BE 887 Empirical Exercise

Paul R. Organ

October 10, 2018

Contents

1 HMR Replication

1.1 Trade Volume in 1986

My replication results are in Table 1. These match closely the results in HMR (2008).

Note that "Island" is 1 if at least one country in a pair is not an island. "Landlock" is defined similarly. Why is this? I do not know.

Table 1: Replication of HMR Table I, Column 1

	(1)	(2)
	ln_trade	
VARIABLES	coef	se
dist	-1.176***	0.031
border	0.458***	0.148
island	-0.391***	0.121
landlock	-0.561***	0.188
legal	0.486***	0.050
lang	0.176***	0.061
colonial	1.299***	0.121
cu	1.364***	0.252
fta	0.759***	0.222
religion	0.102	0.097
Constant	14.581***	0.522
Observations	11,146	
R-squared	0.709	

*** p<0.01, ** p<0.05, * p<0.1

1.2 Any Trade in 1986

My replication results are in Table 2. For most variables, these match closely to the paper, but not quite as closely as my replication for column 1.

[My FTA estimate is way off, and I am not sure why (I am using the raw FTA variable from the dataset).]

Table 2: Replication of HMR Table I, Column 2

VARIABLES	(1) y1 coef	(2) se
dist	-0.263***	0.012
border	-0.152***	0.051
island	-0.137***	0.033
landlock	-0.072	0.045
legal	0.038***	0.014
lang	0.113***	0.017
colonial	0.130	0.122
cu	0.196***	0.057
fta	0.791***	0.126
religion	0.104***	0.025
Observations	24,649	

*** p<0.01, ** p<0.05, * p<0.1

Note: All predictors at their mean value

1.3 Trade Volume in the 1980s

My replication results are in Table 3, and again my results closely match HMR (2008).

Table 3: Replication of HMR Table I, Column 3

	(1)	(2)
	ln_trade	
VARIABLES	coef	se
dist	-1.201***	0.025
border	0.366***	0.132
island	-0.381***	0.096
landlock	-0.582***	0.148
legal	0.406***	0.040
lang	0.207***	0.047
colonial	1.321***	0.110
cu	1.395***	0.185
fta	0.996***	0.214
religion	-0.018	0.077
Constant	14.989***	0.403
Observations	110,697	
R-squared	0.682	

*** p<0.01, ** p<0.05, * p<0.1

1.4 Probit - Using Regulation Costs

My replication results are in Table 4.

[Again I am pretty close, but FTA is way off. Something is going wrong with the FTA variable in my probit reqressions.]

Table 4: Replication of HMR Table II, Column 1

	(1)	(2)
	y1	
VARIABLES	coef	se
dist	-0.219***	0.016
border	-0.090	0.068
island	-0.195***	0.066
landlock	-0.060	0.050
legal	0.054***	0.019
lang	0.102***	0.023
colonial	-0.009	0.130
cu	0.276***	0.068
fta	0.977***	0.189
religion	0.147***	0.034
reg_costs	-0.107***	0.035
reg_costs_procdays	-0.067**	0.030
Observations	12,420	

*** p<0.01, ** p<0.05, * p<0.1

Note: All predictors at their mean value

1.5 Heckman Selection Correction

My replication results are in Table 5.

[How are regulation variables included for the second stage? What's going on here?]

Table 5: Heckman Selection (Similar to HMR Table II, Column 2)

VARIABLES	(1) ln_trade	(2) ln_trade	(3) pos_trade	(4) pos_trade	(5) /	(6) /
dist	-1.163***	0.039	-0.591***	0.043	-0.591***	0.043
border	0.616***	0.164	-0.246	0.184	-0.246	0.184
island	-0.570**	0.271	-0.528***	0.178	-0.528***	0.178
landlock	-0.427**	0.187	-0.162	0.134	-0.162	0.134
legal	0.542***	0.063	0.147***	0.051	0.147***	0.051
lang	0.132*	0.075	0.275***	0.061	0.275***	0.061
colonial	0.911***	0.154	-0.024	0.352	-0.024	0.352
cu	1.493***	0.322	0.745***	0.184	0.745***	0.184
fta	0.968***	0.243	2.625***	0.514	2.625***	0.514
religion	0.290**	0.116	0.396***	0.091	0.396***	0.091
reg_costs			-0.289***	0.095	-0.289***	0.095
reg_costs_procdays			-0.180**	0.080	-0.180**	0.080
Constant	14.365***	0.612	2.863***	0.376	2.863***	0.376
Observations	12,535		12,535		12,535	

*** p<0.01, ** p<0.05, * p<0.1

1.6 Polynomial Regression

[My code for this is not working. How do we construct the Mills ratio and the z variables to include in the regression?]

