Augmented Gravity Models: Measurement and Trade impact of Policies: day 3

Luca De Benedictis (University of Macerata)

Luca Salvatici (Roma Tre University)

AGRODEP Training – Dakar, 2015

ROMA TRE
UNIVERSITÀ DEGLI STUDI

Maria Pina CIPOLLINA, Luca SALVATICI: «Reciprocal trade agreements in gravity models: a meta-analysis"

Review of International Economics, Volume 18 Issue 1, February 2010, pp. 63 - 80.

1. We use point estimates of the relevant parameters from various studies as the individual observations for the meta-regression analysis (MRA) models, adopting weighted least squares (WLS) and checking the robustness and sensitivity of our results.
2. We then focus on the effect of specific trade agreements on bilateral trade.
3. Finally, we run a probit regression in order to identify which factors account for the positive (and significant) impact of RTAs on bilateral trade flows.

ROMA TRE
UNIVERSITÀ DEGLI STUDI

Meta-Analysis (MA)

- MA is a set of quantitative techniques for evaluating and combining the empirical: the main focus is to test the null hypothesis that different point estimates are equal to zero

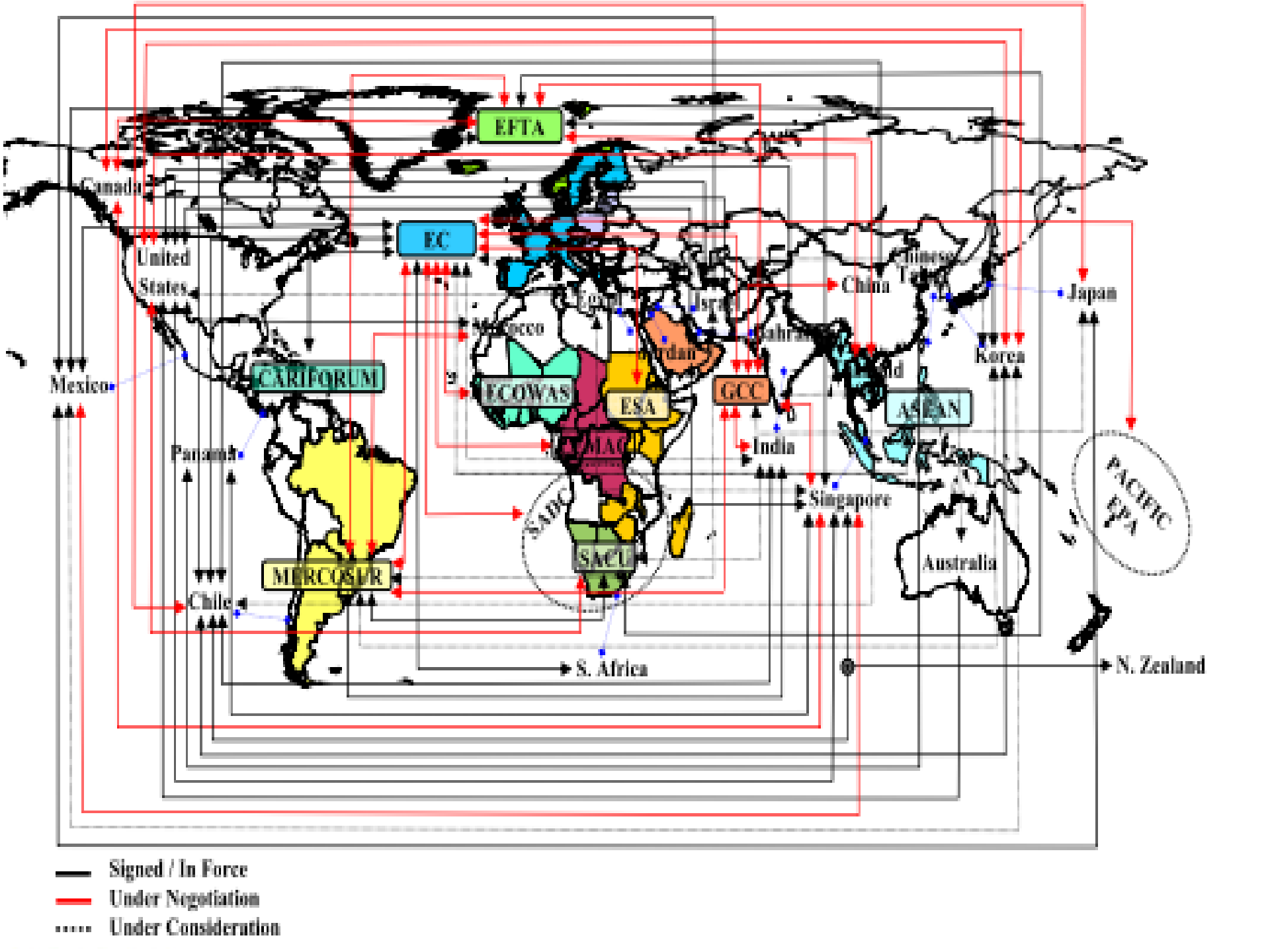
Journal of Economic Surveys special issue (2005)

- MA is a methodology of literature reviewing, NOT an alternative approach to studying the effect of interest
- Our goal is not to find out what is the "true" value of the parameter under investigation, rather to explain why there is so much variation among the reported empirical results of economic studies purportedly investing the same phenomenon.
- Such a regression analysis of regression analyses offer a quantitative methodology for combining all of the estimates, investigating the sensitivity of the overall estimate to variations in underlying assumptions, identifying and filtering out possible biases and explaining the diversity in the study results in relation to the heterogeneity of study features.

Reciprocal (preferential) agreements

- The world trading system is characterized by two major types of preferential agreements:
  - reciprocal (bilateral), entailing symmetric trade liberalization
  - nonreciprocal (unilateral), entailing asymmetric trade liberalization which is aimed at providing support to the country that gains improved market access without needing to open up its own domestic market
- The world has witnessed a veritable explosion of reciprocal preferential trade agreements (RTAs) in the past 15 years:
  - As of June 2009, 438 RTAs have been notified to the GATT/WTO of which 247 are currently in force
  - „Approximately 100 RTAs in the pipeline (signed, not yet in force/under negotiation)
  - more than half of world trade now occurs within actual or prospective trading blocs and nearly every country in the world is a member of one or more agreements

