README and Guidance

for

Local Elites as State Capacity: How City Chiefs Use Local Information to Increase Tax Compliance in the D.R. Congo

By Pablo Balán, Augustin Bergeron, Gabriel Tourek, and Jonathan L. Weigel[†]

October 21, 2021

[*] AEA RCT Registry ID: AEARCTR-0003308.

[†] Affiliations: Harvard University (pbalan@g.harvard.edu), Harvard University (augustinbergeron@fas.harvard.edu), MIT (gztourek@mit.edu), LSE and CEPR (j.weigel@lse.ac.uk).

Contents

1 Overview

The code in this replication package constructs the analysis datasets for "Local Elites as State Capacity" using Stata. The master code file 0_Master.do runs all code to generate the data for the main tables and figures in the paper as well as tables and figures in the Online Data Appendix. Replicators can expect the code to run in 30 minutes.

2 Data Availability and Provenance Statements

2.1 Statement about Rights

We certify that the authors have legitimate access to and permission to use the data used in this manuscript.

2.2 Statement of Data Availability

Some data **cannot be made** publicly available.

2.3 Details on Data Sources

The paper uses data obtained from the Provincial Government of Kasaï Central and survey data collected by the authors. All data and their documentation can be downloaded from http://doi.org/10.3886/E147561V1 (Balán et al., 2021).

1. Administrative Tax Data: Property registration and property tax payment data for the 2018 property tax campaign were obtained from the Provincial Government of Kasaï Central. De-identified copies of the data are included in the replication directory in Data/01_base/admin_data.

2. Survey Data: Primary data were collected by the authors from households at baseline, property registration, midline, and endline and from tax collectors at baseline and endline. De-identified survey datasets are included in the replication directory in Data/01_base/survey_data. Survey instruments are available upon request.

3 Dataset List

The table on the following page lists the base datasets included in the replication directory or produced by replication code. All datasets are located in or saved to the "Data" folder of the replication directory.

Data File	Source	Notes	Provided
Data/01_base/survey_data/baseline_noPII.dta	Primary	Public	Yes
Data/01_base/survey_data/midline_noPII.dta	Primary	Public	Yes
Data/01_base/survey_data/endline_round1_noPIII.dta	Primary	Public	Yes
Data/01_base/survey_data/endline_round2_noPII.dta	Primary	Public	Yes
Data/01_base/admin_data/property_values_MLestimates.csv	Tax Admin	Public	Yes
Data/01_base/survey_data/endline_2016campaign_noPII.dta	Primary	Public	Yes
Data/01_base/admin_data/chief_collector_candidates.dta	Tax Admin	Public	Yes
Data/01_base/survey_data/chief_survey_randbasis_noPII.dta	Primary	Public	Yes
Data/01_base/admin_data/tax_payments_neighborhoods.dta	Tax Admin	Public	Yes
Data/01_base/admin_data/randomization_assignment.dta	Tax Admin	Public	Yes
Data/01_base/survey_data/chief_consultations.dta	Primary	Public	Yes
Data/01_base/admin_data/adjacent_neighborhoods.dta	Tax Admin	Public	Yes
Data/01_base/admin_data/hh_distances.dta	Tax Admin	Public	Yes
Data/01_base/admin_data/campaign_collector_info.dta	Tax Admin	Public	Yes
Data/01_base/survey_data/collector_baseline_noPII.dta	Primary	Public	Yes
Data/01_base/survey_data/chief_survey_noPII.dta	Primary	Public	Yes
Data/01_base/admin_data/neighborhood_centroids.dta	Tax Admin	Public	Yes
Data/01_base/admin_data/neighborhood_transport_cost.dta	Tax Admin	Public	Yes
Data/01_base/admin_data/tax_payment_timing_noPII.dta	Tax Admin	Public	Yes
Data/01_base/admin_data/tax_payments_noPII.dta	Tax Admin	Public	Yes
Data/01_base/survey_data/chief_knowledge.dta	Primary	Public	Yes
Data/01_base/survey_data/resident_info_quiz	Primary	Public	Yes
Data/01_base/survey_data/collector_chars.dta	Primary	Public	Yes
Data/01_base/survey_data/collector_endline_noPII.dta	Primary	Public	Yes
Data/01_base/admin_data/chief_collector_candidates_campaignupdated.dta	Tax Admin	Public	Yes
Data/01_base/survey_data/chief_knowledge_neighborhoods.dta	Primary	Public	Yes
Data/01_base/admin_data/chief_tribe_info.dta	Tax Admin	Public	Yes
Data/01_base/survey_data/collector_knowledge_clean.dta	Primary	Public	Yes
Data/01_base/survey_data/chief_info.dta	Primary	Public	Yes
Data/01_base/admin_data/concessions_chefferies.csv	Tax Admin	Public	Yes
Data/01_base/admin_data/campaign_2016_neighborhoods.dta	Tax Admin	Public	Yes
Data/01_base/admin_data/randomization_schedule.dta	Tax Admin	Public	Yes

