The *CEPII Gravity Dataset*

Information and Codebook

Table of content

1. Introduction

CEPII makes available a "square" gravity dataset for all world pairs of countries (224), for the period 1948 to 2015.

This dataset was originally generated by Keith Head, Thierry Mayer and John Ries[1] for the period 1984-2006. Most variables are described in the data appendix of the paper which you should turn to for details and reference when using this data.

The main variables relating to trade costs come from the CEPII distance datasets to which we added new covariates, and that we arranged in a way such that it can be easily merged with any matrix of bilateral flow (usually trade, but many gravity applications use FDI, migrations, or other types of bilateral flows) using standard ISO codes for countries and for any year between 1948 and 2015 (the original paper and dataset stopped in 2006, the current version has been updated by the successive efforts of Julia Jauer, Jules Hugot, and Eve Sihra).

We also provide for replication purposes of Head et al. (2010) quoted above a lighter dataset, including trade flows (cleaned and treated as described in the data appendix) but restricted to observations where those trade flows are non-missing.

Both datasets are available in Stata13 format. Questions should be sent to thierry.mayer@sciencespo.fr .

[1] Head, K., T. Mayer and J. Ries, 2010, "The erosion of colonial trade linkages after independence" Journal of International Economics, 81(1):1-14. (formerly CEPII discussion paper # 2008-27)

2. Data

This dataset contains the following variables:

variable name	variable label
iso3_o	ISO3 alphanumeric
iso2_o	ISO2 alphanumeric
iso3_d	ISO3 alphanumeric
iso2_d	ISO2 alphanumeric
year	
contig	1 for contiguity
comlang_off	1 for common official or primary language
comlang_ethno	1 if a language is spoken by at least 9% of the population in both countries
comcol	1 for common colonizer post 1945
col45	1 for pairs in colonial relationship post 1945
distw	weighted distance (pop-wt, km)
pop_o	Population, total in mn
pop_d	Population, total in mn
gdp_o	GDP (current US$)
gdp_d	GDP (current US$)
gdpcap_o	GDP per cap (current in US$)
gdpcap_d	GDP per cap (current in US$)
gdp_ppp_o	GDP, PPP (current international $)
gdp_ppp_d	GDP, PPP (current international $)
gdpcap_ppp_o	GDP per cap, PPP (current international $)
gdpcap_ppp_d	GDP per cap, PPP (current international $)
gdp_ppp_pwt_o	GDP, current PPP (2011 US$)
gdp_ppp_pwt_d	GDP, current PPP (2011 US$)
pop_pwt_o	Population, total in mn (PWT)
pop_pwt_d	Population, total in mn (PWT)
area_o	Area in sq. kms
area_d	Area in sq. kms
tdiff	nb of hours difference between ex and im
heg_o	1 if origin is current or former hegemon of destination
heg_d	1 if destination is current or former hegemon of origin
conflict	1 if war

indepdate	Independence date if colony == 1
colony	1 for pair ever in colonial relationship
curcol	1 for pair currently in colonial relationship
sibling	1 for pair ever in sibling relationship
cursib	1 for pair currently in sibling relationship
empire	hegemon if cursib==1
sever	severance year for pairs if sibling == 1
sib_conflict	1=Pair ever in sibling relationship and conflict
comcur	1 for common currency
comrelig	common religion
comleg_pretrans	1 if common legal origins before transition
comleg_posttrans	1 if common legal origins after transition
transition_legalchange	1 if common legal origin changed since transition
legold_o	origin legal system before transition
legold_d	destination legal system before transition
legnew_o	origin legal system after transition
legnew_d	destination legal system after transition
gatt_o	1 if origin is GATT/WTO member
gatt_d	1 if destination is GATT/WTO member
pta_bb	1=Non-reciprocal PTA ; 2=PTA (Source: Baier & Bergstrand, 2009)
fta_wto	1=RTA (Source: WTO, 2015)
fta_bb	1=FTA; 2=Cust. Union; 3=Common Market; 4=Economic union (Source: Baier & Bergstrand, 2009)
fta_hmr	1=FTA (Source: Head, Mayer and Ries, 2010)
acp_to_eu	1 for ACP to EU
eu_to_acp	1 for EU to ACP
gsp_o_d	1 if origin is donator
gsp_d_d	1 if destination is donator
flaggsp_o_d	report of changes in Rose data
flaggsp_d_d	report of changes in Rose data
entry_cost_o	Cost of business start-up procedures (% of GNI per capita)
entry_cost_d	Cost of business start-up procedures (% of GNI per capita)
entry_proc_o	Start-up procedures to register a business (number)
entry_proc_d	Start-up procedures to register a business (number)
entry_time_o	Time required to start a business(days)
entry_time_d	Time required to start a business(days)
entry_tp_o	Days+Procs to start a business

