

## Overview

### Statement about Rights

- I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.
- I certify that the author(s) of the manuscript have documented permission to redistribute/publish the data contained within this replication package. Appropriate permission are documented in the LICENSE.txt file.

### Summary of Availability

- All data **are** publicly available.
- Some data **cannot be made** publicly available. In particular, administrative VAT microdata cannot be made available.
- No data can be made** publicly available.

### Details on each Data Source

The structure of this section is as follows. The first table details the VAT declaration data, which forms the core of the analysis for the paper. The second table details publicly available data. The third table details data used to calculate the share of VAT collected at the border (for Figure A13B in the Online Appendix, which is referenced in the text of the main paper), which is a mix of publicly and non-publicly available data.

Both data sources for the graphs in the main paper and the Online Appendix are provided.

#### VAT declaration data (not publicly available)

With the exception of French VAT administrative data, the data accessed in this project was accessed via existing collaborations or personal connections of the authors. We are not able to share any administrative data due to its confidentiality. VAT administrative data is used for all of the analyses in the paper.

For those interested in replication of the data, we have provided the contact person for each data source. We cannot guarantee that the relevant authorities will approve access to the data for the purposes of replication by external parties, but we will provide our assistance for two years following publication.

Name	Citation	Access information
“VAT Declaration Data for Costa Rica, 2007-2014”	Ministerio de Hacienda [Costa Rica] (2015)	The data is owned by the Costa Rican Ministry of Finance. This data was made available to select World Bank staff in the framework of a broader collaboration between the World Bank and the Costa Rican General Tax Administration (DGT), a dependency of the Ministry of Finance. Please contact Anne Brockmyer (ab r ockmeyer@worldbank.org) for more information.
“VAT Declaration Data for eSwatini, 2012-2017”	eSwatini Revenue Service (2017)	This data is owned by the eSwatini Revenue service. This data was made available to the ICTD via their wider collaboration with the Research, Strategy and Statistics Department of the eSwatini Revenue Service. ICTD affiliated staff may be able to get access to the data. Please contact Fabrizio Santoro (F.Santoro@ids.ac.uk) for more information.

Name	Citation	Access information
“VAT Declaration Data for Ethiopia, 2011-2017”	Ministry of Revenue [Ethiopia] (2018)	This data is owned by the Ethiopian Ministry of Revenue. The data was made available to the the Institute for Fiscal Studies (IFS) as part of the their ongoing collaboration with the Tax Policy Directorate of the Ministry of Finance. IFS affiliated staff may be able to get access to the data. Contact Vedanth Nair ( vedanth.nair@ifs.org.uk) for more information.
” TVA annuel: Données détaillées déclaratives annuelles de TVA” [ 2015-2021].	Ministère des Finances (DGFIP) [France] (2021)	This data is available to researchers. The request process can take several months and cost several thousand euros. Contact [service@casd.eu] (mailto:service@casd.eu) for more information. The data provider’s website also provides more detail.
“VAT Declaration Data for Guatemala, 2006-2015”	Superintendencia de Administración Tributaria [Guatemala] (2015).	This data is owned by the Guatemalan tax administration. The data was made available to select World Bank staff in the framework of the broader agreement that the World Bank has with the Guatemalan tax administration. Please contact Anne Brockmeyer ( abrockmeyer@worldbank.org) for more information.
“VAT Declaration Data for Honduras, 2018-2020”.	Servicio de Administración de Rentas (2020)	This data is owned by the Honduran tax administration. The data was made available to select World Bank staff in the framework of the broader agreement that the World Bank has with the Honduran tax authority. Please contact Anne Brockmeyer ( abrockmeyer@worldbank.org) for more information.
“VAT Declaration Data for Hungary, 2017-2019”.	Pénzügyminisztérium [Hungary] (2021)	The data is owned by the Hungarian Ministry of Finance. The data was made available to the IFS through a collaboration between the IFS and the Ministry of Finance, which is no longer ongoing. Please contact Bálint Van (balint.van.work@gmail.com), who formerly worked at the research department of the Ministry of Finance, and now works in the TaxDev programme jointly run by the IFS, for more information.
“VAT Declaration Data for Pakistan, 2012-2018”	Federal Board of Revenue [Pakistan] (2018)	This data is owned by the Federal Board of Revenue of Pakistan. This data was provided to Mazhar Waseem’s broader research collaboration with the Directorate General of Revenue Analysis, Federal Board of Revenue of Pakistan. Please contact Mazhar Waseem (mazhar.waseem@manchester.ac.uk) for more information.