Data/01_base/admin_data/fliers_pilot_set1.xlsx	Tax Admin	Public	Yes
Data/01_base/admin_data/fliers_pilot_set2.xlsx	Tax Admin	Public	Yes
Data/01_base/admin_data/fliers_pilot_set3.xlsx	Tax Admin	Public	Yes
Data/01_base/admin_data/fliers_campaign.dta	Tax Admin	Public	Yes
Data/01_base/admin_data/registration_noPII.dta	Tax Admin	Public	Yes
Data/01_base/admin_data/taxroll_noPII.dta	Tax Admin	Public	Yes
Data/01_base/admin_data/stratum.dta	Tax Admin	Public	Yes
Data/03_clean_combined/combined_data.dta	Coded	Public	Yes
Data/03_clean_combined/analysis_data.dta	Coded	Public	Yes
Data/03_clean_combined/analysis_data_neighborhoods.dta	Coded	Public	Yes
Data/02_intermediate/concessions_chefferies.dta	Coded	Public	Yes
Data/02_intermediate/2016_tmt.dta	Coded	Public	Yes
Data/02_intermediate/assignment.dta	Coded	Public	Yes
Data/02_intermediate/flier_mailmerge.dta	Coded	Public	Yes
Data/02_intermediate/registration_cleaned.dta	Coded	Public	Yes
Data/02_intermediate/taxroll_cleaned.dta	Coded	Public	Yes
Data/02_intermediate/midline_cleaned.dta	Coded	Public	Yes
Data/02_intermediate/tax_payments_cleaned.dta	Coded	Public	Yes

All data sources are in one of the following formats: .dta, .csv, .xls, .xlsx

4 Computational Requirements

4.1 Software Requirements

Stata (code was last run with Stata/SE version 14.2 for Mac), with additional packages:

- estout
- outtable
- mmat2tex
- geodist
- center
- grstyle
- palettes
- balancetable
- winsor
- revrs
- distplot
- blindschemes
- cem
- GSSU

Program "1_Package_Setup.do" will install all additional required packages locally and needs only to be run once.

4.2 Memory and Runtime Requirements

Reproducing the analysis requires approximately half and hour on a standard desktop machine with Stata version 14.2 or higher.

The code was last run on a 2 GHz Intel® Core™ i5, 8 GB 1867 MHz LPDDR3, on macOS Sierra Version 10.12.6.

The guidelines provided below outline the computational requirements for

Stata/MP and Stata/SE (sourced from `https://www.stata.com/products/compatible-operating-systems/`):

4.2.1 Platform Requirements

Stata for Windows®

- Windows 7, 8, or 10
- Windows Server 2008 R2, 2012, 2016, 2019

* Stata requires 64-bit Windows for x86-64 processors made by Intel® or AMD

Stata for Mac®

- Mac with Apple Silicon or Intel® processor (Core™ i3 or better)
- macOS 11.0 (Big Sur) or newer for Macs with Apple Silicon and macOS 10.12 (Sierra) or newer for Macs with 64-bit Intel® processors

Stata for Linux

- Any 64-bit (x86-64 compatible) running Linux
- Minimum requirements include the GNU C library (glibc) 2.17 or better
- For xstata, GTK 2.24 must be installed

4.2.2 Hardware Requirements

Package	Memory	Disk Space
Stata/MP	4 GB	1 GB
Stata/SE	2 GB	1 GB
Stata/BE	1 GB	1 GB

Stata for Linux requires a video card that can display thousands of colors or more (16-bit or 24-bit color)

5 Description of Programs and Code

The following code files are located in the Dofiles folder of replication directory.

