Web Appendix to

"Do Borders Really Slash Trade? A Meta-Analysis"*

Tomas Havranek[a,b] and Zuzana Irsova[b]

[a]Czech National Bank
[b]Charles University, Prague

January 7, 2016

Abstract

This appendix provides statistical tests of publication bias and robustness checks of the baseline specification for the examination of heterogeneity in the estimated border coefficients. We also provide additional summary statistics for data used in the paper and a list of studies included in the meta-analysis. Moreover, we list potential problems with conducting meta-analysis in economics and discuss how we address them in the paper.

1 Statistical Tests of Publication Bias

Publication selection creates a systematic relationship between estimates and their standard errors (Stanley, 2008):

$$HOME_{ij} = HOME_0 + \beta \cdot SE(HOME_{ij}) + u_{ij}, \tag{1}$$

where $HOME_{ij}$ are i-th estimates of the semi-elasticity reported in j-th study, $SE(HOME_{ij})$ are the reported standard errors of the semi-elasticity estimates, $HOME_0$ is the mean semi-elasticity corrected for potential publication bias, β measures the extent of publication bias, and u_{ij} is a

---

[*] The manuscript is available at `meta-analysis.cz/border`. Corresponding author: Zuzana Irsova, `zuzana.irsova@ies-prague.org`.

normal disturbance term. For example, if the true mean semi-elasticity was zero (implying no border effect) but all researchers reported the 5% of estimates that are positive and statistically significant, the estimated β would be close to two: the researchers would need their t-statistics, $HOME/SE(HOME)$, to equal at least two.

Equation (1) can be interpreted as a test of funnel asymmetry, because it follows from rotating the axes of the funnel plot and inverting the values on the new horizontal axis to show standard errors instead of precision. Note that the test has low power if the true underlying value of the effect is close to zero and the only source of publication bias is selection for statistical significance: when $HOME_0$ is zero and insignificant estimates, positive or negative, are omitted, β is zero, even though publication selection may be substantial (the funnel plot gets hollow, but not asymmetrical). Nevertheless, such a symmetrical selection does not create a bias in the mean of the reported estimates, so it is usually not a source of concern (Stanley, 2005).

In examinations of publication bias it is common to assume, as we have done so far in this section, that the selection criteria leading to the bias are based on the sign and statistical significance of the estimate in question. In the literature estimating the border effect, however, potential publication selection need not be driven by the sign and significance of the resulting coefficients, because negative and insignificant estimates are difficult to obtain due to the relatively large underlying border effect. Instead, researchers are likely to use the well-known results reported by McCallum (1995) as a benchmark, and in this case publication selection could assume the following two forms:

First, researchers may discard estimates inconsistent with McCallum (1995). The benchmark semi-elasticity presented by McCallum (1995) is 3.09 with a standard error of 0.13. Estimates close to McCallum's are reported frequently: those lying within one standard error from McCallum's central estimate account for 12% of all the estimates in the literature, twice the number we would expect if the estimates were normally distributed (given that the literature reports a mean estimate of 3.03 with a standard deviation of 1.6). The over-reporting of estimates similar to McCallum's might reflect the fact that researchers simply try to replicate his results as a part of their analysis, or it could point to genuine publication selection. In any case, because such a selection criterion is symmetrical (both small and large estimates inconsistent with McCallum are omitted), it does not create a bias. Note that the mean of all the semi-elasticities reported in the literature is very close to McCallum's central estimate, and that the mean would only change from 3.03 to 3.02 if we discarded all results lying inside the 95% confidence interval of McCallum's estimate.

Second, researchers may want to shrink the border effect reported by McCallum (1995) and preferentially select small estimates for reporting. Such a selection criterion is asymmetrical, and would result in a downward bias in the literature. Suppose, for example, that researchers would strive to report estimates significantly smaller than McCallum's result. They would need the ratio $(3.09 - \mathit{HOME})/\mathit{SE}$, the relevant t-statistic, to be as large as possible, which would again give rise to a correlation between the nominator and denominator of the ratio and would show as a negative and statistically significant coefficient β in (1). In other words, the corresponding funnel plot would become asymmetrical because large estimates would be reported less often than small estimates with the same precision. Equation (1) measures the degree of asymmetry of the funnel plot and so it is able to detect any selection process that causes a systematic bias in the literature.

We present the results of the funnel asymmetry tests in Table 1. Because regression (1) is heteroskedastic, we report robust standard errors, which are clustered at the level of individual studies and data sets. The first column of panel A shows estimates of the parameters from (1) using all 1,271 semi-elasticities in our sample. The coefficient corresponding to the extent of publication bias is statistically insignificant and close to zero, while the estimated semi-elasticity beyond publication bias is 2.9, close to the mean and median semi-elasticity reported in the literature. Therefore, neither visual nor formal tests show any evidence of publication selection, and the potential selection does not create any bias in the mean reported estimate of the border effect.

The second column of panel A in Table 1 estimates equation (1) using only the semi-elasticities reported in published studies. Perhaps editors or referees prefer coefficients that are significantly smaller than the central estimate of McCallum (1995), which would pull the mean reported semi-elasticity down. Indeed, in a meta-analysis of vertical productivity spillovers from foreign direct investment, Havranek & Irsova (2011) find that studies published in refereed journals show substantially more publication bias than unpublished manuscripts. Our results concerning the border effect, however, show little difference between published and unpublished

Table 1: Funnel asymmetry tests show no publication bias

Panel A: unweighted regressions	All estimates	Published	Fixed effects	Instrument
SE (publication bias)	0.604	0.599	0.383	-0.797
	(0.514)	(0.522)	(0.534)	(2.020)
Constant (effect beyond bias)	2.852^{***}	2.932^{***}	2.918^{***}	3.270^{***}
	(0.321)	(0.339)	(0.159)	(0.724)
Studies	61	48	61	61
Observations	1,271	1,144	1,271	1,271
Panel B: weighted regressions	Precision	Study	Impact	Citations
SE (publication bias)	0.246	1.489	3.062	5.073
	(1.964)	(1.170)	(2.024)	(4.272)
Constant (effect beyond bias)	2.959^{***}	2.204^{***}	1.634^{***}	1.235^{**}
	(0.723)	(0.395)	(0.424)	(0.501)
Studies	61	61	53	49
Observations	1,271	1,271	1,124	1,069

