-small advantageous misreporters, and the remaining 13 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.[12]

However, we also find that 5,111 firms (27 percent) misreport in a *disadvantageous* way. A substantial share of firms thus simply make reporting errors, perhaps due to poor accounting practices or lack of understanding of tax incentives.[13] This result underscores the importance of accounting for heterogeneity in firm sophistication in theory and policy design.

Advantageous and disadvantageous misreporting occurs with comparable frequency among smaller, medium-sized, and somewhat larger VAT-registered firms. In particular, as seen in Figure A.1, the average Q_f measure is similarly distributed across most of the distribution of firm size. However, the figure 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 two types of firms is in Appendix A.2.

4.3 Robustness analysis

The methodology we develop—starting with the two-way fixed effects regression (1) and thereafter classifying firm types using the resulting Q_f measure—allows us to shed new light on the sophistication of firms' decision-making, but has limitations.

A first limitation is that we restrict attention to *unilateral* misreporting. A second is that we define strategic behavior to include only systematic misreporting that benefits the firm financially. If some systematic reporting errors lower the firm's VAT liability, our estimate of disadvantageous misreporting is a lower bound of the extent of erroneous reporting behavior.

[12]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).

[13]Given that negative liabilities can be carried over to later months, one example of disadvantageous misreporting is not bothering to include all input purchases in the firm's tax declaration when its true liability is negative. We find, in fact, that firms classified as disadvantageous misreporters—especially those with a negative buyer fixed effect—are 19 percent less likely to file a VAT return with a negative liability, but 21 percent more likely to file a null return (Table A.2). This is of course just one example of disadvantageous misreporting behavior.

A third limitation is that we do not observe misreporting of sales to final consumers. If firms we classify as disadvantageous misreporters—those that overreport their firm-to-firm sales or underreport their inputs—also underreport a large enough share of sales to final consumers, their total misreporting may in principle be advantageous. To address this concern, we re-estimate our model assuming that all firms underreport a given proportion of their sales to final consumers. As seen in Table 2, the proportion of advantageous firms increases to 76 percent when we assume that all firms underreport final sales by 10 percent. Even assuming an implausibly high[14] degree of misreporting of sales to final consumers—50 percent—the share of disadvantageous firms is about 19 percent.

A final concern is the potential influence of sampling error on fixed effect estimates of objects of interest (Lancaster, 2000). First note that, in our setting, each additional firm observed yields more observations in both of the two fixed effects dimensions (buyer and seller), since each firm trades with others and is itself both a seller and buyer. Therefore the estimated fixed effects are arguably asymptotic both in N and T, instead of only in T, as is usually the case.[15]

To investigate sampling error concerns empirically, we estimate 95% confidence intervals around each of our point estimates for Q_f, using a bootstrapping routine to estimate the covariance between each firm's seller and buyer fixed effect. Then we classify firms into two groups based on their reporting variability. In order to define low variability, we take as a benchmark the subset of firms that display a discrepancy of zero in 60 percent or more of their trading partner-month observations in the raw data. Firms with a confidence interval smaller than the average of this subset of firms are considered to have low reporting variability (7,900 firms), and the rest are considered to have high reporting variability (11,237 firms). For these two subgroups, we reassess the fraction of firms that are advantageous or disadvantageous misreporters, and we label as "neutral" any firm whose confidence interval includes zero (see Table A.3). The share of neutral firms is large in both subgroups (about 88 percent), but the relative proportion of advantageous and disadvantageous firms is notably stable around 3 in the two groups—close to the 2.74 proportion found in our main estimates in Table 1.

As an alternative robustness test, we 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 72 (resp., 63) percent of firms classified as advantageous (disadvantageous) mis-

[14] 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).

[15] This distinguishes our setting from traditional applications of the Abowd *et al.* (2002, 1999) methodology to employer-employee data, where the two fixed effects dimensions are units of different nature.

reporters in year t stay within that classification also in the subsequent year. This suggests that we primarily capture systematic components of firms' reporting behavior that persist across years. It also suggests that disadvantageous behavior is somewhat less persistent over time than advantageous behavior.

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 60 percent of such mismatch transactions involve a "seller shortfall". This suggests that, in aggregate, the revenue consequences of VAT misreporting for the Ugandan government 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 each reporting discrepancy $d_{ff't}$ between the two firms using the 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 in proportion to these. 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.

Our estimates imply that the Ugandan government would have lost USD 138 million in tax revenues during 2013-2016 if (only) disadvantageous misreporting were eliminated, as seen in the bottom rows of Table 3. If (only) advantageous misreporting were eliminated, our estimates imply a revenue *gain* of about USD 522 million. If both forms of misreporting were eliminated, our estimates imply a revenue gain of USD 383 million, or about 28 percent of the total VAT collected.[16] These estimates 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.

[16]Many Ugandan firms have positive outstanding balances with the URA. This 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 gain from eliminating respectively disadvantageous and advantageous misreporting.

5 Enhanced Enforcement Capacity and VAT Evasion by Strategic and Confused Firms

In this section we show that both sophisticated and less sophisticated firms misreport less when the state's tax enforcement capacity is greater, but that less sophisticated firms adjust their behavior to a lesser extent. To do so, we leverage the fact that imports are subject to greater oversight than domestic transactions.

Our empirical exercise is closely related to the discrepancy analysis in Section 4. When Ugandan firms file for customs clearance of an import transaction, they are required to pay the VAT on the imported goods plus tariffs. Later, they should declare the input VAT paid on imports on their VAT "schedules" in order to obtain the corresponding tax credit. We thus compare *a given firm's* double reports of the same transaction amount, one at customs and one in domestic VAT returns. We do so in the sample of firms that report at least one import transaction between and for which we estimate seller and buyer fixed effects in the analysis of domestic transactions (N=9,318). Observations are at the firm-month level.

We find that the same amount is reported at customs and in the firm's VAT declaration in 53 percent of observations. In 14 percent of cases, the firm claims a larger amount in VAT credit than what it reported at customs, thus reducing the firm's VAT liability. This self-advantageous misreporting is less frequent than occurences of seller shortfall in domestic transactions, in line with the intuition that many firms adjust their behavior to the state's enforcement capacity.