entry_tp_d	Days+Procs to start a business
eu_o	1=Origin is a EU member
eu_d	1=Destination is a EU member

3. Codebook

3.1. Countries

iso3_o
Standard ISO code for exporting country (three letters).

iso3_d
Standard ISO code for importing country (three letters).

iso2_o
Standard ISO code for exporting country (two letters).

iso2_d
Standard ISO code for importing country (two letters).

3.2. Period

year
Numeric, from 1948 to 2015. The dataset is a complete dyad*year matrix.

3.3. Geography

area_o
Area of origin in square kilometers Source: CEPII Distance Dataset

area_d
Area of destination in square kilometers. Source: CEPII Distance Dataset

distw
Weighted bilateral distance between origin and destination in kilometer (population weighted). Source: CEPII Distance Dataset

tdiff
Time difference between origin and destination, in number of hours. For countries which stretch over more than one time zone, the respective time zone is generated via the mean of all its time zones (for instance: Russia, Canada, USA)

3.4. Common language, common history

colony
Dummy for origin and destination ever in colonial relationship. Source: Head et al. (2010)

col45
Dummy for origin and destination in colonial relationship post 1945. Source: Head et al. (2010)

col_to
Dummy for origin and destination in colonial relationship post 1945. Source: Head et al. (2010)

col_fr
Dummy for origin and destination in colonial relationship post 1945. Source: Head et al. (2010)

comcol
Dummy for common colonizer of origin and destination post 1945. Source: Head et al. (2010)

comlang_off
Dummy for common official or primary language. Source: Head et al. (2010)

comlang_ethno
Dummy for language spoken by at least 9% of the population in both countries. Source: Head et al. (2010)

conflict
Dummy if war between origin and destination Source: Head et al. (2010)

curcol
Dummy if origin and destination currently in colonial relationship. Source: Head et al. (2010)

cursib
Dummy if origin and destination currently in sibling relationship, i.e. two colonies of the same empire. Source: Head et al. (2010)

empire
ISO code (three letters) of hegemon if cursib == 1. Source: Head et al. (2010)

heg_o
Dummy if origin is current or former hegemon of destination. Source: Head et al. (2010)

heg_d
Dummy if destination is current or former hegemon of origin. Source: Head et al. (2010)

indepdate
Independence date if colony == 1. Source: Head et al. (2010)

sever
Severance year for pairs if sibling == 1. Source: Head et al. (2010)

sibling
Dummy for origin and destination ever in sibling relationship, i.e. two colonies of the same empire. Source: Head et al. (2010)

sib_conflict
Dummy for origin and destination ever in sibling relationship and conflict. Source: Head et al. (2010)