Notes: The table presents the results of regression $HOME_{ij} = HOME_0 + \beta \cdot SE(HOME_{ij}) + u_{ij}$. $HOME_{ij}$ and $SE(HOME_{ij})$ are the i-th estimates of the home coefficient (the coefficient estimated in a gravity equation on the dummy variable that equals one for within-country trade flows) and their standard errors reported in the j-th studies. The standard errors of the regression parameters are clustered at both the study and data set level and shown in parentheses (the implementation of two-way clustering follows Cameron *et al.*, 2011). Published = we only include published studies. Fixed effects = we use study dummies. Instrument = we use the number of observations in the gravity equation as an instrument for the standard error. The regressions in Panel B are estimated by weighted least squares. Precision = we take the inverse of the reported estimate's standard error as the weight. Study = in addition to "Precision" the inverse of the number of estimates reported per study is taken as the weight. Impact = in addition to "Study" the RePEc recursive discounted impact factor of the outlet where the study was published is taken as the weight. Citations = in addition to "Impact" the number of Google Scholar citations received per year is taken as the weight. ***, **, and * denote statistical significance at the 1%, 5%, and 10% level.

studies both in the extent of publication bias and in the mean underlying semi-elasticity beyond any potential bias. Next, in the third column we include fixed effects for individual studies to control for method or other quality characteristics specific to individual studies. The fixed-effects estimation represents another advantage of collecting multiple estimates per study. The results are very similar to the baseline specification reported in the first column; we get no evidence of publication bias, and the mean estimated semi-elasticity is still 2.9.

Specification (1) only includes one explanatory variable, the standard error. It is possible that some method choices affect both the estimated semi-elasticity and the corresponding standard error, which would cause the error term u_{ij} to be correlated with $SE(HOME_{ij})$. In the last column of panel A in Table 1 we use the logarithm of the number of observations in the gravity equation as an instrument for $SE(HOME_{ij})$: the number of observations is correlated with the reported standard errors of the semi-elasticities, but little related to the methods of estimation. The instrumental variable estimation is less precise, but still reports the mean underlying semi-elasticity close to 3 and no evidence of publication bias.

In panel B of Table 1 we weight all the estimates by their precision. We have noted that equation (1) is heteroskedastic, and the explanatory variable directly captures the variance of the response variable. To achieve efficiency, many applications of meta-analysis divide (1) by the corresponding standard error; that is, they multiply the equation by the precision of the estimates. Such an approach has the additional allure of giving more importance to precise results. The first column of panel B shows that precision weights do not change our results.

The second column of panel B adds weighting by the inverse of the number of estimates reported in studies to the precision weights. In line with the summary statistics from the previous section, the mean semi-elasticity decreases when each study gets the same weight. Next, in column 3 we add weighting by the discounted recursive RePEc impact factor of the publication outlet. The estimated semi-elasticity decreases to 1.6: better journals seem to publish smaller estimates. Finally, we also weight the estimates by the number of Google Scholar citations the study receives each year. The semi-elasticity decreases to 1.2, implying a border effect of 3.4. Thus, when we give more weight to highly-cited papers published in good journals, we are able to shrink the mean border effect more than five times. In the next section we explore how these differences between studies can be explained by variation in data and methodology.

2 Robustness Checks and Additional Data Description

We present two additional sets of results. First, we use alternative priors for Bayesian model averaging. Second, we employ unweighted regressions in the BMA exercise. We show that the results are similar to the baseline in terms of the estimated effects of the different aspects of study design on the estimated semi-elasticities, and that the resulting "best practice" estimates of the border effect are close to those reported in the previous section.

In the baseline specification we use the unit information prior for Zellner's g-prior, which means that the prior (each regression coefficient equals zero) provides the same amount of information as one observation in the data set. Because we have 1,271 observations, the prior does not drive the posterior results. The second important choice is the model prior, which determines the prior probability of each model. In the baseline specification we employ the uniform model prior, which gives each model the same prior probability. Eicher *et al.* (2011)

show that these intuitive priors yield the best predictive performance. Nevertheless, there are obviously many other ways of choosing the priors, and the choice could influence our results.

The disadvantage of the uniform model prior is that it gives more weight to models with the mean number of variables, which is $32/2 = 16$ in our case. Such models appear most frequently among the subsets of all the 2^{32} possible models. Nevertheless, the true model may only contain a few variables, so the emphasis on large models may be counterproductive. An alternative is the beta-binomial prior advocated by Ley & Steel (2009), which gives the same prior probability to each *model size*, and thus does not prefer large models. An often-used alternative to the unit information prior is the BRIC g-prior (for example, Fernandez *et al.*, 2001).

Table 2 summarizes the results of Bayesian model averaging with the alternative priors; we provide more details and diagnostics in Table 6 and Figure 2. The results are very similar to our baseline specification concerning the estimated posterior inclusion probabilities for the explanatory variables, the signs of the regression coefficients, and their magnitude. The semi-elasticity conditional on best practice is 1.67, implying a partial equilibrium border effect of 5.3, slightly below the estimate presented in the last section. The region-specific semi-elasticities are also similar: 2.04 for Canada, 0.52 for the US, 1.41 for the EU, 0.40 for the OECD, and 3.06 for emerging countries.