Name	Citation	Access information
“VAT Declaration Data for Rwanda, 2016-2020”	Rwanda Revenue Authority (2020)	This data is owned by the Rwanda Revenue Authority. The data was made available to the ICTD as part of their wider collaboration with the research department of the Rwanda Revenue Authority. ICTD affiliated staff may be able to access to the data. Please contact Adrienne Lees (a.lees@ids.ac.uk) for more information.
“VAT Declaration Data for Senegal, 2009-2015”	Direction Générale des Impôts et des Domaines (2015)	This data is owned by the Senegalese tax authorities. This data was provided in the framework of a collaboration between the World Bank, the tax administration in Senegal, and the Paris School of Economics. Please contact Anne Brockmeyer (abrockmeyer@worldbank.org) for more information.
“VAT Declaration Data for Uganda, 2011-2018”	Uganda Revenue Authority (2018)	This data is owned by the Uganda Revenue Authority. This data was accessed as part of Miguel Almunia’s collaboration with the Research and Innovation Lab of the Ugandan Revenue Authority. Please contact Miguel Almunia (miguel.almunia@cunef.edu) for more information More recently, the data has been made available to researchers through the Secure Research Lab, a joint initiative launched by the Uganda Revenue Authority and the University of North Carolina at Chapel Hill (UNC-CH). More information about how to access the data can be found at <a href="https://www.wider.unu.edu/about/secure-research-data-lab-uganda">https://www.wider.unu.edu/about/secure-research-data-lab-uganda</a> , including contact details of the persons in charge.

### Publicly available data

Data .Name	Da ta.Files	Notes and instructions for access	Citation
“Doing Business”	input data/WB Refund data 2019.xlsx	Note that the Doing Business project has been discontinued due to allegations of manipulation. The data for the last round (2020) can be found on < <a href="https://archive.doingbusiness.org/en/doingbusiness">https://archive.doingbusiness.org/en/doingbusiness</a> >. Select ‘historical data - with scores (Excel)’. The specific variable used is ‘Time taken to comply with a VAT refund (weeks), DB17-20 methodology’.	World Bank (2020)
“GTED Tax Expenditure Database 2022”	input data /IMF_VAT_thresholds.xlsx	The dataset is open access can be accessed using this DOI: 10. 5281/Z ENODO.6334212.	GTED (2022).
“IMF VAT Rate Database”	input data /IMF_VAT_rates.xlsx	Note that the IMF have not provided the date when the dataset was created, but the last available year is 2020.	IMF (undated)
“IMF VAT Threshold Database”	input data /IMF_VAT_thresholds.xlsx	Note that the IMF have not provided the date when the dataset was created, but reforms that took place in 2020 are referred to. The dataset is open access and can be accessed at < <a href="https://www.imf.org/external/np/fad/tf/pages/vat.htm">https://www.imf.org/external/np/fad/tf/pages/vat.htm</a> >, and clicking VAT base > What is the registration threshold >scrolling to the bottom and clicking on ‘here’.	IMF (undated)

Data .Name	Da ta.Files	Notes and instructions for access	Citation
" World Development Indicators"	input da ta/ WB Macro Da ta.xlsx	The specific variables used are GDP per capita, PPP (current international \$), [World Bank identifier = NY .GD P.PCAP.PP.CD]; Final consumption expenditure (current USD) [WB identifier = NE .CON.TOTL.CD]; Imports of goods and services (current US\$) [WB identifier = BM .GSR.GNFS.CD]; Official exchange rate [WB identifier = PA.NUS.FCRF]. The specific version is from February 2022. Newer releases may have revised data. To get the older data, you can go to  < nk. wor ldb ank.org/source /wdi -da tab ase-archives> and select Version = '2022'.	World Bank (2022)
"How many countries have VAT or GST 2023"	input da ta /vat_coun t ries.xlsx	The data is not available as a downloadable file. We copied and pasted the table on the web page into Excel.	Caragher (2023)
"UNU WIDER G overnment Re venue Data base"	input data /GRD _VAT_ Re ve nue.xlsx	The data was accessed using the explorer tool for the 2023 version of the GRD dataset: « <https : //w ww.wider.u» nu. edu /pr oject/grd-gove rnme nt- rev enue-dataset> To reproduce the data used, select 'VAT' in the 'Indicators' section.	UNU-WIDER (2023).

