Estimating the Armington Elasticity: The Importance of Data Choice and Publication Bias

Bajzik, Jozef and Havranek, Tomas and Irsova, Zuzana and Schwarz, Jiri

Charles University, Prague, Czech National Bank

12 July 2019

Estimating the Armington Elasticity: The Importance of Data Choice and Publication Bias*

Josef Bajzik[a,b], Tomas Havranek[a], Zuzana Irsova[a], and Jiri Schwarz[a,b]

[a]Charles University, Prague

[b]Czech National Bank

July 12, 2019

Abstract

A key parameter in international economics is the elasticity of substitution between domestic and foreign goods, also called the Armington elasticity. Yet estimates vary widely. We collect 3,524 reported estimates of the elasticity, construct 34 variables that reflect the context in which researchers obtain their estimates, and examine what drives the heterogeneity in the results. To account for inherent model uncertainty, we employ Bayesian and frequentist model averaging. We present the first application of newly developed non-linear techniques to correct for publication bias. Our main results are threefold. First, there is publication bias against small and statistically insignificant elasticities. Second, differences in results are best explained by differences in data: aggregation, frequency, size, and dimension. Third, the mean elasticity implied by the literature after correcting for both publication bias and potential misspecifications is 3.

Keywords: Armington, trade elasticity, meta-analysis, publication bias, Bayesian model averaging

JEL Codes: C83, D12, F14

*An online appendix with data and codes is available at `meta-analysis.cz/armington`. Corresponding author: Zuzana Irsova, `zuzana.irsova@ies-prague.org`. The views expressed here are ours and not necessarily those of the Czech National Bank.

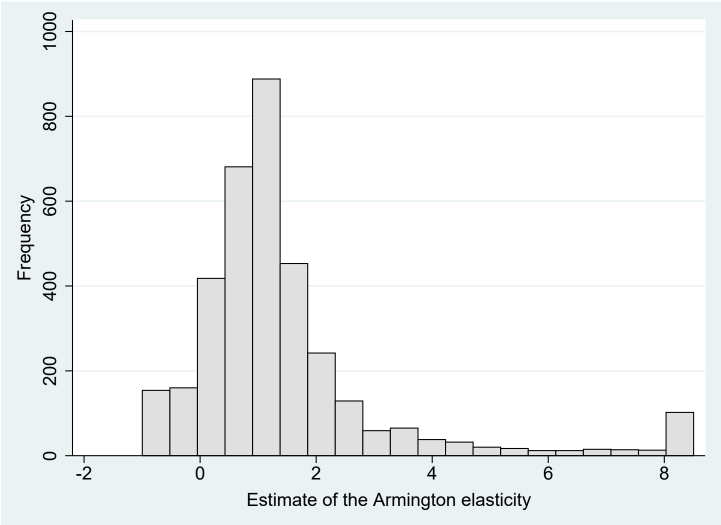
1 Introduction

How does the demand for domestic versus foreign goods react to a change in relative prices? The answer is central to a host of research and policy problems in international trade and macroeconomics: the welfare effects of globalization (Costinot & Rodriguez-Clare, 2014), trade balance adjustments (Imbs & Mejean, 2015), and the exchange rate pass-through of monetary policy (Auer & Schoenle, 2016), to name but a few. Any attempt to evaluate the effect of tariffs in particular depends crucially on the assumed reaction of relative demand to relative prices. In most models, the reaction is governed by the (constant) elasticity of substitution between domestic and foreign goods. The size of the elasticity used for calibration often drives the conclusions of the model, as shown by Schurenberg-Frosch (2015), who recomputes the results of 50 previously published models using different values of the elasticity. She finds that, with plausible changes in the elasticity, the results change qualitatively in more than half of the cases. As Hillberry & Hummels (2013, p. 1217) put it, "it is no exaggeration to say that [*the elasticity*] is the most important parameter in modern trade theory."

Yet no consensus on the magnitude of the elasticity exists. In different contexts, researchers tend to obtain substantially different estimates, as observed by Feenstra *et al.* (2018) and many commentators before them. In this paper we assign a pattern to these differences, a pattern that we hope will be useful for calibrating models in international trade and macroeconomics. The elasticity of substitution between domestic and foreign goods is commonly called the Armington elasticity, in honor of Armington (1969), who first formulated a theoretical model featuring goods distinguished solely by the place of origin. The first estimates of the elasticity followed soon afterward, and many thousand have been published since. As the Armington-style literature turns 50, the time is ripe for taking stock. We collect 3,524 estimates of the elasticity of substitution between domestic and foreign goods and construct 34 variables that reflect the context in which researchers produce their estimates.

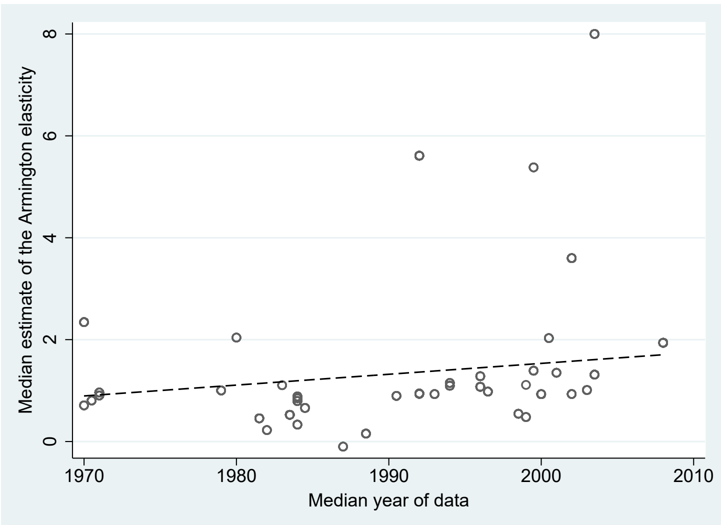
A bird's-eye view of the literature (Figure 1 and Figure 2) shows four stylized facts, three of which corroborate the common knowledge in the field. First, the estimates of the elasticities vary substantially. A researcher wishing to calibrate her policy model has plenty of degrees of freedom; she can easily find empirical evidence for any value of the elasticity between 0 and 8. Such plausible (that is, justifiable by some empirical evidence) changes in the elasticity can

Figure 1: The reported elasticities are often around 1 but can vary widely



Notes: The figure shows the histogram of the estimates of the macro-level Armington elasticity reported in individual studies. Large values are winsorized for ease of exposition.

Figure 2: The mean and variance of reported elasticities increase over time



Notes: The vertical axis measures median estimates of the macro-level Armington elasticity reported in individual studies. The horizontal axis measures the median year of the data used in the corresponding study.

have decisive effects on the results of the model. For example, Engler & Tervala (2016) show that changing the elasticity from 3 to 8 more than doubles the estimated welfare gains from the Transatlantic Trade and Investment Partnership. Second, the median estimated elasticity in the literature is 1, and many estimates are close to that value. Third, the reported elasticity seems to be increasing in time, but it is not clear whether the apparent trend reflects fundamental changes in preferences or improved data and techniques used by more recent studies.

Finally, the fourth stylized fact is that newer studies show more disagreement on the value of the elasticity of substitution. That is, instead of converging to a consensus value, the literature diverges. The increased variance in the estimated elasticities provides additional rationale for a systematic evaluation of the published results. For this evaluation we use the methods of meta-analysis, which were originally developed in (or inspired by) medical research. Recent applications of meta-analysis in economics include Card *et al.* (2018) on the effectiveness of active labor market programs, Anderson *et al.* (2018) on the impact of government spending on poverty, and Havranek & Irsova (2017) on the border effect in international trade. An important problem inherent in meta-analysis is model uncertainty because for many control variables capturing the study design, little theory exists that can help us determine whether they should be included in the baseline model. To address this issue, we use both Bayesian (Raftery *et al.*, 1997; Eicher *et al.*, 2011) and frequentist (Hansen, 2007; Amini & Parmeter, 2012) methods of model averaging (Steel, 2019, provides an excellent description of these techniques).

Meta-analysis also allows us to correct for potential publication bias in the literature. Publication bias arises when, holding other aspects of study design constant, some results (for example, those that are statistically insignificant at standard levels or have the "wrong" sign) have a lower probability of publication than other results (Stanley, 2001). For example, in the context of the elasticity of substitution, it is safe to assume that its sign is positive: a negative value is not compatible with any commonly applied model of preferences. Similarly, it is difficult to interpret a zero elasticity. Thus, from the point of view of an individual study, it makes sense not to report such unintuitive estimates—or find a specification where the elasticity is positive—because non-positive elasticity suggests that something is wrong with the data or the estimation technique. Nevertheless, non-positive estimates will occur from time to time simply because of sampling error; for the same reason, researchers will sometimes obtain estimates much larger than the true value. If large estimates (which are still intuitive) are kept but non-positive ones are omitted, an upward bias arises. Paradoxically, publication bias can thus improve inferences drawn from individual studies (if they avoid making central conclusions based on negative or zero elasticities) but inevitably bias inferences drawn from the literature as a whole. Ioannidis *et al.* (2017) shows that, in economics, the effects of publication selection are dramatic and exaggerate the mean reported estimate twofold.

To correct for publication bias, we use meta-regression techniques based on Egger *et al.* (1997) and their extensions and three new non-linear techniques developed specifically for meta-analysis in economics. The first one is due to Ioannidis *et al.* (2017) and relies on estimates that are adequately powered. The second technique was developed by Andrews & Kasy (2019) and employs a selection model that estimates the probability of publication for results with different p-values. The third non-linear technique is the so-called stem-based method by Furukawa (2019), a non-parametric estimator that exploits the variance-bias trade-off. As far as we know, the latter two estimators have not been applied so far apart from illustrative examples outlined by Andrews & Kasy (2019) and Furukawa (2019).

In all the models we run, linear or non-linear, Bayesian or frequentist, we find evidence of strong publication bias in the estimates of the long-run Armington elasticity. The bias results in an exaggeration of the mean estimate by more than 50%. In contrast, we find no publication bias among the estimates of the short-run elasticity. One explanation consistent with these results is that the short-run elasticity is commonly believed to be small and less important for policy questions, so there are few incentives to discriminate against insignificant (and even potentially negative) estimates of the elasticity. Large estimates of the long-run elasticity, in contrast, appear intuitive and desirable to many researchers (see, for example, the discussion in McDaniel & Balistreri, 2003; Hillberry *et al.*, 2005).

Our findings indicate that the study characteristics are systematically associated with the reported results. Among the 34 variables we construct, the most important are the ones related to the data used in the estimation. We find that, ceteris paribus, using more aggregated data yields smaller estimates of the elasticity. Annual data bring substantially smaller elasticities compared to monthly and quarterly data. If a study uses cross-sectional data, it is more likely to report larger estimates of the elasticity than if time-series data are used. Our results also suggest that employing a small number of observations and ignoring endogeneity in the estimation yields a downward bias. Finally, we find systematic correlation between measures of quality and the magnitude of the reported elasticity. Studies of higher quality (as measured by the number of citations, publication in a refereed journal, and the RePEc impact factor of the outlet) tend to report larger estimates.

Therefore, while publication selection creates an upward bias, many questionable method choices seem to create a downward bias. We exploit the relationships unearthed by Bayesian model averaging to compute a mean effect corrected for publication bias, misspecification biases, and conditional on the maximum quality defined based on the peer-review status, the publication outlet, and the number of citations. The resulting elasticity reaches 3, and we interpret the number as our best guess (based on the available empirical literature published during the last five decades) for how to calibrate a model that allows for only one parameter to govern the aggregate elasticity of substitution between domestic and foreign goods—for example, an open economy dynamic stochastic general equilibrium model of the type used in many central banks. We also report these aggregate elasticities for individual countries and provide information in the online appendix that allows other researchers to use our data to compute the elasticities for individual industries.

The remainder of the paper is structured as follows. Section 2 briefly describes how the Armington elasticity is estimated and how we collect data from primary studies. Section 3 tests for publication bias in the literature. Section 4 explores heterogeneity and computes the aggregate elasticity corrected for publication and misspecification biases. Section 5 concludes the paper. An online appendix at meta-analysis.cz/armington provides the data and codes.