ROMA TRE
UNIVERSITÀ DEGLI STUDI

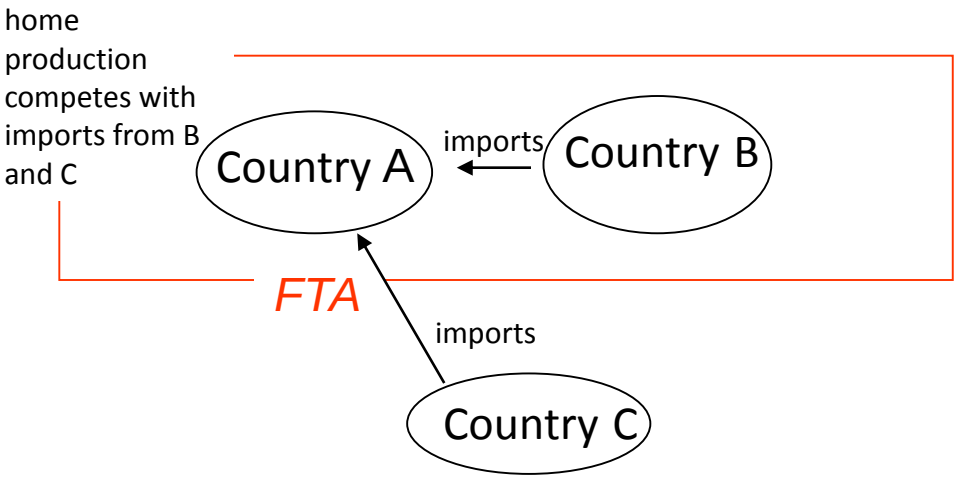

EFTA
Canada
EC
United States
China
Japan
Korea
Mexico
CARIFORUM
ECOWAS
ESA
GCC
ASEAN
India
Panama
SADC
SACU
Singapore
PACIFIC EPA
Australia
MERCOSUR
Chile
S. Africa
N. Zealand
Signed / In Force
Under Negotiation
Under Consideration

A bit of theory. Do PTAs necessarily raise welfare if they raise trade values?

We know that full trade liberalization raises welfare.
But is it the case for partial trade liberalization as well?

General principle: Theorem of the Second Best (can go either way)
Specific answer: Viner's concepts of trade creation and trade diversion

Trade creation and trade diversion: Assumed trade pattern



Trade creation and trade diversion: numerical example

Case 1: Neither creation nor diversion

Producers	A	B	C
Domestic price	15.0	18.0	20.0
Duty-paid price in A			
With 20% MFN tariff	15.0	21.6	24.0
With tariff only on C	15.0	18.0	24.0

Case 2: Trade creation

Imports from partner displace inefficient domestic production

Producers	A	B	C
Domestic price	17.0	15.0	16.0
Duty-paid price in A			
With 20% MFN tariff	17.0	18.0	19.2
With tariff only on C	17.0	15.0	19.2

Case 3: Trade diversion

Imports from partner displace efficient imports from rest of the world

Producers	A	B	C
Domestic price	13.0	11.0	10.0
Duty-paid price in A			
With 20% MFN tariff	13.0	13.2	12.0
With tariff only on C	13.0	11.0	12.0

ROMA TRE
UNIVERSITÀ DEGLI STUDI

Trade creation and trade diversion: gravity estimates

Regressors	coeff	*t-stat*
ln GDP i,t	1.00	*96.7*
ln GDP j,t	1.15	*104.9*
ln D ij	-1.01	*-44.7*
ln RER ij,t	-0.01	*-3.8*
EU		
i,j belong to bloc	0.25	*5.1*
exporter outside	0.15	*2.2*
importer outside	0.36	*15.7*
Mercosur		
i,j belong to bloc	-0.32	*-0.6*
exporter outside	-1.43	*-14.8*
importer outside	0.01	*0.2*
NAFTA		
i,j belong to bloc	0.78	*1.0*
exporter outside	-0.43	*-4.1*
importer outside	0.73	*6.1*
ASEAN		
i,j belong to bloc	1.27	*5.6*
exporter outside	-0.21	*-2.7*
importer outside	1.14	*13.2*
Obs.	240'691	
R-2	0.65	

trade creation

trade diversion

Source: Carrère (2004)

Using the gravity equation to assess the effect of PTAs

$$\ln V_{ijt} = \underbrace{\beta_0 + \beta_1 \ln Y_{it} + \beta_2 \ln Y_{jt} + \beta_3 \ln \delta_{ij} + \mathbf{X}_{ijt}\boldsymbol{\beta}_4 + \mathbf{t}\boldsymbol{\beta}_6}_{W_{ijt}} + \beta_7 PTA_{ijt} + u_{ijt}$$

$$PTA_{ijt} = \begin{cases} 1 & \text{if PTA in force between i and j at t} \\ 0 & \text{otherwise} \end{cases}$$

Predicted trade

$$\ln \hat{V}_{ijt} = \hat{W}_{ijt} + \hat{\beta}_7 PTA_{ijt}$$

Case 1: PTA between i and j; predicted trade (in log)

$$\ln \hat{V}_{ijt} = \hat{W}_{ijt} + \hat{\beta}_7$$

Case 2: No PTA between i and j; predicted trade (in log)

$$\ln \hat{V}_{ijt} = \hat{W}_{ijt}$$

Difference (effect of PTA)

$$\ln \hat{V}_{ijt}^{\text{with}} - \ln \hat{V}_{ijt}^{\text{without}} = \ln\left(\frac{\hat{V}_{ijt}^{\text{with}}}{\hat{V}_{ijt}^{\text{without}}}\right) = \hat{\beta}_7 \qquad \text{so} \qquad \frac{\Delta \hat{V}_{ijt}}{\hat{V}_{ijt}^{\text{without}}} = e^{\hat{\beta}_7} - 1$$

ROMA TRE
UNIVERSITÀ DEGLI STUDI

Selection problems

- The choice to limit the review of RTAs impact to gravity models could be questioned, since some of the literature argue that the pervasive trade creation effect of most RTAs is picking up a "fragile" relationship generated from an individual researcher's specification of the gravity model equation: if gravity models tend to be biased in a particular direction due to a common misspecification, our meta-analysis estimates are going to include the average of this systematic bias.
- We cannot pretend (and don't want) to establish professional consensus or to identify a clear and uncontroversial approach to the evaluation of RTAs impact using MA: our goal is (only) to provide an assessment of the methodological choices and possible (relative) biases induced by model specifications within the large and growing field of the literature using the gravity model

Collected effect size

- Papers written in English.
- Papers were selected via extensive search in Google and in databases, such as EconLit and Web of Science: keywords searched for were “trade agreements” (focus on bilateral flows) and “gravity equation or gravity model” (trade agreements among the control variables) in the title, abstract or subject.
- We also traced some specific papers cross-referenced in other works.

The final sample includes *85 papers* (38 published in academic journals, 47 are working papers or unpublished studies) providing *1827 point estimates* of the impact of RTAs on bilateral trade

ROMA TRE
UNIVERSITÀ DEGLI STUDI

Methodological Problems: ***Dependence of study results***

- A study may quite often provide multiple estimates of the effect being considered: the assumption that multiple observations from the same study are independent becomes too strong.
- Use a single observation, identifying a "preferred" estimate, taking averages or medians of the estimates from each paper or randomly selecting one estimate
- In the MRA, we adopt a "robust with cluster" procedure (each cluster identifies the study the estimate belongs to), adjusting standard errors for intra-study correlation: this changes the variance-covariance matrix and the standard errors of the estimators but not the estimated coefficients themselves

Pooling different estimates into a large sample for meta-analysis raises the question of within-study versus between-study heterogeneity.

