# BE 887 Empirical Exercise

Paul R. Organ

October 11, 2018

## Contents

# 1 HMR Replication

## 1.1 Table 1: Benchmark Gravity and Selection into Trading Relationships

My results are in Table 1. I exactly match the paper in columns (1) and (3). In column (2), I am close on every variable, but my coefficient estimate for $fta$ is off. My observation counts match, and I am not sure what else could be driving this difference.

Code for this replication can be found in Appendix A

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, Columns 1, 2, and 3

| | (1) | (2) | (3) |
|---|---|---|---|
| VARIABLES | ln_trade | Probit | ln_trade |
| dist | -1.176*** | -0.263*** | -1.201*** |
| | (0.031) | (0.012) | (0.024) |
| border | 0.458*** | -0.152*** | 0.366*** |
| | (0.147) | (0.051) | (0.131) |
| island | -0.391*** | -0.137*** | -0.381*** |
| | (0.121) | (0.033) | (0.096) |
| landlock | -0.561*** | -0.072 | -0.582*** |
| | (0.188) | (0.045) | (0.148) |
| legal | 0.486*** | 0.038*** | 0.406*** |
| | (0.050) | (0.014) | (0.040) |
| lang | 0.176*** | 0.113*** | 0.207*** |
| | (0.061) | (0.017) | (0.047) |
| colonial | 1.299*** | 0.130 | 1.321*** |
| | (0.120) | (0.122) | (0.110) |
| cu | 1.364*** | 0.196*** | 1.395*** |
| | (0.255) | (0.057) | (0.187) |
| fta | 0.759*** | 0.791*** | 0.996*** |
| | (0.222) | (0.126) | (0.213) |
| religion | 0.102 | 0.104*** | -0.018 |
| | (0.096) | (0.025) | (0.076) |
| Constant | 14.581*** | | 14.989*** |
| | (0.520) | | (0.402) |
| Observations | 11,146 | 24,649 | 110,697 |
| R-squared | 0.709 | | 0.682 |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Probit: All predictors at mean values

## 1.2 Probit - Using Regulation Costs

My replication results are in Table 2.

Again I am pretty close, but the coefficient estimate on $fta$ is way off.

Table 2: 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

## 1.3 Heckman Selection Correction

The results from my Heckman Selection Correction model are in Table 3.

How are regulation variables included for the second stage? What's going on here? Code is below.

```
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, ///
select(pos_trade = `fs_vars' i.expcode i.impcode) twostep

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

Table 3: Heckman Selection (Similar to HMR Table II, Column 2)

| VARIABLES | (1)<br>ln_trade | (2)<br>pos_trade | (3)<br>/mills |
|---|---|---|---|
| dist | -1.165*** | -0.591*** | -0.591*** |
| | (0.037) | (0.032) | (0.032) |
| border | 0.614*** | -0.243* | -0.243* |
| | (0.137) | (0.126) | (0.126) |
| island | -0.577** | -0.525** | -0.525** |
| | (0.228) | (0.205) | (0.205) |
| landlock | -0.428** | -0.161 | -0.161 |
| | (0.169) | (0.118) | (0.118) |
| legal | 0.543*** | 0.147*** | 0.147*** |
| | (0.054) | (0.043) | (0.043) |
| lang | 0.136* | 0.274*** | 0.274*** |
| | (0.071) | (0.056) | (0.056) |
| colonial | 0.911*** | -0.025 | -0.025 |
| | (0.188) | (0.335) | (0.335) |
| cu | 1.499*** | 0.745*** | 0.745*** |
| | (0.279) | (0.176) | (0.176) |
| fta | 0.970*** | 2.635*** | 2.635*** |
| | (0.214) | (0.494) | (0.494) |
| religion | 0.295*** | 0.396*** | 0.396*** |
| | (0.110) | (0.088) | (0.088) |
| reg_costs | | -0.288*** | -0.288*** |
| | | (0.078) | (0.078) |
| reg_costs_procdays | | -0.180** | -0.180** |
| | | (0.074) | (0.074) |
| Constant | 14.359*** | 2.858*** | 2.858*** |
| | (0.460) | (0.352) | (0.352) |
| Observations | 12,535 | 12,535 | 12,535 |

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

## 1.4 Polynomial Regression

The results from my replication of the polynomial regression using constructed values for $z$ and $\eta$ are shown in Table 4.

Somewhat close on the main coefficients, but my lambda and z's are off. Here's the code I'm running:

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

probit pos_trade `vars' i.expcode i.impcode, vce(cluster pair)

* predict probability of selection (z_ij)
predict psel

* generate z variables
gen z = invnormal(psel)
gen z2 = z^2
gen z3 = z^3

* calculate inv mills ratio
gen lambda = normalden(z)/psel

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

* running regression, fixed effects for exporter, importer
quietly reg ln_trade `vars' i.expcode i.impcode
```

Table 4: Replication of Table II, Column 4 (Polynomial Reg)

| | (1) | (2) |
|---|---|---|
| | ln_trade | |
| VARIABLES | coef | se |
| dist | -0.824*** | 0.165 |
| border | 0.940*** | 0.163 |
| island | -0.185 | 0.278 |
| landlock | -0.332* | 0.181 |
| legal | 0.419*** | 0.069 |
| lang | -0.054 | 0.106 |
| colonial | 0.861*** | 0.203 |
| cu | 1.115*** | 0.342 |
| fta | 0.178 | 0.793 |
| religion | 0.119 | 0.155 |
| lambda | -0.124 | 0.712 |
| z | 0.350 | 0.495 |
| z2 | 0.220** | 0.097 |
| z3 | -0.032*** | 0.008 |
| Constant | 12.681*** | 1.076 |
| Observations | 6,403 | |
| R-squared | 0.667 | |