2 Collecting the Elasticity Dataset

The derivation of the Armington elasticity follows a two-stage optimization process (please refer to Hillberry & Hummels, 2013; Feenstra *et al.*, 2018, for a more detailed treatment than we have the space to offer here): in the first stage, the consumer with a CESutility function $u(Q_D, Q_M) = \left(\beta \cdot Q_D^{(\sigma-1)/\sigma} + (1-\beta) \cdot Q_M^{(\sigma-1)/\sigma}\right)^{(\sigma/(\sigma-1)}$ allocates her total spending to various product categories following her budget constraint with a given general price index. The consumer thus chooses a quantity of the composite good $Q_D + Q_M$, her aggregate demand for goods produced in her home country (D) and foreign countries (M). In the second stage, the consumer decides what proportion of domestic and foreign goods to consume while minimizing her expenditures $Q_D \cdot P_D + Q_M \cdot P_M$ or maximizing her utility. Utility maximization subject to the budget constraint or cost minimization subject to the utility function both imply that the marginal rate of substitution between domestic and foreign goods should equal the corresponding price ratio (Welsch, 2008). The first-order condition follows:

$$\frac{Q_M}{Q_D} = \left[\frac{\beta}{1-\beta} \cdot \frac{P_D}{P_M}\right]^{\sigma}, \tag{1}$$

where the quantity of domestic goods Q_D and foreign goods Q_M is related to the corresponding domestic price P_D and import price P_M. β is a distribution parameter between the domestic and the foreign good, and σ denotes the Armington elasticity. For estimation, the first-order condition is commonly log-linearized:

$$\log\left(\frac{Q_M}{Q_D}\right) = \underbrace{\sigma \log\left(\frac{\beta}{1-\beta}\right)}_{\text{Constant}} + \underbrace{\sigma}_{\text{Armington elasticity}} \log\left(\frac{P_D}{P_M}\right) + e. \tag{2}$$

As the main building block of our dataset, we collect estimates of σ from the literature. Several recent papers, such as Aspalter (2016) or Feenstra *et al.* (2018), call this type of Armington elasticity a *macro-elasticity*. A macro-elasticity governs the substitution between home and foreign goods, where varieties from different foreign countries are aggregated into one composite good. A *micro-elasticity*, on the other hand, governs the substitution among the varieties of foreign goods and thus differentiates among the specific countries of origin (Balistreri *et al.*, 2010). For comparability, in this paper, we focus on macro-elasticities.

We need each study to report a measure of uncertainty of its estimates. Such a measure, which is necessary to test for the potential presence of publication bias in the literature, can be either the standard error or other metrics recomputable to the standard error. This requirement prevents us from using a dozen empirical papers, including the highly cited contribution by Broda & Weinstein (2006). For similar reasons, we drop a few estimates for which uncertainty measures are incorrectly reported (for example, when the reported standard errors are negative or when the reported confidence intervals do not include the point estimate). The final dataset is an unbalanced one because some studies report more estimates than other studies. We

Table 1: Studies included in the meta-analysis

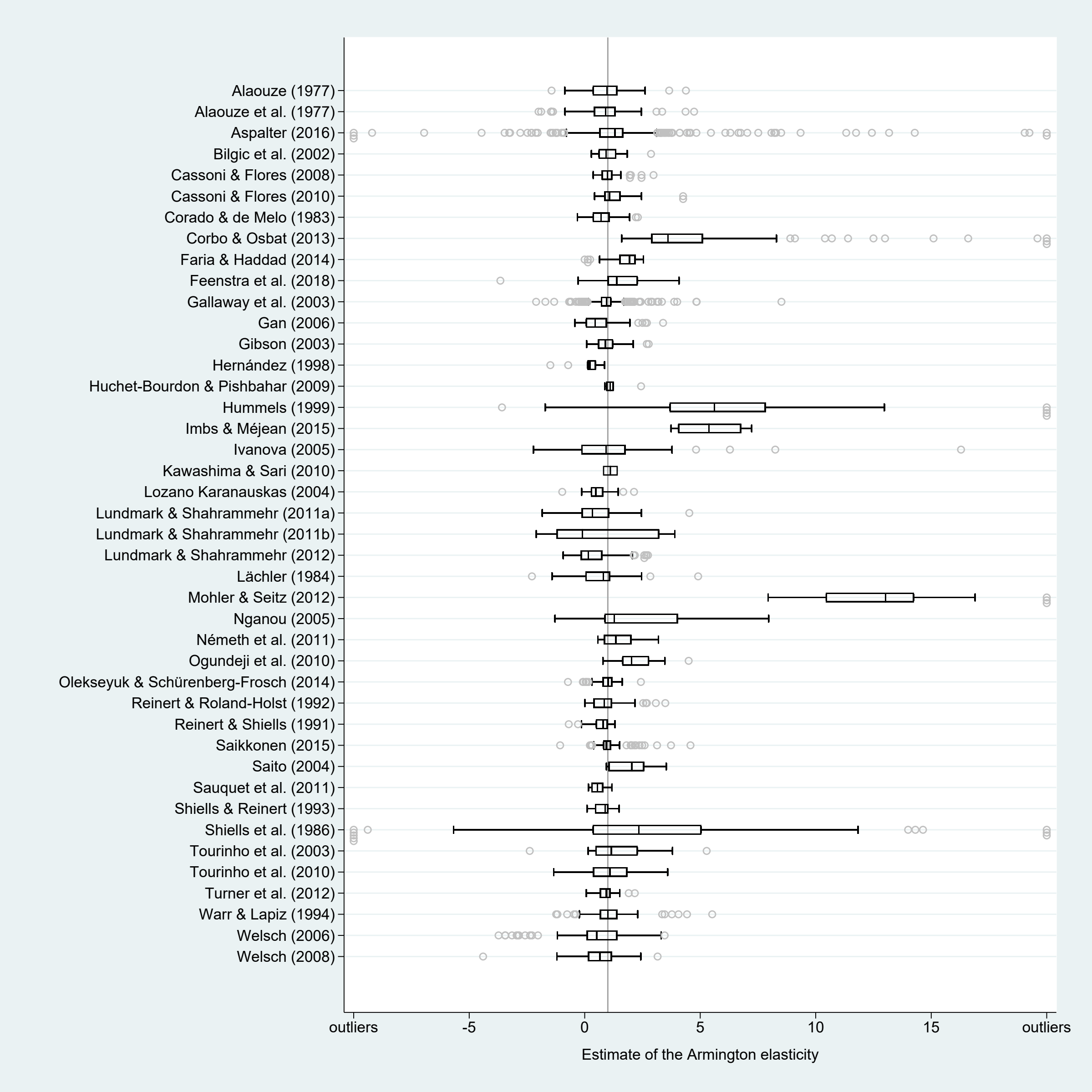
Alaouze (1977)	Huchet-Bourdon & Pishbahar (2009)	Olekseyuk & Schurenberg-Frosch (2016)
Alaouze *et al.* (1977)	Hummels (1999)	Reinert & Roland-Holst (1992)
Aspalter (2016)	Ivanova (2005)	Reinert & Shiells (1991)
Bilgic *et al.* (2002)	Kawashima & Sari (2010)	Saikkonen (2015)
Cassoni & Flores (2008)	Lachler (1984)	Saito (2004)
Cassoni & Flores (2010)	Lozano Karanauskas (2004)	Sauquet *et al.* (2011)
Corado & de Melo (1983)	Lundmark & Shahrammehr (2011a)	Shiells & Reinert (1993)
Corbo & Osbat (2013)	Lundmark & Shahrammehr (2011b)	Shiells *et al.* (1986)
Faria & Haddad (2014)	Lundmark & Shahrammehr (2012)	Tourinho *et al.* (2003)
Feenstra *et al.* (2018)	Imbs & Mejean (2015)	Tourinho *et al.* (2010)
Gallaway *et al.* (2003)	Mohler & Seitz (2012)	Turner *et al.* (2012)
Gan (2006)	Nemeth *et al.* (2011)	Warr & Lapiz (1994)
Gibson (2003)	Nganou (2005)	Welsch (2006)
Hernandez (1998)	Ogundeji *et al.* (2010)	Welsch (2008)

choose to include all the reported estimates because it is often unclear which estimate is the one preferred by the author; moreover, including more estimates obtained using alternative methods or datasets increases the variation we can exploit by meta-analysis.

The first step in a meta-analysis is the search for relevant studies. Building on the comprehensive surveys by McDaniel & Balistreri (2003) and Cassoni & Flores (2008), we design our search query in Google Scholar in a way that shows the well-known studies estimating the Armington elasticity among the first hits. The final query along with the dataset is available online at meta-analysis.cz/armington. We also go through the references of the most recent studies and obtain other papers that might provide empirical estimates of the elasticity. While the keywords we use are specified in English, we do not exclude any study based on the language of publication: several papers written in Spanish (e.g. Hernandez, 1998; Lozano Karanauskas, 2004) and Portuguese (Faria & Haddad, 2014) are included. We add the last study in March 2018 and terminate the literature search. The final set of studies that fulfill all requirements for meta-analysis is reported in Table 1; our sample consists of 3,524 estimates from 42 papers.

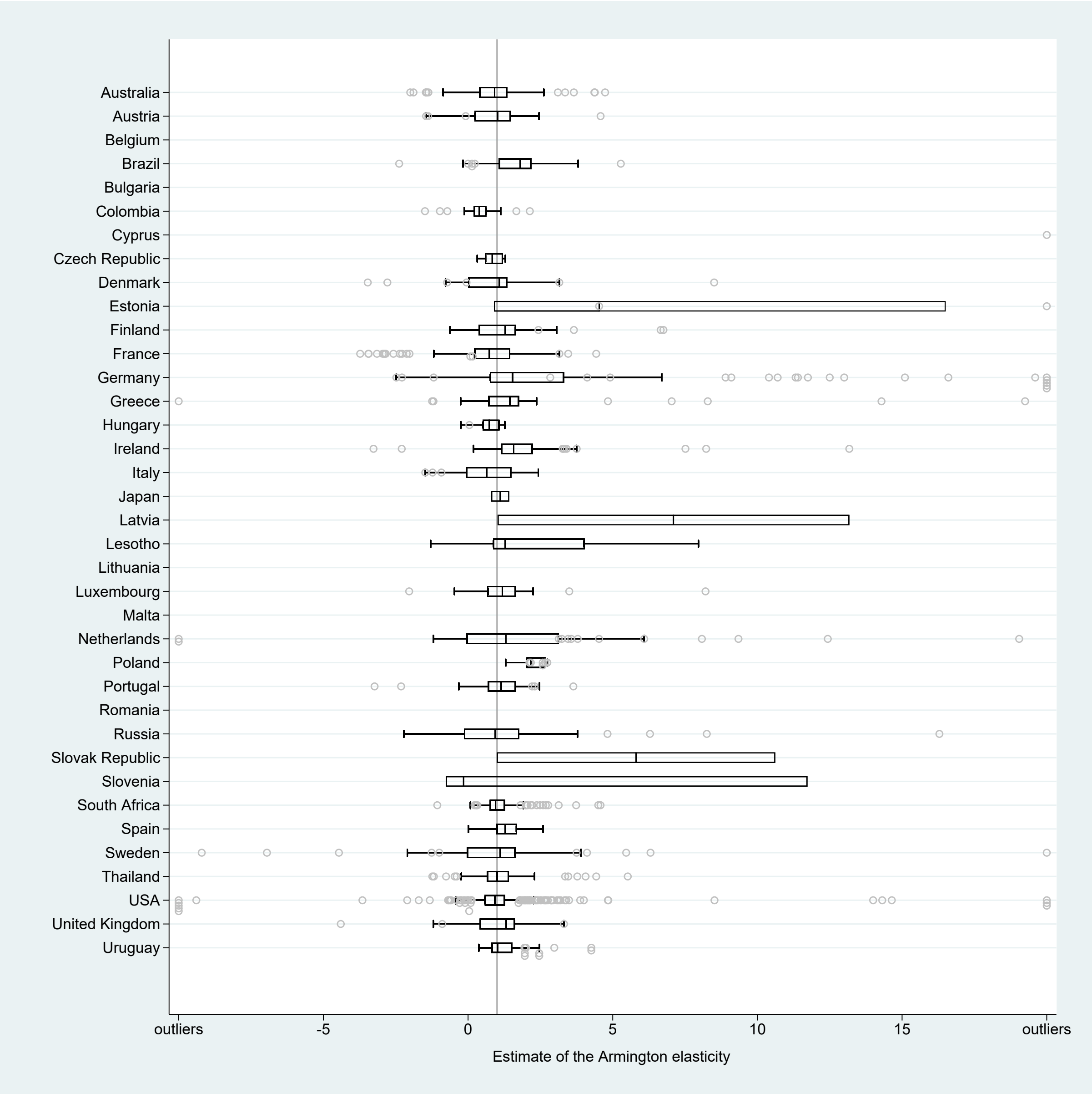
The oldest study in our sample was published in 1977 and the most recent one in 2018, thereby covering more than 40 years of research. The mean reported elasticity is 1.5. Given that there are a few dramatic outliers in our data (their values climb to approximately 50 in absolute value), we winsorize the estimates at the 2.5% level; the mean is not affected by winsorization, and our results hold with alternative winsorizations at the 1% and 5% levels. Approximately 10% of the estimates are negative and commonly believed to occur due to misspecifications in the demand function and problems with import prices (Shiells *et al.*, 1986). More than half

Figure 3: Estimates vary both within and across studies



Notes: The figure shows a box plot of the estimates of the Armington elasticity reported in individual studies. The length of each box represents the interquartile range (P25-P75), and the dividing line inside the box is the median value. The whiskers represent the highest and lowest data points within 1.5 times the range between the upper and lower quartiles. The dots show the outlying estimates with extreme values stacked at the values denoted as 'outliers.' The solid vertical line denotes unity.