	Pooled (published – unpublished)	Lower Bound of 95% CI	Upper Bound of 95% CI
Fixed-effects	0.10 (0.055 – 0.218)	0.097	0.101
Random-effects	0.50 (0.475 – 0.510)	0.482	0.515

$$\hat{\theta} = \frac{\sum_{i=1}^{n} \hat{\theta}_i w_i}{\sum_{i=1}^{n} w_i}$$

$$w_i = \frac{1}{Se(\hat{\theta}_i)^2}$$

Θ = individual estimate of the RTA effect , *w* = weights. Fixed effects:

The fixed-effects estimate assumes that differences across studies are only due to within-variation: RTAs raise trade by 10% ($e^{0.10}$-1=0.10)

We test and reject the null hypothesis of homogeneity among estimates

Random effects: $w_i^* = (w_i^{-1} + \hat{\tau}^2)^{-1}$ $\hat{\tau}^2$ = estimate of the between-study heterogeneity

- The random-effects estimate considers both between- and within-study variability: RTAs raise trade by 65% ($e^{0.50}$-1=0.65)
- Published papers show significantly lower impacts especially in the case of fixed effects estimate (since random effects estimation reduces the relative weighting given to more precise results that are more likely among published studies)

ROMA TRE UNIVERSITÀ DEGLI STUDI

Meta-Analysis of estimates of specific RTAs

- The largest effect is registered for the Baltic Free Trade Area: the fixed effects estimate suggests an increase in trade of around 2000%!
- Other agreements presenting exceedingly high estimates are
  - the Commonwealth of Independent States Customs Union (CISCU) – 1581% – and
  - the Caribbean Community – 400%
- Looking at the most widely studied agreements:
  - the largest impact is for NAFTA (131%)
  - whereas the European agreements register much lower, but possibly more realistic, values: 27% for EFTA and 41% for the EU.
- It is also worth noting that custom unions – EU, CARICOM, MERCOSUR, Central American Common Market, CISCU – do not seem to outperform the free trade areas consistently in terms of trade impact.

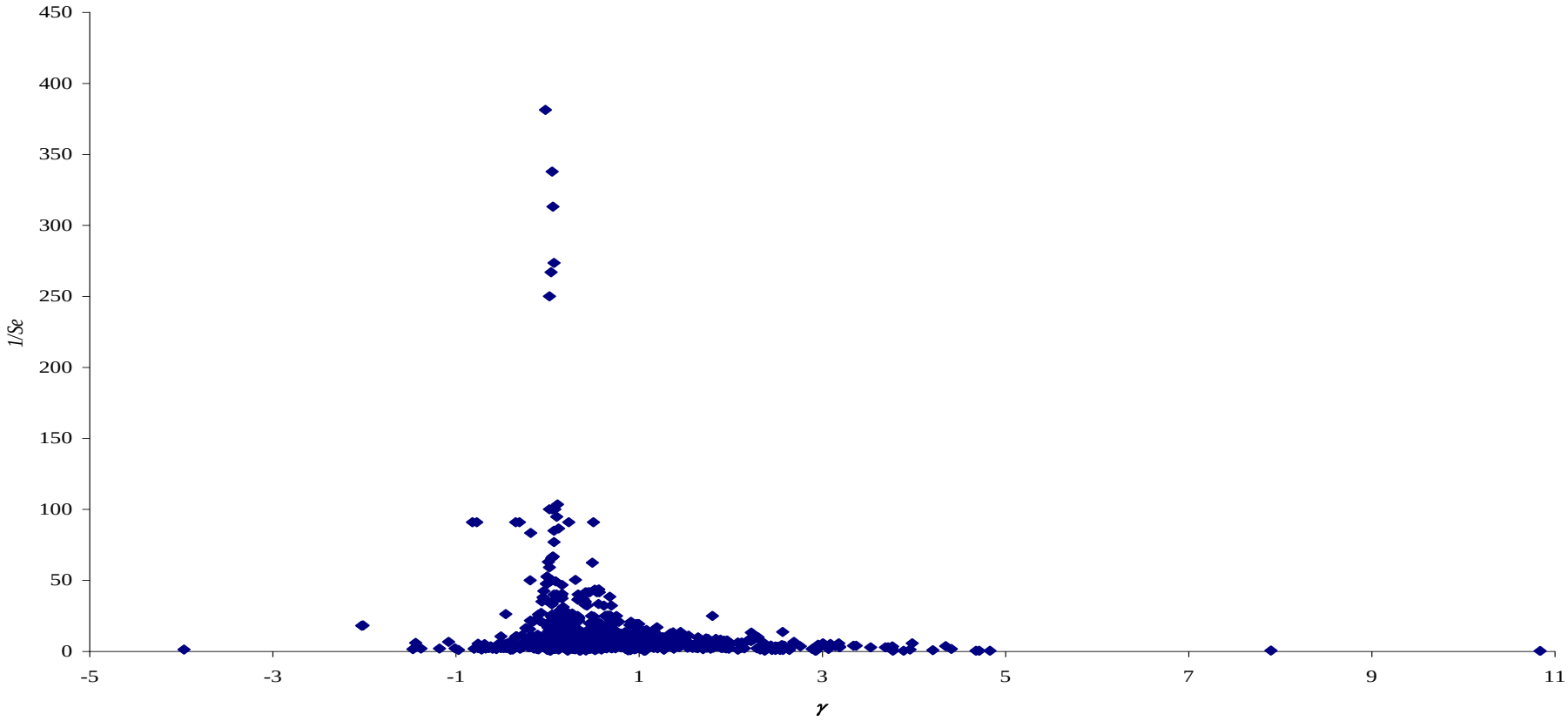
Publication bias?

One of the criticisms of meta-analysis is that the quality of studies included in the dataset can vary considerably and thus papers that have strong methodological or empirical analysis are lumped together with studies that have serious methodological or empirical limitations ("garbage in, garbage out criticism" or… "sausages").

- Any alternative selection schemes might be considered arbitrary and subjective.
- However, the more substantial reason provided by the proponents of MA for the inclusion in the meta-database of both published and unpublished studies is the reduction in the so-called "publication bias": "… it is the result of selection for statistical significance. Researchers, reviewers, and editors are predisposed to treat 'statistically significant' results more favorably; hence, they are more likely to be published. Studies that find relatively small and 'insignificant' effects tend to remain in the 'file drawer'" (Stanley, 2005)

Since in a meta-analysis, notwithstanding the wide variation in the quality of the point estimates included in the study, each estimate in the sample is weighted equally, it may be argued that there is a *non-publication bias* due to the lower quality of unpublished research.

Funnel graph diagram



Estimates of RTAs effects seem to indicate a positive effect on trade (the average of the top six points on the graph, that is, the estimates associated with the smallest standard errors, is equal to 0.04, implying a 4.1% increase in trade), but Figure clearly shows that the plot is overweighted on right side (the simple average of all 1827 estimates is 0.59 implying a 80% increase in trade): is there a bias?