3.5. Legal and religion similarities

comleg_pretrans
Dummy if origin and destination share common legal origins before transition. Source: LaPorta, Lopez-de-Silanes and Shleifer (2008)

comleg_posttrans
Dummy if origin and destination share common legal origins after transition. Source: LaPorta, Lopez-de-Silanes and Shleifer (2008)

comrelig
Religious proximity (Disdier and Mayer, 2007) is an index calculated by adding the products of the shares of Catholics, Protestants and Muslims in the exporting and importing countries. It is bounded between 0 and 1, and is maximum if the country pair has a religion which (1) comprises a vast majority of the population, and (2) is the same in both countries. Source of religion shares: LaPorta, Lopez-de-Silanes, Shleifer and Vishny (1999), completed with the CIA world factbook.

legnew_o
Legal system of origin after transition. This variable takes the values: “fr” for French, “ge” for German, “sc” for Scandinavian, “so” for Socialist and “uk” for British legal origin. Source: LaPorta, Lopez-de-Silanes and Shleifer (2008)

legnew_d
Legal system of destination after transition. For values of the variable, see legnew_o. Source: LaPorta, Lopez-de-Silanes and Shleifer (2008)

legold_o
Legal system of origin before transition. For values of the variable, see legnew_o. Source: LaPorta, Lopez-de-Silanes and Shleifer (2008)

legold_d
Legal system of destination before transition. For values of the variable, see legnew_o. Source: LaPorta, Lopez-de-Silanes and Shleifer (2008)

transition_legalchange
Dummy if common legal origin changed since transition. Source: LaPorta, Lopez-de-Silanes and Shleifer (2008)

3.6. Gross Domestic Product, Population

gdp_o
Gross Domestic Product of origin (current US$) Source: From 1960 to 2015, the data comes from the World Development Indicators, World Bank. For Taiwan, the data comes from the Directorate-General of Budget, Accounting and Statistics (DGBAS). For data before 1960 and for some countries for which WDI does not keep track (for instance, Russia), we complete the data using the Maddison[2] and Barbieri databases[3].

gdp_d
Gross Domestic Product of destination (current US$) Source: see gdp_o

gdpcap_o
Gross Domestic Product per capita of origin (current US$) Source: computed from gdp_o and pop_o

gdpcap_d
Gross Domestic Product per capita of destination (current US$) Source: computed from gdp_d and pop_d

[2] Data available on http://www.ggdc.net/maddison/maddison-project/home.htm.
[3] Data available on http://www.correlatesofwar.org.

gdp_ppp_o
Gross Domestic Product of origin (PPP, current international $) Source: World Development Indicators, World Bank For Taiwan, Penn World Table

gdp_ppp_d
Gross Domestic Product of destination (PPP, current international $) Source: see gdp_ppp_o

gdpcap_ppp_o
Gross Domestic Product per capita of origin (PPP, current international $) Source: see gdp_ppp_o

gdpcap_ppp_d
Gross Domestic Product per capita of destination (PPP, current international $) Source: see gdp_ppp_o

gdp_ppp_pwt_o
Gross Domestic Product of origin (current PPP, 2011 US$) Source: Penn World Table (Feenstra, Inklaar and Timmer, 2015)[4]

gdp_ppp_pwt_d
Gross Domestic Product of destination (current PPP, 2011 US$) Source: see gdp_ppp_pwt_o

pop_o
Population of origin, total in million Source: see gdp_o

pop_d
Population of destination, total in million Source: see gdp_o

[4] Data available on www.ggdc.net/pwt.

pop_pwt_o
Population of origin, total in million (Penn World Table) Source: see gdp_ppp_pwt_o

pop_pwt_d
Population of destination, total in million (Penn World Table) Source: see gdp_ppp_pwt_o

3.7. Currency Unions

comcur
Dummy for common currency Updated with dataset from De Sousa (2015)