Note that the only datasets required for the main paper are the World Development Indicators and the Doing Business data. The rest are for the online appendix.

## Data used for the calculation of counterfactual RST / TOT rates

The section ‘**Should the Value-Added Tax be Replaced by a Retail Sales Tax or a Turnover Tax?**’ states that ‘the typical revenue-neutral turnover tax rate in our sample is about 5-8 percent, when applied to the sales of firms at each step in the supply chain’. A key data point for this analysis is the share of VAT that is collected at the border. The sources for this information are provided below. Publicly available files are stored under `documentation/Documentation for TOT simulations`.

Country	Name of source	Source	Notes
Costa Rica	“Customs administrative data, 2017”.	Ministerio de Hacienda [Costa Rica] (2017).	See information on Costa Rica VAT declaration data. The code that processes this to derive VAT revenue collected at the border is saved under <code>documentation/Files/condary/financial_import_vat_CRI.R</code>
Ethiopia	“General Revenue Statistics 1993-2017”	Ministry of Revenue, Ethiopia.	Unpublished report, which is available as part of the Institute for Fiscal Studies’ collaboration with the Tax Policy Department of the Ministry of Finance of Ethiopia. Please contact Vedanth Nair (vedanth.nair@ifs.org.uk) for more information.
Guatemala	“Ingresos Tributarios de la Administración Central 2002-22”	Ministero de Finanzas Publicas [Guatemala]	Publicly available.
Honduras	“Untitled news article”	Poder Popular, 2023.	Publicly available.
Honduras	“Revenue report”	Servicio de Administración de Rentas (2022)	Saved on <code>documentation/Documentation for TOT simulations/HND Revenue Report.xlsx</code>

Country	Name of source	Source	Notes
Senegal	“Customs administrative data, 2015”	Direction Générale des Impôts et des Domaines (2015)	See information on Senegalese VAT declaration data above. The code that processes this to derive VAT revenue collected at the border is saved under <code>do files /secondary/ind_import_vat_SEN.R</code>
Pakistan	“Year Book 2021-22”	Ministry of Finance [Pakistan], 2022	Publicly available.
Pakistan	“Transaction-level VAT data”	Ministry of Finance [Pakistan]	See information on Pakistan VAT declaration data above.
Rwanda	“Annual Activity Report 2018/19”	Rwanda Revenue Authority (2019)	Publicly available

See also `input data/import_vat_share.xlsx` for more information on how the calculation is done.

## Computational requirements

### Software Requirements

R 4.2.3

- `here` (1.01)
- `tidyverse` (2.0.0)
- `reshape2` (1.4.4)
- `DescTools` (0.99.48)
- `readxl` (1.4.3)
- `broom` (1.0.5)
- `fixest` (0.11.1)
- `ggrepel` (0.9.3)
- `ggpubr` (0.6.0)
- `xtable` (1.8.4)
- `tinytex` (0.47)
- `kableExtra` (1.3.4)
- `rdrobust` (0.99.4)
- `countrycode` (1.5.0)

The code automatically installs all required packages.

### Memory and Runtime Requirements

**Summary** Approximate time needed to reproduce the analyses on a standard (2023) desktop machine:

☒ <10 minutes

Note that <10 minutes only refers to the code that creates graphs from the aggregates statistics. The full code (conditional on having access to all the data) takes 2-8 hours to run.