The second robustness check involves unweighted regressions, which means that studies presenting many estimates wield more influence in the meta-analysis. Table 3 shows that the posterior inclusion probabilities differ from the baseline specification for some variables. Concerning data characteristics, the age of the data seems to be important: the reported semi-elasticity decreases each year by about 0.025. Studies that do not have direct data on within-country trade flows report larger estimates of the border effect. Adding one to zero trade flows typically yields lower semi-elasticities (by about 0.7). Moreover, the impact factor of the journal and the number of citations of the study seem to be important: better journals tend to report smaller estimates, while broadly cited studies usually report larger estimates. Nevertheless, the best practice estimates of the border effect for the entire world and for individual regions are again very close to our baseline results. The overall mean semi-elasticity is 1.82, implying a partial equilibrium border effect of 6.2.

Table 2: Robustness check—alternative priors for BMA

Response variable:	Bayesian model averaging			Frequentist check (OLS)		
Estimate of Home	Post. mean	Post. SD	PIP	Coef.	Std. er.	p-value
Data characteristics						
Mid-year of data	0.003	0.003	0.466	-0.001	0.012	0.926
Panel data	0.004	0.062	0.102			
Disaggregated	0.745	0.143	1.000	0.545	0.306	0.075
Obs. per year	0.000	0.008	0.060			
No. of years	0.113	0.082	0.738	0.100	0.098	0.310
Countries examined						
Canada	0.724	0.126	1.000	0.823	0.317	0.010
US	-1.183	0.133	1.000	-1.131	0.227	0.000
EU	-0.518	0.161	0.995	-0.548	0.383	0.152
OECD	-0.975	0.176	1.000	-0.902	0.343	0.009
Emerging	0.868	0.268	0.990	0.602	0.322	0.062
Design of the analysis						
No internal trade	0.184	0.209	0.508	0.361	0.389	0.354
Inconsistent dist.	0.754	0.145	1.000	0.521	0.304	0.087
Actual distance	-0.907	0.155	1.000	-0.716	0.331	0.030
Total trade	-0.001	0.062	0.041			
Asymmetry	0.518	0.121	0.999	0.492	0.246	0.045
Instruments	-0.008	0.054	0.055			
Treatment of multilateral resistance						
Remoteness	-0.016	0.066	0.090			
Country fixed eff.	0.362	0.334	0.601	0.214	0.272	0.431
Ratio estimation	0.628	0.491	0.721	0.738	0.506	0.145
Anderson est.	0.389	0.376	0.579	0.162	0.308	0.599
No control for MR	0.961	0.314	1.000	0.641	0.297	0.031
Treatment of zero trade flows						
Zero plus one	0.004	0.033	0.050			
Tobit	-0.640	0.155	0.998	-0.600	0.321	0.062
PPML	-0.726	0.155	1.000	-0.860	0.529	0.104
Zeros omitted	-0.007	0.035	0.074			
Control variables						
Adjacency control	0.125	0.156	0.453	0.341	0.245	0.163
Language control	-0.001	0.022	0.046			
FTA control	-0.253	0.167	0.778	-0.466	0.321	0.147
Publication characteristics						
Published	0.346	0.103	0.986	0.276	0.272	0.311
Impact	0.021	0.045	0.230			
Citations	0.003	0.014	0.077			
Publication year	0.074	0.011	1.000	0.055	0.032	0.083
Constant	0.081	NA	1.000	1.267	1.135	0.264
Studies	61			61		
Observations	1,271			1,271		

Notes: Home = the coefficient estimated in a gravity equation on the dummy variable that equals one for within-country trade flows. PIP = posterior inclusion probability. SD = standard deviation. In the frequentist check we only include explanatory variables with PIP > 0.3. The standard errors in the frequentist check are clustered at both the study and data set level (the implementation of two-way clustering follows Cameron *et al.*, 2011). In this specification we use the beta-binomial prior advocated by Ley & Steel (2009) (the prior model probabilities are the same for all possible model sizes) and set Zellner's g prior following Fernandez *et al.* (2001). More details on the BMA estimation are available in Table 6 and Figure 2. A detailed description of all variables is available in Table 4.

Table 3: Robustness check—unweighted regressions

Response variable:	Bayesian model averaging			Frequentist check (OLS)		
Estimate of Home	Post. mean	Post. SD	PIP	Coef.	Std. er.	p-value
Data characteristics						
Mid-year of data	-0.025	0.003	1.000	-0.027	0.006	0.000
Panel data	0.215	0.165	0.695	0.283	0.155	0.069
Disaggregated	0.619	0.120	1.000	0.537	0.235	0.022
Obs. per year	0.060	0.054	0.617	0.105	0.127	0.407
No. of years	0.022	0.050	0.195			
Countries examined						
Canada	0.996	0.137	1.000	0.940	0.293	0.001
US	-1.655	0.181	1.000	-1.730	0.285	0.000
EU	-1.317	0.114	1.000	-1.313	0.258	0.000
OECD	-1.069	0.159	1.000	-1.062	0.263	0.000
Emerging	0.870	0.164	1.000	0.810	0.233	0.001
Design of the analysis						
No internal trade	1.239	0.164	1.000	1.128	0.283	0.000
Inconsistent dist	0.016	0.071	0.074			
Actual distance	-0.655	0.215	0.970	-0.722	0.301	0.016
Total trade	0.005	0.056	0.030			
Asymmetry	0.001	0.023	0.028			
Instruments	-0.007	0.055	0.038			
Treatment of multilateral resistance						
Remoteness	-0.001	0.028	0.026			
Country fixed eff.	-0.002	0.044	0.040			
Ratio estimation	0.035	0.111	0.125			
Anderson est.	0.001	0.039	0.026			
No control for MR	0.489	0.131	0.990	0.470	0.177	0.008
Treatment of zero trade flows						
Zero plus one	-0.686	0.181	0.986	-0.571	0.308	0.064
Tobit	-0.131	0.221	0.309	-0.436	0.252	0.084
PPML	-0.969	0.174	1.000	-1.024	0.388	0.008
Zeros omitted	-0.001	0.025	0.028			
Control variables						
Adjacency control	0.093	0.147	0.336	0.294	0.221	0.184
Language control	-0.001	0.021	0.029			
FTA control	-0.015	0.062	0.083			
Publication characteristics						
Published	-0.001	0.032	0.031			
Impact	-0.186	0.055	0.979	-0.188	0.125	0.131
Citations	0.182	0.047	0.992	0.173	0.106	0.103
Publication year	0.097	0.015	1.000	0.089	0.039	0.023
Constant	2.750	NA	1.000	2.678	0.974	0.006
Studies	61			61		
Observations	1,271			1,271		

Notes: Home = the coefficient estimated in a gravity equation on the dummy variable that equals one for within-country trade flows. PIP = posterior inclusion probability. SD = standard deviation. In the frequentist check we only include explanatory variables with PIP > 0.3. The standard errors in the frequentist check are clustered at both the study and data set level (the implementation of two-way clustering follows Cameron *et al.*, 2011). In this specification we do not weight the regressions by the inverse of the number of estimates reported per study. More details on the BMA estimation are available in Table 7 and Figure 3. A detailed description of all variables is available in Table 4.