2 Alternative Specifications and Tests

2.1 Poisson Quasi-MLE

Similar to HMR Table I, Column 1. Here I present the alternative specification, using Poisson regression (left) and my replication of Table I, Column 1 (right).

[Add discussion of distance coefficients or other geographic coefficients]

(a) Alternate Specification (Poisson)			(b) Replication of Table II, Column 1		
	(1)	(2)		(1)	(2)
	trade			y1	
VARIABLES	coef	se	VARIABLES	coef	se
dist	-0.643***	0.062	dist	-0.219***	0.016
border	0.736***	0.153	border	-0.090	0.068
island	0.429*	0.240	island	-0.195***	0.066
landlock	-0.542*	0.290	landlock	-0.060	0.050
legal	0.286***	0.087	legal	0.054***	0.019
lang	-0.163	0.114	lang	0.102***	0.023
colonial	0.245	0.153	colonial	-0.009	0.130
cu	-0.683	0.560	cu	0.276***	0.068
fta	0.250**	0.108	fta	0.977***	0.189
religion	-0.243	0.190	religion	0.147***	0.034
reg_costs	-0.241	0.151	reg_costs	-0.107***	0.035
reg_costs_procdays	-0.290	0.184	reg_costs_procdays	-0.067**	0.030
Constant	14.408***	0.668			
			Observations	12,420	
Observations	13,572				
*** p<0.01, ** p<0.05, * p<0.1			*** p<0.01, ** p<0.05, * p<0.1		
			Note: All predictors at their mean value		

2.2 Poor Man's Zero Correction

Here we include zero-trade observations by adding 1 to all flows, so that for flows of zero, we have $\log(flow + 1) = \log(0 + 1) = 0$.

Again I include my replication of Table II, Column 1 for comparison. Note that the coefficient on distance grows in magnitude under this alternate specification.

(a) Alternate Specification (Poor Man's Zero-Correction)

VARIABLES	(1) ln_trade_poorman coef	(2) se
dist	-1.091***	0.054
border	0.477	0.293
island	-0.329	0.241
landlock	-0.716***	0.148
legal	0.268***	0.067
lang	0.578***	0.079
colonial	0.890***	0.240
cu	0.744**	0.336
fta	0.263	0.341
religion	0.588***	0.128
reg_costs	-0.131	0.104
reg_costs_procdays	-0.280**	0.109
Constant	10.342***	0.622
Observations	13,572	
R-squared	0.752	

*** p<0.01, ** p<0.05, * p<0.1

(b) Replication of Table II, Column 1

VARIABLES	(1) y1 coef	(2) se
dist	-0.219***	0.016
border	-0.090	0.068
island	-0.195***	0.066
landlock	-0.060	0.050
legal	0.054***	0.019
lang	0.102***	0.023
colonial	-0.009	0.130
cu	0.276***	0.068
fta	0.977***	0.189
religion	0.147***	0.034
reg_costs	-0.107***	0.035
reg_costs_procdays	-0.067**	0.030
Observations	12,420	

*** p<0.01, ** p<0.05, * p<0.1

Note: All predictors at their mean value

2.3 Fixed Effects

Finally we test various combinations of fixed effects in linear models, including only FTA as a regressor.

[The third specification with im*yr, ex*yr, and im*ex doesn't run on my computer, says too many covariates?]