## Description of programs/code

### Creation of aggregated statistics from administrative microdata - not available to reproduce.

- The program in the folder `dofiles/Country-Specific Cleaning/XXX` contain R and STATA files imports the raw administrative data for country `XXX`. These were last run in their entirety between April and June 2023, depending on the country.
  - Unless the replicator has access to the exact copy of the raw data used in this paper (e.g. if they work at the World Bank and are able to get access to countries where there is a sharing agreement for the World Bank, or if they use CASD to get access to the raw microdata), then it is highly unlikely that the file that cleans the raw data will work. This is because changes in IT systems, or even changes in the individual providing the data, often result in different data formats.
  - To address this, we provide the file `documentation/VAT form to variable mapping.xlsx`. This details how the line items in the VAT form are combined to get at the main variables used in this study: total sales, exports, exempt sales, reduced rates/zero rated sales, purchases, output VAT, import VAT, and credits carried forward from previous periods. Even if the format of the VAT administrative data changes, it is generally possible to link the variables in the data to the VAT form.
- The program `dofiles/1.Common Cleaning.R` applies further cleaning steps that are common to all countries, such as replacing NAs with 0 and collapsing the data to the (fiscal) yearly level.
  - This was last run in their entirety between April and June 2023, depending on the country.
- The program `dofiles/2. Create statistics from clean data.R` creates aggregate statistics from the cleaned microdata. The aggregate statistics are stored in `output/XXX`, where `XXX` are relevant country codes.
  - This last run in their entirety between April and June 2023, depending on the country.

### Created of graphs from aggregated statistics - available to reproduce

- The graphs in the JEP paper are created by `dofiles/3. Create JEP graphs from statistics.R`, using the aggregated statistics in `output/`.
  - This was last run in its entirety in November 2023.

## Instructions to Replicators

**For the figures in the main paper:** run the file `dofiles/3. Create JEP graphs from statistics.R`. There is no need to edit the file, provided the file is opened in the RProject space. It will produce all the figures in the folder `output/JEP graphs`

**For the figures in the Online Appendix:** run the file `dofiles/3.1. Create appendix graph from statistics.R`. There is no need to edit the file, provided the file is opened in the RProject space. It will produce all of the tables and figures in the folder `output/Online Appendix graphs`. The exception is for Figure A13A (RST simulations), which is created by `dofiles/RST_Simulation_Pakistan.do` (requires access to confidential data to run and therefore this file is not replicated by the package).

**Instructions to replicators with access to the exact copy of the confidential data used by the authors in this study.**

- Edit the file `0. Global metadata and functions_replication.R`, line 50-61. In line 50, edit the countries for which you have the raw microdata for (e.g c(“CRI”, “GTM”) for Costa Rica and Guatemala.
  - If the file path in lines 53-63 is different to the one supplied, change it.
- Run the file `00_RUN_ALL_00.R`. This will 1) process the raw microdata 2) create statistics from raw microdata and 3) create graphs
  - Note that step 1 for Costa Rica, Ethiopia and Pakistan is written in STATA. Therefore, you need to manually run the relevant files (`dofiles/Country Specific Cleaning/VAT_cleaning_XXX.do`) for those countries if they are being replicated.



## List of tables and programs

The provided code reproduces:

- ☒ All numbers provided in text in the paper\*
- ☒ All tables and figures in the paper\*

\*Note that none of the figures, except for Figure 4A, can be RST\_Simulation\_Pakistanrecreated from the raw data.

All of the graphs are created using the file 'dofiles/3. Create JEP graphs from statistics.R'. The line references in the file where the aggregated statistics were created from microdata 'dofiles/2. Create statistics from clean data.R' is also given.

### Figures in the main paper

Note that XXX refers to the names of the countries (e.g. `output/CRI/3.csv`) stores the results for Figure 1A and 1B for Costa Rica. Note that only Figure 4A can be run fully from scratch without any confidential data.