Figure 4: Estimates vary both within and across countries



Notes: The figure shows a box plot of the estimates of the Armington elasticity reported for individual countries. The length of each box represents the interquartile range (P25-P75), and the dividing line inside the box is the median value. The whiskers represent the highest and lowest data points within 1.5 times the range between the upper and lower quartiles. The dots show the outlying estimates with extreme values stacked at the values denoted as 'outliers.' The solid vertical line denotes unity.

of the estimates are larger than unity, which suggests that domestic and foreign goods can often be expected to form gross substitutes. Nevertheless, estimates differ greatly both within and between individual studies and home countries, as Figure 3 and Figure 4 demonstrate. To assign a pattern to this variance, for each estimate, we collect 43 explanatory variables describing various characteristics of data, home countries, methods, models, and quality; these sources of heterogeneity are examined in detail in Section 4.

Table 2: Armington elasticities for different subsets of data

		Unweighted			Weighted		
	No. of obs.	Mean	95% conf. int.		Mean	95% conf. int.	
Temporal dynamics							
Short-run effect	556	0.88	0.83	0.93	0.91	0.85	0.98
Long-run effect	2,968	1.56	1.49	1.63	1.74	1.65	1.82
Data characteristics							
Monthly data	488	1.04	0.97	1.11	1.18	1.12	1.24
Quarterly data	745	1.22	1.09	1.34	2.64	2.41	2.87
Annual data	2,291	1.62	1.54	1.70	1.32	1.25	1.40
Structural variation							
Primary sector	366	0.83	0.70	0.95	0.73	0.61	0.85
Agriculture, forestry, and fishing	*260*	*0.92*	*0.77*	*1.06*	*0.77*	*0.63*	*0.91*
Mining and quarrying	*103*	*0.58*	*0.33*	*0.84*	*0.38*	*0.14*	*0.62*
Secondary sector	3,044	1.46	1.40	1.52	1.40	1.34	1.46
Manufacturing	*2,963*	*1.46*	*1.40*	*1.52*	*1.40*	*1.34*	*1.46*
Utilities	*54*	*1.85*	*1.29*	*2.40*	*1.84*	*1.39*	*2.28*
Construction	*24*	*0.60*	*0.10*	*1.10*	*0.67*	*0.15*	*1.19*
Tertiary sector	75	1.42	1.13	1.71	1.25	0.90	1.61
Trade, catering, and accommodation	*23*	*0.97*	*0.65*	*1.28*	*0.84*	*0.53*	*1.16*
Transport, storage, and communication	*16*	*1.92*	*0.75*	*3.09*	*2.10*	*0.71*	*3.50*
Finance, insurance, real estate, and business	*8*	*1.07*	*0.43*	*1.72*	*0.57*	*0.03*	*1.10*
Services	*21*	*1.63*	*1.35*	*1.92*	*1.47*	*1.19*	*1.76*
Developing countries	856	1.83	1.69	1.96	1.54	1.43	1.66
Developed countries	738	1.24	1.16	1.32	1.24	1.15	1.34
Publication status							
Published papers	1,385	1.23	1.13	1.32	1.65	1.52	1.78
Unpublished papers	2,139	1.60	1.53	1.68	1.61	1.53	1.68
All estimates	3,524	1.45	1.40	1.51	1.64	1.56	1.71

Notes: The definitions of subsets are available in Table 4. Weighted = estimates weighted by the inverse of the number of estimates reported per study. Several elasticities in our dataset are estimated for all industries or across more sectors; these observations are excluded from the table.

Table 2 provides a first indication of the potential causes of heterogeneity. We compute the mean values of the Armington elasticity estimates for different groups of data based on temporal dynamics (short- or long-run), data frequency, structural variation, and publication characteristics. To account for the unbalancedness of our dataset, we also compute mean estimates weighted by the inverse of the number of estimates reported per study so that each study gets the same weight. The table shows that the long-run elasticities are approximately twice as large as the short-run elasticities, which corroborates the arguments of Gallaway *et al.* (2003)

and the common notion that short-run elasticities are smaller. In fact, Cassoni & Flores (2008) argue that smaller short-run estimates are given by the estimation design itself, unless overshooting occurs. Quarterly and annual data are typically used to capture the long-run effects (Gallaway *et al.*, 2003) and thus can be expected to produce larger elasticities than monthly data, which is supported by the statistics shown in the table.

The smaller elasticities reported for the primary sector (with respect to other sectors) suggest that the products of agriculture, forestry, fishing, mining, and quarrying are more difficult to substitute with their foreign alternatives. Concerning agriculture, this finding can be explained, as Kuiper & van Tongeren (2006) point out, by a common, explicit or implicit, support of domestic (or even local) produce. In contrast, the largest elasticities are typically found for utilities (approximately 1.85) and transport, storage, and communication (1.92). The elasticity also tends to be 50% larger for developing countries than for developed countries. Finally, although the means suggest a difference between the typical results of published and unpublished papers, the weighted means, in which each study has the same weight, suggest that the publication process is not associated with the magnitude of the estimates of the Armington elasticity. This simple analysis suggests there is potential for systematic differences among the reported elasticities, but any particular conclusion can be misleading without accounting for the correlation between individual aspects of data and methodology, which we address in Section 4. It can also be misleading without correcting for publication bias, and we turn to this problem in the following section.

3 Testing for Publication Bias

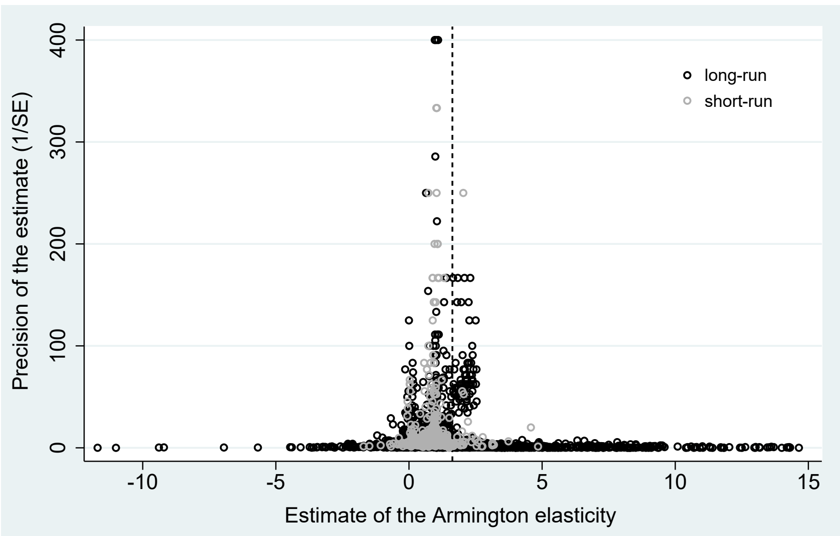
Publication bias is widespread in science, and economics is no exception: Ioannidis *et al.* (2017) document that the typical estimate reported in economics is exaggerated twofold because of publication selection. Publication selection arises because of the general preference of authors, editors, and referees for estimates that have the "right" sign and are statistically significant. Of course, this is not to say that publication selection equals cheating: in contrast, it makes sense for (and improves the value of) an individual study not to focus on estimates that are evidently wrong. But when most authors follow the strategy of ignoring estimates that have the "wrong" sign or are statistically insignificant, our inference from the literature as a whole (and also from many individual studies) becomes distorted. Given the degrees of freedom available to researchers in economics, estimates with the "right" sign and statistical significance at the 5% level are almost always possible to obtain after a sufficiently large number of specifications have been tried. A useful analogy provided by McCloskey & Ziliak (2019) is the Lombard effect, in which speakers increase their vocal effort in the presence of noise: given noisy data or estimation techniques, the researcher has more incentives to search through more specifications for a significant effect. When statistical significance becomes the implicit requirement for publication, significance will be produced but will no longer reflect what the statistical theory expects of it.

A conspicuous feature of the Armington elasticity is that it must be positive if both domestic and foreign goods are useful to the consumer. Therefore, from the very beginning, the literature has shunned negative and zero estimates as clear artifacts of data or method problems. One of the first studies, Alaouze (1977, p. 8), notes, "we shall concentrate on the ...[*industries*]... for which the elasticity of substitution has the correct [*positive*] sign." Among the latest studies, Feenstra *et al.* (2018, p. 144) find that the estimated elasticity is negative for some varieties and isolate them from the dataset: "these data are faulty or incompatible with our model." As we have noted, this approach can improve the inference drawn from an individual study but generally creates a bias. Given the inherent noise in trade data, estimated elasticities for some industries or specifications will always be insignificant, negative, or both. For other industries or specifications, the same noise produces estimates that are much larger than the true effect. However, no upper bound exists that would immediately deem elasticities implausible; some domestic and foreign goods can be perfectly substitutable in theory. Therefore, the large estimates will be kept in the paper and interpreted. This psychological asymmetry between zero and infinity coupled with inevitable imprecision in data and estimation creates publication bias. One apparent solution is symmetrical trimming: when the authors ignore 10 negative or insignificant estimates, they should also ignore the 10 largest positive estimates. Winsorizing would be better still, but it is rarely employed in practice.

A common tool used to assess the extent of publication bias is the so-called funnel plot (Egger *et al.*, 1997). The funnel plot shows the magnitude of the estimated effect on the horizontal axis and the precision of the estimate (the inverse of the standard error) on the vertical axis. There should be no relation between these two quantities because virtually all techniques used by the researchers to estimate the Armington elasticity guarantee that the ratio of the estimate to its standard error has a symmetrical distribution (typically a t-distribution). Therefore, regardless of their magnitude and precision, the estimates should be symmetrically distributed around the true mean effect. With decreasing precision, the estimates become more dispersed around the true effect and thus form a symmetrical inverted funnel. In the presence of publication bias, the funnel becomes either hollow (because insignificant estimates are omitted), asymmetrical (because estimates of a certain sign or size are excluded), or both.

The funnel plot in Figure 5 gives us a mixed message, as we show short- and long-run estimates of the Armington elasticity separately. The short-run elasticities are symmetrically distributed around their most precise estimates, which are slightly less than 1. The long-run elasticities, in contrast, form an asymmetrical funnel: the most precise estimates are also close to 1, but among imprecise estimates, there are many more that are much larger than 1 compared to those that are smaller than 1. This finding is consistent with no publication selection among short-run elasticities and publication selection against negative and insignificant elasticities among long-run elasticities. Nevertheless, the funnel plot is only a simple visual test, and the dispersion of the long-run estimates could suggest heterogeneity in data and methods, the other systematic factor driving the estimated coefficients. Regression-based funnel asymmetry tests

Figure 5: Funnel plot suggests publication bias among long-run elasticities



Notes: In the absence of publication bias, the funnel should be symmetrical around the most precise estimates of the elasticity. The dashed vertical line denotes the simple mean of the full sample of elasticities. Outliers are excluded from the figure for ease of exposition but included in all statistical tests.

provide a more concrete way to test for publication bias. As we have noted, if publication selection is present, the reported estimates and standard errors are correlated (Stanley, 2005; Stanley & Doucouliagos, 2010; Havranek, 2015):

$$\sigma_{ij} = \sigma_0 + \delta \cdot SE(\sigma_{ij}) + \mu_{ij}, \tag{3}$$

where σ_{ij} denotes i-th estimate of the Armington elasticity with the standard error $SE(\sigma_{ij})$ estimated in the j-th study; μ_{ij} is the error term. σ_0 is the mean underlying effect beyond publication bias (that is, conditional on maximum precision), and the coefficient δ of the standard error $SE(\sigma_{ij})$ represents the strength of publication bias. If $\delta = 0$, no publication bias is present. If $\delta \neq 0$, σ's and their standard errors are correlated, the correlation can arise either because researchers discard negative estimates of the elasticity (in which case the correlation occurs due to the apparent heteroskedasticity) or because researchers compensate for large standard errors with large estimates of the elasticity (the Lombard effect).

Table 3 presents the results of (3) using various estimation techniques run for three samples: the pooled set of elasticities, short-run elasticities, and long-run elasticities. Panel A uses unweighted data. In the baseline OLS model, the coefficient δ from (3) is not statistically significant for the short-run sample, and the estimated corrected mean is the same as the simple mean of 0.9. In the sample of long-run elasticities, in contrast, we find strong publication bias that decreases the underlying mean from 1.56 (the uncorrected mean) to 0.9 (the mean corrected for publication bias). The result for a pooled sample of short- and long-run elasticities is close to that of long-run elasticities because long-run elasticities dominate the dataset.

