1. Data

Bilateral trade and production data for 25 industrial sectors of the world's 75 largest economies in 2006 are sourced from CEPII. The industrial sectors are classified in the International Standard Industrial Classification (ISIC) Revision 2. The number of goods is measured by the extensive margin data constructed by Hummels and Klenow (2005). The fixed trade cost is measured by the regulation costs of firm entry collected by Djankov et al. (2002). The quality-adjusted prices are sourced by Feenstra and Romalis (2014). All of the data are collected into two Stata data files:

_DataClean.dta: main data including trade, production, tariff, gravity variables, etc.
industrylist.dta: the data of industry list, names, ISIC codes, etc.

The other data files are for temporary use during the estimation:

_DataReg.dta: the data cleaned for regressions.
_DataReged.dta: the data after regression for counterfactuals.
counter1.dta: the data used to display counterfactual results.
ZERO.txt: the data used to display zero trade flow distributions.
ZERO_fit.txt: the data used to display model fitness.

2. Variables

country_o	country of origin
country_d	country of destination
iso_o	iso of origin
iso_d	iso of destination
distw	weighted distance (pop-wt, km)
pop_o	population of origin, total in mn
gdp_o	GDP of origin (current mn US$)
gdpcap_o	GDP per cap of origin (current in US$)
pop_d	population of destination, total in mn
gdp_d	GDP of destination (current mn US$)
gdpcap_d	GDP per cap of destination (current in US$)
totflow	total trade flow
flow_1-flow_25	trade flow of sector 1-25
tari_1-tari_25	tariffs of sector 1-25
price_FR	quality adjusted export price index by Feenstra and Romalis (2014)
EM	extensive margin data by Hummels and Klenow (2005)
ex_cost	regulation costs of firm entry of origin by Djankov et al. (2002) (% of income per capita)
im_cost	regulation costs of firm entry of destination by Djankov et al. (2002) (% of income per capita)

3. Codes and software

The main.do file conducts most analysis for our paper using our data.
The Aplot.m generates the relevant graphs that show the zero trade flow distribution.

Stata codes are run on Stata 15, while Matlab codes are on Matlab 2021a.

4. Tables and figures

Below are places of codes for each Table or Figure in our paper.

Table 1. main.do lines 164-180
Table 2. main.do lines 262-339
Table 3. main.do lines 400-440
Table 4. main.do lines 464-512
Table 5. main.do lines 535-588

Figure 1. main.do lines 113-140
Figure 2. main.do lines 142-153, and Aplot.m lines 1-22
Figure 3. by hand
Figure 4. main.do lines 442-461, and Aplot.m lines 25-43