STATA Syntax. Dofile

```
/* Country dummies */

tab country1, gen(importer_)

tab country0, gen(exporter_)

/*Time dummies*/

tab year, gen(year_)

/*Country-time dummies*/

tab gdp_d, gen(importeryear_)

tab gdp_o, gen(exporteryear_)

/*Pair dummies*/

egen pairid = group(country0 country1)

tab pairid, gen(pair_)

xtset pairid year
```

Logarithm form of major variables

```
gen LOGTRADE=ln(flow)

gen LOGDIS=ln( distw)

gen LOGPOPO=ln( pop_o)

gen LOGPOP1=ln( pop_d)

gen LOGAREA0=ln(area_o)

gen LOGAREA0=ln(area_d)

gen loggdpc0=ln(gdpcap_o)

gen loggdpc1=ln(gdpcap_d)
```

Where, inflow=trade flow from exporter to inporter, distw=distance between countries' capital, pop_o =population of exporting countries, pop_d= population of importer countries, areo_o =areal size of exporter countries, area_d=areal size of importer countries, gdpcap_o=GDP per capita of exporter countries and gdpcap_d= GDP per capita of importer countries.

/*to create dummy variable*/

tabulate colony, generate(COL)

tabulate comlang_off , generate(LAN)

tabulate gatt_o , generate(WTO0)

tabulate gatt_d , generate(WTO12)

tabulate comcol , generate(COMCOL2)

tabulate accesstosea , generate(SEA)

tabulate fta_wto, generate(RTA)

where, COL=colonial relationship, LAN=common language, WTO0=WTO membership (exporter country), WTO12=WTO membership (importer country), COMCOL2=common colony, SEA=access to sea and RTA=regional trade agreements.

Lagged values of time varying explanatory variables

gen llagloggdpc0=loggdpc0[_n-1]

gen llagloggdpc1=loggdpc1[_n-1]

gen llaglogpop0=logpop0[_n-1]

gen llaglogpop1=logpop1[_n-1]

/*principal component analysis*/

physical_communication infrastructure index

pca intrntusers100prsn qualityinf qualityair qualityroad tel

estat kmo

predict pc1 pc2, score

where, pc1=physical_communication1, pc2=physical_communication2

gen physical_communication4=physical_communication1+4 (to make values positive)

gen physical_communication24=physical_communication2+4

gen logphysical_communication1=ln(physical_communication4)

gen logphysical_communication2=ln(logphysical_communication24)

gen loghardfinal=logphysical_communication2*logphysical_communication1

entry cost index

pca entry_cost_o entry_proc_o entry_time_o

estat kmo

predict pc1, score

where, pc1=ccc

gen ccc1=ccc+4

gen logccc1=ln(ccc1)

lnccc1=entry cost

border_transport efficiency index

documenttoexport documenttoimport timetoimport timetoexportdays

estat kmo

predict pc1, score

gen border_transport4=pc1+4

gen logsoft=ln(border_transport4)

logsoft=border_transport index

econ_institutions index

pca vanew rlnew ccnew genew rqnew stabnew economicfreedomsummaryindex

where, vanew=voice and accountability, rlnew= rule of law, ccnew= control of corruption, genew=government effectiveness, rqnew= regulatory quality, stabnew= absence violence and political instability and economicfreedomsummaryindex= economic freedom

estat kmo

predict pc1, score

where, pc1=pc1inseconfinal

gen inseconfinal4=pc1inseconfinal+4

gen loginseconfinal=ln(inseconfinal4)

loginsecnfinal=institutional quality index

/*regression heckman*/

heckman LOGTRADE loggdpc0 loggdpc1 LOGPOPO LOGPOP1 MR1DIS MRREL COMCOL22 COL2 WTO122 WTO02 LOGAREA1 LOGAREA0 LAN2 SEA2 RTA2 loginseconfinal year_15 year_14 year_13 year_12 year_11 year_10 year_9 year_8 year_7 year_6 year_5 year_4 year_3 year_2 , select(loggdpc0 loggdpc1 LOGPOPO LOGPOP1 MR1DIS MRREL COMCOL22 COL2 WTO122 WTO02 LOGAREA1 LOGAREA0 LAN2 SEA2 RTA2 loginseconfinal logccc1 year_15 year_14 year_13 year_12 year_11 year_10 year_9 year_8 year_7 year_6 year_5 year_4 year_3 year_2) twostep