In the next model, we add study-level fixed effects to the baseline specification, which slightly deepens the difference between the mean short- and long-run effects beyond bias. Finally, for Panel A, we use a multilevel estimation technique that implements partial pooling at the study level and uses the data to influence the pooling weights. Given that the estimated elasticities are nested within each study, hierarchical modeling is a convenient choice to analyze the variance in the elasticities: one can expect that the stochastic term of (3) depends on the design of each individual study and therefore does not have the same dispersion across individual studies. It follows that the regression coefficients δ are probably not the same across studies. Nevertheless, δ's should be related, and the hierarchical modeling treats them as random variables of yet another linear regression at the study level. We apply a hierarchical Bayes model and implement

Table 3: All tests indicate publication bias among long-run Armington elasticities

	All	Short-run	Long-run
PANEL A: Unweighted estimations			
OLS			
SE (publication bias)	0.808***	0.0791	0.805***
	(0.0652)	(0.0826)	(0.0630)
Constant (effect beyond bias)	0.873***	0.867***	0.901***
	(0.133)	(0.0249)	(0.168)
Fixed effects			
SE (publication bias)	0.621***	-0.00578	0.627***
	(0.0588)	(0.104)	(0.0580)
Constant (effect beyond bias)	1.007***	0.883***	1.047***
	(0.0423)	(0.0192)	(0.0476)
Hierarchical Bayes			
SE (publication bias)	0.500**	-0.0810	0.630***
	(0.190)	(0.480)	(0.190)
Constant (effect beyond bias)	1.200***	0.887**	1.250***
	(0.240)	(0.310)	(0.0476)
PANEL B: Weighted OLS estimations			
Weighted by the inverse of the number of estimates reported per study			
SE (publication bias)	1.017***	0.0975*	1.033***
	(0.249)	(0.0514)	(0.251)
Constant (effect beyond bias)	1.011***	0.893***	1.046***
	(0.254)	(0.0694)	(0.303)
Weighted by the the inverse of the standard error			
SE (publication bias)	1.559	2.698	0.906**
	(0.969)	(2.213)	(0.431)
Constant (effect beyond bias)	0.761***	0.510	0.922***
	(0.217)	(0.325)	(0.205)
PANEL C: Non-linear estimations			
Weighted average of adequately powered (Ioannidis *et al.*, 2017)			
Effect beyond bias	1.049***	0.872***	1.101***
	(0.017)	(0.024)	(0.021)
Selection model (Andrews & Kasy, 2019)			
Effect beyond bias	0.911***	0.863***	0.943***
	(0.015)	(0.018)	(0.021)
Stem-based method (Furukawa, 2019)			
Effect beyond bias	0.992***	1.031***	0.994***
	(0.024)	(0.070)	(0.042)
Observations	3,524	556	2,968

Notes: The uncorrected mean of the estimates of the long-run Armington elasticity is 1.56. Panels A and B report the results of regression $\sigma_{ij} = \sigma_0 + \delta \cdot SE(\sigma_{ij}) + \mu_{ij}$, where σ_{ij} denotes i-th Armington elasticity estimated in the j-th study and $SE(\sigma_{ij})$ denotes the corresponding standard error. All = the entire dataset, Short-run = short-run Armington elasticities, Long-run = long-run Armington elasticities, SE = standard error. Standard errors, clustered at the study and country level, are reported in parentheses (except Hierarchical Bayes, which has posterior standard deviation in parentheses). The available number of observations is reduced for Ioannidis *et al.* (2017)'s estimation (all 3,440; short-run 555; long-run 2,885) and Furukawa (2019)'s estimation (all 1,850; short-run 105; long-run 965). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Stars for hierarchical Bayes are presented only as an indication of the parameter's statistical importance to keep visual consistency with the rest of the table.

the Gibbs sampler for hierarchical linear models with a standard prior, following Rossi *et al.* (2005). The hierarchical model corroborates the evidence presented earlier but finds slightly weaker publication bias among the estimates of the long-run elasticity.

Panel B of Table 3 presents weighted alternatives to the baseline OLS model of Panel A. First, the regression is weighted by the inverse of the number of estimates reported by each study, so that both small and large studies are all assigned the same importance. Second, the regression is weighted by the inverse of the standard error so that more precise estimates are assigned greater importance. Panel B shows results that support the conclusions from Panel A. Finally, Panel C shows the latest alternatives to linear meta-analysis models. The problem with the linear regression that we have used so far is the implicit assumption that publication bias is a linear function of the standard error. If the assumption does not hold, our conclusion concerning publication bias can be misleading. Here, we apply three non-linear techniques that relax this assumption. The corrected means of both the short- and long-run Armington elasticity remain close to unity in all three alternative approaches: the weighted average of adequately powered estimates by Ioannidis *et al.* (2017), the stem-based method by Furukawa (2019), and the selection model by Andrews & Kasy (2019).

Based on a survey involving more than 60,000 estimates, Ioannidis *et al.* (2017) document that the median statistical power among the published results in economics is 18%. They show how low power is associated with publication bias and then propose a simple correction procedure that focuses on the estimates with power above 80%. Monte Carlo simulations presented in Ioannidis *et al.* (2017) suggest that this simple technique outperforms the commonly used meta-regression estimators. The intuition of the model presented by Furukawa (2019) rests on the fact that the most precise estimates suffer from little bias: with very small standard errors, the authors can easily produce estimates that are statistically significant. While previous authors have recommended meta-analysts to focus on a fraction of the most precise estimates in meta-analysis (for example, Stanley *et al.*, 2010), Furukawa (2019) finds a clever way to estimate this fraction based on exploiting the trade-off between bias and variance (omitting studies increases variance). Andrews & Kasy (2019) use the observation reported by many researchers (for instance, Havranek, 2015; Brodeur *et al.*, 2016) that standard cut-offs for the p-value (0.01, 0.05, 0.1) are associated with jumps in the distribution of reported estimates.

Andrews & Kasy (2019) build on Hedges (1992) and construct a selection model that estimates publication probability for each estimate in the literature given its p-value. They show that, in several areas, the technique gives results similar to those of a large-scale replication.

Several important findings can be distilled from the estimations reported in Table 3. First, we find publication bias among long-run elasticities but not among short-run elasticities. One explanation consistent with this result is that short-run elasticities are typically deemed less important than long-run elasticities, especially for policy purposes. They are often reported only as complements to the central findings of the paper. It can take time before consumers shift their demand between domestic and foreign goods; consequently, insignificant estimates of the short-run elasticity are more likely to survive the publication process than insignificant estimates of the long-run elasticity. Second, publication bias inflates the mean estimate of the long-run Armington elasticity by at least 50%, which can have a strong impact on the results of a model informed by the empirical literature in terms of the calibration of the elasticity. Third, the large difference between the short- and long-run elasticities reported in Table 2 (and observed in many studies, see Gallaway *et al.*, 2003) is all but erased once publication bias is taken into account. In sum, we find robust evidence of publication bias in this literature. However, some of the apparent correlations between the estimated elasticities and their standard errors can be due to data and method heterogeneity. We turn to this issue in the next section.

4 Why Elasticities Vary

4.1 Potential Factors Explaining Heterogeneity

Three reasons for the systematic differences in the estimates of the Armington elasticity have been frequently discussed in the literature. First, studies using disaggregated data are often observed to yield larger estimates than studies using aggregate data (Imbs & Mejean, 2015). Second, cross-sectional studies tend to yield larger estimates than time-series studies (Hillberry & Hummels, 2013). Third, multi-equation estimation techniques typically give larger estimates than single-equation techniques (Goldstein & Khan, 1985). Many literature reviews (including Cassoni & Flores, 2008; Marquez, 2002; McDaniel & Balistreri, 2003), moreover, stress other characteristics of estimates and studies that can significantly influence the results. We present the first attempt to shed light on the sources of heterogeneity in Table 2. To investigate the heterogeneity among the estimates of the Armington elasticity more systematically, we codify 43 characteristics of the study design and augment equation (3) by adding these characteristics as explanatory variables. Given that publication bias affects only the long-run elasticity, we replace the standard error in the equation by an interaction term between the standard error and a dummy variable that equals one if the estimate corresponds to a long-run elasticity.

Table 4 lists all the codified variables, their definitions and summary statistics, including the simple mean, standard deviation, and mean weighted by the inverse of the number of observations reported in a study. For ease of exposition, we divide the variables into groups reflecting data characteristics (11 aspects), structural variations (11 aspects), estimation techniques (14 aspects), and publication characteristics potentially related to quality that are not captured by data and estimation characteristics (3 aspects). The distinction between short- and long-run elasticities is among the most important factors stressed in the literature (Gallaway *et al.*, 2003). Nevertheless, in the previous section, we find that publication bias plagues the estimates of long-run elasticities and that beyond publication bias, short- and long-run elasticities have comparable magnitudes. In this section, we will examine whether the claim still holds when other possible systematic influences on the estimates of the Armington elasticity are taken into account.

Table 4: Description and summary statistics of the regression variables

Variable	Description	Mean	SD	WM
Armington elasticity	The reported estimate of the Armington elasticity.	1.45	1.78	1.64
Standard error (SE)	The reported standard error of the Armington elasticity estimate.	0.72	1.18	0.61
SE * Long-run effect	The interaction between the standard error and the estimated long-run Armington elasticity.	0.69	1.19	0.59
Temporal dynamics				
Short-run effect	=1 if the estimated Armington elasticity is short-term (reference category for the group of dummy variables describing temporal dynamics).	0.16	0.36	0.12
Long-run effect	=1 if the estimated Armington elasticity is long-term.	0.84	0.36	0.88
Data characteristics				
Data disaggregation	The level of data aggregation according to SIC classification (min = 1 if fully aggregated, max = 8 if disaggregated).	6.49	1.58	6.20
Results disaggregation	The level of results aggregation according to SIC classification (min = 1 if fully aggregated, max = 8 if disaggregated).	5.06	1.21	5.34
Monthly data	=1 if the data are in monthly frequency.	0.14	0.35	0.08

Continued on next page

Table 4: Description and summary statistics of the regression variables (continued)

Variable	Description	Mean	SD	WM
Quarterly data	=1 if the data are in quarterly frequency (reference category for the group of dummy variables describing data frequency).	0.21	0.41	0.25
Annual data	=1 if the data are in yearly frequency.	0.65	0.48	0.67
Panel data	=1 if panel data are used (reference category for the group of dummy variables describing time and cross-sectional dimension of data).	0.34	0.47	0.27
Time series	=1 if time-series data are used.	0.58	0.49	0.65
Cross-section	=1 if cross-sectional data are used.	0.08	0.27	0.08
Data period	The length of the time period in years.	14.24	9.76	17.08
Data size	The logarithm of the total number of observations used to estimate the elasticity.	4.64	1.93	4.55
Midyear	The median year of the time period of the data used to estimate the elasticity.	23.45	11.54	22.48
Structural variation				
Primary sector	=1 if the estimate is for the primary sector (agriculture and raw materials; reference category for the group of dummy variables describing sectors).	0.10	0.31	0.31
Secondary sector	=1 if the estimate is for the secondary sector (manufacturing).	0.86	0.34	0.58
Tertiary sector	=1 if the estimate is for tertiary sector (services).	0.02	0.14	0.03
Developing countries	=1 if the estimate is for a developing country (reference category for the group of dummy variables describing the level of development).	0.24	0.43	0.28
Developed countries	=1 if the estimate is for a developed country.	0.79	0.41	0.74
Market size	The logarithm of the market size of the home country (GDP in billions of USD, 2015 prices).	6.45	1.86	5.94
Tariffs	The tariff rate of the home country (weighted mean, all products, %).	6.78	7.15	6.07
Non-tariff barriers	Additional cost to import of the home country (USD per container).	0.94	0.26	0.97
FX volatility	The volatility of the exchange rate using the DEC alternative conversion factor (home country currency unit per USD).	0.58	0.55	0.69
National pride	Home bias captured by the percentage of "I am very proud of my country" answers from the World Values Survey.	0.53	0.22	0.51
Internet usage	The number of fixed broadband subscriptions of the home country (per 100 people).	2.91	5.02	1.23
Estimation technique				
Static model	=1 if a static model is used for estimation.	0.23	0.42	0.30
Distributed lag and trend model	=1 if a distributed lag or trend model is used.	0.10	0.30	0.27
Partial adjustment model	=1 if a partial adjustment model is used for estimation.	0.15	0.35	0.11
First-difference model	=1 if a first-difference model is used.	0.09	0.29	0.05
Error-correction model	=1 if an error-correction model is used.	0.04	0.20	0.04
Nonlinear model	=1 if a nonlinear model is used.	0.28	0.45	0.13
Other models	=1 if another model is used (reference category for the group of dummy variables describing models used).	0.11	0.31	0.10
OLS	=1 if the OLS or GLS estimation method is used.	0.48	0.50	0.67
CORC	=1 if the Cochrane-Orcutt or FGLS estimation method is used.	0.16	0.37	0.13
TSLS	=1 if the instrumental method is used.	0.09	0.28	0.06
GMM	=1 if the GMM estimation method is used.	0.24	0.43	0.10

Continued on next page

Table 4: Description and summary statistics of the regression variables (continued)

Variable	Description	Mean	SD	WM
Other methods	=1 if other types of estimation are used (reference category for the group of dummy variables describing the estimation method used).	0.03	0.17	0.05
Import constraint	=1 if the study includes some measure of import restriction.	0.03	0.18	0.06
Seasonality	=1 if the study controls for seasonality.	0.20	0.40	0.12
Publication characteristics				
Impact factor	The recursive discounted impact factor from RePEc.	0.12	0.24	0.17
Citations	The logarithm of the number of Google Scholar citations normalized by the number of years since the first draft of the paper appeared in Google Scholar.	1.26	1.01	1.00
Published	=1 if a study is published in a peer-reviewed journal.	0.39	0.49	0.65

Notes: SD = standard deviation, WM = mean weighted by the inverse of the number of estimates reported per study, SIC = Standard Industrial Classification system for classifying industries by a four-digit code. Market size, tariff and non-tariff barriers, FX volatility, and internet usage have been collected from the World Bank database (WB, 2017), data on national pride from the World Values Survey (Inglehart *et al.*, 2014). The impact factor is downloaded from RePEc and the number of citations from Google Scholar. The rest of the variables are collected from studies estimating the Armington elasticity.