In the remaining 34 percent of observations, firms report a lower amount in their VAT declaration than at customs, thus leaving input tax credit unclaimed. This behavior—we label it *seemingly anomalous*—is analogous to buyer shortfall discrepancies in domestic VAT transactions, with the difference that here, the same firm makes both tax declarations. Seemingly anomalous underclaiming of input tax credit from imported goods may reflect disadvantageous behavior. Suggestive evidence that this is part of the explanation comes from cross-sectional and time variation in such cases. First, monthly VAT returns reporting a null tax liability are 22 percentage points more likely to display seemingly anomalous import reporting than returns with a positive VAT liability, perhaps because some firms with a null VAT liability do not bother claiming input VAT credits from imports (see Table 4). (Footnote 13 discusses a similar form of disadvantageous misreporting in domestic VAT reporting). Second, seemingly anomalous reporting is less frequent in the early and final months of each fiscal year, when tax matters may be more salient to taxpayers (see Table A.4). However, seemingly anomalous reporting may also represent strategic behavior. There is for example anecdotal evidence that some goods are imported

into Uganda by businesses even though they are destined for final consumption by individuals. Because these are not actual business inputs, they do not generate input VAT credits and are legitimately not reported as such.

To investigate, we compare transaction amounts reported at customs and in domestic VAT declarations separately for firms that we classified as advantageous and disadvantageous misreporters based on (purely) domestic VAT transactions in Section 4. In Table 4 the outcome variable is a dummy variable that is equal to one for monthly observations with seemingly anomalous reporting as defined above.[17]

We find that disadvantageous misreporters and firms with a negative buyer fixed effect are respectively 4.5 and 8.2 percentage points (13 and 24 percent) more likely to engage in seemingly anomalous reporting of imports than other firms.[18] These estimates, shown in columns 1 and 3 of Table 4, help validate the classification procedure in Section 4 and imply financially irrational behavior by firms, especially those that engage in disadvantageous VAT (mis)reporting.

We find no statistically significant difference in self-advantageous misreporting of imports between advantageous and disadvantageous misreporters (see Table A.5). Both types of firms appear to adjust their behavior to the verifiable nature of imported inputs and engage in less self-advantageous misreporting of imports than self-advantageous misreporting of domestic transactions.

Overall, the results in this section indicate that strategic firms misreport less when the state's tax enforcement capacity is greater, while confused firms do so to a lesser extent.

6 Conclusion

In this paper we study the sophistication of firms' decision-making, using tax reporting behavior in a low-enforcement setting—Uganda—as a lab for analyzing the extent to which firms make decisions that benefit themselves. We document widespread discrepancies between seller and buyer VAT 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—about a quarter—consistently misreports such that their tax liability increases.

[17] We allow for rounding errors and pure timing mismatches, in the same way as in Section 4. We also control for firm size (deciles of reported annual turnover) and sector in all specifications.

[18] These estimates remain essentially unchanged when we control for null VAT reported or include dummies for the type of goods being imported, as seen in columns 2 and 4. Note also that the coefficient for firms with a negative seller fixed effect is negative and smaller in magnitude, although statistically significant.

In the second part of the paper, we show that firms classified as strategic and confused—advantageous and disadvantageous misreporters—respond differently to the state's tax enforcement capacity. All firms misreport less at customs where goods are subject to greater monitoring, but confused firms are more likely to underreport their input tax credit for imported goods on their VAT returns, leaving tax credits on the table. We interpret these findings as indicating that (i) the proportion of firms that do not engage in sophisticated optimization as usually assumed is high—with important implications for theory and policy—but (ii) the majority of firms nevertheless respond to low state capacity by evading taxes. Together, these two conclusions call into question the self-enforcement properties of the VAT in limited enforcement contexts.

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*. Census Bureau Technical paper No. TP 2002-06.

ACEMOGLU, DARON, GARCIA-JIMENO, CAMILO, & ROBINSON, JAMES A. 2015. State Capacity and Economic Development: A Network Approach. *American Economic Review*, **105**(8), 2364–2409.

AGHION, PHILIPPE, AKCIGIT, UFUK, , 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.

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

BENZARTI, YOUSSEF. forthcoming. How Taxing Is Tax Filing? Using Revealed Preferences to Estimate Compliance Costs. *American Economic Journal: Economic Policy*.

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.

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. 2019. *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, KROFT, KORY, & LOONEY, ADAM. 2009. Salience and Taxation: Theory and Evidence. *American Economic Review*, **99**(4), 1145–1177.

DELLAVIGNA, STEFANO, & GENTZKOW, MATTHEW. 2019. Uniform Pricing in US Retail Chains. *Quarterly Journal of Economics*, **134**(4), 2011–2084.

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*, **86**(6), 2123–2160.

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.

FAN, HAICHAO, QIAN, NANCY, WEN, JAYA, & YU, LIU. 2019. *The Dynamic Effects of Computerized VAT Invoices on Chinese Manufacturing Firms*. NBER Working Paper No. 24414.

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*, **17**(3), 876–916.

HJORT, JONAS, LI, XUAN, & SARSONS, HEATHER. 2020. *Across-Country Wage Compression in Multinationals*. NBER working paper no. 26788.

HORTACSU, ALI, LUCO, FERNANDO, PULLER, STEVEN L., & ZHU, DONGNI. 2019. Does Strategic Ability Affect Efficiency? Evidence from Electricity Markets. *American Economic Review*, **109**(12), 4302–4342.

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*. International Monetary Fund Fiscal Affairs Department.

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, HENRIK. 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.

LANCASTER, TONY. 2000. The Incidental Parameter Problem since 1948. *Journal of Econometrics*, **95**, 391–413.

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.

NARITOMI, JOANA. 2019. Consumers as Tax Auditors. *American Economic Review*, **109**(9), 3031–3072.

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.

RECK, DANIEL. 2016. *Taxes and Mistakes: What?s in a Sufficient Statistic?* Mimeo LSE.

REES-JONES, ALEX, & TAUBINSKY, DMITRY. 2018. Attention Variation and Welfare: Theory and Evidence from a Tax Salience Experiment. *Review of Economic Studies*, **85**(4), 317–330.

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.

