This file provides information on the code and dataset used in the tests of Table 2. For detailed information on the sample, variables or tests see Section I.D.1 of the Internet Appendix.

This folder contains the following files:

1. "Table 2 Replication.do": This Stata do file provides the code used to replicate Table 2. The variables and statistical method used in this file are described as follows:
    a. Dependent variables:
        i. xrdw_atw: defined as the firm's annual R&D expenditures divided by total assets. Both R&D expenditures and total assets are obtained from Compustat Annual Fundamentals file and are winsorized at the 1% and 99% levels. This potentially identifying variable is rounded to the second decimal place in the pseudo data.
        ii. patnumc: the citation-weighted patent count of a firm computed for each year. Patents and citations are obtained from the Kogan et al (2017) data.
    b. Independent variables:
        i. liu: The ratio of an SIC industry's annual US imports from China to US market volume in 1991. All Chinese import data is obtained from the Autor et al (2013) data.
        ii. liusq: the square of liu
        iii. lio: The ratio of comparison countries' total imports to US market volume in 1991.
        iv. liosq: the square of lio
    c. Control Variables:
        i. cr4: used as a control variable and is the four-firm concentration ratio obtained from the Census Bureau. The four-firm concentration ratio is the percentage of Value of Shipments accounted for by the four largest firms in an SIC industry and includes all establishments in the industry, both privately held and public.
        ii. sic4: used as a fixed effect (FE) and is the 4-digit code assigned to an industry.
        iii. year: used as a FE and reflects the corresponding year of the observation.
        iv. segnum: used as a FE and is the number of segments of a firm in each year. Obtained from Compustat Segment data.
        v. gvkey: standard errors are clustered using this variable. This is the firm identifier in the Compustat files. This identifying variable is anonymized.
    d. Partitioning variables: The variable d2vscopeq2 (d2vscopeq3, d2vscopeq4) is used to partition the sample by the median (tercile, quartile) level of the d2vscope variable in the full sample. This variable measures a firm's scope in a given year and is obtained from the Hoberg and Phillips (2024) data.
    e. Statistical Model: a 2sls model is used. The Stata "ivreghdfe" function by Baum et al (2010) and Correia (2017).
2. "pseudo_chinaimp_comp.dta": This pseudo data includes a randomized 1% of the original sample used in the analysis of Table 2. Detailed information on the construction of this sample is provided in Section I.D.1 of the Internet Appendix. Note that this pseudo dataset is for the purpose of illustrating data structure and testing the code and will not reproduce the results reported in Table 2.

References

Autor, D.H., Dorn, D. and Hanson, G.H., 2013. The China syndrome: Local labor market effects of import competition in the United States. American economic review, 103(6), pp.2121-2168.

Baum, C.F., Schaffer, M.E., Stillman, S. 2010. ivreg2: Stata module for extended instrumental variables/2SLS, GMM and AC/HAC, LIML and k-class regression. http://ideas.repec.org/c/boc/bocode/s425401.html

Correia, Sergio. 2017. Linear Models with High-Dimensional Fixed Effects: An Efficient and Feasible Estimator (Working Paper) https://github.com/sergiocorreia/reghdfe/#citation

Hoberg, G. and Phillips, G.M., 2024. Scope, scale and concentration: The 21st century firm. Journal of Finance.

Kogan, L., Papanikolaou, D., Seru, A. and Stoffman, N., 2017. Technological innovation, resource allocation, and growth. Quarterly Journal of Economics, 132(2), pp.665-712.