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

# 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 II, Column 1 (right).

(a) Alternate Specification (Poisson)

| | (1) | (2) |
|---|---|---|
| | trade | |
| VARIABLES | coef | se |
| dist | -0.643*** | 0.062 |
| border | 0.736*** | 0.153 |
| island | 0.429* | 0.240 |
| landlock | -0.542* | 0.290 |
| legal | 0.286*** | 0.087 |
| lang | -0.163 | 0.114 |
| colonial | 0.245 | 0.153 |
| cu | -0.683 | 0.560 |
| fta | 0.250** | 0.108 |
| religion | -0.243 | 0.190 |
| reg_costs | -0.241 | 0.151 |
| reg_costs_procdays | -0.290 | 0.184 |
| Constant | 14.408*** | 0.668 |
| Observations | 13,572 | |

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

(b) Replication of 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

## 2.2 Poor Man's Zero Correction

Here we include zero-trade observations by adding one 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)

| | (1) | (2) |
|---|---|---|
| | ln_trade_poorman | |
| VARIABLES | coef | 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

| | (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

## 2.3 Fixed Effects

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

Table 5: Testing Different Combinations of Fixed Effects

| | (1) | (2) | (3) |
|---|---|---|---|
| VARIABLES | ln_trade | ln_trade | ln_trade |
| fta | 3.919*** | 3.949*** | 0.392*** |
| | (0.090) | (0.093) | (0.084) |
| Constant | 8.315*** | 8.315*** | 8.375*** |
| | (0.008) | (0.008) | (0.005) |
| Observations | 65,973 | 65,973 | 65,605 |
| R-squared | 0.564 | 0.580 | 0.842 |
| Importer and Exporter FE | Y | N | N |
| Im*Year and Ex*Year FE | N | Y | Y |
| Importer*Exporter FE | N | N | Y |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

# A Stata Code

```
*******************************************************************************
* Author: Paul R. Organ
* Purpose: BE 887 Econometric Exercise
* Last Update: Oct 11, 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 "data\data1980s_share.dta"

*******************************************************************************
*** Data Cleaning
*******************************************************************************

* data cleaning for HMR Rep parts 1, 2, and 3 and Alt Specs part 3

* 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

** create id we can use for clustering
* convert to strings
gen impcode_str = impcode
gen expcode_str = expcode
tostring impcode_str, replace
tostring expcode_str, replace

* create id for pairs (need exp-imp and imp-exp to be same, so we do conditional)
gen newid = cond(impcode_str <= expcode_str, impcode_str, expcode_str) ///
        + cond(impcode_str >= expcode_str, impcode_str, expcode_str)

* create pair variable we will use for clustering
egen pair = group(newid)

* drop temporary variables
drop newid impcode_str expcode_str

* define indicator for country pairs with positive trade
gen pos_trade = !missing(ln_trade)
```

```
* save data
save "data\cleaned_part1.dta", replace

*******************************************************************************
* data cleaning for HMR Rep parts 4, 5, and 6, and Alt Specs parts 1 and 2

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

* 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

save "data\cleaned_part2.dta", replace

*******************************************************************************
*** HMR Replication
*******************************************************************************

** Replication of Table 1, Columns 1, 2, and 3

clear all
use "data\cleaned_part1.dta"

* Column 1
* 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)

* write to file for use in tex
outreg2 using "tables/t1.tex", keep(`vars') ///
 stats(coef se) dec(3) replace

*******************************************************************************
* Column 2
* local to list variables for inclusion in regression
local vars = "dist border island landlock legal lang colonial cu fta religion"

* regression (exclude Congo)
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/t1.tex", keep(`vars') ///
 stats(coef se) dec(3) append ///
 addnote(Note: All predictors at their mean value)

*******************************************************************************
* Column 3
* 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)

* write to file for use in tex
outreg2 using "tables/t1.tex", keep(`vars') ///
 stats(coef se) dec(3) append

*******************************************************************************
** Replication of Table II, Column 1

clear all
use "data\cleaned_part2.dta"

keep if year==1986

* 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, vce(cluster pair)

margins, dydx(`vars') atmeans post
```

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

 ********************************************************************************
** Heckman Correction Model and Table II, Column 4
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, ///
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') ///
 stats(coef se) dec(3) replace

 ********************************************************************************
** Table 2, Column 4
* local to list variables for inclusion in regression
local vars = "dist border island landlock legal lang colonial cu fta religion reg_costs reg

* regression
probit pos_trade `vars' i.expcode i.impcode, vce(cluster pair)

* predict probability of selection (z_ij)
predict psel

* generate z variables
gen z = invnormal(psel)
gen z2 = z^2
gen z3 = z^3

* calculate inv mills ratio
gen lambda = normalden(z)/psel

* define variables
local vars = "dist border island landlock legal lang colonial cu fta religion lambda z z2 z

* running regression
* fixed effects for exporter, importer
quietly reg ln_trade `vars' i.expcode i.impcode

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

********************************************************************************
*** Alternative Specifications and Tests
********************************************************************************

** Santos-Silva and Tenreyro (2006) Approach
clear all
use "data\cleaned_part2.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)

* 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 "data\cleaned_part2.dta"

set matsize 11000

* importer, exporter, year FE
quietly reghdfe ln_trade fta, a(expcode impcode) vce(r)
outreg2 using "tables/alt3.tex", keep(fta) stats(coef se) 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)

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

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