- 0_Master.do will run the entire replication from start to finish, setting dependencies, installing necessary packages, and calling all other programs in the replication directory
- 1_Package_Setup.do will install all packages required for replication if not already installed

- 2_Data_Construction.do will clean and append datasets to produce intermediate datasets and analysis datasets used to produce output
- 3_Main_Tables_Figures.do will create all tables and figures in the main paper, calling on individual dofiles in Dofiles/Tables_Figures
- 4_Appendix_Tables_Figures.do will create all tables and figures in the appendix, calling on individual dofiles in Dofiles/Tables_Figures

6 Instructions for Replicators

To run entire replication:

1. Enter default path in Dofiles/0_Master.do, line 23
2. Run Dofiles/0_Master.do to run all code files in sequence
3. Compile outputs using output.tex
    - Note that the formatting of outputs in output.tex will not align precisely with outputs in paper due to manually formatting undertaken by authors. However, all results should match.
    - The Documents folder in the replication directory contains outputs not generated using data.

If running files individually:

- 0_Master.do lines 1–33 must be run at the beginning of each session
- 1_Package_Setup.do need only be run once in a computing environment to install packages
- Code files to clean and combine datasets (2_Data_Construction) must be run before any used to produce paper exhibits (those listed in Section 7)

7 List of Paper Exhibits and Programs

Code files listed below reproduce paper tables and figures. Note that descriptive text added to paper tables and figures — e.g., fixed effects and controls included in regressions — may not appear in replication output but are observable in the code that produces specific outputs.

Exhibit	Program	Line(s)	Output Files
Table 1	Table1.do	81	campaign_components.tex
Table 2	Table2.do	23	treatment_allocation.tex
Table 3	Table3.do	474	main_balance_baseline.csv
		605	main_balance_midline.csv
		763	main_attrition.tex
Table 3 Note: Table 3 in the paper organizes measures by the entity they describe whereas the replication output is grouped by the data from which measures are drawn. Recreating the format of Table 3 in the paper requires reorganizing replication output by entity. This can be done by following the variable names and superscripts indicating from which dataset each measure is drawn in the paper table version. Measures in Panel A are drawn from both main_balance_baseline.csv and main_balance_midline.csv; measures in Panel B from main_balance_midline.csv; measures in Panel C from main_balance_baseline.csv; and measures in Panel D from main_attrition.tex. The same applies to Table A3.			
Table 4	Table4.do	114	main_compliance_results.tex
		220	main_revenues_results.tex
Table 5	Table5.do	181	assessment_bribes.tex
		465	attitudes.tex
Table 6	Table6.do	103	main_visits_results.tex
Table 7	Table7.do	137	main_centralwinfo_results.tex
Table 8	Table8.do	370	chiefs_info_payease.tex
		458	chiefs_info_wtp.tex
Figure 1	Figures1_A9_A10_A11_A14.do	554	chars_visited.pdf
		768	chars_PEXHQ.pdf
Table 9	Table9.do	485	main_incidence_results.tex
Figure A1	N.A. (no data)		
Figure A2	N.A. (no data)		
Figure A3	N.A. (no data)		
Table A1	TableA1.do	170	collector_summary.tex
Figure A4	FigureA4.do	97	taxes_paid_DGRKOC_educ_lvl.pdf
		213	taxes_paid_chief_educ_lvl.pdf
		107	taxes_paid_DGRKOC_educ_yrs.pdf
		223	taxes_paid_chief_educ_yrs.pdf
		117	taxes_paid_DGRKOC_possessions_nb.pdf