Data characteristics. Many studies (Feenstra *et al.*, 2018; McDaniel & Balistreri, 2003; Welsch, 2008, among others) argue that because intra-industry diversity decreases with an increasing level of sectoral aggregation, more aggregated data should yield smaller elasticities. Feenstra *et al.* (2018) note that some recent macro-studies (Bergin, 2006; Heathcote & Perri, 2002) estimate the aggregate elasticities around unity, while studies focusing on individual product groups (Broda & Weinstein, 2006; Imbs & Mejean, 2015) imply much stronger responses. McDaniel & Balistreri (2003) compare two articles on US data that use 3-digit SIC level (Reinert & Roland-Holst, 1992) and 4-digit SIC level (Gallaway *et al.*, 2003) aggregations and come to the same conclusion: higher disaggregation brings higher substitutability. We codify the *data disaggregation* variable according to the SIC classification. Fully aggregated, whole-economy data acquire the value of 1; in contrast, fully disaggregated product-level data acquire the value of 8. Given the consensus in the literature, we expect the variable to show a positive association with the reported elasticities. Furthermore, in some papers (such as Aspalter, 2016; Mohler & Seitz, 2012), the level of aggregation of the input data differs from the level of aggregation of the reported results. Imbs & Mejean (2015) argue that a pooled estimate that ignores heterogeneity across sectors tends to be biased downwards. To reflect the problem of aggregating the results, we create an additional variable based on the same principles as the variable for data aggregation.

Another commonly discussed issue is data frequency. It is related to the short- or long-run nature of the elasticity, but we control for this feature separately. Cassoni & Flores (2008) show that aggregation from monthly to quarterly data removes short-term adjustment patterns, such as overshooting (Cassoni & Flores, 2010) or J-curve effects (Backus *et al.*, 1994). They also note that *monthly data* often contain atypical observations that could misrepresent the underlying trade data. Gallaway *et al.* (2003), on the other hand, estimate long-run elasticities based on monthly and quarterly data and find no systematic difference in the estimates. Given that quantity measures are notoriously noisy, Hillberry & Hummels (2013) state that the measurement error often becomes exacerbated with monthly or quarterly data and high product disaggregation. The use of *quarterly* instead of *yearly data* may be necessary to gain a sufficiently large dataset, but Hertel *et al.* (1997) argue that problems associated with quarterly data could lead to overly inelastic estimates. A number of studies, including Aspalter (2016), Olekseyuk & Schurenberg-Frosch (2016), and Feenstra *et al.* (2018), use annual data, especially when the authors want to identify both micro and macro elasticities. Aspalter (2016) also suggests that the annual frequency of data often leads to a more consistent cross-country dataset.

We further distinguish among *time series*, *cross-section*, and *panel data*, using panel data as the reference category. The survey by McDaniel & Balistreri (2003) reports that cross-sectional data are associated with larger reported elasticities because cross-sectional estimates also consider supply conditions. Cassoni & Flores (2008), however, argue that the conclusion of McDaniel & Balistreri (2003) stems from comparing results based on heterogeneous analyses and data and point out that the impact of data cross-sectionality depends on the correct specification of the model and the estimation technique employed. The variable *data period* reflects how estimates differ when obtained over longer time periods, while the variable *data size* captures the potential effects of small-sample bias. We also control for the age of the data by including a variable that reflects the midpoint year of the sample (variable *midyear*) with which the Armington elasticity is estimated. Figure 2 suggests that the elasticity is increasing in time (and some studies, for example Schurenberg-Frosch, 2015; Welsch, 2008, observe a similar pattern). In this vein, Hubler & Pothen (2017) argue that globalization might have increased the Armington elasticity by decreasing the heterogeneity of products and reducing the market power of individual countries.

Structural variation. The elasticity of substitution might depend systematically on the characteristics of the product, industry, or country in question. Blonigen & Wilson (1999) suggest that with greater physical differences, the elasticity of substitution between products decreases. Shiells *et al.* (1986) and more recent papers such as Faria & Haddad (2014), Nemeth *et al.* (2011), and Saikkonen (2015) provide evidence of how the Armington elasticity differs across industries. Moreover, Saito (2004) shows that heterogeneous goods (e.g., final products such as automobiles or medical equipment) are more difficult to substitute across countries than more homogenous goods (e.g., intermediate products such as glass or metals). Because we do not have enough variation in our dataset to control for the many individual product categories or industries (if all these controls were included, collinearity would skyrocket), we control for sectoral differences by dividing the sample into three groups: the *primary sector* with industries related to raw materials, the *secondary sector* with manufacturing industries, and the *tertiary sector* of services. Nevertheless, the data that we provide in the online appendix include more details and researchers can use these data and codes to construct implied elasticities for the individual industries in which they are interested.

We also control for the characteristics of the country for which the elasticity is estimated (the home country). Developing countries can be expected to face a larger pool of substitutable products abroad because the rest of the world encompasses the production of all levels of technology. In contrast, for developed countries with better production technologies, it might be more difficult to find adequate substitutes abroad. Moreover, Kapuscinski & Warr (1999) note that developing countries often provide poor data, and the resulting biases could lead to larger elasticities. We divide the countries into two categories: a group of *developed* countries, which includes Central and Western Europe, North America, Australia, New Zealand, and Japan; and a group of *developing countries*, which covers the rest of Asia, Latin America, and Africa.

It has been shown in the literature that even physically identical goods can be differentiated by aspects such as availability, customer service, and perception of quality. Linder (1961) suggests that countries with similar income per capita should trade more because their consumers have similar tastes, as reflected in the production of goods in each country (more details are provided in Francois & Kaplan, 1996). Ideally, to capture these features of consumers' preferences, we follow the study on the border effect by Havranek & Irsova (2017) and create a variable representing the income dissimilarity of the home country and the corresponding foreign country. Because this bilateral approach is not feasible for the Armington elasticity literature, we use another representation of consumer preferences: we include a proxy variable *national pride* to capture consumer bias for home goods over foreign ones (Trefler, 1995; Kehoe *et al.*, 2017). The variable is constructed as the percentage of 'very proud' answers to the question 'How proud are you of your country?' from the World Values Survey (Inglehart *et al.*, 2014). Wolf (2000), for example, shows that the home bias could go beyond the influence of typical quantifiable trade barriers and also exist on a sub-national level.

Several potential country-level determinants of the Armington elasticity have a strong connection to the border effect first presented by McCallum (1995). One of the common border effect determinants is *market size*: any border barrier in a small economy increases the ratio of within-country trade more than in a large economy. We thus expect this variable to have a positive association with the reported elasticity. To proxy for market size, we use GDP for the midpoint of the data period used in the study. Moreover, trade barriers and other extra transaction costs associated with crossing the border have also been considered an important determinant of the Armington macro elasticity (Lopez & Pagoulatos, 2002). These trade frictions are captured by variables *tariff* (representing the tariff rate) and *non-tariff barriers* (representing the cost to import); all these data are obtained from WB (2017).

According to Parsley & Wei (2001), contracting costs and insecurity represent other potential determinants that affect cross-country trade and possibly the Armington elasticity. We approximate these additional trade frictions by the volatility of the exchange rate in the home country versus the US dollar (variable *FX volatility*). Parsley & Wei (2001) suggest that the exchange rate volatility may not only contribute to cross-border market insecurities but also explain the price dispersion of similar goods across the border. Finally, we account for information barriers and use the number of broadband subscriptions per 100 people as a measure of *internet usage*. The expansion of internet use creates new types of tradable services and is believed to have increased cross-border trade (IBRD, 2009).

Estimation technique. A large variety of models and methods exist to estimate the Armington elasticity. To simplify, denoting the expression $\log(Q_M/Q_D)$ in (2) as y , $\log(P_D/P_M)$

as x, and $\log(\beta/1-\beta)$ as σ_0, we obtain the *static model* $y_t = \sigma_0 + \sigma_1 x_t + e_t$, where σ_1 is the Armington elasticity and e is the error term. Static models constitute approximately 23% of our dataset. Another category labeled *distributed lag and trend model* includes elasticities estimated using distributed lag models (Tourinho *et al.*, 2003) and models with a time trend variable added to achieve data stationarity (Lundmark & Shahrammehr, 2012): $y_t = \sigma_0 + \sum_{l=0}^{\tau} \sigma_{l+1} x_{t-l} + \sigma_{\tau+1} t + e_t$, $\tau \geq 0$. The *partial adjustment model*, on the other hand, allows for a non-instantaneous adjustment of the demand structure to the changes of the relative prices (for example Ogundeji *et al.*, 2010) by adding the lagged dependent variable y_{t-1} among the explanatory variables and reads $y_t = \sigma_0 + \sigma_1 x_t + \sigma_2 y_{t-1} + e_t$ (Alaouze, 1977, shows that the omission of the lagged dependent variable in cases where it is significant biases the estimates downwards).

If the corresponding levels of time series are not stationary or cointegrated, the authors take *first differences* (see Gibson, 2003, for example). In some cases, the lagged value of the level of the explanatory variable is also included, and the authors end up with $\Delta y_t = \sigma_0 + \sigma_1 \Delta x_t + \sigma_2 x_{t-1} + e_t$. When the time series are cointegrated, authors also use an *error-correction model* to estimate the elasticity (such as Gan, 2006, does); then, the model reads $\Delta y_t = \sigma_0 + \sigma_1 \Delta x_t + \sigma_2 y_{t-1} + \sigma_3 x_{t-1} + e_t$. Several studies, including Corado & de Melo (1983), Feenstra *et al.* (2018), and Saikkonen (2015), employ different forms of *non-linear models*. The non-linear model category constitutes 28% of our dataset. There is no unifying specification presentable in this case, as the individual approaches differ. The reference category for the group of dummy variables describing the models used to estimate the Armington elasticity is the variable *other models*, which covers the rest of the used approaches that do not fall under any of the above-mentioned categories.

Shiells & Reinert (1993) use the GLS technique, ML estimation, and simultaneous equation estimator that employs a distributed lag model to estimate the elasticities. They find the estimates to be relatively insensitive to the three alternative estimation procedures. Not all studies, however, come to the same conclusion of methodological indifference. To account for the potential effect of estimation techniques, we group the most frequently used methods of estimation into five categories: OLS estimation together with the GLS estimator (variable *OLS*), Cochrane-Orcutt estimation together with the FGLS (variable *CORC*), two-stage least squares and related techniques (variable *TSLS*), a separate group of *GMM* estimates, and all *other methods*, which represent the reference category for this group of dummies. We also include a control that equals one if the specification includes some measures of *import constraints*. Alaouze (1977) stresses that quantitative and tariff quota restrictions could bias the estimates of the elasticity because importers cannot fully utilize the advantages of price changes or must pay a fee when exceeding a certain amount of imported goods. Another coded aspect of the data is whether the authors control for *seasonality* in the demand function (Tourinho *et al.*, 2010), which is a particularly important characteristic of agricultural products. Seasonality is commonly captured by quarterly dummies (see, for example, Ogundeji *et al.*, 2010).

