Replication files for "China's Low-Productivity Innovation Drive: Evidence from Patents"

By Yuen Yuen Ang, Nan Jia, Bo Yang, and Kenneth Huang

Summary

This folder includes the materials required to replicate all results in the paper "China's Low-Productivity Innovation Drive: Evidence from Patents."

The datasets and programs were prepared using Stata 17.0.

The folder includes two datasets:

- *main.dta* This is the main panel dataset. The unit-of-analysis is city-year. It is used to generate most results in the main texts or the Appendix.
- *national_year.dta* This is a time series dataset at the national level. It is used to generate Figure 1 and Figure 3 in the main texts of the paper.

The programs required to run the above two datasets are stored in *replication codes.do*.

The table below provides a mapping between variable names in the paper and variable names in the *main.dta* dataset.

Variables in the paper	Variable names in the *main.dta*
1. Number of patents (log)	lnpatent_count
2. Share of novel patents (%)	patent_novelshare
3. Intensity of political competition (IPC)	govrevenueg_cls
4. Post-Campaign	post
5. GDP per capita (log)	lngdppc
6. Population growth	popugrowth
7. FDI/GDP	fdishare
8. Gender	gender
9. Age	age
10. Education	edu
11. Ethnic minority	minority
12. Provincial GDP (log)	lngdp_province

Note: the dataset contains two additional variables. The *govrevenueg_cls_nosd* refers to the intensity of political competition (IPC) variable before standardization. It is used to replicate the results in Figure A1 and Table A2. The *govrevenueg_rank* refers to a city's ranking within the same province in a given year, based on the growth rate of fiscal revenue. It is used to replicate the results in Table A3 and A4.

Original Data Sources

- Patent variables were constructed using data sourced from China's State Intellectual Property Office (SIPO), renamed CNIPA in 2018.
- City-level economic statistics were collected from the China City Statistical Yearbooks.
- Characteristics of city party secretaries were collected from the China Stock Market & Accounting Research (CSMAR) Database.
- Provincial GDP data were collected from the website of China National Bureau of Statistics. This information is only available starting from 1992. We used the 1992 value to fill in the blanks for 1990 and 1991.