

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

¹⁹For 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).

²⁰This 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.

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

²²With 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.

²³LTOs 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). STO 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.

²⁴The 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 shown 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 Disadvantageous 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 downstream. 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:

$$d_{ff't} = \delta_{fy}^b + \delta_{f'y}^s + \delta_t + r_{ff't}, \quad (\text{A.1})$$

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:

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

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 shown 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., $Q_f \equiv \delta_f^s + \delta_f^b$; 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

²⁵Uganda is divided up into a total of 1,403 sub-counties (Electoral Commission, 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:

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

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:

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

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.

²⁶For 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, $\hat{\delta}_1^b$ 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:

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

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

²⁷In 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.

²⁹When 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.

³⁰We 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.

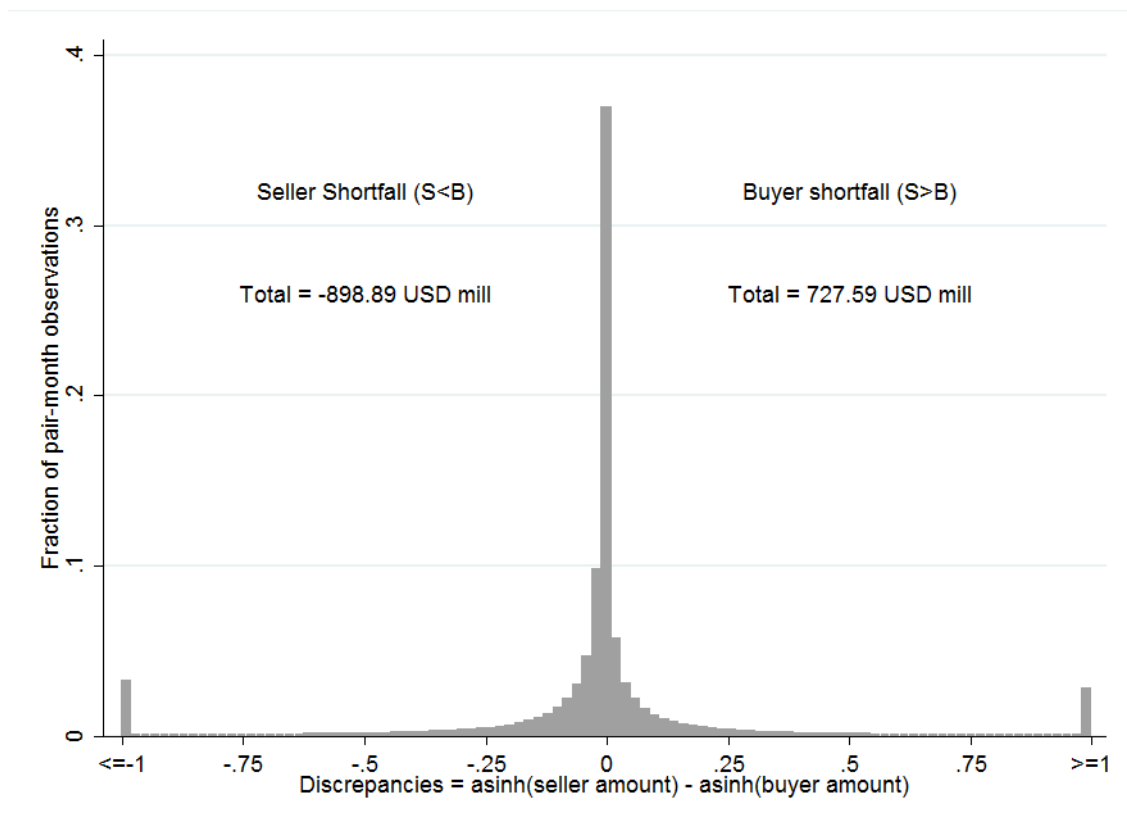
the sample of Advantageous misreporters, while the effect is close to zero in magnitude and not statistically significant for Disadvantageous misreporters (Table 3). The corresponding OLS estimates are shown in Table A.11, and are, respectively 7.8 percent; 8.4 percent; and 6 percent. The larger IV estimates 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).