		233	taxes_paid_chief_possessions_nb.pdf
Table A2	TableA2.do	405	balance_Ftest_nbhdvars.tex
		472	balance_Ftest_baselinevars.tex
		539	balance_Ftest_midlinevars.tex
Table A3	Table A3	491	balance_baseline_wcontrol.csv
		642	balance_midline_wcontrol.csv
		800	attrition_wcontrol.tex
Table A3 Note: See *Table 3 Note* above for guidance on how replication outputs map to paper table structure.			
Table A4	TableA4.do	452	balance_midline_missing.tex
Figure A5	FigureA5.do	48	compliance_over_time.pdf
Table A5	TablesA5_A18_FigureA6.do	238	compl_CvL_results_timeimbal.tex
		484	rev_CvL_results_timeimbal.tex
Figure A6	TablesA5_A18_FigureA6.do	125	shiftFE_compl_CvL.pdf
		371	shiftFE_rev_CvL.pdf
		619	shiftFE_compl_CvCLI.pdf
		863	shiftFE_rev_CvCLI.pdf
Table A6	TableA6.do	79	compl_results_saturated.tex
		140	rev_results_saturated.tex
Table A7	TableA7.do	322	compl_results_controls.tex
		394	rev_results_controls.tex
Table A8	TablesA8_A14_A15_A27	573	cvl_collector_differences_control.tex
Table A9	TableA9.do	182	exemptions.tex
Table A10	TableA10.do	138	awareness_other_tmts.tex
Table A11	TableA11.do	183	fiscal_externalities.tex
Table A12	TableA12.do	178	salongo_tax_actual.tex
		436	salongo_tax_predicted.tex
Table A13	TableA13.do	31	cvl_total_tax_burden.tex
Table A14 (Rows 1–3)	TablesA14_A43_A44	537	bribe_chief_worried_sanctions_col1-3.tex
Table A14 (Rows 4–6)	TablesA8_A14_A15_A27	593	bribe_chief_worried_sanctions_col4-6.tex
Table A15	TablesA8_A14_A15_A27	510	bribe_chief_het_condensed_edited.tex
Table A16	TableA16.do	372	predicted_bribe_p75.tex
		413	predicted_bribe_p90.tex
Table A17	TableA17.do	107	visits_results_nohouseFE.tex
Table A18	TablesA5_A18_FigureA6.do	732	compl_CvCLI_results_timeimbal.tex

		976	rev_CvCLI_results_timeimbal.tex
Table A19	TableA19.do	290	centralwinfo_controls.tex
		290	centralwinfo_nohouseFE.tex
Table A20	TableA20.do	114	owner_present_reg.tex
Table A21	TableA21.do	332	chiefs_info_payease_nohouseFE.tex
		410	chiefs_info_wtp_nohouseFE.tex
Table A22	TableA22.do	349	chiefs_info_payease_interact.tex
		396	chiefs_info_wtp_interact.tex
Figure A7	FigureA7_TableA23.do	113	visits_chefknowindex_L_binned.pdf
		117	taxes_paid_chefknowindex_L_binned.pdf
		132	visits_chefknowindex_CwI_binned.pdf
		136	taxes_paid_chefknowindex_CwI_binned.pdf
		153	visits_chefknowindex_C_binned.pdf
		157	taxes_paid_chefknowindex_C_binned.pdf
Table A23	FigureA7_TableA23.do	234	tablechefknowindex.tex
Table A24	TablesA24_A25.do	46	dist_collectors.tex
Table A25	TablesA24_A25.do	80	collectors_C_near_home.tex
		110	collectors_C_CLI_near_home.tex
Table A26	TableA26.do	132	pay_registration.tex
Figure A8	N.A. (no data)		
Table A27	TablesA8_A14_A15_A27	510	taxes_paid_chief_het_condensed_edited.tex
Table A28	TableA28_A29_A30.do	64	flier_effects_tax_timeFE.tex
Table A29	TableA28_A29_A30.do	121	flier_PanelA_timeFE.tex
	TableA28_A29_A30.do	159	flier_PanelB_timeFE.tex
	TableA28_A29_A30.do	197	flier_PanelC_timeFE.tex
	TableA28_A29_A30.do	235	flier_PanelD_timeFE.tex
	TableA28_A29_A30.do	273	flier_PanelE_timeFE.tex
Table A30	TableA28_A29_A30.do	315	flier_PanelA_timeFE_Lhet.tex
	TableA28_A29_A30.do	353	flier_PanelB_timeFE_Lhet.tex
	TableA28_A29_A30.do	391	flier_PanelC_timeFE_Lhet.tex
	TableA28_A29_A30.do	429	flier_PanelD_timeFE_Lhet.tex
	TableA28_A29_A30.do	467	flier_PanelE_timeFE_Lhet.tex
Figure A9	Figures1_A9_A10_A11_A14.do	674	chars_visited_bytmt.pdf
		890	chars_PEXHQ_bytmt.pdf