ROMA TRE
UNIVERSITÀ DEGLI STUDI

Methodological Problems: **publication 'impact'**

- Meta-regression tests, as funnel asymmetry test (FAT), allows us to provide an objective assessment of publication bias:

$$\gamma_i = \beta_1 + \beta_0 Se_i + \varepsilon_i$$

In the absence of publication impact the magnitude of the reported effect will be independent of its standard error, then β_0 will be zero.

- Since the studies in the literature may differ greatly in data sets, sample sizes, independent variables, variances of these estimated coefficients may not be equal (as a result, meta-regression errors are likely to be heteroscedastic).

$$\frac{\gamma_i}{Se_i} = t_i = \beta_0 + \beta_1 (1/Se_i) + e_i$$

The weighted least squares is obtained by dividing regression equation by the individual estimated standard errors

ROMA TRE
UNIVERSITÀ DEGLI STUDI

Meta Regression model

Empirical research suggests the following meta regression model including a set of explanatory variables (X) to integrate and explain its diverse findings:

$$\frac{\gamma_{ji}}{Se_{ji}} = t_{ji} = \beta_0 + \beta_1(1/Se_{ji}) + \sum_{k=1}^{K}(\alpha_k X_{jik}/Se_{ji}) + e_{ji}$$

β_1= 'true' value of the parameter conditional on the impact of two groups of explanatory variables:

(1) Dummies explaining the diversity in the results from a methodological point of view: based on a recent survey of the errors in the empirical literature applying gravity equations carried out by Baldwin and Taglioni (*Gravity for Dummies and Dummies for Gravity Equations*, 2006))

(2) Dummies regarding structural features of the studies considered.

ROMA TRE
UNIVERSITÀ DEGLI STUDI

Errors: ***bronze medal***

The bronze medal mistake refers to a common practice in the literature, that of deflating the nominal trade values by some aggregate price indexes: such a procedure may create biases via spurious correlations

- *no-time effects* dummy is equal to "1" when time fixed effects are not included in the regression to control the global trends existing in the data

ROMA TRE
UNIVERSITÀ DEGLI STUDI

Errors: ***silver medal***

The silver medal mistake arises from the fact that the most frequently used measure of bilateral trade flows is the average of bilateral trade and models are usually estimated in log since the log of the average is larger than the average of the logs this leads to an overestimation

- *Log of average bilateral trade flows* dummy

Another problem related to the log specification is due to the existence of zero trade flows:

- different methodological dummies (*Heckman*, *Tobit*, *Poisson*) for the estimation methods dealing with the selection: no expected sign

ROMA TRE
UNIVERSITÀ DEGLI STUDI

MRA: silver and bronze medals dummies

Variables	*Coefficient*
Intercept	3.34^{***}
$1/Se_i$	0.43^{***}
Log of average trade	0.13^{**}
No-Time effects	0.15^{**}

*significant at :***1%; **:5%; *:10%*

Results show:

- A significant *RTA effect* on trade around 40%
- The confusion between the *log of the average* and the average of the logs tends to inflate the gravity estimates: this result confirms and provides a quantitative assessment of the silver medal mistake pointed out by Baldwin
- The *no-time effects dummy* is expected to offset the bronze medal error implied by the mistaken deflation procedure: the positive sign associated with this variable shows that uncorrected studies tend to overestimate the RTAs impact on trade

ROMA TRE
UNIVERSITÀ DEGLI STUDI

Errors: ***gold medal***

The gold medal of classic gravity model mistakes arises from the correlation between omitted variables and trade-cost terms (estimation bias)

- *ols* dummy since these estimates are likely to be biased and inconsistent
- Databases: *cross-section* and *pooled* (no *panel* to avoid collinearity)
- *random effects* (panel data) dummy since these estimation methods would be unbiased only with zero correlation between unobserved variables and RTAs
- *Anderson-van Wincoop* dummy signalling the failure to take into account the multilateral trade resistance term (in panel studies)
- *no-country effect* dummy for the studies not using country-fixed in cross-section models

MRA (1827 RTAs Effects): gold medal dummies

The exclusion or mismeasurement of trading pair–specific variables (the gold medal mistake) can seriously affect the estimation of the RTAs trade impact in both directions:

- We find positive and significant coefficients both for the *ols* and *random effects* dummies
- Results are negative for both *cross-section* and *pooled* dummies: these results support the claim by Baier and Bergstrand (2005) that cross-section estimates are downward biased
- studies that do not include fixed-effects to characterize trade flows involving a particular country (*no-country effects* dummy) show the largest negative coefficient

No significant impact for the *Anderson-van Wincoop* dummy but overall there is some evidence of a significant downward bias on the estimated impacts.

Variables	***Coefficient***
Ols	0.21***
Random effects	0.13*
Cross section	-0.22***
Pooled data	-0.19***
No-Country effects	-0.26***

*significant at :***1%; **:5%; *:10%*

ROMA TRE
UNIVERSITÀ DEGLI STUDI

Structural variables

1. The *agreement* dummy takes the value "1" if the original paper used a variable for each type of agreement
2. An *outliers* dummy handles the 38 extreme values in the sample (Grubbs test in order to detect them)
3. A *dynamic* dummy refers to the few (most recent) panel studies using dynamic techniques
4. Different dummies – *before 1970*, *the 70s*, *the 80s* and *the 90s* – in order to collect studies using data only related to a specific time period

Since we believe that published and specifically interested researches tend to perform more accurate econometric analyses, we introduce

5. a dummy *unpublished* equal to "1" for not published papers
6. a dummy *control* equal to "1" for papers that insert the variable RTA simply as a control variable

ROMA TRE
UNIVERSITÀ DEGLI STUDI

MRA: structural variables

- The estimated coefficient of the *outliers* dummy is clearly positive since most extreme values are positive
- Significant and negative coefficients are associated with the dummies *period ranges* (except for the 1970s): however, their increase is consistent with the often noted evolution from "shallow" to "deep" regional integration agreements

Variables	***Coefficient***
Outliers	3.00***
Before 1970s	-0.35***
1970s	0.04
1980s	-0.21***
After 1990	-0.19***

*significant at :***1%; **:5%; *:10%*

ROMA TRE
UNIVERSITÀ DEGLI STUDI

MRA: other structural variables

Variables	*Coefficient*
Agreements	-0.11***
Control	-0.31***
Unpublished	0.10***
Obs *No of Clusters* *R-squared* *Prob > F* *S.E. of regression*	1827 85 0.25 0.00 5.41

- Negative and highly significant coefficient for the *agreements* dummy: studies focusing on specific RTAs tend to estimate lower impacts on trade.
- The positive coefficient of the *unpublished* dummy suggest that editors tend to exclude the highest (and possibly less realistic) results
- The dummy *control* is strongly negative, hinting to the existence of a downward bias in studies that are not primarily interested in estimating the RTA effect, but use such a dummy as a control variable
- There remains quite a large unexplained variation

*significant at :***1%; **:5%; *:10%*

Probit Significance Equation

$$s_{\mu} = a + \sum_{k=1}^{K} b_k x_{\mu k} + e_{\mu},$$

where the dependent variable is a dummy that takes the value 1 if the estimated effect size is positive and statistically significant.