Table 4: Description and summary statistics of country-level variables

Variable	Description	Mean	SD	WM
Relative size	The ratio of the GDP of the country for which the border effect is estimated to the average GDP of the other countries included in the estimation.	3.15	7.59	3.54
Tariffs	The average of effectively applied tariff rates by the country for which the border effect is estimated weighted by the product import shares corresponding to each partner country.	10.47	75.59	6.53
Non-tariff barriers	Cost to import excluding tariffs. All the costs associated with completing the procedures to import goods are included.	1.15	0.20	1.15
Financial dev.	Domestic credit to private sector relative to GDP.	1.07	0.41	1.01
ER volatility	Volatility of the exchange rate relative to USD (annual; in %).	9.97	8.19	8.27
Income dissimilarity	The absolute value of the distance between 1 and the ratio of GPD per capita (2011 constant $) of the country for which the border effect is estimated to the average GDP per capita of the other countries included in the estimation.	0.27	0.71	0.28
National pride	The percentage of answers "very proud" to the question: How proud are you of your country?	0.55	1.43	0.54
Internet usage	Fixed broadband subscriptions per 100 people.	2.52	4.37	3.18
Rule of law	The extent to which agents have confidence in the rules of society, and in particular the quality of contract enforcement (relative to those of the other countries included in the estimation).	1.02	0.13	0.99

Notes: SD = standard deviation. WM = mean weighted by the inverse of the number of estimates reported per study. Data are collected from World Value Survey (National pride) and the World Bank's Global Governance Indicators (Rule of law) and World Development Indicators (all other variables). When more countries are pooled in the estimation, we use an average of country-specific values weighted by each country's GDP. To each estimated border effect we assign the value of the country-specific variable that is the closest available to the year for which the border effect was estimated.

3 Criticisms of Meta-Analysis

1. *Studies of low quality should be excluded.* Our data set includes estimates from studies published in top journals, but also from studies not published in good outlets. As an alternative to meta-analysis, Slavin (1995) proposes "best evidence synthesis," which would only take into account good studies. The obvious problem is where to draw the line between good and bad ones. We prefer to include as many papers as possible and give weight to different aspects of study design according to what we believe is the consensus on best practice methodology. In this way we can explore the influence of different methods on the estimated border effects. We also control for the impact factor of the publication outlet and for the number of citations each study gets.

2. *The analysis omits some studies.* We try to include as many studies as possible, but may still miss some. To allow other researchers to replicate our analysis, we use the query described in Section 2 to search for studies estimating the border effect. We believe it is

not a problem to miss some studies, as long as their results do not differ systematically from the results of the studies included. With 1,271 estimates taken from 61 studies, our paper ranks among the largest meta-analyses conducted in economics (according to the survey by Doucouliagos & Stanley, 2013).

3. *Studies reporting many estimates dominate the meta-analysis.* When each estimate gets the same weight, the unbalanced nature of data in meta-analysis means that studies with many estimates drive the results. One remedy involves the mixed-effects multilevel model, which gives each study approximately the same weight if the within-study correlation of the estimates is large (Havranek & Irsova, 2011). The problem is that the method introduces study-level random effects, which may be correlated with explanatory variables. With so many explanatory variables defined at the study level, we prefer to simply weight the regressions by the inverse of the number of estimates reported per study.

4. *Authors' preferred estimates should get more weight.* Studies examining the border effect usually present many estimates, and often prefer a subset of these estimates (many results are shown as robustness checks). Some authors make it clear what their preference is, but for many studies it is impossible to select the preferred estimates. We control for data and methodology instead, which is easier to code and should capture most of the authors' preferences, for example, the control for multilateral resistance.

5. *Individual estimates are not independent, because authors use similar data.* Meta-analysis was originally designed for synthesizing medical research, where individual clinical trials can be considered approximately independent. In contrast, the regression results reported in economics are not independent, but neither are the observations in most economics data sets. To account for the dependence among observations we cluster the standard errors at the level of individual studies and data sets.

6. *Weighting by precision is inappropriate in economics because some methods underestimate standard errors.* Meta-analysts often use precision weights to remove heteroskedasticity in the regression estimating publication bias. We find no evidence of publication bias, so we can exclude the standard error from the equation and do not have to weight the estimates

by precision to yield efficiency. Section 3 also illustrates that weighting by precision has little effect on the estimated border effect.

7. *Standard errors are not exogenous to the estimated coefficients.* When the choice of method systematically affects both the magnitude of the estimated border effect and its standard error, the explanatory variable in (2) will be correlated with the error term. Our solution is to use the number of observations as an instrument for the standard error: studies with more observations yield more precise estimates, but the number of observations is little correlated with the choice of methodology.

8. *The analysis omits some factors that may cause heterogeneity in the reported estimates.* We collect 32 aspects of data, methodology, and studies that may affect the estimated border effects. More specifics of study design could be included: for example, the exact method for computing internal distance (we only include a dummy variable which equals one if the method differs from the computation of external distance); but we have to draw a line somewhere for the data collection to be feasible. Still, we collect more variables than most meta-analyses in economics. Nelson & Kennedy (2009) review 140 meta-analyses and report that a median analysis uses 12 explanatory variables; the largest meta-analysis has 41 variables.