Table 6: Testing Different Combinations of Fixed Effects

VARIABLES	(1) ln_trade	(2) ln_trade
fta	3.376***	3.370***
	0.062	0.063
Constant	8.726***	8.730***
	0.109	0.326
Observations	110,697	110,697
R-squared	0.589	0.603
Importer and Exporter FE	Y	N
Im*Year and Ex*Year FE	N	Y
Importer*Exporter FE	N	N

*** p<0.01, ** p<0.05, * p<0.1

A Stata Code

```
*******************************************************************************
* Author: Paul R. Organ
* Purpose: BE 887 Econometric Exercise
* Last Update: Oct 9, 2018
*******************************************************************************
clear all
set more off
capture log close

cd "C:\Users\prorgan\Box\Classes\BE 887\Exercise"
log using organ_exercise.log, replace

* load data
use "data1980s_share.dta"


*******************************************************************************
*** Replication of Helpman, Melitz, and Rubenstein (2008)
*******************************************************************************
** Table 1, Column 1 for 1986

* define landlock if we don't have two landlocked countries
gen landlock = n_landlock!=2

* define island if we don't have two island countries
gen island = n_islands!=2

* rename variables for use in regs
rename ln_distance dist
rename legalsystem_same legal
rename common_lang lang
rename religion_same religion

* define country pairs
gen pair = expcode + impcode

* local to list variables for inclusion in regression
local vars = "dist border island landlock legal lang colonial cu fta religion"

* running regression
* fixed effects for exporter and for importer
* standard errors clustered at the country-pair level
quietly reg ln_trade 'vars' i.expcode i.impcode if year==1986, vce(cluster pair)
estimates table, se keep('vars') stats(N r2)

* write to file for use in tex
outreg2 using "tables/t1c1.tex", keep('vars') ///
 side stats(coef se) noparen dec(3) replace


*******************************************************************************
** Table 1, Column 2 for 1986 (Probit)

* define indicator for country pairs with positive trade
```

```
gen pos_trade = !missing(ln_trade)

* check we get 11,146 with positive trade (footnote 21)
tab pos_trade if year==1986

* local to list variables for inclusion in regression
local vars = "dist border island landlock legal lang colonial cu fta religion"

* regression
quietly probit pos_trade 'vars' i.expcode i.impcode ///
if year==1986 & expcode!=141780, vce(cluster pair)

local controls = "dist border island landlock legal lang colonial cu fta religion i.impcode

margins, dydx('controls') atmeans post

* write to file for use in tex
outreg2 using "tables/t1c2.tex", keep('vars') ///
 side stats(coef se) noparen dec(3) replace ///
 addnote(Note: All predictors at their mean value)

*******************************************************************************
** Table 1, Column 3 for 1980s panel
* local to list variables for inclusion in regression
local vars = "dist border island landlock legal lang colonial cu fta religion"

* running regression
* fixed effects for exporter, importer, and year
* standard errors clustered at the country-pair level
quietly reg ln_trade 'vars' i.expcode i.impcode i.year, vce(cluster pair)
estimates table, se keep ('vars') stats(N r2)

* write to file for use in tex
outreg2 using "tables/t1c3.tex", keep('vars') ///
 side stats(coef se) noparen dec(3) replace

*******************************************************************************
** Merge in Regulation data
* see do file 'clean_data' for small pre-cleaning steps done before this
merge m:1 expcode using reg_costs_exp
drop _merge
merge m:1 impcode using reg_costs_imp
drop _merge

* generate indicators for pairs
gen reg_costs = (exp_ind_cost + imp_ind_cost)==2
replace reg_costs = . if missing(exp_ind_cost) | missing(imp_ind_cost)
gen reg_costs_procdays = (exp_ind_procdays + imp_ind_procdays)==2
replace reg_costs_procdays = . if missing(exp_ind_procdays) | missing(imp_ind_procdays)

save "merged_data.dta", replace

*******************************************************************************
** Table 2, Column 1
```

```
* drop if (1) reg cost data is missing or
* (2) exporter in 8 country list or (3) importer is Japan
* Japan = 413920
* Hong Kong = 453440
* France = 532500
* Germany = 532800
* Italy = 533800
* Netherlands = 535280
* UK = 538260
* Sweden = 557520
drop if exp_ind_cost == .
drop if imp_ind_cost == .