- In our dataset of 1827 effect sizes, 1134 are significantly different from zero at the level of 5% and 1048 of these estimates are positive.
- Comparing these estimates with those provided by the MRA we single out 3 groups of variables:
  1. significant variables in both cases with the same sign
  2. significant variables in both cases with opposite signs
  3. significant variables in the probit regression that were dropped from the MRA.

ROMA TRE
UNIVERSITÀ DEGLI STUDI

Significant variables: same sign

Probit Estimation	*Marginal effect*
Before 1970	-0.35***
1970s	-0.22***
1980s	-0.32***
After 1990	-0.10***
Control	- 0.17***
Agreement	-0.19***
Random effects	0.19***

*significant at :***1%; **:5%; *:10%*

- Older agreements (or first stages of implementation) are less likely to detect a positive impact on trade: using data ***before 1970***, for instance, reduces the probability by 35 percent.
- The use of data on specific ***agreements*** and the use of the variable as a simple ***control*** substantially reduce the probability of estimating a positive impact on trade respectively by 19 and 17 percent, as it could have been expected given that the estimates provided by these studies are generally lower.
- On the contrary, panel estimates through ***random effects*** raise the probability of finding a positive and significant effect.

ROMA TRE
UNIVERSITÀ DEGLI STUDI

Significant variables: opposite sign

Studies that

- do not include fixed-effects to characterize trade flows involving a particular country
- do not use panel data
- use Ols estimates

are more likely to generate fallacious positive estimates while the downward bias indicated by the MRA is mostly due to negative estimates.

In the other cases, the probit estimates show that possible errors and biases previously mentioned tend to decrease the probability to find a significant and positive impact on trade notwithstanding the overestimation highlighted in the MRA.

Probit Estimation	*Marginal effect*
No-Country effects	0.11**
Cross-section	0.16***
Pooled	0.20***
Ols	0.16***
Log of average trade	-0.16**
No-Time effects	-0.10**
Unpublished	- 0.08***

*significant at :***1%; **:5%; *:10%*

ROMA TRE
UNIVERSITÀ DEGLI STUDI

Significant variables: dropped variables

Probit Estimation	*Marginal effect*
Anderson-van Wincoop	-0.21 ***
Dynamic	-0.21***
Heckman	-0.18*
Tobit	-0.32***
Poisson	-0.27***
Intercept	- 0.35***

*significant at :***1%; **:5%; *:10%*

Here we find some methodological dummies that have not a significant impact when we use the larger sample.

- failure to take into account the multilateral trade resistance term (***Anderson-van Wincoop***)
- use of ***dynamic*** estimation methods
- dealing with the selection bias and the presence of zero trade flows (***Heckman, Tobit, Poisson***)

decrease the probability to get a positive and significant estimate ("false positive" result).

ROMA TRE
UNIVERSITÀ DEGLI STUDI

Conclusions: RTAs & gravity models

- The estimated effect of RTAs varies widely from study to study and sometimes even within the same study
- There is evidence that *ex post* empirical evidence of the effects of RTAs on trade flows shows a positive and not trivial impact: after filtering out the publication impact and other biases, the meta-analysis confirms a robust, positive RTA effect, equivalent to an increase in trade of around 40%
- Estimates tend to get larger in recent years and this could be a consequence of the evolution from "shallow" to "deep" trade agreements
- There appears to be evidence of a downward bias due to omitted variables problems (**gold medal** mistake)
- Data measurement (**silver medal** mistake) and specification problems (**bronze medal** mistake) are less likely to get (statistically speaking) "good results" and tend to get estimates biased in the opposite direction

Conclusions: is there a (non)publication bias?

- Estimates obtained from gravity models using the RTA dummy as (just another) control variable are largely downward biased and are much less likely to find significant results.
- Our results fly in the face of the general belief among meta-analysts that referees and editors would be predisposed to treat “large and significant” results more favourably.
- In the literature that we reviewed, as a matter of fact, there is strong statistical evidence of a non-publication bias which favours the reporting of positive trade effects, while the publication process leads to lower, and probably more realistic, estimates.

Caveats

- MA is NOT a substitute for sound, original, applied analysis
- MA does NOT allow to escape reading (and understanding) the relevant literature
- Indeed, MA quantifies the impact of the methodological choices pointed out by qualitative literature reviews: it should be considered a COMPLEMENT rather than a SUBSTITUTE

Maria Cipollina and Filomena Pietrovito

TRADE IMPACT OF EUROPEAN UNION PREFERENTIAL POLICIES: A META-ANALYSIS OF THE LITERATURE

Introduction

- The gravity model is used frequently to estimate the impact of European Union (EU) Preferential Trade Agreements (PTA) on trade flows.
- Because of differences in the datasets, sample sizes and independent variables employed, existing studies report very different estimates.
- Meta-Analysis (MA) to provide pooled estimates of the effect of PTA on bilateral trade, based on fixed and random effects models.

Preference measurement

- Most of the studies in our sample use a dummy variable for PTA which is equal to 1 if there is a preferential arrangement between EU and the country in question
- Using a dummy to capture the impact of PTA on trade is not adequate because:
  1. it captures all the other factors that are specific to the country-pair and contemporaneous with the PTA;
  2. it does not discriminate among the instruments adopted for preferential trade policy;
  3. it does not indicate the level of the trade preferences

Data sources

Papers written in English, via an extensive search in:

- Google Scholar search produced papers published in academic journals as well as working papers and unpublished studies.
- EconLit provides coverage of the economics literature since 1969, and includes 750 journals.
- The Web of Science provides access to current and retrospective multidisciplinary information from approximately 8,700 of the most prestigious, international, high impact research journals (199 journals in the field of economics), covering 1992 to 2010.
- Scopus includes the abstracts and references for 15,000 peer-reviewed journals from more than 4,000 international publishers, which ensures broad interdisciplinary coverage.

We also traced some specific papers that other work cross-referenced.

Data collection

We searched on the keywords “preferential trade agreements”, “gravity equation” and “gravity model” in the title, abstract or text.

These identified papers dealing with PTA, and papers that use a gravity approach:

- from the first group we selected papers analysing PTA and focusing on trade flows to the EU;
- in the second group, we selected those studies that had PTA as a key explanatory variable in the gravity equation.