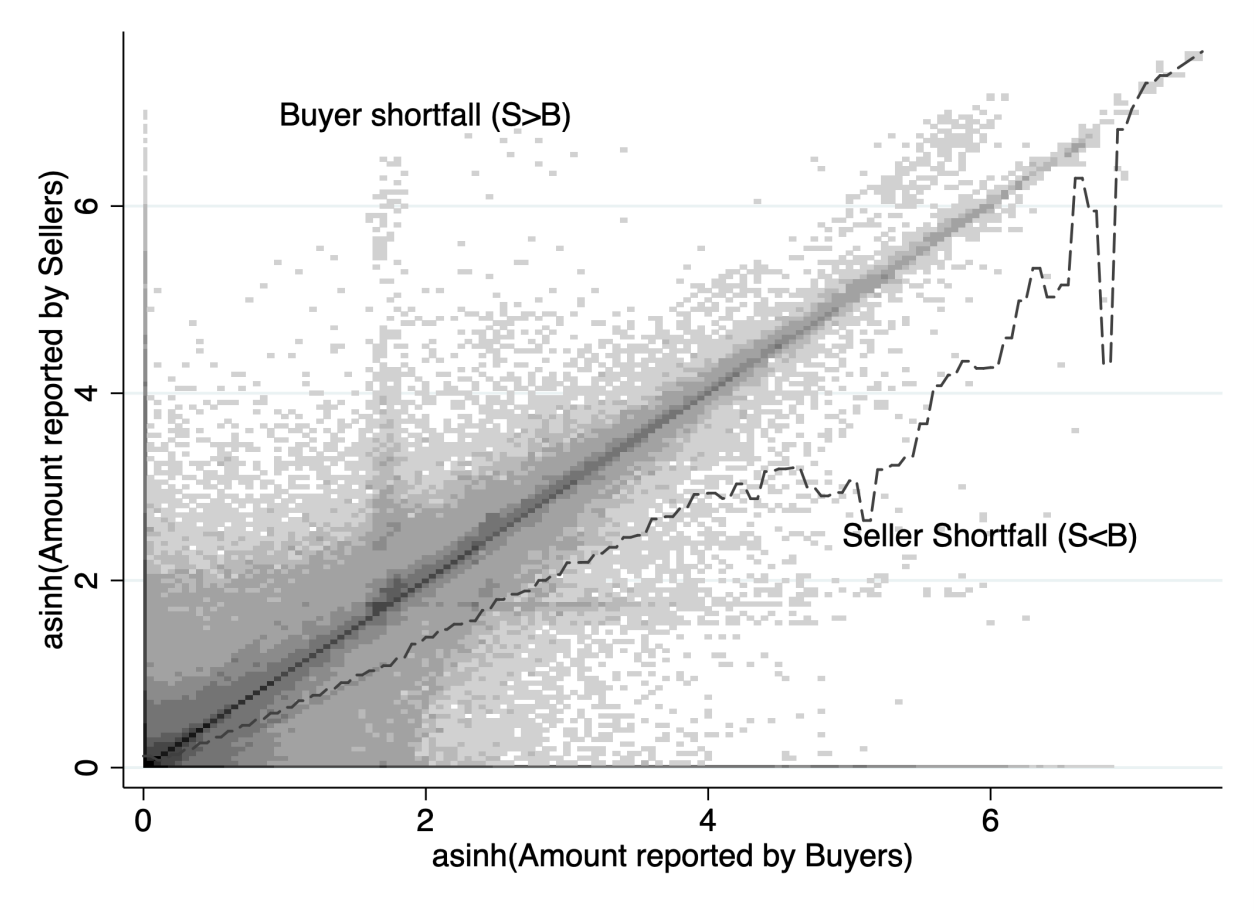
UGANDA REVENUE AUTHORITY. 2016. *Domestic Tax Laws of Uganda: Updated and Tracked Compendium*. Government of Uganda.

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: This figure plots the inverse hyperbolic sine (asinh) transformation of amounts reported by sellers over that by buyers for all monthly transaction data in fiscal years 2013-2016. The data are grouped into a 0.05×0.05 grid and the color represents the number of observations in each square, going from 1 (lightest gray) to more than 50,000 (black). Squares on the 45-degree line correspond to observations where seller and buyer-reported amounts match. Observations above that line correspond to cases of buyer shortfall, while those below indicate cases of seller shortfall. The dashed line represents the conditional mean of asinh(Amount reported by sellers) for the values of asinh(Amount reported by buyers). Data source: VAT Schedules data for fiscal years 2013-2016.

Tables

TABLE 1
FIRM TYPE CLASSIFICATION BASED ON Q STATISTIC

	No. of Firms	Share of firms
Advantageous	14,026	0.73
Conspicuous	10,732	0.56
Looking small	1,503	0.08
Looking big	1,791	0.09
Disadvantageous	5,111	0.27
Ratio of Adv. to Disadv.		2.74
N	19,137	

Notes: Data Source: VAT Schedules and Monthly Summary data for fiscal years 2013-2016. Firm types are defined based on $Q(f)$, which is calculated as the weighted sum of the estimated firm-as-buyer fixed effect and firm-as-seller fixed effect, i.e., : $Q(f) = w_s \cdot \hat{\delta}_b + w_b \cdot \hat{\delta}_s$. w_s (respectively, w_b) is the number of firm-trading partner monthly obervations as a seller (resp., as a buyer). (1) **Advantageous**: $Q(f) > 0$. Advantageous firms are further categorized into: (1a) Conspicuous Advantageous: $w_s \cdot \hat{\delta}_s(f) \geq 0$ and $w_b \cdot \hat{\delta}_b(f) \geq 0$; (1b) Looking small Advantageous: $w_s \cdot \hat{\delta}_s(f) \geq 0$ and $w_b \cdot \hat{\delta}_b(f) < 0$; and (1c) Looking big Advantageous: $w_s \cdot \hat{\delta}_s(f) < 0$ and $w_b \cdot \hat{\delta}_b(f) \geq 0$. (2) **Disadvantageous**: $Q(f) < 0$.