drop if expcode == 413920 | expcode == 453440 | expcode == 532500 | ///
expcode == 532800 | expcode == 533800 | expcode == 535280 | ///
expcode == 538260 | expcode == 557520
drop if impcode == 413920

* local to list variables for inclusion in regression
local vars = "dist border island landlock legal lang colonial cu fta religion reg_costs reg

* regression
quietly probit pos_trade `vars' i.expcode i.impcode if year==1986, vce(cluster pair)

margins, dydx(`vars') atmeans post

* write to file for use in tex
outreg2 using "tables/t2c1.tex", keep(`vars') ///
 side stats(coef se) noparen dec(3) replace ///
 addnote(Note: All predictors at their mean value)

********************************************************************************
** Heckman Selection Correction
set matsize 10000

local fs_vars = "dist border island landlock legal lang colonial cu fta religion reg_costs

local ss_vars = "dist border island landlock legal lang colonial cu fta religion"

heckman ln_trade `ss_vars' i.expcode i.impcode if year==1986, ///
select(pos_trade = `fs_vars' i.expcode i.impcode) vce(cluster pair)

* write to file for use in tex
outreg2 using "tables/t2c2.tex", keep(`fs_vars') ///
 side stats(coef se) noparen dec(3) replace

********************************************************************************
** Table 2, Column 4
* predict probability of selection (z_ij)
predict psel if year==1986, xb

gen mills = exp(-.5*psel^2)/(sqrt(2*_pi)*normprob(psel))

gen z = invnormal(psel)
gen z2 = z^2
```

```
gen z3 = z^3

local vars = "dist border island landlock legal lang colonial cu fta religion mills z z2 z3

* running regression
* fixed effects for exporter, importer, and year
* standard errors clustered at the country-pair level
quietly reg ln_trade 'vars' i.expcode i.impcode if year==1986, vce(cluster pair)
estimates table, se keep('vars') stats(N r2)

* write to file for use in tex
outreg2 using "tables/t2c4.tex", keep('vars') ///
 side stats(coef se) noparen dec(3) replace


********************************************************************************
*** Alternative Specifications and Tests using data from HMR (2008)
********************************************************************************
** Santos-Silva and Tenreyro (2006) Approach
* reload merged data
clear all
use "merged_data.dta"

* replace missing ln_trade with 0, convert to levels
replace ln_trade = 0 if missing(ln_trade)
gen trade = exp(ln_trade)

* run same regression as in part 4
local vars = "dist border island landlock legal lang colonial cu fta religion reg_costs reg

poisson trade 'vars' i.expcode i.impcode if year==1986, vce(cluster pair)

* write to file for use in tex
outreg2 using "tables/alt1.tex", keep('vars') ///
 side stats(coef se) noparen dec(3) replace


********************************************************************************
** Adding one to all trade flows
gen trade_plus_one = trade + 1
gen ln_trade_poorman = ln(trade_plus_one)

local vars = "dist border island landlock legal lang colonial cu fta religion reg_costs reg

reg ln_trade_poorman 'vars' i.expcode i.impcode if year==1986, vce(cluster pair)
* distance coefficient grows in magnitude

* write to file for use in tex
outreg2 using "tables/alt2.tex", keep('vars') ///
 side stats(coef se) noparen dec(3) replace


********************************************************************************
** Fixed Effects
clear all
use "merged_data.dta"
```

13

```
* test different combinations of FE, writing to tex file

* importer, exporter, year FE
quietly reg ln_trade fta i.expcode i.impcode
outreg2 using "tables/alt3.tex", keep(fta) stats(coef se) noparen dec(3) replace ///
 addtext(Importer and Exporter FE, Y, Im*Year and Ex*Year FE, N, Importer*Exporter FE, N)

* importer*year, exporter*year,
egen exp_year = group(expcode year)
egen imp_year = group(impcode year)

set matsize 11000

quietly reg ln_trade fta i.exp_year i.imp_year
outreg2 using "tables/alt3.tex", keep(fta) stats(coef se) noparen dec(3) append ///
 addtext(Importer and Exporter FE, N, Im*Year and Ex*Year FE, Y, Importer*Exporter FE, N)

* importer*year, exporter*year, importer*exporter
set matsize 11000
quietly reg ln_trade fta i.exp_year i.imp_year i.pair
outreg2 using "tables/alt3.tex", keep(fta) ///
 stats(coef se) noparen dec(3) append

* too many interactions

*********************************************************************************
```