

Taxing Identity

Theory and Evidence from Early Islam

Mohamed Saleh* Jean Tirole*

March 20, 2021

Replication Instructions

The replication folder consists of five main datasets that are needed to replicate most of the figures and tables in the main text, the Supplemental Appendix, and the Webpage Appendix:

1. “*Dataset 1 - cadastre1315.dta*”: This dataset consists of the universe of Egyptian villages in the 1315 cadastre ($N = 1,782$ villages). It contains all the variables that are used in the empirical cross-sectional analysis of the two outcomes: (1) =1 if no Coptic church or monastery in village j circa 1200, and (2) the total tax transfer (*ibra*) per unit of land in village j in 1375.
2. “*Dataset 2 - papyripolltax641–1100.dta*”: This dataset contains the individual-level papyrological poll-tax sample in 641–1100 ($N = 408$ taxpayers). It contains all the variables that are used in the empirical cross-sectional analysis of the two other outcomes: (3) the poll tax rate of taxpayer h in *kura i* in 641–1100, and (4) the imputed poll tax revenue per capita in *kura i* circa 1200.
3. “*Dataset 3 - 1848_1868_population_census_samples_restricted.dta*”: This dataset is a sub-sample of the 1848 and 1868 individual-level population census samples. It is restricted to Coptic and Muslim men of rural district of origin, who are above 15 years of age, with non-missing age, occupational title, and district of origin. It

*Toulouse School of Economics and Institute for Advanced Study in Toulouse, 1 Esplanade de l'Université, Toulouse 31080, France.

is used in one of the robustness checks in the cross-sectional analysis of conversions: See columns (5)–(6) in Supplemental Appendix Table B.3, and Webpage Appendix Table D.4.

4. “*Dataset 4 - Caliphs_641_1170.dta*”: This is the Caliph-level dataset that is employed in the empirical time-series analysis of poll tax hikes and conversion waves in 641–1170.
5. “*Dataset 5 - Governors_641_1170.dta*”: This is the governor-level dataset that is employed as a robustness check in the empirical time-series analysis of poll tax hikes and conversion waves in 641–1170.

In addition, there are four supplementary datasets that are used in Supplemental Appendix Figures B.1–B.3:

6. “*Dataset 6 - cadastre1315_Map.dta*”: This dataset maps villages in the 1315 cadastre using the 2006 population census third-level administrative division (Supplemental Appendix Figure B.1).
7. “*Dataset 7 - EGY_Subdistricts_coord.dta*”: This dataset contains the coordinates of the 2006 population census third-level administrative division (Supplemental Appendix Figure B.1).
8. “*Dataset 8 - longtermtrendoutcomes.dta*”: This dataset contains the long-term trends of poll tax hikes and conversion waves in 641–1170, expanded to the year level, that are used in Supplemental Appendix Figure B.2.
9. “*Dataset 9 - longtermtrendtaxrevenue.dta*”: This dataset contains the long-term trend of poll and land tax revenues in 641–1170 that is used in Supplemental Appendix Figure B.3.

There are five STATA do-files:

1. “*Dofile 1 - Tables.do*”: This do-file contains the replication coding of Tables 1–4.
2. “*Dofile 2 - Supplemental_Appendix_Figures.do*”: This do-file contains the replication coding of Supplemental Appendix Figures B.1–B.3.
3. “*Dofile 3 - Supplemental_Appendix_Tables.do*”: This do-file contains the replication coding of Supplemental Appendix Tables B.1–B.6.
4. “*Dofile 4 - Webpage_Appendix_Figures.do*”: This do-file contains the replication coding of Webpage Appendix Figures D.2–D.3.

5. “*Dofile 5 - Webpage_Appendix_Tables.do*”: This do-file contains the replication coding of Webpage Appendix Tables D.3–D.20.

Note: Webpage Appendix Tables D.1–D.2 are based on the full 1848 and 1868 individual-level population census samples that will be released on IPUMS-International in 2021.