9. *There are too many potential explanatory variables and it is not clear which should be included.* With so many aspects of study design one cannot find a theory that motivates the inclusion of all of them. For example, we would like to give more weight to large studies published in good journals, but it is not obvious why they should report systematically different results. We prefer to collect as many variables as possible and use Bayesian model averaging to resolve the resulting model uncertainty. The variables picked by BMA contain the ones that we feel should be included, such as the control for multilateral resistance and the measurement of internal distance.

10. *Meta-analysis compares apples with oranges.* Meta-analysis in economics examines heterogeneous estimates. Different estimates are produced using different methods, and we try to control for the differences in the design of primary studies. We also provide separate results for the regions examined in the literature. To increase the comparability of

the estimates in our data set, we choose to only include the results concerning the effect of *international* borders on trade and omit the large literature on intranational border barriers.

11. *Meta-analysis may disagree with large primary studies.* The major reason for conducting meta-analyses in medical science is to increase statistical power by combining small but costly clinical trials. Because individual clinical trials use similar methods, a comparison of a meta-analysis with a later, large clinical trial provides a viable test of the reliability of the meta-analysis. In economics the methods differ, and meta-analysis can be thought of as a weighted average of many different approaches. It would be difficult to construct a primary study reflecting all recent advances in the methodology of the gravity equation and all possible aspects of our definition of best practice. Moreover, the advantage of meta-analysis is the ability to evaluate the systematic effects of various method choices, and discuss the consequences of changing the definition of the best practice. Unlike primary studies, meta-analysis does not rely on a particular data set of trade flows. We can evaluate the mean effects of the choice of different methods (the effects may differ for different data sets, which is unobservable for the authors of primary studies), and thus hope to obtain more robust results.

12. *Mistakes in data coding are inevitable.* The collection of data for meta-analysis involves months of reading and coding the data. We do not use research assistants for this work, because it is too tempting to jump directly to regression tables and code the data without reading much of the primary studies. We cannot exclude errors, but we do our best to minimize their number by collecting the data independently and then comparing and correcting the data sets.

13. *Publication bias invalidates meta-analysis.* When researchers prefer to report estimates showing a particular sign or statistical significance, the mean reported estimate will get biased. We test for publication bias in Section 3 and find little evidence of preferential selection. When we correct for any potential publication bias, we obtain a border effect close to the simple mean and median estimates. In general the file drawer problem matters for any type of literature synthesis, but meta-analysis can correct for the bias.

4 Diagnostics of BMA

Table 5: Summary of BMA estimation, baseline specification

Mean no. regressors	*Draws*	*Burn-ins*	*Time*
18.5374	$2 \cdot 10^6$	$1 \cdot 10^6$	6.914583 minutes
No. models visited	*Modelspace*	*Visited*	*Topmodels*
$311,863$	$4.3 \cdot 10^9$	0.0073%	98%
Corr PMP	*No. Obs.*	*Model Prior*	*g-Prior*
0.9994	$1,271$	uniform	UIP
Shrinkage-Stats			
Av= 0.9992			

Notes: In this specification we employ the priors suggested by Eicher *et al.* (2011) based on predictive performance: the uniform model prior (each model has the same prior probability) and the unit information prior (the prior provides the same amount of information as one observation of data).

Figure 1: Model size and convergence, baseline specification

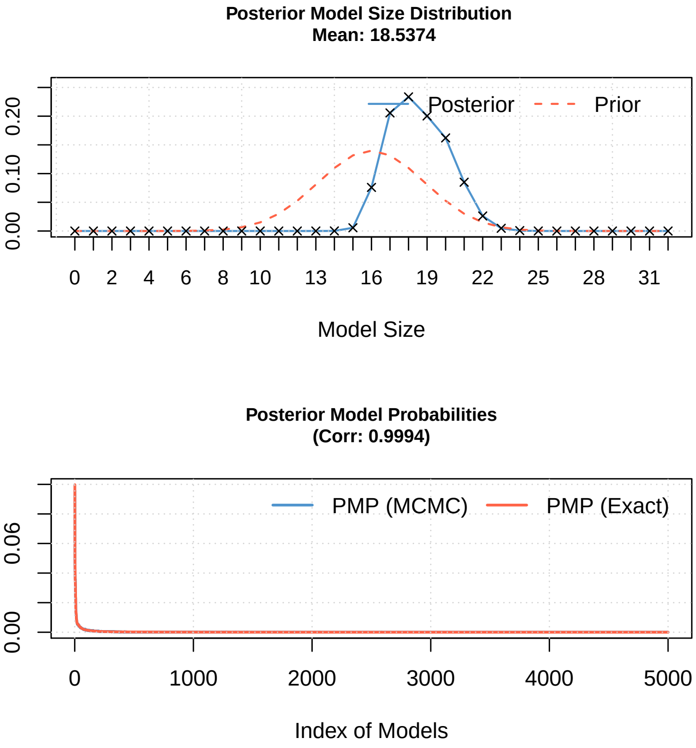


Table 6: Summary of BMA estimation, alternative priors

Mean no. regressors	*Draws*	*Burn-ins*	*Time*
19.6891	$2 \cdot 10^6$	$1 \cdot 10^6$	7.2395 minutes
No. models visited	*Modelspace*	*Visited*	*Topmodels*
394, 789	$4.3 \cdot 10^9$	0.0092%	96%
Corr PMP	*No. Obs.*	*Model Prior*	*g-Prior*
0.9993	1, 271	random	BRIC
Shrinkage-Stats			
Av= 0.9992			

Notes: The "random" model prior refers to the beta-binomial prior advocated by Ley & Steel (2009): the prior model probabilities are the same for all possible model sizes. In this specification we set Zellner's g prior following Fernandez *et al.* (2001).