heckman LOGTRADE loggdpc0 loggdpc1 LOGPOPO LOGPOP1 MR1DIS MRREL COMCOL22 COL2 WTO122 WTO02 LOGAREA1 LOGAREA0 LAN2 SEA2 RTA2 loginseconfinal logsoft loghardfinal year_15 year_14 year_13 year_12 year_11 year_10 year_9 year_8 year_7 year_6 year_5 year_4 year_3 year_2 , select(loggdpc0 loggdpc1 LOGPOPO LOGPOP1 MR1DIS MRREL COMCOL22 COL2 WTO122 WTO02 LOGAREA1 LOGAREA0 LAN2 SEA2 RTA2 loginseconfinal logsoft loghardfinal logccc1 year_15 year_14 year_13 year_12 year_11 year_10 year_9 year_8 year_7 year_6 year_5 year_4 year_3 year_2) twostep

heckman LOGTRADE loggdpc0 loggdpc1 LOGPOPO LOGPOP1 MR1DIS MRREL COMCOL22 COL2 WTO122 WTO02 LOGAREA1 LOGAREA0 LAN2 SEA2 RTA2 logsoft year_15 year_14 year_13 year_12 year_11 year_10 year_9 year_8 year_7 year_6 year_5 year_4 year_3 year_2 , select(loggdpc0 loggdpc1 LOGPOPO LOGPOP1 MR1DIS MRREL COMCOL22 COL2 WTO122 WTO02 LOGAREA1 LOGAREA0 LAN2 SEA2 RTA2 logsoft logaventry year_15 year_14 year_13 year_12 year_11 year_10 year_9 year_8 year_7 year_6 year_5 year_4 year_3 year_2) twostep

heckman LOGTRADE loggdpc0 loggdpc1 LOGPOPO LOGPOP1 MR1DIS MRREL COMCOL22 COL2 WTO122 WTO02 LOGAREA1 LOGAREA0 LAN2 SEA2 RTA2 loghardfinal year_15 year_14 year_13 year_12 year_11 year_10 year_9 year_8 year_7 year_6 year_5 year_4 year_3 year_2 , select(loggdpc0 loggdpc1 LOGPOPO LOGPOP1 MR1DIS MRREL COMCOL22 COL2 WTO122 WTO02 LOGAREA1 LOGAREA0 LAN2 SEA2 RTA2 loghardfinal logccc1 year_15 year_14 year_13 year_12 year_11 year_10 year_9 year_8 year_7 year_6 year_5 year_4 year_3 year_2) twostep

/*regression IV*/

ivregress liml LOGTRADE loggdpc0 loggdpc1 LOGPOPO LOGPOP1 MR1DIS MRREL COMCOL22 WTO122 WTO02 LOGAREA1 LOGAREA0 COL2 LAN2 SEA2 RTA2

(loghardfinal=govvfrac checkandbalance civilliberty llaglogpop1 llaglogpop0 llagloggdpc1 llagloggdpc0), vce(robust)

ivregress liml LOGTRADE loggdpc0 loggdpc1 LOGPOPO LOGPOP1 MR1DIS MRREL COMCOL22 WTO122 WTO02 LOGAREA1 LOGAREA0 COL2 LAN2 SEA2 RTA2 (logsoft=govvfrac checkandbalance civilliberty llaglogpop1 llaglogpop0 llagloggdpc1 llagloggdpc0), vce(robust)

ivregress liml LOGTRADE loggdpc0 loggdpc1 LOGPOPO LOGPOP1 MR1DIS MRREL COMCOL22 WTO122 WTO02 LOGAREA1 LOGAREA0 COL2 LAN2 SEA2 RTA2 (loginseconfinal=leg2new leg3new llagloggdpc0 llagloggdpc1 llaglogpop0 llaglogpop1), vce(robust)

/*simulation*/

margins, at ((asobserved) loginseconfinal) at (loginseconfinal =2.26)

margins, at ((asobserved) loghardfinal) at (loghardfinal =3.7)

margins, at ((asobserved) logsoft) at (logsoft =2.34)

Note: for Africa regional trade we did same after we excluded other countries from sample.

If you need any further explanation, do not hesitate to request.