Figure	Line Number (3. Create JEP graphs from statistics.R)	Line Number (2. Create statistics from clean data.R)	Name of aggregate statistics file that is input to graph	Name of output graph
Figure 1A, Figure 1B	15	187-287	<code>output/ XXX/3.csv</code>	JEP graphs/ Figure 1A.pdf , JEP graphs/ Figure 1B.pdf
Figure 2	81	474-610	<code>output/XXX/7_1.csv</code>	JEP graphs/ Figure 2.pdf
Figure 3A	112	721-774	<code>output/XXX/7_35.csv</code>	JEP graphs/ Figure 3A.pdf
Figure 3B	128	885-994	<code>output/XXX/12.csv</code>	JEP graphs/ Figure 3B.pdf
Figure 4A	166	NA	<code>input data/WBR efund 2019.xlsx</code>	JEP graphs/ Figure 4A.pdf
Figure 4B	205	985-1119	<code>output/XXX/14.csv</code>	JEP graphs/ Figure 4B.pdf
Figure 4C	240	1157-1223	<code>output/XXX/14_3.csv</code>	JEP graphs/ Figure 4C.pdf

### Figures in the Online Appendix

Note that only Figures A3, A6 and A13 can be created from scratch without any confidential data.

Figure	Line Number (3.1 Create appendix graphs from sta t istics.R)	Line Number (2. Create st atistics from clean data.R)	Name of aggregate s tatistics file that is input to graph	Name of output graph
Table A1	12	155-185	output/ XXX /1_ 4 .csv; ou tput/XXX/ 7_35.csv	Onl ine A ppendi x g ra phs / T able1.tex
Figure A1A, Figure A1B, Figure A2A, Figure A2B	73	187-287	output/X XX/3.csv	Onl ine Append ix graph s/ Figure A1A .png; On li ne Appe nd i x gra phs/Fi gur e A1B.p n g; nline Append ix raphs /F igu re 2A .png ; Onl ine Ap pen di x gra p hs/2B.png
Figure A3	115	NA	NA	Onlin e Ap pendix gra ph s/Fig u re A3.png
Figure A4	178	719-824	out put/XXX/7 _1R.csv	Onlin e Ap pendix gra ph s/Fig u re A4.tex
Figure A5	209	474-616	out put/XXX/7 _15.csv	Onlin e Ap pendix gra ph s/Fig u re A5.tex
Figure A6	241	NA	NA	Onlin e Ap pendix gra ph s/Fig u re A6.tex
Figure A7	314	346-465	outp ut/XXX /7 _05.csv	Onlin e Ap pendix gra ph s/Fig u re A7.tex
Figure A8	348	719-821	outp ut/XXX/ 7 _35R.csv	Onlin e Ap pendix gra ph s/Fig u re A8.tex
Figure A9	378	115-185	ou tput/XXX/ 1_4.csv	Onlin e Ap pendix gra ph s/Fig u re A9.tex

Figure	Line Number (3.1 Create appendix graphs from statistics.R)	Line Number (2. Create statistics from clean data.R)	Name of aggregate statistics file that is input to graph	Name of output graph
Figure A10	403	983-1155	output/XXX /14.csv	On line Appendix graphs /Figure A10.tex
Figure A11	432	983-1155	output/XXX /14.csv	On line Appendix graphs /Figure A11.tex
Figure A12	472	1155-1225	output/XXX /14_3.csv	On line Appendix graphs /Figure A12.tex
Figure A13	578	NA	NA	On line Appendix graphs /Figure A14.tex
Figure A14A	Created in the file: dofiles/RST_Simulation_Pakistan.do, based on confidential data.	Created in the file: dofiles/RST_Simulation_Pakistan.do, based on confidential data.	NA	On line Appendix graphs /Figure A13A.tex
Figure A14B	502	946-983	output/XXX /13.csv	On line Appendix graphs /Figure A13B.tex

## References

- Caragher, J., 2023. “How many countries have VAT or GST? 175”. <https://www.vatcalc.com/global/how-many-countries-have-vat-or-gst-174/>. Accessed on 2023/11/08
- Direction Générale des Impôts et des Domaines (2015), “Customs administrative microdata”. Unpublished confidential dataset.
- Direction Générale des Impôts et des Domaines, 2015. “VAT Declaration Data for Senegal, 2009-2015”. Unpublished confidential dataset.
- eSwatini Revenue Service, 2018. “VAT Declaration Data for eSwatini, 2012-2017”. Unpublished confidential dataset.
- Federal Board of Revenue, 2018. “VAT Declaration Data for Pakistan, 2012-2018”. Unpublished confidential dataset.
- International Monetary Fund (undated). “IMF historical VAT rate database”. <https://www.imf.org/external/np/fad/tpaf/pages/vat.htm>. Accessed on 2023/11/08.
- International Monetary Fund (undated). “IMF VAT threshold database”. <https://www.imf.org/external/np/fad/tpaf/pages/vat.htm>. Accessed on 2023/11/08

Ministerio de Hacienda, 2015. “VAT Declaration Data for Costa Rica, 2007-2014”. Unpublished confidential dataset.