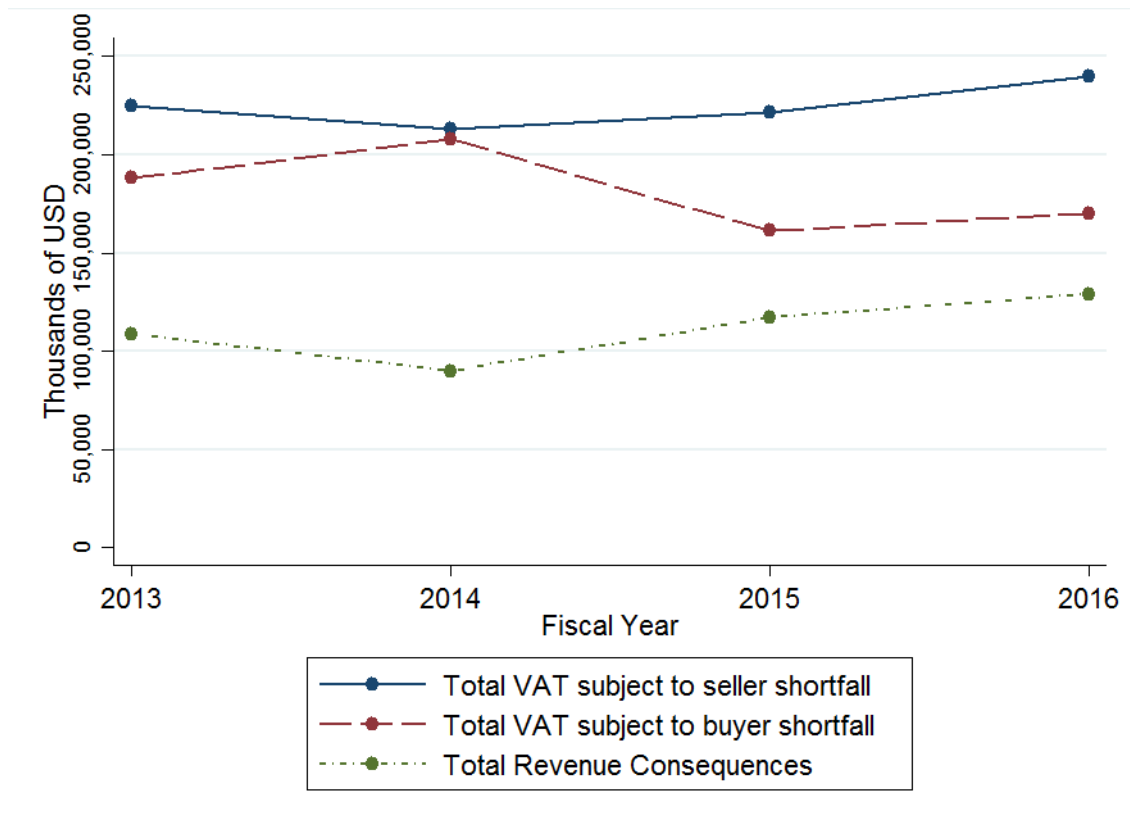
Figures

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



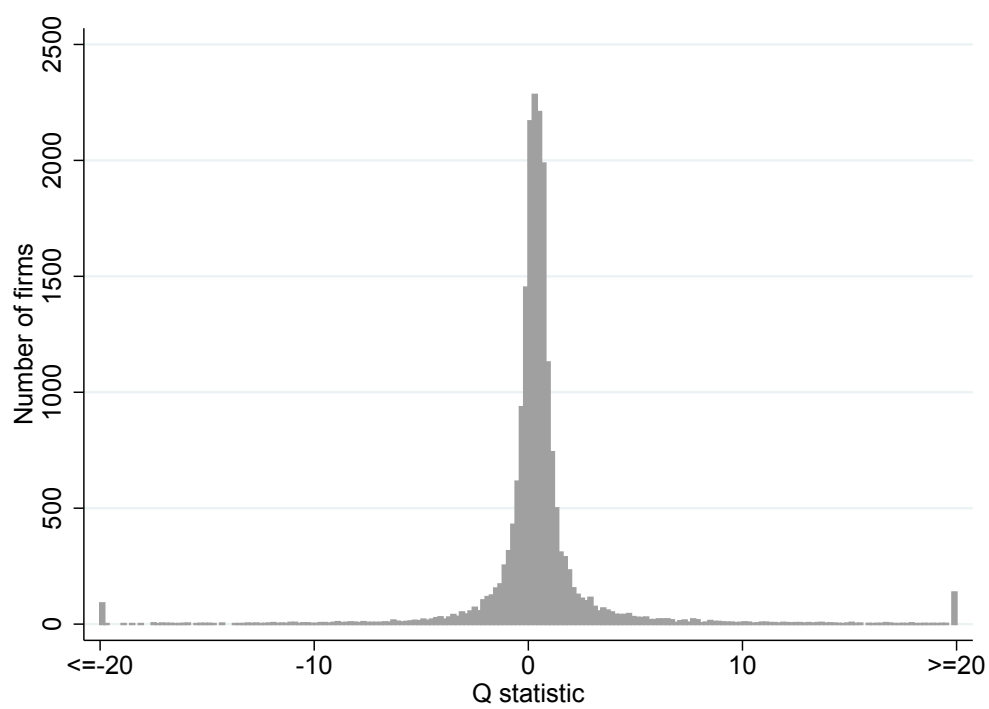
Notes: In this Figure, we show the distribution of discrepancies in the reporting of transactions by sellers and buyers for fiscal years 2013-2016. Data source: VAT Schedules data. Calculated by taking the difference between VAT charged in VS1 and VAT paid in VS24. We use the inverse hyperbolic sine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



