Replication Package for "Industrialization and the Big Push: Theory and Evidence from South Korea" *

Jaedo Choi
University of Texas at Austin

Younghun Shim
International Monetary Fund

February 2026

*E-mail: jaedo.choi@utexas.edu

1. Overview

- *SHELL.do* in the *CODE* folder creates the dataset and regression results.
- *SHELL.m* in the *MATLAB/CODE* folder implements quantitative exercises.

2. Raw data

The *RAW* folder contains the raw datasets:

- Exchange rates and GDP deflators information is in the *RAW/AGG* folder.
- Industrial complex data are in *RAW/COMPLEX* folder.
- KLEMS data is in *RAW/EUKLEMS* folder.
- Firm-level data (balance sheet and region sector information in the *RAW/FIRM* folder.
- Geographic information is in the *RAW/GEO* folder.
- Business group information is in the *RAW/GROUP* folder.
- Input-output tables are in the *RAW/IO* folder.
- KOTRA international trade fair information is in the *RAW/KOTRA* folder.
- Population information is in the *RAW/POP* folder.
- STAN industrial database is in the *RAW/STAN* folder.
- Credit information is in the *RAW/SUBSIDY* folder.
- Tariff information is in the *RAW/TARIFF* folder.
- Firm-level technology adoption and cancellation episodes are in the *RAW/TECH* folder.
- USPTO patent datasets are in the *RAW/USPTO*.
- Trade datasets from Feenstra et al. (2005) is in the *RAW/TRADE*.

3. Empirical Analysis

STATA codes for empirical analysis are in the *CODE* folder. *SHELL.do* in the *CODE* folder creates the dataset and regression results.

- *cleaning_data.do* cleans datasets used for regression input.
- *stylized_facts_summary_stat.do* produces descriptive statistics and figures related to the historical background (Figure 1; Appendix Figures A1, A2, and A3; and Appendix Tables C1, C2, and C3).
- *tfp_estimation.do* estimates production function coefficients.
- *fact1_direct.do* produces main results and robustness checks for Fact 1 on direct effects (Figure 2; Appendix Table C4, C5, C6, C7, C8, C9, C10, and C11; and Appendix Figure C1 and C2).
- *ma_construction.do* creates controls for market size access used for spillover/complementarity regression.
- *IV_construction.do* creates the IV used for spillover/complementarity regression.

- *IV_size_construction.do* creates the IV used for business group size expansion, used as a control for spillover/complementarity regression in a reduced-form fashion.
- *facts2_3_spillover_complementarity.do* conducts the empirical analysis related to Facts 2 and 3 on local spillovers and complementarity (Table 2; and Appendix Table C12, C13, C14, C15, C16, C17, C18, C19, C20, C21, C22, C23, C24, and C25).
- *to_matlab.do* creates inputs used for the quantitative exercises. Created inputs are saved in the *MATLAB/INPUT* folder.
- *non_targeted_moments.do* creates non-targeted moments using the data input created by *MATLAB/CODE/targeted_non_targeted.m* (Appendix Figure F1; and Appendix Table F3)

4. Quantitative Exercises

The *MATLAB* folder contains Matlab codes for the quantitative exercises. The codes are in *MATLAB/CODE*. The *MATLAB/INPUT* folder contains data moments used for the quantification. The output results are saved in the *MATLAB/OUTPUT* folder. *SHELL.m* in the *MATLAB/CODE* folder implements quantitative exercises.

- *MM_shell.m* creates the output. *mm_inner.m* is a subcode that evaluates moments from the model.
- *cft.m* runs counterfactual analysis. *cfft_solve_eq.m* is a subcode that solves the equilibrium.
- *optimal_subsidy.m* searches for the optimal subsidy rate.
- *targeted_non_targeted.m* creates estimation results and non-targeted moments.
- *graphs.m* creates graphs based on the quantitative results (Figure 4; and Appendix Figures F2 and F3).
- *rand_spatial_subsidy.m* implement a robustness exercise which randomizes subsidized regions.
- *cft_results.m* creates counterfactual results (Table 3; and Appendix Tables F4 and F5).

References

Feenstra, Robert C, Robert E Lipsey, Haiyan Deng, Alyson Ma, and Hengyong Mo, "World trade flows: 1962-2000," 2005.