Our final sample includes 36 papers (10 published in an academic journal and 26 working papers or unpublished studies), providing 638 point estimates based on a dummy variable for PTA and 388 point estimates based on quantitative variables for PTA

Structure of the database

Trade\Policy	Specific agreements		PTA	
(number of estimates)	Dummy variable	Preference Margin	Dummy variable	Preference Margin
Sectoral trade	164	217	83	0
Total trade	330	38	144	50

Cap5.do: descriptive statistics and outliers generation

```
bysort dummy_policy: sum gamma, d
gen outlier = 1 if gamma < -21 & dummy_policy == 0
replace outlier= 1 if gamma > 8 & dummy_policy == 0
replace outlier= 1 if gamma < -5 & dummy_policy == 1
replace outlier= 1 if gamma > 3 & dummy_policy == 1
replace outlier = 0 if outlier ==
```

Cap5.do: variable generation

We generate the following variables normalized by the standard errors.

```
gen tratio=gamma/segamma
gen invse=1/segamma
gen nocountry_se=nocountry/segamma
gen notime_se=notime/segamma
gen cross_se=cross/segamma
gen year70s_se=year70s/segamma
gen year80s_se=year80s/segamma
gen year90s_se=year90s/segamma
gen year2000s_se=year2000s/segamma
gen aggr_data_se=aggr_data/segamma
gen agg_eu_se=agg_eu/segamma
gen agriculture_se=agriculture/segamma
gen nozero_se=nozero/segamma
gen GMM_se=GMM/segamma
```

```
gen HTM_se=HTM/segamma
gen Heck_se=Heckman/segamma
gen Poisson_se=Poisson/segamma
gen Tobit_se=Tobit/segamma
gen over_se=overdispertion/segamma
gen unpublished_se=unpublished/segamma
gen outlier_se=outlier/segamma
gen pta_se=pta/segamma
gen acp_se=acp/segamma
gen eba_se=eba/segamma
gen euromed_se=euromed/segamma
gen gsp_se=gsp/segamma
gen gspplus_se=gspplus/segamma
```

Cap5.do: labels definition

```
label var GMM_se "GMM"
label var HTM_se "Hausman-Taylor"
label var Heck_se "Heckman"
label var Poisson_se "Poisson"
label var Tobit_se "Tobit"
label var over_se "ZIP/Negative
Binomial"
label var unpublished_se "Unpublished"
label var outlier_se "Outliers"
label var pta_se "PTAs"
label var acp_se "ACP"
label var eba_se "EBA"
label var euromed_se "Euro-Med"
label var gsp_se "GSP"
label var gspplus_se "GSP-Plus"
label var tratio "t-statistic for gamma"
label var invse "1/Se"
label var nocountry_se "No-country effects"
label var notime_se "No-time effects"
label var cross_se "Cross-section"
label var year70s_se "1970s"
label var year80s_se "1980s"
label var year90s_se "1990s"
label var year2000s_se "2000s"
label var aggr_data_se "Aggregated data"
label var agg_eu_se "Aggregated EU"
label var agriculture_se "Agriculture"
label var nozero_se "No-zero Treatment“
```

ROMA TRE
UNIVERSITÀ DEGLI STUDI

Test

We perform the Fisher test of the null hypothesis of no effect of PTAs on trade.

gen pgamma=ttail(obs,tratio)

replace pgamma=1.e-20 if pgamma==0.

label var pgamma "P-value for gamma"

metap pgamma if dummy_policy==1, method(f)

metap pgamma if dummy_policy==0, method(f)

“The Fisher test suggests that the null hypothesis of no effect of PTA on trade should be rejected at any standard level of significance (χ^2 is equal to 6,509 and to 7,536 respectively for estimates from studies using dummies and preference margins for trade policy)”

ROMA TRE
UNIVERSITÀ DEGLI STUDI

Meta-analysis

meta gamma segamma if dummy_policy==1

meta gamma segamma if dummy_policy==0

Sample	Effects	Pooled Estimate	Lower Bound of 95% CI	Upper Bound of 95% CI	p-value for H_0: no effect	Q-test (p-value)
Dummy for PTAs	Fixed effects	0.02	-0.01	0.02	0.00	0.00
	Random effects	0.20	0.17	0.23	0.00	0.00
Preference margin	Fixed effects	0.06	0.06	0.06	0.00	0.00
	Random effects	0.07	0.06	0.07	0.00	0.00

We perform the same meta-analysis by considering different PTAs.

meta gamma segamma if dummy_policy==1 & acp==1
meta gamma segamma if dummy_policy==1 & eba==1
meta gamma segamma if dummy_policy==1 & euromed==1
meta gamma segamma if dummy_policy==1 & gsp==1
meta gamma segamma if dummy_policy==1 & gspplus==1
meta gamma segamma if dummy_policy==0 & acp==1
meta gamma segamma if dummy_policy==0 & eba==1
meta gamma segamma if dummy_policy==0 & euromed==1
meta gamma segamma if dummy_policy==0 & gsp==1
meta gamma segamma if dummy_policy==0 & gspplus==1

Sample	Random effects				
	ACP	EBA	Euro-Med	GSP	GSP-Plus
Dummy for PTAs	0.66***	-0.33**	0.05	0.14***	1.32
Preference margin	0.03***	0.02**	0.01	0.02***	-0.01

ROMA TRE
UNIVERSITÀ DEGLI STUDI

MRA: preference dummy

eststo clear

eststo: reg tratio invse nocountry_se notime_se cross_se year70s_se year80s_se year90s_se year2000s_se aggr_data_se agg_eu_se agriculture_se nozero_se GMM_se HTM_se Heck_se Poisson_se Tobit_se over_se unpublished_se outlier_se pta_se if dummy_policy==1, robust cluster(paper)

eststo: reg tratio invse nocountry_se notime_se cross_se year70s_se year80s_se year90s_se year2000s_se aggr_data_se agg_eu_se agriculture_se nozero_se GMM_se HTM_se Heck_se Poisson_se Tobit_se over_se unpublished_se outlier_se acp_se eba_se euromed_se gsp_se gspplus_se if dummy_policy==1, robust cluster(paper)

esttab using “name”. doc, title (Table - MRA of PTAs effects for papers using dummies for PTAs) se ar2 label replace rtf b(2) star(* 0.10 ** 0.05 *** 0.01) se(2) mti

MRA: preference margin

eststo clear

eststo: reg tratio invse nocountry_se notime_se cross_se year2000s_se agg_eu_se Heck_se Poisson_se over_se unpublished_se outlier_se pta_se if dummy_policy==0, robust cluster(paper)

eststo: reg tratio invse nocountry_se notime_se cross_se year2000s_se agg_eu_se Heck_se Poisson_se over_se unpublished_se outlier_se acp_se eba_se euromed_se gsp_se gspplus_se if dummy_policy==0, robust cluster(paper)

esttab using “name”.doc, title (Table - MRA of PTAs effects for papers using preference margins for PTAs) se ar2 label replace rtf b(2) star(* 0.10 ** 0.05 *** 0.01) se(2) mti append

Figures

We perform the Funnel and the Egger test.

```
funnel segamma gamma if dummy_policy == 1, xlab title (Figure 1.1) saving (Figure_5.1.gph, replace)
funnel segamma gamma if dummy_policy == 0, xlab title (Figure 1.2) saving (Figure_5.2.gph, replace)

metabias gamma segamma if dummy_policy == 1, graph (egger) saving (Figure_2.1_Egger.gph, replace)
metabias gamma segamma if dummy_policy == 0, graph (egger) saving (Figure_2.2_Egger.gph, replace)

end
```

Hands-on gravity estimation with STATA

IFPRI Guide

Data files

There are two Data files:

- dataset_all.dta: it contains all the essential variables used in the regressions using panel data (Part 1). The dataset covers the period from 1996 to 2006 and includes 154 developed and developing countries.
- dataset_us.dta: it contains all the essential variables used in the regressions using cross-section data (Part 2 and Part 3). It refers to year 2004 US agricultural imports from 226 countries. Data are disaggregated at the most detailed level allowed by the international Harmonized System (HS) classification (6 digits) and include 689 products.