Figure 2: Model size and convergence, alternative priors

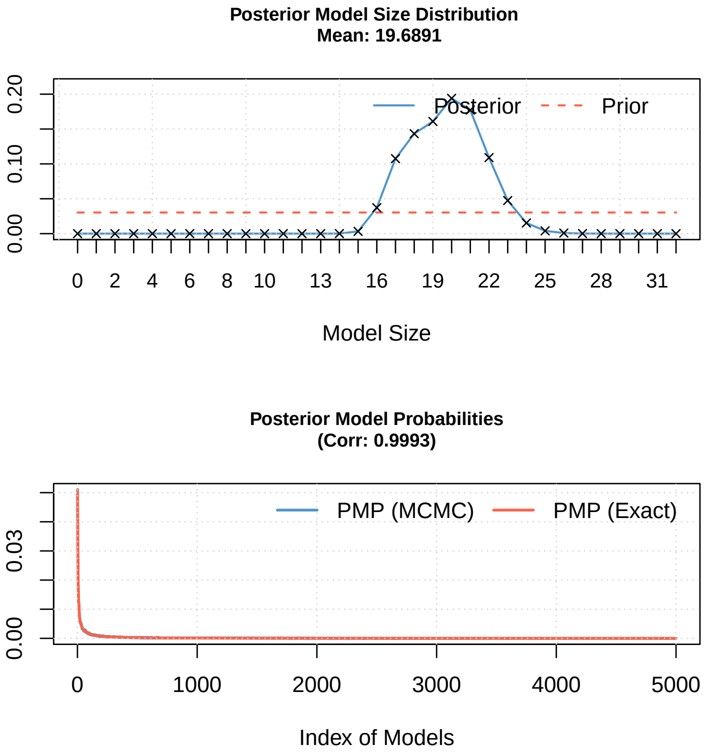
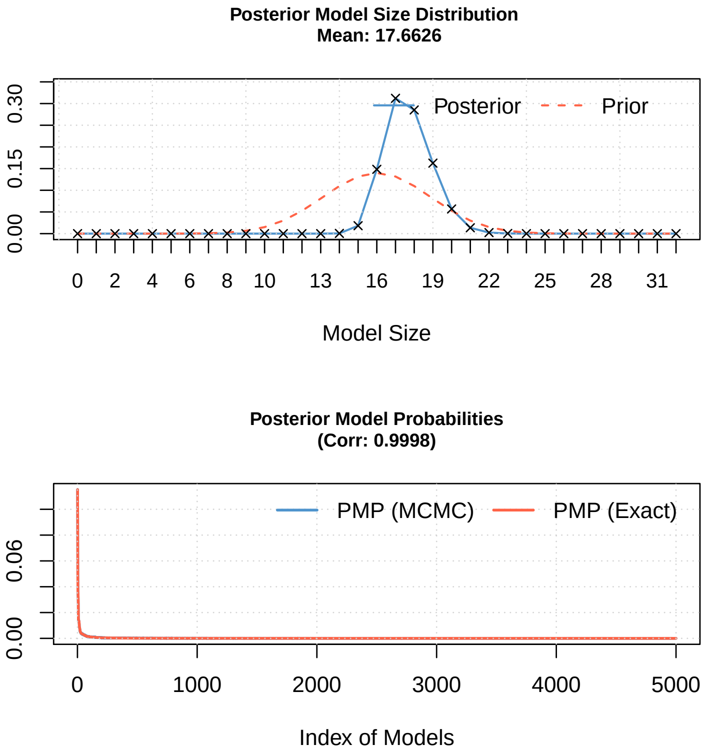


Table 7: Summary of BMA estimation, unweighted regressions

Mean no. regressors	*Draws*	*Burn-ins*	*Time*
17.6626	$2 \cdot 10^6$	$1 \cdot 10^6$	7.121633 minutes
No. models visited	*Modelspace*	*Visited*	*Topmodels*
350, 260	$4.3 \cdot 10^9$	0.0082%	98%
Corr PMP	*No. Obs.*	*Model Prior*	*g-Prior*
0.9998	1, 271	uniform	UIP
Shrinkage-Stats			
Av= 0.9992			

Notes: In this specification we employ the priors suggested by Eicher *et al.* (2011) based on predictive performance: the uniform model prior (each model has the same prior probability) and the unit information prior (the prior provides the same amount of information as one observation of data).

Figure 3: Model size and convergence, unweighted regressions



References

Cameron, A. C., J. B. Gelbach, & D. L. Miller (2011): "Robust Inference With Multiway Clustering." *Journal of Business & Economic Statistics* **29(2)**: pp. 238–249.

Card, D. & A. B. Krueger (1995): "Time-Series Minimum-Wage Studies: A Meta-Analysis." *American Economic Review* **85(2)**: pp. 238–43.

Doucouliagos, H. & T. D. Stanley (2013): "Are All Economic Facts Greatly Exaggerated? Theory Competition and Selectivity." *Journal of Economic Surveys* **27(2)**: pp. 316–339.

Eicher, T. S., C. Papageorgiou, & A. E. Raftery (2011): "Default Priors and Predictive Performance in Bayesian Model Averaging, with Application to Growth Determinants." *Journal of Applied Econometrics* **26(1)**: pp. 30–55.

Fernandez, C., E. Ley, & M. F. J. Steel (2001): "Benchmark priors for Bayesian model averaging." *Journal of Econometrics* **100(2)**: pp. 381–427.

Havranek, T. & Z. Irsova (2011): "Estimating Vertical Spillovers from FDI: Why Results Vary and What the True Effect Is." *Journal of International Economics* **85(2)**: pp. 234–244.

Ley, E. & M. F. Steel (2009): "On the effect of prior assumptions in Bayesian model averaging with applications to growth regressions." *Journal of Applied Econometrics* **24(4)**: pp. 651–674.

McCallum, J. (1995): "National Borders Matter: Canada-U.S. Regional Trade Patterns." *American Economic Review* **85(3)**: pp. 615–23.