Publication characteristics. Despite the large number of variables we collect, the list of aspects potentially related to quality is unlimited. Therefore, we also employ several publication characteristics that can be expected to be correlated with the unobserved features of the quality of the paper. To see if published studies yield systematically different results, we include a dummy variable that equals one if the study is *published* in a peer-reviewed journal. To take into account the differences in the quality of publication outlets, we include the discounted recursive RePEc *impact factor* of the respective study (this impact factor is available for both journals and working paper series). Finally, for each study, we create a variable reflecting the logarithm of the number of Google Scholar *citations* normalized by the number of years since the first draft of the study appeared in Google Scholar.

4.2 Estimation

To relate the variables introduced above to the magnitude of the estimated Armington elasticities, one could run a standard regression with all the variables. But such an estimation would ignore the model uncertainty inherent in meta-analysis: while we have a strong rationale to include some of the variables, others are considered mainly as controls for which there is no theory on how they could affect the results of studies estimating the Armington elasticity. To address model uncertainty, we employ Bayesian model averaging (BMA). BMA runs many regressions with different subsets of the 2^{34} possible combinations of explanatory variables. We do not estimate all possible combinations but employ Monte Carlo Markov Chain (specifically, the Metropolis-Hastings algorithm of the `bms` package for R by Zeugner & Feldkircher, 2015), which walks through the most likely models. In the Bayesian setting, the likelihood of each model is represented by the posterior model probability. The estimated BMA coefficients for each variable are represented by posterior means and are weighted across all models by their posterior probability. Each coefficient is then assigned a posterior inclusion probability that reflects the probability of the variable being included in the underlying model and is calculated as the sum of posterior model probabilities across all the models in which the variable is included. Further details on BMA can be found in, for example, Raftery *et al.* (1997) or Eicher *et al.* (2011). BMA has been used in meta-analysis, for example, by Havranek *et al.* (2015).

In the baseline specification, we employ the priors suggested by Eicher *et al.* (2011), who recommend using the uniform model prior (giving each model the same prior probability) and the unit information g-prior prior (giving the prior the same weight as one observation of the data). These priors reflect the lack of prior knowledge regarding the probability of individual specifications, model size, and parameter values. We use unweighted data to estimate the baseline but later provide weighted alternatives to evaluate the robustness of our results. Furthermore, as a robustness check, we follow Ley & Steel (2009) and apply the beta-binomial random model prior, which gives the same weight to each model size, as well as Fernandez *et al.* (2001), who advocate for the so-called BRIC g-prior. In addition, to avoid using priors entirely, we also apply frequentist model averaging (FMA). Following Hansen (2007), we use Mallow's criterion for model averaging and the approach of Amini & Parmeter (2012) towards the orthogonalization of the covariate space. Amini & Parmeter (2012) provide a comprehensive comparison of different averaging techniques, including Mallow's weights and other frequentist alternatives.

4.3 Results

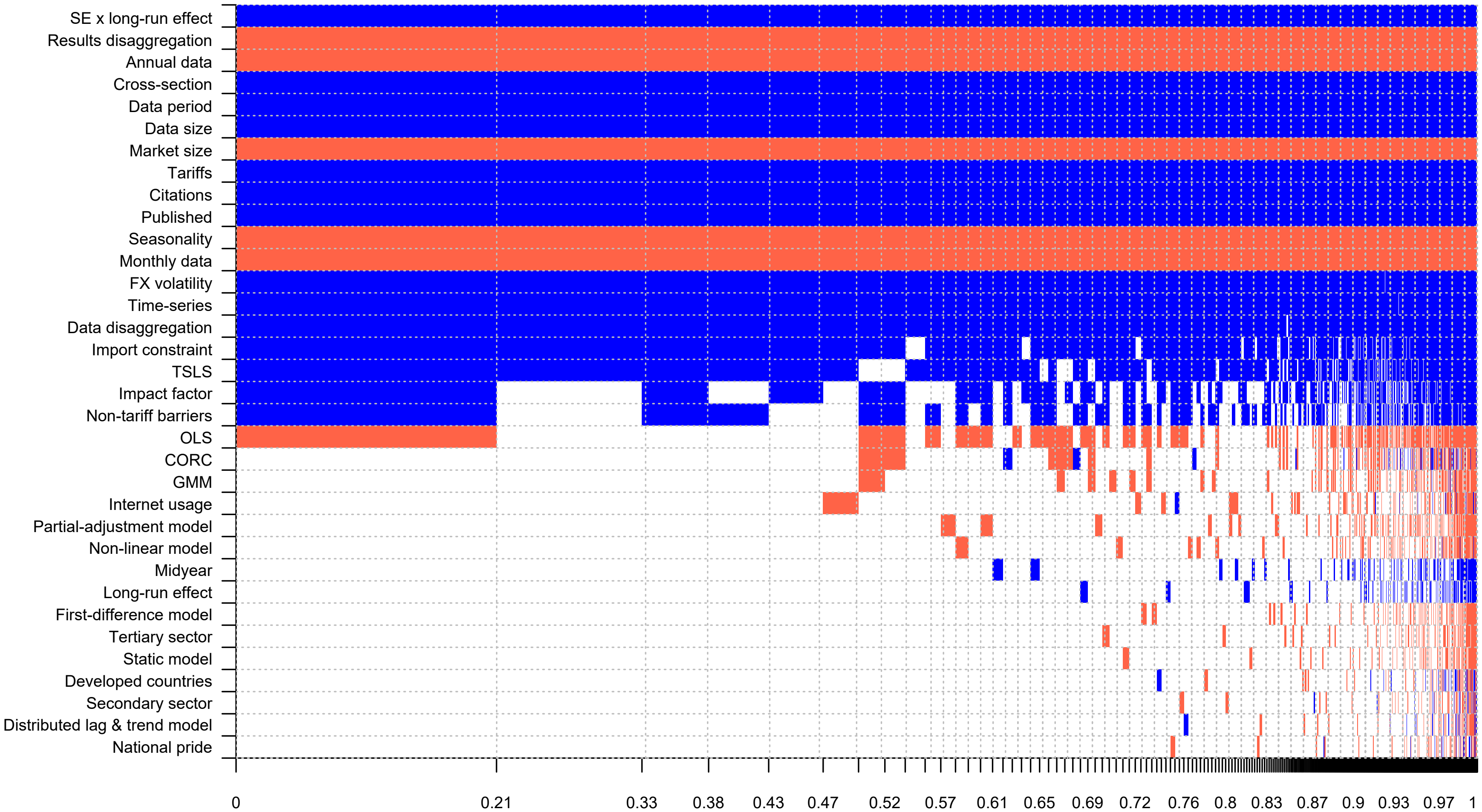
Figure 6 visualizes the results of Bayesian model averaging. The columns of the figure denote the individual regression models, and the column widths indicate the posterior model probability. The columns are sorted by posterior model probability from left to right. The rows of the figure denote individual variables included in each model. The variables are ordered by their posterior inclusion probability from top to bottom in descending order. If a variable is excluded from the model, the corresponding cell is left blank. Otherwise, the blue color (darker in grayscale) indicates a positive sign of the variable's coefficient in the particular model; the red color (lighter in grayscale) indicates a negative sign. Figure 6 shows that approximately half of our variables are included in the best models, and the signs of these variables are robust across specifications.

The numerical results of the BMA exercise with priors according to Eicher *et al.* (2011) are reported in Table 5. Additionally, we show two alternative estimations. First, we estimate simple OLS, which excludes the 13 variables that were deemed unimportant by the BMA exercise (according to Eicher *et al.*, 2011, the effect of a variable is considered *decisive* if the posterior inclusion probability is between 0.99 and 1, *strong* between 0.95 and 0.99, *substantial* between 0.75 and 0.95, and *weak* between 0.5 and 0.75). OLS results mostly correspond with the results of BMA: the coefficients display the same signs and similar magnitudes, and their p-values typically correspond to the information extracted from the respective posterior inclusion probabilities (with the exception of country-level variables, which will be discussed below). Second, we estimate frequentist model averaging, which includes all variables used in the BMA model. FMA conclusions are mostly in line with the baseline; except that, unlike BMA, it considers estimation techniques to be important factors driving the magnitude of the Armington elasticity.

The complete set of robustness checks, including BMA exercises with alternative priors and weights, can be found in Table A2. When using alternative priors (according to Fernandez *et al.*, 2001; Ley & Steel, 2009), we obtain evidence that supports the conclusions of our baseline model. BMA weighted by the inverse of the number of estimates reported per study confronts our baseline model on estimation techniques, and we will discuss the differences later on. We also report BMA with precision weights, although such an estimation is problematic in our case because weighting by precision introduces artificial variation to the study-level variables. BMA results from Table 5 testify to the decisive importance of the effects caused by *data* and *results disaggregation*, the usage of *monthly* and *annual data*, *time-series* and *cross-section* type of input data, *data period* and *data size* of a study, the country's *market size*, imposed *tariffs*, *FX volatility*, a control for *seasonality*, the number of *citations*, and *published* studies. The results further point to substantial evidence of effects caused by imposed *import constraints* and weak evidence of effects caused by imposed *non-tariff barriers* and *nonlinear model* choice. We will concentrate on the variables for which we have the most robust evidence.

The presence of publication bias in the estimates of the long-run Armington elasticity is supported by evidence across all the models that we run. The reported long-run elasticities,

Figure 6: Model inclusion in Bayesian model averaging



Notes: The figure depicts the results of Bayesian model averaging with a uniform model prior and unit information prior (Eicher *et al.*, 2011). On the vertical axis, the explanatory variables are ranked according to their posterior inclusion probabilities from the highest at the top to the lowest at the bottom. The horizontal axis shows the values of cumulative posterior model probability for each model ranked from the highest on the left to the lowest on the right. All variables are described in Table 4. Numerical results are reported in Table 5. The blue color (darker in grayscale) means that the estimated parameter of the corresponding explanatory variable is positive. The red color (lighter in grayscale) indicates that the estimated parameter is negative. No color denotes that the corresponding explanatory variable is not included in the model. The results are based on the unweighted specification. The robustness checks in which the specification is weighted by the number of estimates reported per study and by the standard error of the estimate are provided in Table A2 in the Appendix. Detailed diagnostics are provided in Table A1 and Figure A1.

Table 5: Why elasticities vary

Response variable:	Bayesian model averaging			Frequentist check (OLS)			Frequentist model averaging		
Estimate of the Armington elasticity	Post. mean	Post. SD	PIP	Coef.	SE	p-value	Coef.	SE	p-value
Constant	-1.38	NA	1.00	-1.53	0.70	0.03	-1.65	0.53	0.00
SE * Long-run effect	0.73	0.02	1.00	0.73	0.05	0.00	0.74	0.03	0.00
Long-run effect	0.00	0.02	0.03				0.17	0.12	0.13
Data characteristics									
Data disaggregation	0.19	0.04	1.00	0.18	0.10	0.07	0.20	0.05	0.00
Results disaggregation	-0.23	0.04	1.00	-0.22	0.10	0.03	-0.25	0.04	0.00
Monthly data	-0.59	0.13	1.00	-0.50	0.27	0.06	-0.48	0.16	0.00
Annual data	-1.16	0.12	1.00	-1.15	0.39	0.00	-0.89	0.16	0.00
Time series	0.54	0.13	1.00	0.59	0.46	0.20	0.83	0.17	0.00
Cross-section	2.21	0.19	1.00	2.34	0.40	0.00	2.09	0.22	0.00
Data period	0.03	0.00	1.00	0.03	0.01	0.00	0.03	0.01	0.00
Data size	0.32	0.02	1.00	0.32	0.09	0.00	0.33	0.02	0.00
Midyear	0.00	0.00	0.05				0.01	0.01	0.13
Structural variation									
Secondary sector	0.00	0.01	0.02				0.02	0.08	0.82
Tertiary sector	0.00	0.03	0.02				-0.07	0.17	0.68
Developed countries	0.00	0.02	0.02				0.22	0.15	0.15
Market size	-0.11	0.02	1.00	-0.12	0.06	0.05	-0.14	0.03	0.00
Tariffs	0.03	0.01	1.00	0.03	0.01	0.02	0.04	0.01	0.00
Non-tariff barriers	0.21	0.20	0.58	0.40	0.21	0.05	0.38	0.15	0.01
FX volatility	0.31	0.07	1.00	0.28	0.15	0.07	0.18	0.07	0.01
National pride	0.00	0.02	0.01				-0.17	0.19	0.37
Internet usage	0.00	0.00	0.07				0.01	0.01	0.46
Estimation technique									
Static model	0.00	0.01	0.02				-0.26	0.11	0.02
Distributed lag and trend model	0.00	0.01	0.01				-0.34	0.15	0.02
Partial adjustment model	-0.01	0.04	0.06				-0.29	0.12	0.02
First-difference model	0.00	0.02	0.03				-0.17	0.13	0.17
Nonlinear model	-0.01	0.07	0.05				-0.65	0.29	0.03
OLS	-0.14	0.18	0.47	-0.23	0.14	0.09	-0.39	0.18	0.03
CORC	-0.04	0.13	0.13				-0.25	0.18	0.15
TSLS	0.42	0.19	0.89	0.40	0.16	0.01	0.29	0.19	0.13
GMM	-0.03	0.11	0.09				-0.10	0.21	0.63
Import constraint	0.48	0.18	0.94	0.58	0.24	0.02	0.51	0.17	0.00
Seasonality	-0.57	0.11	1.00	-0.52	0.34	0.13	-0.37	0.13	0.00
Publication characteristics									
Impact factor	0.23	0.22	0.59	0.43	0.12	0.00	0.52	0.18	0.00
Citations	0.55	0.05	1.00	0.53	0.17	0.00	0.46	0.08	0.00
Published	0.57	0.09	1.00	0.61	0.30	0.04	0.58	0.15	0.00
Studies	42			42			42		
Observations	3,524			3,524			3,524		