The variable names are largely self-explanatory and are described when the labels are created: their generation and construction can thus be directly inspected.

Data Source

1. *Dataset_all.dta*: Dataset is building on extraction from WITS (http://wits.worldbank.org/wits/index.html) and on information provided by the Cepii dataset (http://www.cepii.fr/).

- The WITS application gives access to international trade statistics of UN COMTRADE (The United Nations Commodity Trade Statistics database) and tariff database of UNCTAD-TRAINS (Trade Analysis and Information System).
- The Cepii dataset includes data on GDP and distances between countries and dummies for contiguity, common language, and former colonial links.

2. *Dataset_us.dta*: Data on trade and tariffs at the HS6 level of detail are taken from the MAcMapHS6-V2 database (http://www.cepii.fr/). MAcMap provides a consistent worldwide assessment of protection, including ad valorem equivalent rates of specific duties and tariff rate quotas (including those introduced at the end of the Uruguay Round), for 2004. Data for the remaining explanatory variables are from the Cepii dataset.

Do files

There are three Do files:

- regressions aggregated data.do: it runs regressions using panel, aggregated data (Part 1).
- regressions disaggregated data.do: it runs regressions using cross-section, disaggregated as well as aggregated data (Part 2).
- regressions zeroes treatment.do: it runs regressions using non-linear estimators (Heckman or Poisson) dealing with 'zero' trade flows (Part 3).

regressions aggregated data.do: variable generation

We use the data file dataset_all

use dataset_all.dta

We take the logs of all continuous variables included in the regressions:

g limports=ln(imports)
g lgdp_o=ln(gdp_o)
g lgdp_d=ln(gdp_d)
g ldist= ln(distw)
g ltariff=ln(1+s_average)

We label the variables to be included in the tables.

la var limports "Ln(Imports)"
la var colony "Colonial link"
la var comlang_off "Common language"
la var contig "Border"
la var ldist "Ln(distance)"
la var lgdp_d "Ln(GDP_importer)"
la var lgdp_o "Ln(GDP_exporter)"
la var rta "Regional Trade Agreement"
la var ltariff "Ln(1+Tariff)"

Finally, we generate the different fixed effects.

qui tab imp, g(dimp)
qui tab exp, g(dexp)
qui tab pair, g(dpair)
qui tab year, g(dyear)

regressions aggregated data.do: regression specifications

We start by declaring data to be panel.

tsset pair year

In order to show the consequences of ignoring the multilateral resistance term, we firstly estimate equation (2) without fixed effects

eststo: reg limports lgdp_d lgdp_o ldist contig colony comlang_off rta, robust

Then, we introduce the different types of fixed effects:

eststo: reg limports lgdp_d lgdp_o ldist contig colony comlang_off rta dyear, robust*

eststo: reg limports lgdp_d lgdp_o ldist contig colony comlang_off rta dimp dexp*, robust*

eststo: reg limports lgdp_d lgdp_o ldist contig colony comlang_off rta dimp dexp* dyear*,*

robust

eststo: reg limports lgdp_d lgdp_o ldist contig colony comlang_off rta dpair dyear*, robust*

The command “esttab” creates the regression table in a file regressions1.doc

esttab using regressions1.doc, title (aggregate-dummy policy) se ar2 label replace rtf b(2) star (0.10 ** 0.05 *** 0.01) se(2) mti drop (dexp* dimp* dyear* dpair*)*

Panel results with different fixed effects

Variables	1. Without FE	2. With FE			
	(1.1)	(2.1)	(2.2)	(2.3)	(2.4)
Ln(GDP_importer)	0.75***	0.74***	0.70***	0.91***	1.05***
	(0.02)	(0.02)	(0.14)	(0.20)	(0.15)
Ln(GDP_exporter)	1.20***	1.22***	0.01	0.08	0.18
	(0.01)	(0.01)	(0.12)	(0.14)	(0.12)
Ln(distance)	-1.50***	-1.55***	-1.47***	-1.47***	-1.59***
	(0.04)	(0.04)	(0.05)	(0.05)	(0.23)
Dummy: Border	0.61***	0.54***	0.54***	0.54***	2.96***
	(0.10)	(0.10)	(0.11)	(0.11)	(0.85)
Dummy: Colonial link	-0.51**	-0.61***	0.09	0.08	6.39***
	(0.23)	(0.24)	(0.25)	(0.25)	(0.62)
Dummy: Common language	1.10***	1.21***	0.78***	0.78***	6.88***
	(0.06)	(0.06)	(0.09)	(0.09)	(0.40)
Dummy: Regional trade agreement	0.54***	0.71***	0.64***	0.62***	0.28
	(0.08)	(0.09)	(0.10)	(0.10)	(0.18)
Constant	6.51***	6.69***	22.37***	20.25***	6.63***
	(0.38)	(0.38)	(2.34)	(2.80)	(0.33)

regressions aggregated data.do: continuous treatment

reg limports lgdp_d lgdp_o ldist contig colony comlang_off ltariff dimp dexp* dyear*, robust*

Then we create the results table

esttab using regressions1.doc, title (aggregate-tariff) se ar2 label replac rtf b(2) star (0.10 ** 0.05 *** 0.01) se(2) mti drop (dexp* dimp* dyear*) append*

Panel results with continuous policy variable

Ln(Gdp_importer)	0.61***
	(0.20)
Ln(Gdp_exporter)	0.12
	(0.14)
Ln(distance)	-1.43***
	(0.04)
Dummy: Border	0.62***
	(0.11)
Dummy: Colonial link	0.05
	(0.24)
Dummy: Common language	0.74***
	(0.09)
Ln(1+tariff)	-0.57***
	(0.05)
Constant	22.87***
	(2.84)
Time fe	Si
Exporter fe	Si
Importer fe	Si
Observations	7797
Adjusted R^2	0.743

Exercises

1. Continuous and discrete policy variables
2. Sample splits:

- by region/continent
- by year (before/after 2000)

3. Additional dummies

- time varying importer
- time varying exporter

Part 2

1. In Part 2, cross-sections estimations show the importance of working with disaggregated data.
2. *dataset_us.dta*: it refers to year 2004 US agricultural imports from 226 countries. Data are disaggregated at the most detailed level allowed by the international Harmonized System (HS) classification (6 digits) and include 689 products. Data on trade and tariffs at the HS6 level of detail are taken from the MAcMapHS6-V2 database (http://www.cepii.fr/). Data for the remaining explanatory variables are from the Cepii dataset.

regressions disaggregated data.do: Variable Generation

We use the data file dataset_us

- *use dataset_us.dta*

We take the logs of all continuous variables included in the regressions

- *g limports=ln(trade)*
- *g lgdp_o=ln(gdp_o)*
- *g lgdp_d=ln(gdp_d)*
- *g ldist= ln(distw)*
- *g ltariff=ln(1+tariff)*

We label the variables to be included in the tables.