Nelson, J. & P. Kennedy (2009): "The Use (and Abuse) of Meta-Analysis in Environmental and Natural Resource Economics: An Assessment." *Environmental & Resource Economics* **42(3)**: pp. 345–377.

Slavin, R. E. (1995): "Best evidence synthesis: an intelligent alternative to meta-analysis." *Journal of Clinical Epidemiology* **48(1)**: pp. 9–18.

Stanley, T. D. (2005): "Beyond Publication Bias." *Journal of Economic Surveys* **19(3)**: pp. 309–345.

Stanley, T. D. (2008): "Meta-Regression Methods for Detecting and Estimating Empirical Effects in the Presence of Publication Selection." *Oxford Bulletin of Economics and Statistics* **70(1)**: pp. 103–127.

Studies Included in the Meta-Analysis

Anderson, J. E., C. A. Milot, & Y. V. Yotov (2014): "How Much Does Geography Deflect Services Trade? Canadian Answers." *International Economic Review* **55(3)**: pp. 791–818.

Anderson, J. E. & E. van Wincoop (2003): "Gravity with Gravitas: A Solution to the Border Puzzle." *American Economic Review* **93(1)**: pp. 170–192.

Anderson, J. E. & Y. V. Yotov (2010a): "Specialization: Pro- and Anti-globalizing, 1990-2002." *NBER Working Papers 16301*, National Bureau of Economic Research, Inc.

Anderson, J. E. & Y. V. Yotov (2010b): "The Changing Incidence of Geography." *American Economic Review* **100(5)**: pp. 2157–86.

Anderson, M. & S. Smith (1999a): "Canadian Provinces in World Trade: Engagement and Detachment." *Canadian Journal of Economics* **32(1)**: pp. 22–38.

ANDERSON, M. A. & S. L. S. SMITH (1999b): "Do National Borders Really Matter? Canada-US Regional Trade Reconsidered." *Review of International Economics* **7(2)**: pp. 219–27.

BAIER, S. L. & J. H. BERGSTRAND (2009): "Bonus vetus OLS: A simple method for approximating international trade-cost effects using the gravity equation." *Journal of International Economics* **77(1)**: pp. 77–85.

BALISTRERI, E. J. & R. H. HILLBERRY (2007): "Structural estimation and the border puzzle." *Journal of International Economics* **72(2)**: pp. 451–463.

BAS, M. & I. LEDEZMA (2009): "Trade integration in manufacturing: the Chilean experience." *The EU and Emerging Markets* **12**: pp. 167–186.

BEHRENS, K., C. ERTUR, & W. KOCH (2012a): "'Dual' Gravity: Using Spatial Econometrics To Control For Multilateral Resistance." *Journal of Applied Econometrics* **27(5)**: pp. 773–794.

BEHRENS, K., G. MION, Y. MURATA, & J. SUEDEKUM (2012b): "Trade, Wages, and Productivity." *CESifo Working Paper Series 4011*, CESifo Group Munich.

BENZ, S. (2013): "Gravity with Google Maps: the Border Puzzle Revisited." *Economics Bulletin* **33(3)**: pp. 2414–2421.

BERGSTRAND, J. H., P. EGGER, & M. LARCH (2013): "Gravity Redux: Estimation of gravity-equation coefficients, elasticities of substitution, and general equilibrium comparative statics under asymmetric bilateral trade costs." *Journal of International Economics* **89(1)**: pp. 110–121.

BRACONIER, H. & M. PISU (2013): "Road Connectivity and the Border Effect: Evidence from Europe." *OECD Economics Department Working Papers 1073*, OECD.

BRÜLHART, M. & F. TRIONFETTI (2009): "A test of trade theories when expenditure is home biased." *European Economic Review* **53(7)**: pp. 830–845.

CAFISO, G. (2011a): "Rose effect versus border effect: the Euro's impact on trade." *Applied Economics* **43(13)**: pp. 1691–1702.

CAFISO, G. (2011b): "Sectoral border effects and the geographic concentration of production." *Review of World Economics* **147(3)**: pp. 543–566.

CEGLOWSKI, J. (2000): "Has the Border Narrowed?" *The North American Journal of Economics and Finance* **11(1)**: pp. 61–75.

CHEN, N. (2004): "Intra-national versus international trade in the European Union: why do national borders matter?" *Journal of International Economics* **63(1)**: pp. 93–118.

COUGHLIN, C. C. & D. NOVY (2013): "Is the International Border Effect Larger than the Domestic Border Effect? Evidence from US Trade." *CESifo Economic Studies* **59(2)**: pp. 249–276.

DEALMEIDA, F. M. & O. M. DA SILVA (2007): "Comércio e integracao dos estados Brasileiros." *Revista de Economia e Agronegocio* **5(4)**: pp. 487–502.

EVANS, C. L. (2003): "The Economic Significance of National Border Effects." *American Economic Review* **93(4)**: pp. 1291–1312.

EVANS, C. L. (2006): "Border effects and the availability of domestic products abroad." *Canadian Journal of Economics* **39(1)**: pp. 211–246.

FEENSTRA, R. C. (2002): "Border Effects and the Gravity Equation: Consistent Methods for Estimation." *Scottish Journal of Political Economy* **49(5)**: pp. 491–506.

FONTAGNÉ, L., T. MAYER, & S. ZIGNAGO (2005): "Trade in the Triad: how easy is the access to large markets?" *Canadian Journal of Economics* **38(4)**: pp. 1401–1430.

Fukao, K. & T. Okubo (2011): "Why Has the Border Effect in the Japanese Machinery Sectors Declined?: The Role of Business Networks in East Asian Machinery Trade." *Journal of Economic Integration* **26**: pp. 651–671.

Gallego, N. & C. Llano (2011): "Revisiting the border effect in Europe and Spain: New results using region-to-region intra-national and inter-national flows." *Working paper*, Universidad Autónoma de Madrid.

Ghemawat, P., C. Llano, & F. Requena (2010): "Competitiveness and interregional as well as international trade: The case of Catalonia." *International Journal of Industrial Organization* **28(4)**: pp. 415–422.