Notes: SD = standard deviation. SE = standard error. PIP = posterior inclusion probability. Bayesian model averaging (BMA) employs the priors suggested by Eicher *et al.* (2011). The frequentist check (OLS) includes the variables recognized by BMA to comprise the best model and is estimated using standard errors clustered at the study and country levels. Frequentist model averaging (FMA) employs Mallow's weights (Hansen, 2007) using the orthogonalization of the covariate space suggested by Amini & Parmeter (2012). All variables are described in Table 4. Additional details on the BMA exercise can be found in the Appendix.

therefore, are found to be systematically exaggerated due to publication bias even if we control for various data and method characteristics of the individual studies. The inclusion of these controls lowers the estimated magnitude of publication bias reported in Table 3, but only slightly (the coefficient decreases from 0.8 to approximately 0.7). At the same time, our results suggest that, after controlling for publication bias and other aspects of study design, the difference between the estimated short-run and *long-run elasticity* is, on average, close to zero.

Data characteristics. The evidence on the effect of *data disaggregation* is consistent with the prevalent opinion in the literature following mostly Hummels (1999): higher disaggregation of data leads to more homogenous products and brings higher international substitutability. The bias is thus believed to originate in the heterogeneity of goods included in aggregated categories. Our results suggest that the effect is statistically important; still, the economic importance of the effect seems relatively low (the coefficient equals 0.2 in Table 5) in comparison to other sources of heterogeneity. In the majority of the studies in our dataset, data disaggregation and *results disaggregation* have the same value, but some of the studies use disaggregated data while reporting aggregated elasticities. Imbs & Mejean (2015) show that if elasticities are heterogeneous, the aggregate elasticity of substitution is given by an adequately weighted average of good-specific elasticities. We find that, contrary to Imbs & Mejean (2015), the output data granularity (disaggregation of resulting elasticities) is negatively associated with the reported elasticities.

Data frequency is another systematic factor that influences the estimates of the Armington elasticity. Table 2 shows that elasticities estimated using datasets with annual and quarterly frequencies tend to be larger than when monthly data are employed for estimation. Hertel *et al.* (1997) states that, in general, with lower data frequencies, more inelastic estimates are to be expected, as adjustment patterns become lost in aggregation. When we control for publication bias and other aspects of study design, the elasticities estimated with *quarterly data* appear to be robustly higher—by approximately 1.5—than what any other data frequencies produce.

Our results also corroborate the importance of using *cross-sectional* data versus time-series data. When the time dimension of the data is accounted for, the estimated elasticities tend to be smaller by at least 1.5 (observation-weighted BMA in Table A2 puts the cross-sectional coefficient at 1.6 and the unweighted baseline at 2.2), although the length of the time series does not seem to play a substantial additional role. Studies with a small number of observations produce small estimates of the elasticity, which might reflect small-sample bias. Although some commentators in the literature note that the estimates of the Armington elasticity are increasing in time (Schurenberg-Frosch, 2015; Welsch, 2008; Hubler & Pothen, 2017), we argue that once the study design is controlled for, no such pattern remains.

Structural variation. Given that the majority of studies deal with either the United States or Europe (and the economies of the United States, Germany, and France alone account for approximately 1,500 observations in our sample), our data sample suffers from a lack of cross-country variation, and the conclusions concerning the country-level variables should be taken with a grain of salt. Indeed, most of the country-level variables lose statistical significance in the frequentist check, where standard errors are clustered at the country level. With that disclaimer in mind, we briefly describe the results. Zhang & Verikios (2006) argues that small countries feature relatively low Armington elasticities because they are rather import-dependent and tend to boast highly specialized industries. The negative coefficient of variable *market size* across all models, albeit small, is not in line with this argument. Our results suggest that larger markets tend to have rather smaller Armington elasticities; some evidence from our weighted specification suggests that developed countries also feature smaller elasticities. Zhang & Verikios (2006), on the other hand, argue that developing countries have underdeveloped domestic industries that are often unable to compete with imports, which should contribute to smaller Armington elasticities.

Blonigen & Wilson (1999) find evidence that barriers to entry lower the elasticity of substitution between domestic and foreign goods. Our results indicate that barriers to trade, tariff or non-tariff related, have either economically unimportant or statistically insignificant effects on the reported Armington elasticity. This evidence is, however, not entirely conclusive because the baseline and alternative prior specification (first panel of Table A2) offer an unintuitive sign for the coefficient of non-tariff barriers, even though the evidence for this coefficient is rather weak. Volatility in the exchange rate, moreover, shows a statistically and economically important positive effect. Finally, we do not find our proxy for home bias or the spread of Internet use important for the magnitude of the elasticity of substitution.

Estimation techniques. The evidence on the systematic importance of model and estimation techniques is rather mixed. The baseline unweighted specification does not offer a strong case for any of the model or method choices to have a systematic impact on the estimated elasticity. The baseline specification suggests that lower reported elasticities are associated with OLS and that larger elasticities are associated with a control for endogeneity. In the study-weighted specification, the usage of the *static model*, *nonlinear model*, and *GMM* seem to have not only statistically but also economically important effects. The static model is often used to capture the long-run effect using OLS. Non-linear models also typically apply GMM to capture the long-run Armington elasticity. Goldstein & Khan (1985) argue that single-equation estimation techniques commonly generate price elasticities biased downward because they constitute a weighted average of the actual demand and supply elasticity. GMM is also commonly applied to help with endogeneity issues in the estimation procedures (Aspalter, 2016). The non-linear estimation technique is applied differently in different studies, but many follow Feenstra *et al.* (2018), which is currently considered the best practice in the literature. Next, Huchet-Bourdon & Pishbahar (2009) show that estimation ignoring *import tariffs*, for example, may produce biased results. Our results suggest that if a control for tariffs is not included in the estimation, elasticities indeed tend to be systematically larger. The control for *seasonality* in the estimation model, on the other hand, seems to diminish the estimated elasticities.

Publication characteristics. Our results indicate a remarkably strong association between publication characteristics (publication in a peer-reviewed journal, the impact factor of the outlet, and the number of citations) and the reported results of a study. We interpret this association as the effect of quality on the results: higher-quality studies tend to report substantially larger Armington elasticities. However, a qualification is in order. Publication bias can influence this association, for example, if peer-reviewed journal and generally better outlets prefer larger elasticities. Moreover, if researchers calibrating their models also prefer large elasticities, they may preferentially cite studies that deliver such estimates. We experimented with adding additional interactions of the standard error and the publication characteristics, but none proved important. Therefore, we find no evidence that the association between quality and the size of the reported Armington elasticity is driven by publication bias.

The results presented so far suggest that publication bias exaggerates the mean Armington elasticity but that many questionable data and method choices may result in a downward bias (and some may also do so in an upward bias). Finally, studies of higher quality tend to report larger elasticities. In the remainder of this section, we attempt to put all of this information together and derive the mean Armington elasticity implied by the literature and corrected for all biases related to publication selection, data and methods, and quality. To this end, we construct a synthetic study that uses all 3,524 estimates but gives each estimate a weight based on our baseline BMA results and our definition of a "best practice" study. Such best practices are inherently subjective, depending on our decision about the best choices for data, method, and publication choices. We execute several robustness checks to ensure that the important results hold in different but plausible settings.

The best-practices estimate is a result of a linear combination of the BMA coefficients from the baseline specification in Table 5 and our chosen values for the respective variables. We prefer the most precise estimates (and, as a consequence, no publication bias), so we plug in zero for the standard error. We focus on the long-run elasticity because the long-run effect is the area in which most policy makers are interested. We prefer the full disaggregation of data and results. We also prefer panel data, the maximum size of the dataset, and the maximum length of the data period. We plug in the maximum for the midyear of data used in individual studies because we want to give more weight to recent information. We also prefer studies published in peer-reviewed journals with a large impact factor and those with a high number of citations.[1]

To estimate the best-practice mean elasticity for our entire sample, we evaluate all the structural variables, including the country-specific variables at the sample mean. The estimates for individual countries in Table 6, on the other hand, are estimated using country-specific values for the cross-country variables (these include *developed* economies, *market size*, *tariffs*, *non-tariff barriers*, *FX volatility*, *national pride*, and *internet usage*). We prefer estimates obtained using nonlinear models and the GMM estimator and estimations controlling for import constraints and seasonality. We also prefer annual data because they abstract from short-term fluctuations that might obscure the estimates of the elasticity; in meta-analysis, there is no lack of power that would force us to move to a higher (and noisier) frequency.

[1]Three variables display large outliers: the number of *citations*, *data size*, and *impact factor*. To ensure that our estimates are not driven by the outliers, we take the 95^{th} percentile of the value of these variables in our dataset. If we took the maximum, our resulting estimate of the elasticity would be larger.

Table 6: Armington elasticities implied for individual countries

	Mean	95% conf. int.	
Australia	3.2	1.8	4.6
Austria	2.8	1.5	4.0
Belgium	2.8	1.4	4.1
Brazil	3.2	1.4	5.0
Bulgaria	3.1	1.6	4.6
Colombia	3.4	1.4	5.4
Cyprus	3.0	1.5	4.6
Czech Republic	3.6	2.3	4.9
Denmark	2.7	1.3	4.0
Estonia	3.0	1.6	4.5
Finland	2.7	1.4	4.0
France	2.7	1.4	4.0
Germany	2.7	1.5	4.0
Greece	2.8	1.5	4.0
Hungary	3.1	1.8	4.4
Ireland	2.8	1.5	4.1
Italy	2.7	1.4	4.0
Japan	3.2	2.1	4.3
Latvia	3.0	1.6	4.4
Lesotho	3.9	1.8	5.9
Lithuania	3.0	1.6	4.3
Luxembourg	3.0	1.5	4.6
Malta	3.1	1.5	4.8
Netherlands	2.6	1.4	3.8
Poland	3.0	1.7	4.3
Portugal	2.9	1.5	4.4
Romania	3.1	1.7	4.5
Russia	3.4	1.6	5.1
Slovak Republic	3.1	1.6	4.5
Slovenia	2.9	1.5	4.4
South Africa	3.4	1.6	5.3
Spain	2.7	1.4	4.0
Sweden	2.7	1.5	3.9
Thailand	3.1	1.2	5.0
United Kingdom	2.9	1.8	4.1
United States	2.4	1.1	3.7
Uruguay	3.4	1.6	5.2
Euro area	2.6	1.5	3.7
All countries	2.9	1.3	4.4

Notes: The table presents the mean estimates of the Armington elasticity implied by the Bayesian model averaging exercise and our definition of best practices. The confidence intervals are approximate and constructed using the standard errors estimated by OLS.

Table 6 reports the results of our best-practice exercise. The elasticities implied for different countries after correction for publication and other biases range from 2.7 to 3.4; the mean estimate for the entire world is 2.9. The mean elasticities would be even larger if we preferred quarterly data instead of annual data, pushing the corrected mean to 4 for the overall sample (3.7 for the European Union and 3.6 for the United States). The elasticity would also be larger if we took the maxima instead of the 95% percentiles for data size, the impact factor of the outlet, and the number of citations, and if we preferred TSLS instead of GMM and cross-sectional data instead of panel data. The 95% confidence intervals, although quite wide, imply that the aggregate Armington elasticity of substitution is above 1.3 with a 95% probability. This finding resonates with Imbs & Mejean (2015) and their call for elasticity optimism.

5 Concluding Remarks

We present the first quantitative synthesis of the vast empirical literature on the elasticity of substitution between domestic and foreign goods, also known as the macro-level Armington elasticity (Feenstra *et al.*, 2018). The elasticity is a key parameter for both international trade and international macroeconomics. In computable general equilibrium models commonly used to evaluate trade policy, the elasticity of substitution governs the effects of newly introduced tariffs, among other things. In open-economy dynamic stochastic general equilibrium models used by many central banks to evaluate and plan monetary policy, the elasticity of substitution governs the strength and speed of the exchange rate pass-through.