Ministerio de Hacienda, 2017. “Customs administrative data, 2017”. Unpublished confidential dataset.

Ministère des Finances (DGFIP), 2021, “TVA\_annuel : Données détaillées déclaratives annuelles de TVA”. Confidential dataset. DOI: <https://doi.org/10.34724/CASD.429.4194.V3>.

Ministero de Finanzas Publicas, 2022. “Ingresos Tributarios de la Administración Central 2002-22”. <https://www.minfin.gob.gt/images/daf/documentos/doc108.pdf>. Accessed on 2023/11/11.

Ministry of Finance [Pakistan], 2022, “Year Book 2021-22”; [https://download1.fbr.gov.pk/Docs/20229301594336726RevenueDivisionYearBook2021-22\(30September,2022\)-Final.pdf](https://download1.fbr.gov.pk/Docs/20229301594336726RevenueDivisionYearBook2021-22(30September,2022)-Final.pdf). Accessed on 2023/11/11.

Ministry of Revenue, 2018. “General Revenue Statistics 1993-2017”. Unpublished dataset.

Ministry of Revenue, 2018. “VAT Declaration Data for Ethiopia, 2011-2017”. Unpublished confidential dataset.

Pénzügyminisztérium, 2021. “VAT Declaration Data for Hungary, 2017-2019”. Unpublished confidential dataset.

Poder Popular, 2023. “Untitled newspaper article”. [https://www.poderpopular.hn/vernoticias.php?id\\_noticia=756](https://www.poderpopular.hn/vernoticias.php?id_noticia=756). Accessed on 2023/11/11.

Redonda, A.; von Haldenwag, C. and Aliu, F., 2022. “Global Tax Expenditure Database (GTED)”. Version 1.1.3. DOI: 10.5281/ZENODO.6334212

Rwanda Revenue Authority, 2019. “Annual Activity Report 2018/19”. [https://www.rra.gov.rw/fileadmin/user\\_upload/rra\\_annual\\_activity\\_report\\_2018-19.pdf](https://www.rra.gov.rw/fileadmin/user_upload/rra_annual_activity_report_2018-19.pdf). Accessed 2023/11/11.

Rwanda Revenue Authority, 2020. “VAT Declaration Data for Rwanda, 2016-2020”. Unpublished confidential dataset.

Servicio de Administración de Rentas, 2020. “VAT Declaration Data for Honduras, 2018-2020”. Unpublished confidential dataset.

Superintendencia de Administración Tributaria, 2020. “VAT Declaration Data for Guatemala, 2006-2015”. Unpublished confidential dataset.

Uganda Revenue Authority, 2018. “VAT Declaration Data for Uganda, 2011-2018”. Unpublished confidential dataset.

Uganda Revenue Authority, 2018. “Revenue Performance Report 2017/18”. [https://ura.go.ug/wp-content/uploads/2022/10/3-Annual-Revenue-Report\\_2017\\_18.pdf](https://ura.go.ug/wp-content/uploads/2022/10/3-Annual-Revenue-Report_2017_18.pdf). Accessed on 2023/11/11.

UNU-WIDER, 2023. “UNU-WIDER Government Revenue Dataset”. Version 2023. <https://doi.org/10.35188/UNU-WIDER/GRD-2023>.

World Bank, 2020. “Doing Business Data”. Dataset. Available at <https://archive.doingbusiness.org/en/data>. Accessed 2023/11/08.

World Bank, 2023. “World Development Indicators”. Dataset. Available at: <https://datacatalog.worldbank.org/search/dataset/0037712/World-Development-Indicators>. Accessed 2023/11/08.

---

## Acknowledgements

Some content on this page was copied from Hindawi. Other content was adapted from Fort (2016), Supplementary data, with the author’s permission.