TABLE 2
FIRM TYPES ASSUMING UNDERREPORTING OF SALES TO FINAL CONSUMERS

	Panel A 10% of sales to FC		*Panel B* 30% of sales to FC		*Panel C* 50% of sales to FC	
	No. of Firms	Share of firms	No. of Firms	Share of firms	No. of Firms	Share of firms
Disadvantageous	4,508	0.24	3,959	0.21	3,605	0.19
Advantageous	14,629	0.76	15,178	0.79	15,532	0.81
Conspicuous	11,568	0.60	12,120	0.63	12,445	0.65
Looking small	1,861	0.10	2,226	0.12	2,481	0.13
Looking big	1,200	0.06	832	0.04	606	0.03

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. Firm types are defined based on $Q(f)$, which is calculated as the weighted sum of the estimated firm-as-buyer fixed effect and firm-as-seller fixed effect, i.e., : $Q(f) = w_s \cdot \hat{\delta}_b + w_b(\cdot\hat{\delta}_s + FC)$. w_s (respectively, w_b) is the number of firm-trading partner monthly obervations as a seller (resp., as a buyer). FC indicates average monthly unreported sales to final consumers: in Panel A, we consider that sellers do not report 10 percent of their sales to final consumers, in Panel B, 30 percent, in Panel C, 50 percent. (1) **Advantageous**: $Q(f) > 0$. Advantageous firms are further categorized into: (1a) Conspicuous Advantageous: $w_s \cdot \hat{\delta}_s(f) \geq 0$ and $w_b \cdot \hat{\delta}_b(f) \geq 0$; (1b) Looking small Advantageous: $w_s \cdot \hat{\delta}_s(f) \geq 0$ and $w_b \cdot \hat{\delta}_b(f) < 0$; and (1c) Looking big Advantageous: $w_s \cdot \hat{\delta}_s(f) < 0$ and $w_b \cdot \hat{\delta}_b(f) \geq 0$. (2) **Disadvantageous**: $Q(f) < 0$.

TABLE 3
REVENUE CONSEQUENCES BY FIRM TYPE

	All	(1) Disadv.	(2) Adv.	(2a) Conspic.	(2b) Looking Small	(2c) Looking Big
No. of distinct firms	19,137	5,111	14,026	10,732	1,503	1,791
Percentage of all firms	(100%)	(27%)	(73%)	(56%)	(8%)	(9%)
Total net VAT due	1,553,971	669,721	884,250	558,794	113,157	212,299
Seller shortfall						
Number of distinct firms with seller shortfall	17,249	4,258	12,991	9,751	1,482	1,758
Total net VAT due from firms with seller shortfall	1,275,917	575,195	700,722	438,926	95,317	166,479
Total VAT subject to seller shortfall	899,736	101,761	797,975	349,518	398,076	50,381
Buyer shortfall						
Number of distinct firms with buyer shortfall	17,979	4,804	13,175	9,918	1,482	1,775
Total net VAT due from firms with buyer shortfall	1,316,813	612,195	704,618	439,889	96,937	167,791
Total VAT subject to buyer shortfall	727,354	419,678	307,675	148,332	51,837	107,506
Correcting seller shortfall and buyer shortfall						
Impact on total net VAT due	383,730	−138,421	522,151	205,956	326,524	−10,328
Percentage of total VAT collected	28.2%	−10.2%	38.3%	15.1%	24.0%	−0.8%

Notes: Data Source: VAT Schedules and Monthly Summary data for fiscal years 2013-2016. 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. Shortfall is assigned using firms' estimated fixed effects, see Appendix A.3 for details. The first column shows results for the whole sample, while Columns (1) to (2c), firms are divided into sub-types based on their $Q(f)$ statistic. $Q(f) = w_s \cdot \hat{\delta}_b + w_b \cdot \hat{\delta}_s$. w_s (respectively, w_b) is the number of firm-trading partner monthly obervations as a seller (resp., as a buyer). (1) **Disadvantageous**: $Q(f) < 0$. (2) **Advantageous**: $Q(f) > 0$. Advantageous firms are further categorized into: (2a) Conspicuous Advantageous: $w_s \cdot \hat{\delta}_s(f) \geq 0$ and $w_b \cdot \hat{\delta}_b(f) \geq 0$; (2b) Looking small Advantageous: $w_s \cdot \hat{\delta}_s(f) \geq 0$ and $w_b \cdot \hat{\delta}_b(f) < 0$; and (2c) Looking big Advantageous: $w_s \cdot \hat{\delta}_s(f) < 0$ and $w_b \cdot \hat{\delta}_b(f) \geq 0$. All values are in thousands of USD.

TABLE 4
NON SELF BENEFICIAL BEHAVIOR AT CUSTOMS AND FIRM TYPE

	Dep.Var.: non-self beneficial (NSB) reporting			
Firm Type	(1)	(2)	(3)	(4)
Disadvantageous	0.045***	0.041***		
	(0.010)	(0.009)		
Null VAT		0.220***		0.213***
		(0.012)		(0.012)
Negative Buyer FE			0.082***	0.069***
			(0.009)	(0.009)
Negative Seller FE			-0.007	0.001
			(0.009)	(0.008)
Month-Year FE	Yes	Yes	Yes	Yes
Size and Sector FE	Yes	Yes	Yes	Yes
HS Share of Import	No	Yes	No	Yes
N	123303	123303	123303	123303
R2	0.03	0.07	0.03	0.07
Mean of dep.	0.34	0.34	0.34	0.34

Notes: Data source: VAT Schedule 3, MVR and Customs data for fiscal years 2013-2016. This regression analyzes whether Disadvantageous firms, and firms which have a negative seller (buyer) fixed-effect are more likely to behave in a non-self beneficial way at customs. Observations are at the firm-month level. The dependent variable is a dummy equal to one if the firm claims lower VAT amounts incurred on imports in VS3 than VAT paid on imports recorded in the Customs data in the same month. We allow for 1,000 UGX rounding and for pure timing mismatches. In Columns (1) and (2), the explanatory variable of interest is a (time invariant) dummy for firm type, equal to one if the firm is classified as Disadvantageous, based on the value of $Q(f)$, as explained in Section 4.4. In Columns (3) and (4), the explanatory variables of interest are dummies equal to one if the buyer (resp. seller) fixed-effect estimated for the firm as described in Section 4 is negative. In all specifications, we control for firm size as measure by annual decile of reported turnover, and for firm sector. In Columns (2) and (4), we additionally control for null monthly VAT liability as reported in MVR, and for the type of goods imported as measured by dummies for each of the 21 HS Good Code Sections, equal to one if the firm imports at least one good from the corresponding section. 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).

[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.

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 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.

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.

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:

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

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

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. We also restrict the sample to firms that appear at least in two consecutive years. Table A.7 shows the results as a transition matrix laying out firms' classification in year $t+1$ conditional on their year t classification. We find that 72 percent of advantageous firms and 63 percent of disadvantageous firms stay within their classification in the following year.

A.2.3 Robustness

We perform two robustness checks by varying the sample used for the two-way fixed effect estimation. First, we 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 19,137 to 18,629. The results are shown in Panel A of Table A.8. They are similar to what we obtained when running the regression without controls: 75 percent of firms are classified as Advantageous (against 73 percent in the main analysis) and 25 percent are classified as Disadvantageous. Among the Advantageous firms, the respective shares of Conspicuous, Looking-small and Looking-Big are very similar to the ones in the main analysis.

