ReadMe for Replicating

Informal Institutions, Collective Action, and Public Investment in Rural China

Authors:
YIQING XU, Stanford University
YANG YAO, Peking University

Installing Packages

R

1. Install packages from CRAN:

   ```
   install.packages(c("tidyr", "dplyr", "haven",
     "estimatr", "modelsummary", "ggplot2",
     "paneltools", "fect", "fixest", "kableExtra",
     "rdrobust", "panelView", "patchwork"))
   ```

2. Make sure to set your working directory to the root folder of this replication package.[1]

Folders and Their Functionalities

- **data:** Contains two data files (`XuYao2015.dta`, `lineageorg.dta`) and a codebook of variable definitions.

- **code:** Holds `0_install.R`, `1_figure.R` and `2_table.R`, which install the required packages and replicate all main tables and figures in the paper.

- **tutorial:** Contains a condensed tutorial of the original paper (in `.Rmd` and extracted code in `.R` formats) plus example scripts that replicate the analysis for an online tutorial.

Notes

1. Please set your path to the root replication folder in R.

2. Several updates have been made to the original published version; some results may not be exactly replicated.

---

[1]Replication files organized by Jinwen Wu, a predoctoral fellow at Stanford University, under the guidance of Professor Yiqing Xu. For any replication inquiries, please contact: Jinwen Wu.

- **Figure 4**: Updated to focus on the first two largest clans, instead of the top four clans in the published version.

- **Figure 5**: Presented in a regression table rather than a figure, to summarize the results more concisely.

- **Figure 6**: Added confidence intervals to highlight regression discontinuity results more clearly.

- **Figure 7(b)**: Included a trend line of levies for cases with no public project.

- **Table 7**: Introduced an additional specification with a more comprehensive set of controls.

Data and Code Usage

.dta Files for Tables and Figures

.dta File	Figures	Tables	Description / Purpose
XuYao2015.dta	1, 3, 6, 7	1–7	Main dataset (loaded as `df`). Contains annual village data, clan indicators, election info, and public investment.
lineageorg.dta	2, 4	Updated Figure 5	Secondary dataset (loaded as `data`). Contains clan-level variables (lineage hall, ceremonies, genealogies).

R Files and Replication Outputs

0_install.R

- Check and install all required packages for replication.

1_figure.R

- *Reads*: `XuYao2015.dta` (as `df`) and `lineageorg.dta` (as `data`) for plotting.

- *Creates Figures 1–7*:

  - **Figure 1**: Clan population shares, elected VCs over time, and public investment trends.
  - **Figure 2**: Bar chart of lineage halls, ceremonies, and maintaining lineage halls during the Cultural Revolution.
  - **Figure 3**: Event-study plot of large-clan VCs on public investment.

- **Figure 4**: VCs of large clans, clan size, and cohesiveness. Updated to include *only the two largest* clans.
- **Figure 5**: Now replaced by a regression table.
- **Figure 6**: Robustness check with regression discontinuity design. Added confidence intervals for the plot.
- **Figure 7**: Public investment and levies on villagers. Added levies trend for "no project" vs. "has project."

2_table.R

- *Reads*: `XuYao2015.dta` (as `df`) and `lineageorg.dta` (as `data`).

- *Produces Tables 1–7*:

  - **Table 1**: Descriptive statistics. Summaries of key variables (public investment, population, income).
  - **Table 2**: Main two-way fixed effects regressions; public investment vs. largest-/2nd-largest-clan VCs.
  - **Table 3**: Public investment under large clan VCs by project type.
  - **Table 4**: VCs of large clans and levies.
  - **Table 5**: Administrative expenditures.
  - **Table 6**: Effects of VC characteristics on public investment.
  - **Table 7**: Interplay of electoral institutions and large-clan VCs.