- *la var limports "Ln(Imports)"*
- *la var colony "Colonial link"*
- *la var comlang_off "Common language"*
- *la var contig "Border"*
- *la var ldist "Ln(distance)"*
- *la var lgdp_d "Ln(GDP_importer)"*
- *la var lgdp_o "Ln(GDP_exporter)"*
- *la var ltariff "Ln(1+Tariff)"*

We generate the exporter and product fixed effects

- *qui tab exp, g(dexp)*
- *qui tab hs6, g(dhs6)*

regressions aggregated data.do: regression specifications

We firstly use the OLS estimator without fixed effects

eststo: reg limports lgdp_o ldist contig colony comlang_off ltariff, robust

Then, we introduce the different types of fixed effects:

eststo: reg limports lgdp_o ldist contig colony comlang_off ltariff dhs6, robust*

eststo: reg limports contig colony comlang_off ltariff dexp dhs6*, robust*

Finally, we collapse the dataset in order to obtain aggregated data for a robustness analysis.

collapse (sum) trade (mean) tariff gdp_o distw contig colony comlang_off, by(exp)

g limports=ln(trade)

g lgdp_o=ln(gdp_o)

g ldist= ln(distw)

g ltariff=ln(1+tariff)

and we run again the regression to highlight the relevance of the aggregation issue

eststo: reg limports lgdp_o ldist contig colony comlang_off ltariff, robust

The command “esttab” creates the regression table in a file regressions2.doc

esttab using regressions2.doc, title (dati_us_agr) se ar2 label replac rtf b(2) star (0.10 ** 0.05 *** 0.01) se(2) mti drop (dexp* dhs6*) append*

Cross-section results with different fixed effects and levels of aggregation

	(1)	(2)	(3)	(4)
	est1	est2	est3	est4
lgdp_o	0.34***	0.44***		1.07***
	(0.01)	(0.01)		(0.06)
ldist	-0.39***	-0.36***		-1.96***
	(0.05)	(0.05)		(0.40)
(mean) contig	1.79***	1.92***	7.26***	-0.52
	(0.13)	(0.12)	(0.43)	(0.74)
(mean) colony	0.04	0.03	4.24***	-0.23
	(0.08)	(0.07)	(0.54)	(0.87)
(mean) comlang_off	0.29***	0.38***	-0.74	1.09***
	(0.05)	(0.04)	(0.49)	(0.38)
ltariff	0.19	-2.85***	-2.00***	-5.00
	(0.16)	(0.41)	(0.41)	(10.85)
Constant	-4.34***	-11.38***	-5.41***	9.85***
	(0.42)	(0.40)	(0.38)	(3.38)
Observations	20902	20902	21136	176
Adjusted R^2	0.113	0.254	0.303	0.602

Exercises

1. Add one (or more) RTAs
2. Sample split:

- by geography/policy
- by product

Part 3

1. , Part 3 shows how you can solve the 'zero (trade flows) problem' using either Heckman or Poisson estimators
2. *dataset_us.dta*: it refers to year 2004 US agricultural imports from 226 countries. Data are disaggregated at the most detailed level allowed by the international Harmonized System (HS) classification (6 digits) and include 689 products. Data on trade and tariffs at the HS6 level of detail are taken from the MAcMapHS6-V2 database (http://www.cepii.fr/). Data for the remaining explanatory variables are from the Cepii dataset.

The Log of Gravity page:

http://personal.lse.ac.uk/tenreyro/LGW.html

Reference:

Santos Silva, J.M.C. and Tenreyro, Silvana (2006), The Log of Gravity, The Review of Economics and Statistics, 88(4), pp. 641-658.

In this page you can find the data set used in the paper, codes to extend some of the results in the paper, and other useful information on the implementation of the PPML estimator.

regressions disaggregated data.do: Variable Generation

We use the data file dataset_us

- *use dataset_us.dta*

We take the logs of all continuous variables included in the regressions

- *g limports=ln(trade)*
- *g lgdp_o=ln(gdp_o)*
- *g lgdp_d=ln(gdp_d)*
- *g ldist= ln(distw)*
- *g ltariff=ln(1+tariff)*

We label the variables to be included in the tables.

- *la var limports "Ln(Imports)"*
- *la var colony "Colonial link"*
- *la var comlang_off "Common language"*
- *la var contig "Border"*
- *la var ldist "Ln(distance)"*
- *la var lgdp_d "Ln(GDP_importer)"*
- *la var lgdp_o "Ln(GDP_exporter)"*
- *la var ltariff "Ln(1+Tariff)"*

We generate the exporter and product fixed effects

- *qui tab exp, g(dexp)*
- *qui tab hs6, g(dhs6)*

regressions aggregated data.do: regression specifications

We firstly run the regression using the Heckman estimator

eststo: heckman limports contig colony comlang_off ltariff , select(contig colony ltariff) mills(lambda)

Then we run the regression using the Poisson Pseudo-Maximum Likelihood estimator

eststo: ppml trade contig colony comlang_off ltariff

The command “esttab” creates the regression table in a file regressions3.doc

esttab using regressions3.doc, title (treatment of zeros) se ar2 label replace rtf b(2) star (0.10 ** 0.05 *** 0.01) se(2) mti*

Cross-section results with different fixed effects and levels of aggregation

	(1)	(2)
	est1	est2
main		
Border	2.67***	3.20***
	(0.18)	(0.16)
Colonial link	0.76***	1.30***
	(0.11)	(0.24)
Common language	-0.09**	0.12
	(0.04)	(0.15)
Ln(1+Tariff)	0.25	-0.58
	(0.16)	(0.35)
Constant	-3.26***	-0.52***
	(0.18)	(0.09)
select		
Border	1.95***	
	(0.06)	
Colonial link	0.68***	
	(0.03)	
Ln(1+Tariff)	-0.04	
	(0.04)	
Constant	-0.31***	
	(0.01)	
athrho		
Constant	-0.21***	
	(0.07)	
lnsigma		
Constant	1.03***	
	(0.01)	
Observations	52340	52340

ROMA TRE
UNIVERSITÀ DEGLI STUDI

Exercises

1. Add one (or more) RTAs
2. Sample splits

- by geography/policy
- by product