Second, we replicate the analysis on a more consistent sample, keeping only firm-pairs with a number of observations larger than ten, to assess the extent to which this affects the results. The firm classification is displayed in Panel B of Table A.8. The share of Advantageous firms increases to 88 percent. Among advantageous firms, a larger share are classified as conspicuous – 91 percent, against 77 percent in the main analysis.

A.3 Firm type classification and revenue consequence computation

A.3.1 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 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.

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

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.3.2 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:

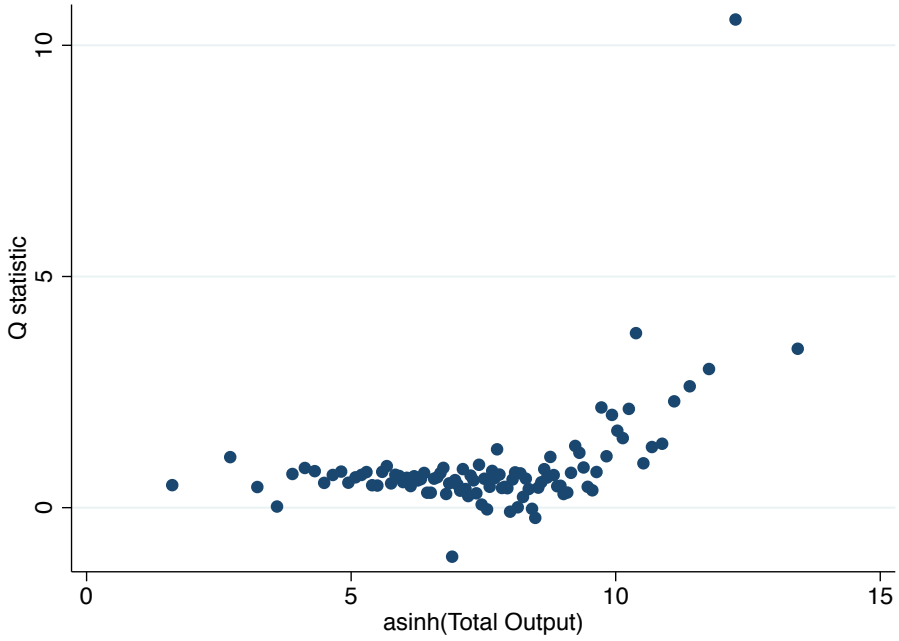
$$d_{1t} = \begin{cases} \frac{d_{12t} + (\hat{\delta}_1^b - \hat{\delta}_2^s)}{2}, & \text{if } \mid d_{12t} \mid > \mid \hat{\delta}_1^b - \hat{\delta}_2^s \mid . \\ d_{12t} \frac{\max(\hat{\delta}_1^b - \hat{\delta}_2^s, 0)}{\hat{\delta}_1^b - \hat{\delta}_2^s}, & \text{if } \mid d_{12t} \mid \leq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

In Column 2 of Table A.9, we report the revenue consequence calculations using the approach described above. The revenue loss due to misreporting remains of the same order of magnitude as in our baseline approach: the adjusted revenue implications amount to 27 percent of VAT revenue over the whole time period, against 32 percent in the main approach.

[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.

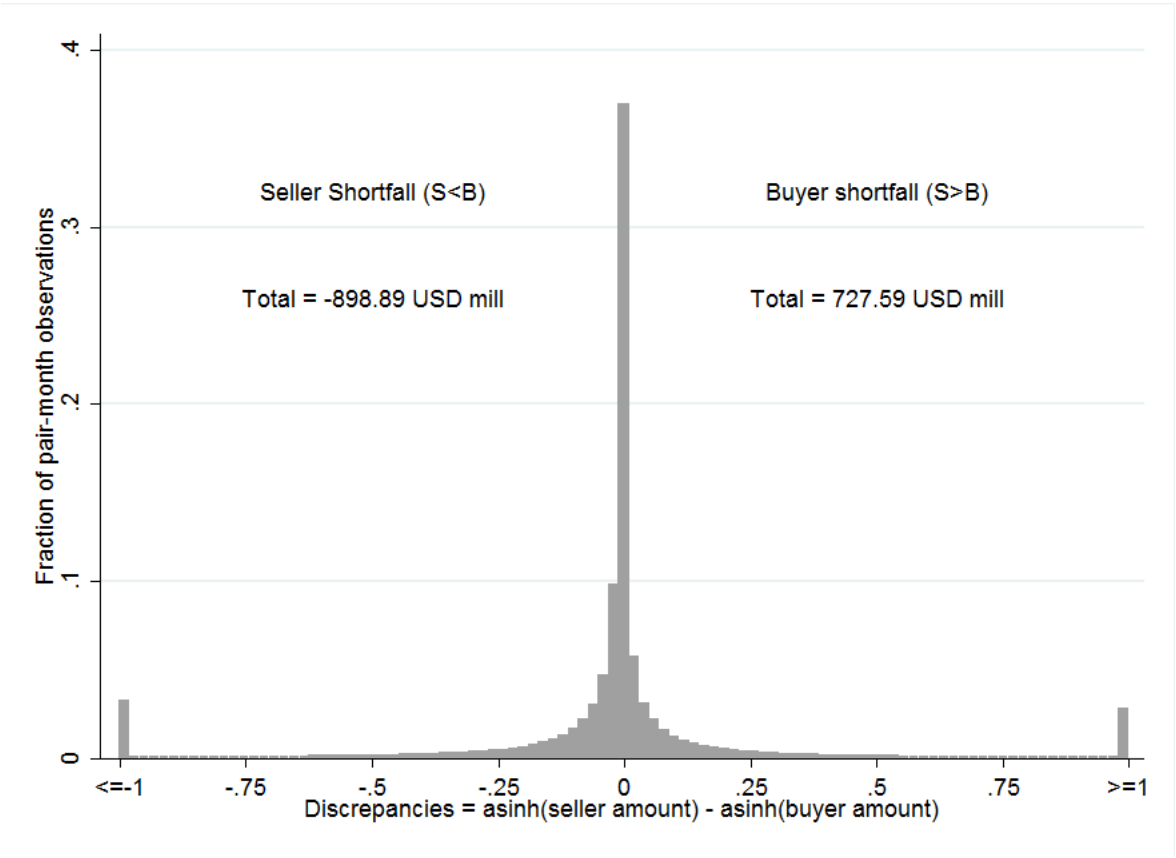
Figures

FIGURE A.1
Q STATISTIC OVER SIZE.