3.8. GATT/WTO Membership

gatt_o
Dummy if origin is GATT/WTO member. Source: WTO

gatt_d
Dummy if destination is GATT/WTO member. Source: WTO

3.9. Regional Trade Agreements

eu_o
Dummy if origin is a member of the European Union. Source: CEPII Distance Dataset

eu_d
Dummy if destination is a member of the European Union. Source: CEPII Distance Dataset

eu_d
Dummy if destination is a member of the European Union. Source: CEPII Distance Dataset

fta_bb
Variable coded as 1 for Free Trade Area; 2 for Customs Union; 3 for Common Market; 4 for Economic Union. Source: Baier and Bergstrand (2009)

fta_hmr
Dummy for Free Trade Agreement. Source: Head, Mayer and Ries (2010)

fta_wto
Dummy for Regional Trade Agreement. Source: WTO (2015)

3.10. Preferential Trade Agreements

acp_to_eu
Dummy for ACP country exporting to EC/EU member. A Preferential Trade Agreement on imports between former colonies and other developing countries (Africa, Caribbean, Pacific), and members of the European Union.

eu_to_acp
Dummy for EC/EU member exporting to ACP country. A Preferential Trade Agreement on imports between former colonies and other developing countries (Africa, Caribbean, Pacific), and members of the European Union.

flaggsp_o_d
Report changes in Rose's data on gsp_o_d. No gsp recorded in Rose; Data directly from Rose; Changes in data from Rose; Assumption that gsp continues after 1999.

flaggsp_d_d
Report changes in Rose's data on gsp_d_d. No gsp recorded in Rose; Data directly from Rose; Changes in data from Rose; Assumption that gsp continues after 1999.

gsp_o_d
Dummy if origin is donator in Generalized System of Preferences (GSP). Source: Rose (2004)

gsp_d_d
Dummy if destination is donator in Generalized System of Preferences (GSP). Source: Rose (2004)

pta_bb
Variable coded as 1 for non-reciprocal Preferential Trade Agreement; 2 for Preferential Trade Agreement. Source: Baier and Bergstrand (2009)

3.11. Entry Costs

entry_cost_o
Cost of business start-up procedures for origin (% of GNI per capita). Source: World Development Indicators, World Bank

entry_cost_d
Cost of business start-up procedures for destination (% of GNI per capita). Source: World Development Indicators, World Bank

entry_proc_o
Start-up procedures to register a business for origin (number) Source: World Development Indicators, World Bank

entry_proc_d
Start-up procedures to register a business for destination (number) Source: World Development Indicators, World Bank

entry_time_o
Time required to start a business for origin (days) Source: World Development Indicators, World Bank

entry_time_d
Time required to start a business for destination (days) Source: World Development Indicators, World Bank

4. References

Baier, S. L., and J. H. Bergstrand. 2009. “Estimating the effects of free trade agreements on international trade flows using matching econometrics”, *Journal of International Economics*, vol. 77(1), pages 63-76.

Barbieri, Katherine. 2005. *The Liberal Illusion: Does Trade Promote Peace?* Ann Arbor: University of Michigan Press.

Disdier, A-C., and T. Mayer. 2007. “Je t'aime, moi non plus: Bilateral Opinions and International Trade”, *European Journal of Political Economy* 23(4): 1140-1159.

De Sousa, J. 2012. “The currency union effect on trade is decreasing over time”, *Economics Letters*, 117(3), 917-920.

Feenstra, R. C., R. Inklaar and M. P. Timmer. 2015. "*The Next Generation of the Penn World Table*", American Economic Review, 105(10), 3150-3182

Head, K., T. Mayer, and J. Ries. 2010. “The erosion of colonial trade linkages after independence”, *Journal of International Economics* 81(1):1-14.

LaPorta, R., Lopez-de-Silanes, F., Shleifer, A., and R. Vishny. 1999. “The Quality of Government”, *Journal of Law, Economics and Organization* 15 (1): 222-279.

LaPorta, R., Lopez-de-Silanes, F., and A. Shleifer. 2008. “The Economic Consequences of Legal Origins”, *Journal of Economic Literature* 46 (2): 285-332.

Rose, A. 2004. “Do We Really Know that the WTO increases Trade?”, *American Economic Review* 94 (1), 98-114.