

Progress Report

Sylvie

February 17, 2025

1 Data Availability

1.1 Variables of Interest

revo_group: newly formed groups or cumulated one?

1.2 Control Variables

Variables such as *rice_price* and *price_level* is useless since it just reflects differences in time (like time-series data). However, we can still use it as part of the interaction term.

* The next step: Find panel data that is suitable as control variables.

→ Cultural Capital, Train, ...

* I want to find peasant rebellion data or civil society data to continue my study.

* Pay attention the the value '0' and remember to make the data 'full' * Present the data with a detailed table.

1.3 Regional Limit

- Check whether the geographical region during 1881-1911 changed.
- Delete unnecessary provinces (exclude provinces with *tele* and *revo_goup* both equals 0)
- If necessary, merge some data (e.g. Fengtian, Jilin and Heilongjiang could be considered as *Dongsansheng*)

1.4 Potential IV

We could use some climate data (such as drought, water precipitation etc.) to run IV regression (to solve endogenous problems)

2 Baseline Regression

2.1 Fixed Effect and Robustness Check

* The current problem is that the result does not comply with **Homoscedasticity** hypothesis - find some solutions

2.2 Regression Coefficient

Calculate the mean, median etc. to check the significance of the coefficient.

3 Mechanism Analysis

- * RDD means all variables are continuous except one.
- * RD regression: Maybe we can use the Southeast Mutual Protection as an IV. (or maybe an event-study)
- * What is related to communication cost? * Short-term and long-term effect: Consider lag effect