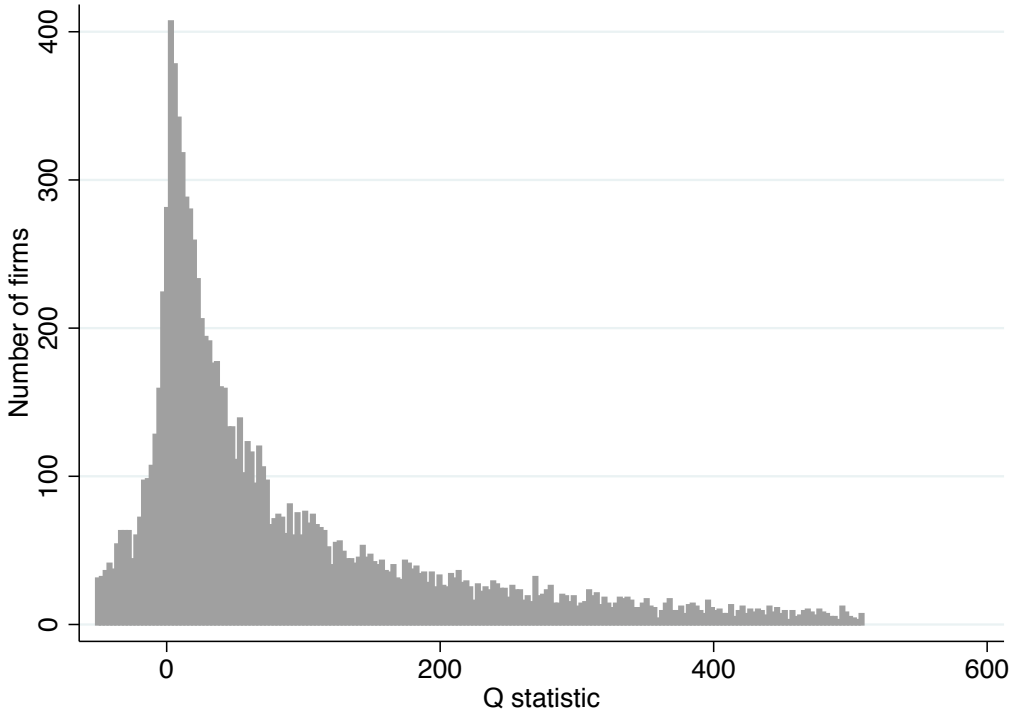
Notes: In this Figure, we plot firms' estimated Q statistic (Q^f in Equation (2)) over the inverse hyperbolic sine transformation of firms' total output in the estimation period. Data source: VAT Schedules and Monthly Summary data for fiscal years 2013-2016.

FIGURE A.2
DISTRIBUTION OF REPORTING DISCREPANCIES IN THE DOMESTIC VAT



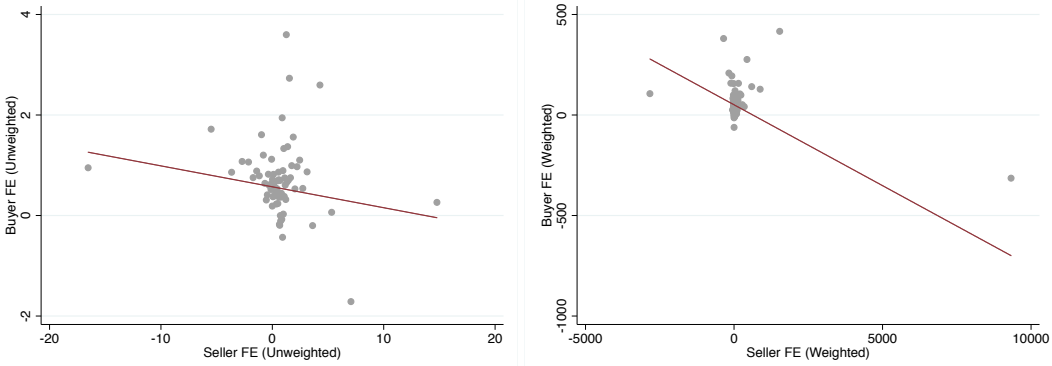
Notes: In this Figure, we show the distribution of discrepancies in the reporting of transactions by sellers and buyers for fiscal years 2013-2016. Data source: VAT Schedules data. Calculated by taking the difference between VAT charged in VS1 and VAT paid in VS24. We use the inverse hyperbolic sine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

FIGURE A.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 firms' estimated buyer fixed effect over their estimated seller fixed effect. Observations are averaged by bins of seller fixed effect. In the left panel, we use the unweighted seller and buyer fixed effects. In the right panel, the fixed effects obtained in the two-way fixed effect estimation are weighted by the number of firm-trading partner monthly observations respectively as a seller and as a buyer. Data source: VAT Schedules data for fiscal years 2013-2016.

Tables

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

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

	Panel A: Low variability		
	No. of Firms	Share of all firms	Share of low var. firms
Advantageous	403	0.02	0.05
Conspicuous	344	0.02	0.04
Looking small	24	0.00	0.00
Looking big	35	0.00	0.00
Disadvantageous	168	0.01	0.02
Neutral	7,329	0.38	0.93
Ratio of Adv. to Disadv.		2.40	
N	7,900		
	Panel B: High variability		
	No. of Firms	Share of all firms	Share of low var. firms
Advantageous	1,347	0.07	0.12
Conspicuous	1,088	0.06	0.10
Looking small	110	0.01	0.01
Looking big	149	0.01	0.01
Disadvantageous	383	0.02	0.03
Neutral	9,507	0.50	0.85
Ratio of Adv. to Disadv.		3.52	
N	11,237		