Gil-Pareja, S., R. Llorca-Vivero, J. A. Martínez-Serrano, & J. Oliver-Alonso (2005): "The Border Effect in Spain." *The World Economy* **28(11)**: pp. 1617–1631.

Gomez, E., B. Martens, & G. Turlea (2013): "The Drivers and Impediments for Cross-border e-Commerce in the EU." *Working paper*, Institute for Prospective Technological Studies.

Head, K. & T. Mayer (2000): "Non-Europe: The magnitude and causes of market fragmentation in the EU." *Review of World Economics* **136(2)**: pp. 284–314.

Head, K. & T. Mayer (2010): "Illusory Border Effects: Distance Mismeasurement Inflates Estimates of Home Bias in Trade." In P. A. G. van Bergeijk & S. Brakman (editors), "The Gravity Model in International Trade: Advances and Applications," pp. 165–192. Cambridge University Press.

Heinemeyer, H. C. (2007): "The treatment effect of borders on trade: The great war and the disintegration of Central Europe." *Cliometrica* **1(3)**: pp. 177–210.

Helble, M. (2007): "Border Effect Estimates for France and Germany Combining International Trade and Intranational Transport Flows." *Review of World Economics* **143(3)**: pp. 433–463.

Helliwell, J. F. (1996): "Do National Borders Matter for Quebec's Trade?" *Canadian Journal of Economics* **29(3)**: pp. 507–22.

Helliwell, J. F. (1997): "National Borders, Trade and Migration." *Pacific Economic Review* **2(3)**: p. 165–185.

Helliwell, J. F. & G. Verdier (2001): "Measuring internal trade distances: a new method applied to estimate provincial border effects in Canada." *Canadian Journal of Economics* **34(4)**: pp. 1024–1041.

Hillberry, R. H. (1998): "Regional trade and "the medicine line": The national border effect in U.S. commodity flow data." *Journal of Borderlands Studies* **13(2)**: pp. 1–17.

Hillberry, R. H. (2002): "Aggregation bias, compositional change, and the border effect." *Canadian Journal of Economics* **35(3)**: pp. 517–530.

Llano, C., A. Minondo, & F. Requena (2011): "Is the Border Effect an Artefact of Geographical Aggregation?" *The World Economy* **34(10)**: pp. 1771–1787.

McCallum, J. (1995): "National Borders Matter: Canada-U.S. Regional Trade Patterns." *American Economic Review* **85(3)**: pp. 615–23.

Minondo, A. (2007): "The disappearance of the border barrier in some European Union countries' bilateral trade." *Applied Economics* **39(1)**: pp. 119–124.

Nitsch, V. (2000): "National borders and international trade: evidence from the European Union." *Canadian Journal of Economics* **33(4)**: pp. 1091–1105.

Nitsch, V. (2001): “It’s Not Right But It’s Okay: On the Measurement of Intra- and International Trade Distances.” *Working paper*, Bankgesellschaft Berlin.

Okubo, T. (2004): “The border effect in the Japanese market: A Gravity Model analysis.” *Journal of the Japanese and International Economies* **18(1)**: pp. 1–11.

Okubo, T. (2007): “Trade bloc formation in inter-war Japan: A gravity model analysis.” *Journal of the Japanese and International Economies* **21(2)**: pp. 214–236.

Olper, A. & V. Raimondi (2008a): “Agricultural market integration in the OECD: A gravity-border effect approach.” *Food Policy* **33(2)**: pp. 165–175.

Olper, A. & V. Raimondi (2008b): “Explaining National Border Effects in the QUAD Food Trade.” *Journal of Agricultural Economics* **59(3)**: pp. 436–462.

Olper, A. & V. Raimondi (2008c): “Market Access Asymmetry in Food Trade.” *Review of World Economics* **144(3)**: pp. 509–537.

Pacchioli, C. (2011): “Is the EU internal market suffering from an integration deficit? Estimating the ‘home-bias effect’.” *Working paper*, Centre for European Policy Studies.

Requena, F. & C. Llano (2010): “The border effects in Spain: an industry-level analysis.” *Empirica* **37(4)**: pp. 455–476.

da Silva, O. M., F. M. de Almeida, & B. M. de Oliviera (2007): “Comércio internacional ”x” intranacional no Brasil: medindo o efeito-fronteira.” *Nova Economia* **17(3)**: pp. 427–439.

de Sousa, J., T. Mayer, & S. Zignago (2012): “Market access in global and regional trade.” *Regional Science and Urban Economics* **42(6)**: pp. 1037–1052.

Straathof, B. (2008): “Gravity with gravitas: comment.” *CPB Discussion Paper 111*, CPB Netherlands Bureau for Economic Policy Analysis.

Taglioni, D. (2004): *Monetary union, exchange rate variability and trade.* Ph.D. thesis, University of Geneve.

Tan, S. (2012): “Structural Estimation of a Flexible Translog Gravity Model.” *Department of Economics - Working Papers Series 1164*, The University of Melbourne.

Vancauteren, M. (2002): “The Impact of Technical Barriers to Trade on Home Bias: An application to EU data.” *Discussion Papers 2002032*, Université catholique de Louvain, Institut de Recherches Economiques et Sociales.

Vancauteren, M. & D. Weiserbs (2011): “Intra-European Trade of Manufacturing Goods: An Extension of the Gravity Model.” *International Econometric Review* **3(1)**: pp. 1–24.

Wall, H. J. (2000): “Gravity model specification and the effects of the Canada-U.S. border.” *Working Papers 2000-024*, Federal Reserve Bank of St. Louis.

Wei, S.-J. (1996): “Intra-National versus International Trade: How Stubborn are Nations in Global Integration?” *NBER Working Papers 5531*, National Bureau of Economic Research, Inc.

Wolf, N. (2009): “Was Germany Ever United? Evidence from Intra- and International Trade, 1885–1933.” *The Journal of Economic History* **69(03)**: pp. 846–881.