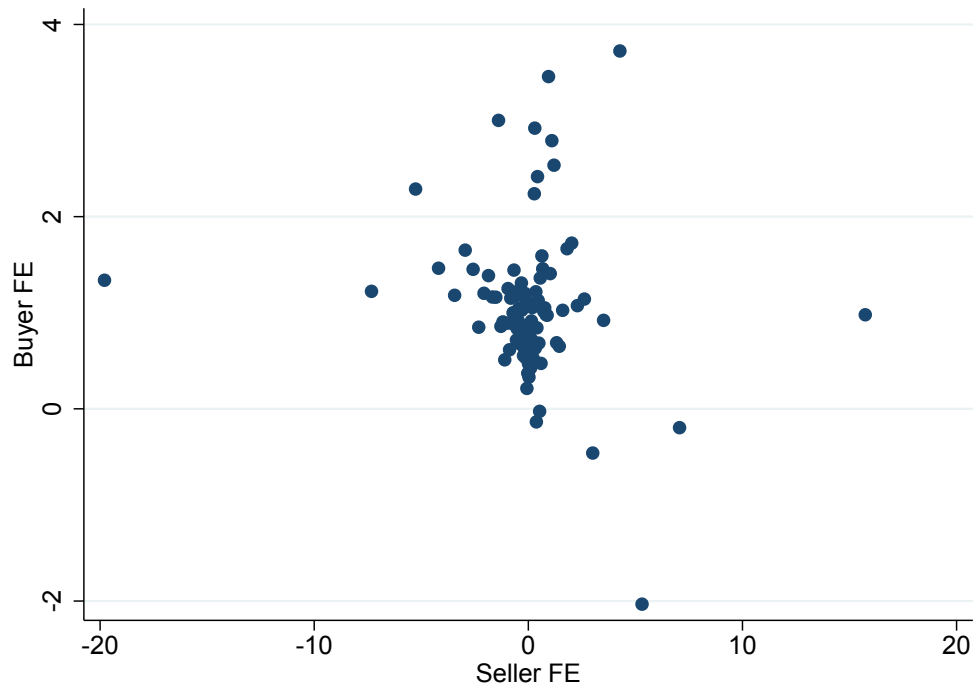
Notes: Data sources: VAT Schedules and Monthly Summary data for fiscal years 2013-2016. In this Figure, we show the evolution over time of total VAT under seller shortfall, total VAT under buyer shortfall, and the resulting revenue consequences. We include firms where either the firm-as-buyer and/or 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)	(2)	(3)	(4)
	Output VAT - Input VAT	VAT offsets from previous year	VAT liability (1) - (2)	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)	(2)	(3)
	Main	Alt.	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
<i>Panel A: Full sample</i>				
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%
<i>Panel B: Study sample</i>				
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
<i>in Kampala</i>	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

<i>Panel A: All firms</i>				
	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
<i>Panel B: Common firm-pairs</i>				
	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
<i>Panel C: Time dummies</i>				
	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 (0.011)	-0.063563** (0.014)	0.028678 (0.015)	0.043828* (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*** (0.001)	-0.006*** (0.001)	-0.005** (0.002)
China	-0.005*** (0.001)	-0.005*** (0.001)	-0.005*** (0.001)
France	-0.006 (0.008)	-0.002 (0.007)	-0.017 (0.015)
United Kingdom	-0.001 (0.001)	-0.001 (0.001)	-0.000 (0.002)
Indonesia	-0.013*** (0.003)	-0.013*** (0.003)	-0.029** (0.014)
India	-0.007*** (0.001)	-0.008*** (0.001)	-0.005*** (0.002)
Japan	-0.009*** (0.001)	-0.010*** (0.001)	-0.007*** (0.002)
Kenya	-0.016*** (0.001)	-0.017*** (0.002)	-0.013*** (0.002)
United States	-0.006*** (0.002)	-0.007*** (0.002)	-0.004 (0.003)
South Africa	-0.006*** (0.002)	-0.005** (0.002)	-0.008*** (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

Dependent variable	OLS		
	<i>asinh(Seller shortfall amounts)</i>		
Sample	(1) Full	(2) Advantageous	(3) Disadvantageous
ImportShare	-0.256*** (0.008)	-0.280*** (0.010)	-0.194*** (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	<i>asinh(Seller shortfall amounts)</i>					
	OLS			IV		
Sample	Full (1)	Adv. (2)	Disadv. (3)	Full (4)	Adv. (5)	Disadv. (6)
ImportShare	-0.242*** (0.008)	-0.261*** (0.010)	-0.194*** (0.013)	-0.604*** (0.161)	-0.764*** (0.172)	-0.243 (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 assignment 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

	<i>Import Share</i>					
	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 specifications. 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.