Figure A10	Figures1_A9_A10_A11_A14.do	1205	chars_visited_nohouseFE.pdf
		1299	chars_PEXHQ_nohouseFE.pdf
Figure A11	Figures1_A9_A10_A11_A14.do	1004	chars_visited_nonbhdmean.pdf
		1091	chars_PEXHQ_nonbhdmean.pdf
Figure A12	FigureA12.do	438	chief_indices_CvL.pdf
		550	chief_indices_LvCLI.pdf
		672	chief_indices_bytmt.pdf
Table A31	TableA31.do	138	ethnicity_interaction.tex
Table A32	TableA32.do	483	incidence_results_nohouseFE.tex
Table A33	TableA33.do	511	incidence_results_nonbhdmean.tex
Table A34	TablesA34_A35.do	452	incidence_interactions_house.tex
Table A35	TablesA34_A35.do	525	incidence_interactions_compl.tex
		572	incidence_interactions_rev.tex
Figure A13	FigureA13.do	435	dist_housequal_visited.tex
		458	dist_housequal_taxpayers.tex
		483	dist_inc_visited.tex
		506	dist_inc_taxpayers.tex
		531	dist_liq_visited.tex
		555	dist_liq_taxpayers.tex
Figure A14	Figures1_A9_A10_A11_A14.do	1413	chars_visited_LvCLI.pdf
Table A36	TableA36.do	299	wellbeing_tax.tex
		358	wellbeing_instrtax.tex
		420	wellbeing_instrtaxbribe.tex
Table A37	TableA37.do	311	views_tax.tex
		343	views_bribe.tex
Figure A15	FigureA15.do	113	return_by_day_central_month1-2_bin.pdf
		117	return_by_visits_central_month1-2_bin.pdf
Figure A16	N.A. (no data)		
Figure A17	N.A. (no data)		
Figure A18	FigureA18_TableA38.do	49	compliance_over_time_CXL.pdf
Table A38	FigureA18_TableA38.do	163	centralxlocal_results.tex
Table A39	TableA39.do	108	CvL_Teamwork_TeamComp_a7.tex
Table A40	TablesA40_A41_A42.do	305	visits_time_deciles.tex

Table A41	TablesA40_A41_A42.do	240	demoralization_checks.tex
Table A42	TablesA40_A41_A42.do	623	centralwinfo_exposure.tex
Table A43	TablesA14_A43_A44	486	endline_collector_traits_small.tex
Table A44	TablesA14_A43_A44	533	collector_amotivation.tex
Figure A19	FigureA19.do	191	KnowsIndex_C_vs_L_NonCollectorChiefs.pdf
Figure A20	FigureA20.do	25	time_collection_CvLvCLI.pdf
Figure A21	FiguresA21_A22.do FiguresA21_A22.do	96 64	costs_by_treatment.pdf marginal_revenue_hypothetical.pdf.pdf
Figure A22	FiguresA21_A22.do	71	scatter_benefit_cost_dist_center_CvsL.pdf
Table A45	TableA45.do	168	bribe_multiplier.csv

References

Balán, Pablo, Augustin Bergeron, Gabriel Tourek, and Jonathan Weigel, "Replication Data for: Local Elites as State Capacity: How City Chiefs Use Local Information to Increase Tax Compliance in the D.R. Congo," Technical Report, American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor] 2021.