Notes: Data Source: VAT Schedules and Monthly Summary data for fiscal years 2013-2016. Firm types are defined based on $Q(f)$, which is calculated as the weighted sum of the estimated firm-as-buyer fixed effect and firm-as-seller fixed effect, i.e., : $Q(f) = w_s \cdot \hat{\delta}_b + w_b \cdot \hat{\delta}_s$. w_s (respectively, w_b) is the number of firm-trading partner monthly obervations as a seller (resp., as a buyer). (1) **Advantageous**: $Q(f) > 0$. Advantageous firms are further categorized into: (1a) Conspicuous Advantageous: $w_s \cdot \hat{\delta}_s(f) \geq 0$ and $w_b \cdot \hat{\delta}_b(f) \geq 0$; (1b) Looking small Advantageous: $w_s \cdot \hat{\delta}_s(f) \geq 0$ and $w_b \cdot \hat{\delta}_b(f) < 0$; and (1c) Looking big Advantageous: $w_s \cdot \hat{\delta}_s(f) < 0$ and $w_b \cdot \hat{\delta}_b(f) \geq 0$. (2) **Disadvantageous**: $Q(f) < 0$. We compute the 95 percent confidence interval around $Q(f)$, for each firm. We run 100 iterations of the two-way fixed-effects model through a bootstrap routine. For each firm, we then compute $VAR(Q(f)) = w_s^2 \cdot VAR(\hat{\delta}_s(f)) + w_b^2 \cdot VAR(\hat{\delta}_b(f)) + 2 \cdot w_s \cdot w_b COV(\hat{\delta}_s(f), \hat{\delta}_b(f))$. We further classify firms based on whether $Q(f)$ displays **Low variability**, with a tight confidence interval (Panel A), or **High variability**, with a wide confidence interval (Panel B). A confidence interval is classified as tight if its length is smaller or equal to a benchmark value for neutral reporting. The benchmark value is the average length of the confidence interval obtained for the subset of neutral reporting firms (identified as firms that display a discrepancy of zero in 60 percent or more of their trading partner-month observations in the raw data). Both low and high variabily firms are classified as (3) **Neutral** if the 95 percent confidence interval includes zero.

TABLE A.2
FIRM-TYPE AND VAT MONTHLY LIABILITY

	Dep. Var.: VAT Liability					
Firm Type	Null (1)	Null (2)	Positive (3)	Positive (4)	Negative (5)	Negative (6)
Disadvantageous	0.042***		0.024***		-0.066***	
	(0.005)		(0.007)		(0.006)	
Negative Buyer FE		0.034***		0.028***		-0.062***
		(0.005)		(0.006)		(0.006)
Negative Seller FE		-0.028***		0.020***		0.008
		(0.005)		(0.007)		(0.007)
Month-Year FE	Yes	Yes	Yes	Yes	Yes	Yes
N	712927	712927	712927	712927	712927	712927
R2	0.02	0.02	0.00	0.00	0.01	0.01
Mean of dep.	0.19	0.19	0.46	0.46	0.35	0.35

Notes: Data source: VAT Schedules and Monthly Summary data for fiscal years 2013-2016. This table shows the results of the regression of monthly VAT liability on firm-type. In Columns (1) and (2) (resp., Columns (3) and (4), resp. Columns (5) and (6)), the dependent variable is a dummy equal to one if the VAT liability is null (resp., positive, resp. negative). In Columns (1), (3), (5), the regressor of interest is $Disadvantageous$, a dummy equal to one if the firm is categorized as Disadvantageous and zero otherwise. In Columns (2), (4), (6), the regressors of interest are two dummies, *Negative Buyer FE* and *Negative Seller FE*, equal to one if the firm's buyer (resp., seller) fixed-effect is equal to zero. We control for firm size in all specifications, with a categorical variable indicating whether a firm is classified as medium taxpayer (MTO), large taxpayer (LTO), or none (STO). Standard errors, clustered at the firm level, are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

TABLE A.4
NON-SELF BENEFICIAL CUSTOMS REPORTING BY VAT LIABILITY AND BY MONTH

	OLS					
Dependent variable:	NSB	NSB Extensive	NSB Intensive	NSB	NSB Extensive	NSB Intensive
Customs behavior	(1)	(2)	(3)	(4)	(5)	(6)
July	-0.013**	-0.020***	0.001			
	(0.006)	(0.005)	(0.007)			
August	-0.015**	-0.016***	-0.003			
	(0.006)	(0.005)	(0.007)			
September	-0.010*	-0.011**	-0.002			
	(0.006)	(0.005)	(0.007)			
October	-0.006	-0.007	-0.003			
	(0.006)	(0.005)	(0.007)			
November	-0.002	-0.003	-0.001			
	(0.006)	(0.005)	(0.007)			
January	-0.011**	-0.007	-0.007			
	(0.006)	(0.005)	(0.007)			
February	-0.040***	-0.017***	-0.032***			
	(0.006)	(0.005)	(0.007)			
March	-0.019***	-0.019***	-0.005			
	(0.006)	(0.005)	(0.007)			
April	-0.024***	-0.021***	-0.010			
	(0.006)	(0.005)	(0.007)			
May	-0.026***	-0.028***	-0.007			
	(0.006)	(0.005)	(0.007)			
June	-0.021***	-0.027***	-0.002			
	(0.006)	(0.005)	(0.007)			
Null VAT				0.220***	0.295***	0.102***
				(0.014)	(0.015)	(0.018)
Year FE	Yes	Yes	Yes			
Month-Year FE				Yes	Yes	Yes
Size and Sector FE	Yes	Yes	Yes	Yes	Yes	Yes
N	123304	123304	76510	123304	123304	76510
R2	0.02	0.06	0.01	0.04	0.09	0.01
Mean of dep.	0.34	0.20	0.23	0.34	0.20	0.23

Notes: Data source: VAT Schedule 3, MVR and Customs data for fiscal years 2013-2016. This regression analyzes whether the likelihood of non-self beneficial behavior at customs varies across months within the year, and with monthly VAT liability as reported in the MVR. Observations are at the firm-month level. The dependent variable in Columns (1) and (4) is a dummy equal to one if the firm claims lower VAT amounts incurred on imports in VS3 than VAT paid on imports as recorded in the Customs data for the same month. We allow for rounding of 1,000 UGX and for pure timing mismatches. In Columns (2) and (5), the outcome variable indicates non-self beneficial reporting on the extensive margin, equal to one if the firm reports nothing in VS3 for a month in which VAT paid on imports at customs is non-zero. In Columns (3) and (6), we restrict the sample to firm-month observations where a positive amount is reported both at Customs and in VS3, and the dependent variable is a dummy indicating non-self beneficial behavior on the intensive margin, equal to one if the VAT claimed in VS3 is lower than the VAT paid on imports as reported in Customs. In Columns (1) to (3), the explanatory variables are dummies for each month. The reference is November. Note that the fiscal year in Uganda runs from July to June. Months are based on invoice dates. In Columns (4) to (6), the explanatory variable of interest is a dummy equal to one if the VAT liability reported in the MVR is zero. In all specifications, we control for firm size as measure by annual decile of reported turnover, and for firm sector. Standard errors, clustered at the firm level, are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

TABLE A.5
SELF BENEFICIAL BEHAVIOR AT CUSTOMS AND FIRM TYPE

	Dep.Var.: Self beneficial (SB) reporting			
Firm Type	(1)	(2)	(3)	(4)
Disadvantageous	0.001	0.005		
	(0.007)	(0.006)		
Null VAT		-0.081***		-0.075***
		(0.006)		(0.006)
Conspicuous			0.006	0.005
			(0.006)	(0.006)
Month-Year FE	Yes	Yes	Yes	Yes
Size and Sector FE	Yes	Yes	Yes	Yes
HS Share of Import	No	Yes	No	Yes
N	123303	123303	100724	100724
R2	0.05	0.06	0.05	0.06
Mean of dep.	0.14	0.14	0.13	0.13

Notes: Data source: VAT Schedule 3, MVR and Customs data for fiscal years 2013-2016. This regression analyzes whether Disadvantageous firms, and within Advantageous firms, Conspicuous firms, are more likely to behave in a self beneficial way at customs. Observations are at the firm-month level. The dependent variable is a dummy equal to one if the firm claims higher VAT amounts incurred on imports in VS3 than VAT paid on imports recorded in the Customs data in the same month. We allow for 1,000 UGX rounding and for pure timing mismatches. In Columns (1) and (2), the explanatory variable of interest is a (time invariant) dummy for firm type, equal to one if the firm is classified as Disadvantageous, based on the value of $Q(f)$, as explained in Section 4.4. In Columns (3) and (4), we restrict the sample to Advantageous firms, and the explanatory variable of interest is a dummy equal to one if the firm is classified as a Advantageous and Conspicuous. In all specifications, we control for firm size as measure by annual decile of reported turnover, and for firm sector. In Columns (2) and (4), we additionally control for null monthly VAT liability as reported in MVR, and for the type of goods imported as measured by dummies for each of the 21 HS Good Code Sections, equal to one if the firm imports at least one good from the corresponding section. Standard errors, clustered at the firm level, are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

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	0.02	0.04**	0.00
Distance to URA office	−0.02	0.03**	0.00
MTO/LTO	−0.04	0.00***	−0.04
VAT Payable	0.02	0.14	0.00
VAT Due	−0.02	0.22	0.00
Total input	0.02	0.22	0.00
Total output	−0.04	0.05**	−0.01
Ratio of sales to FC	0.02	0.04**	0.01
Number of clients	0.02	0.00***	0.01
Number of suppliers	0.00	0.79	0.00
Upstreamness	−0.01	0.17	0.00
Distinct outputs (all good codes)	0.01	0.71	0.01
Distinct outputs (relevant good codes)	0.00	0.94	0.00
Distinct inputs (all good codes)	0.06	0.08*	0.00
Distinct inputs (relevant good codes)	−0.06	0.09*	0.00
Sectors:			
Agriculture, forestry, fishing	−0.01	0.36	0.00
Mining, Quarrying	−0.02	0.00***	−0.02
Manufacturing	0.01	0.11	0.01
Water, Electricity services	−0.01	0.01***	−0.01
Construction	0.01	0.19	0.01
Wholesale and retail	0.00	0.00	0.00
Transportation, accomodation services	−0.02	0.00***	−0.02
Information, communication	−0.01	0.04**	−0.01
Financial services	−0.03	0.00***	−0.03
Real estate	−0.02	0.00***	−0.01
Professional, Admin, Other Services	0.00	0.83	0.00
Public Administration	−0.01	0.60	0.00
Education	−0.01	0.27	0.00
Health and social work	0.00	0.43	0.00
Arts and Entertainment	−0.01	0.27	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.

TABLE A.7
FIRM-TYPE TRANSITION MATRIX

	Firm-pairs observed throughout 2013-2016		
	Advantageous (t)	Disadvantageous (t)	Share
Advantageous (t+1)	43.84	14.60	58.44
	(71.97)	(37.35)	
Disadvantageous (t+1)	17.07	24.49	41.56
	(28.03)	(62.65)	
Share	60.91	39.09	100.00
	(100.00)	(100.00)	

Notes: Data source: VAT Schedules and Monthly Summary data for fiscal years 2013-2016. This table presents the transition matrix for firm classifications. The sample is restricted to firms that appear at least in two consecutive years, and within each year, to the largest connected set.

TABLE A.8
FIRM TYPE CLASSIFICATION BASED ON ROBUSTNESS ESTIMATIONS

	Total	
	No. of Firms	**Share of firms**
	Panel A: Two-way fixed effect estimation with controls	
Advantageous	13,990	0.75
Conspicuous	10,842	0.58
Looking small	1,409	0.08
Looking big	1,739	0.09
Disadvantageous	4,639	0.25
N	18,629	1.00
	Panel B: Sample of firm-pairs with ≥ 10 *observations*	
Advantageous	11,049	0.88
Conspicuous	10,080	0.80
Looking small	351	0.03
Looking big	618	0.05
Disadvantageous	1,516	0.12
N	12,565	1.00

Notes: Data Source: VAT Schedules and Monthly Summary data for fiscal years 2013-2016. *In Panel A:* 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. *In Panel B:* We run the two-way fixed-effects model on the subset of firm-pairs that appear ten times or more in the initial dataset.

TABLE A.9
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	19,137	19,137	19,137
Percentage of all firms	(100%)	(100%)	(100%)
Total net VAT due	1,553,971	1,553,971	1,553,971
Seller shortfall			
Number of distinct firms with seller shortfall	17,249	17,249	13,448
Total net VAT due from firms with seller shortfall	1,275,917	1,275,917	1,133,456
Total VAT subject to seller shortfall	899,736	899,736	899,736
Buyer shortfall			
Number of distinct firms with buyer shortfall	17,979	17,979	17,181
Total net VAT due from firms with buyer shortfall	1,316,813	1,316,813	1,262,499
Total VAT subject to buyer shortfall	727,354	727,354	727,354
Correcting seller shortfall and buyer shortfall			
Impact on total net VAT due	383,730	436,372	492,844
Percentage of total VAT collected	28.2%	32.0%	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.