Consider, for example, two European central banks that, in the wake of the Great Recession, introduced exchange rate floors to limit their currencies' appreciation against the euro: the Swiss National Bank and the Czech National Bank. Currency depreciation (relative to the counterfactual without the currency floor) produces two effects relevant to the aggregate price level. First, imported goods become more expensive, which directly increases inflation. With a large elasticity of substitution between domestic and foreign goods, however, this effect becomes muted and delayed because consumers shift toward relatively cheaper domestic goods. Second, currency depreciation stimulates the economy by encouraging exports and discouraging imports, which raises inflation in the medium term. With a larger elasticity of substitution, this effect strengthens. Because both the Swiss National Bank and the Czech National Bank use open-economy dynamic stochastic general equilibrium models for policy analysis, the assumed size of the Armington elasticity played an important (if implicit) role in the decision on when and how to implement the exchange rate floor.

We collect 3,524 previously reported estimates of the Armington elasticity, which makes our paper one of the largest meta-analyses conducted in economics so far. Ioannidis *et al.* (2017) survey 159 economics meta-analyses and report that the mean analysis uses 400 estimates. We also construct 34 variables that reflect the context in which researchers obtain their estimates. Several characteristics of the studies and individual estimates might affect the results systematically, as was claimed by previous studies: for example, the level of data aggregation (Hummels, 1999), data frequency (Hertel *et al.*, 1997), the distinction between short- and long-run effects (Gallaway *et al.*, 2003), and estimation strategy (Cassoni & Flores, 2008). Other studies stress the potential importance of structural determinants of the Armington elasticity at the industry or country level (Blonigen & Wilson, 1999; Lopez & Pagoulatos, 2002; McDaniel & Balistreri, 2003). Our aim in this paper is to assign a pattern to the great variation observed among the reported estimates of the elasticity.

Our results, based primarily on the Bayesian and frequentist model averaging that address the model uncertainty inherent to meta-analysis, suggest that the single most important variable for the explanation of the variation in the reported elasticities is the standard error. Large standard errors are associated with large estimates, which is inconsistent with the property of almost all techniques used to estimate the elasticity: the ratio of the estimate to its standard error has a t-distribution (or other symmetrical distribution). The property implies that estimates and standard errors should be statistically independent quantities. The violation of independence suggests a preference for large estimates that compensate for large standard errors, which we further corroborate by employing the new non-linear techniques by Ioannidis *et al.* (2017), Andrews & Kasy (2019), and Furukawa (2019). This publication selection results in an exaggeration of long-run estimates by more than 50% on average. After correcting for publication bias, we observe no systematic difference between the reported sizes of short- and long-run elasticities.

We find that a large part of the variation in reported elasticities can be explained by data characteristics. In particular, data aggregation, low frequency, and small samples are typically associated with smaller estimates. In contrast, the use of cross-sectional data tends to result in large estimates: on average, greater by 1.7 than when time-series data are used. After controlling for these data characteristics, we find no association between data age and the size of the reported elasticity. Thus, the larger elasticities reported by more recent studies are typically given by the move from time-series to cross-sectional data analysis. Our results also suggest that study quality (roughly approximated by publication status, the RePEc impact factor of the outlet, and the number of citations) is robustly associated with study results: higher-quality studies tend to report larger elasticities. We use all of this information to construct a synthetic study that draws on all 3,524 estimates but gives more weight to better estimates and controls for publication bias. While defining "better" estimates is inevitably subjective, we argue that, given plausible definitions of best practice, the best possible guess concerning the aggregate Armington elasticity is close to 3—at least based on the empirical research of the last 50 years since Armington (1969).

Three qualifications of our results are in order. First, the 3,524 estimates that we collect are not independent but likely correlated within studies and countries. We try to account for this problem by using Bayesian hierarchical analysis and clustering the standard errors (where possible) at the level of both studies and countries. Second, while we control for 34 aspects of studies and estimates, one could still add more variables, as the pool of potential controls is unlimited. We omit industry-level variables, for example, because their inclusion would cause serious collinearity. But the entire dataset together with the code is provided in the online appendix and allows interested researchers to focus on different subsets of variables. Third, while we do our best to include all studies reporting an estimate of the macro-level Armington elasticity, we might have missed some. This potential omission does not create a bias in meta-analysis as long as it is not conditional on study results.

References

Alaouze, C. M. (1977): "Estimates of the elasticity of substitution between imported and domestically produced goods classified at the input-output level of aggregation." *Centre of Policy Studies/IMPACT Project Working Paper o-13/1977*, Victoria University, Centre of Policy Studies/IMPACT Centre, Melbourne.

Alaouze, C. M., J. S. Marsden, & J. Zeitsch (1977): "Estimates of the elasticity of substitution between imported and domestically produced commodities at the four digit ASIC level." *Centre of Policy Studies/IMPACT Project Working Paper o-11/1977*, Victoria University, Centre of Policy Studies/IMPACT Centre, Melbourne.

Amini, S. M. & C. F. Parmeter (2012): "Comparison of model averaging techniques: Assessing growth determinants." *Journal of Applied Econometrics* **27(5)**: pp. 870–876.

Anderson, E., M. A. J. d'Orey, M. Duvendack, & L. Esposito (2018): "Does Government Spending Affect Income Poverty? A Meta-regression Analysis." *World Development* **103(C)**: pp. 60–71.

Andrews, I. & M. Kasy (2019): "Identification of and correction for publication bias." *American Economic Review* **(forthcoming)**.

Armington, P. S. (1969): "A theory of demand for products distinguished by place of production." *IMF Staff Papers* **16(1)**: pp. 159–178.

Aspalter, L. (2016): "Estimating industry-level armington elasticities for EMU countries." *Working Paper 217/2016*, Department of Economics, Vienna University of Economics and Business, Vienna.

Auer, R. A. & R. S. Schoenle (2016): "Market structure and exchange rate pass-through." *Journal of International Economics* **98(C)**: pp. 60–77.

Backus, D., P. Kehoe, & F. Kydland (1994): "Dynamics of the trade balance and the terms of trade: The j-curve?" *American Economic Review* **84(1)**: pp. 84–103.

Balistreri, E. J., A. Al-Qahtani, & C. A. Dahl (2010): "Oil and petroleum product Armington elasticities: A new-geography-of-trade approach to estimation." *The Energy Journal* **31(3)**: pp. 167–179.

Bergin, P. (2006): "How well can the new open economy macroeconomics explain the exchange rate and current account." *Journal of International Money and Finance* **25(5)**: pp. 675–701.

Bilgic, A., S. King, A. Lusby, & D. F. Schreiner (2002): "Estimates of U.S. regional commodity trade elasticities of substitution." *Journal of Regional Analysis and Policy* **32(2)**: pp. 79–98.

Blonigen, B. A. & W. W. Wilson (1999): "Explaining Armington: What determines substitutability between home and foreign goods?" *The Canadian Journal of Economics / Revue canadienne d'Economique* **32(1)**: pp. 1–21.

Broda, C. & D. E. Weinstein (2006): "Globalization and the gains from variety." *The Quarterly Journal of Economics* **121(2)**: pp. 541–585.

Brodeur, A., M. Le, M. Sangnier, & Y. Zylberberg (2016): "Star Wars: The Empirics Strike Back." *American Economic Journal: Applied Economics* **8(1)**: pp. 1–32.

Card, D., J. Kluve, & A. Weber (2018): "What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations." *Journal of the European Economic Association* **16(3)**: pp. 894–931.

Cassoni, A. & M. Flores (2008): "Methodological shortcomings in estimating Armington elasticities." *MPRA Working Paper 34544/2008*, Munich Personal RePEc Archive, Munich.

Cassoni, A. & M. Flores (2010): "Armington elasticities: Estimates for Uruguayan manufacturing sectors." *Serie Documentos de Trabajo (working papers) 7/2010*, Departamento de Economía, Facultad de Ciencias Sociales, Universidad de la República.

Corado, C. & J. de Melo (1983): "An ex-ante model for estimating the impact on trade flows of a country's accession to a customs union." *Departmental Working Paper DRD67/1983*, Development Research Department, World Bank.

Corbo, V. & C. Osbat (2013): "Trade adjustment in the European Union - A structural estimation approach." *ECB Working Paper 1535/2013*, European Central Bank, Frankfurt a. M.

Costinot, A. & A. Rodriguez-Clare (2014): "Trade Theory with Numbers: Quantifying the Consequences of Globalization." In G. Gopinath, E. Helpman, & K. S. Rogoff (editors), "Handbook of International Economics," volume 4. Elsevier Ltd., Oxford.

Egger, M., G. D. Smith, M. Schneider, & C. Minder

(1997): "Bias in meta-analysis detected by a simple, graphical test." *British Medical Journal* **315(7109)**: pp. 629–634.

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.

Engler, P. & J. Tervala (2016): "Welfare effects of TTIP in a DSGE model." *Discussion Paper 2016/17*, Free University Berlin, School of Business & Economics.

Faria, W. R. & E. A. Haddad (2014): "Estimacao das Elasticidades de Substituicao do comercio regional do Brasil [Estimations of Elasticities of regional trade in Brazil]." *Nova Economia* **24(1)**: pp. 141–168.

Feenstra, R. C., P. Luck, M. Obstfeld, & K. N. Russ (2018): "In search of the Armington elasticity." *The Review of Economics and Statistics* **100(1)**: pp. 135–150.

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

Francois, J. F. & S. Kaplan (1996): "Aggregate Demand Shifts, Income Distribution, and the Linder Hypothesis." *The Review of Economics and Statistics* **78(2)**: pp. 244–250.

Furukawa, C. (2019): "Publication bias under aggregation frictions: Theory, evidence, and a new correction method." *Working paper*, MIT.

Gallaway, M. P., C. A. McDaniel, & S. A. Rivera (2003): "Short-run and long-run industry-level estimates of US Armington elasticities." *The North American Journal of Economics and Finance* **14(1)**: pp. 49–68.

Gan, J. (2006): "Substitutability between US domestic and imported forest products: The Armington approach." *Forest Science* **52(1)**: pp. 282–306.

Gibson, K. L. (2003): "Armington elasticities for South Africa: Long- and short-run industry level estimates." *TIPS Working Paper Series 12/2003*, Trade and Industrial Policy Strategies, Johannesburg.

Goldstein, M. & M. S. Khan (1985): "Income and price effects in foreign trade." In R. W. Jones & P. B. Kenen (editors), "Handbook of International Economics," volume 2, chapter 20, pp. 1041–1105. Amsterdam: North-Holland.

Hansen, B. (2007): "Least Squares Model Averaging." *Econometrica* **75(4)**: pp. 1175–1189.

Havranek, T. (2015): "Measuring Intertemporal Substitution: The Importance of Method Choices and Selective Reporting." *Journal of the European Economic Association* **13(6)**: pp. 1180–1204.

Havranek, T., R. Horvath, Z. Irsova, & M. Rusnak (2015): "Cross-country heterogeneity in intertemporal substitution." *Journal of International Economics* **96(1)**: pp. 100–118.

Havranek, T. & Z. Irsova (2017): "Do borders really slash trade? A meta-analysis." *IMF Economic Review* **65(2)**: pp. 365–396.

Heathcote, J. & F. Perri (2002): "Financial Autarky and International Business Cycles." *Journal of Monetary Economics* **49(3)**: pp. 601–627.

Hedges, L. V. (1992): "Modeling publication selection effects in meta-analysis." *Statistical Science* **7(2)**: pp. 246–255.

Hernandez, G. (1998): "Elasticidades de sustitucion de las importaciones para la Economia Colombiana." *Revista de Economia del Rosario* **1(2)**: pp. 79–89.

Hertel, T. W., E. Ianchovichina, & B. J. McDonald (1997): "Multi-Region General Equilibrium Modeling." In J. F. Francois & K. A. Reinert (editors), "Applied Methods for Trade Policy Analysis: A Handbook," chapter 9, p. 258–299. Cambridge: Cambridge University Press.

Hillberry, R. H., M. A. Anderson, E. J. Balistreri, & A. K. Fox (2005): "Taste parameters as model residuals: Assessing the 'fit' of an Armington trade model." *Review of International Economics* **13(5)**: p. 973–984.

Hillberry, R. H. & D. Hummels (2013): "Trade elasticity parameters for a Computable General Equilibrium model." In P. B. Dixon & D. W. Jorgenson (editors), "Handbook of Computable General Equilibrium Modeling," volume 1, chapter 18, pp. 1213–1269. Elsevier.

Hubler, M. & F. Pothen (2017): "Trade-induced productivity gains reduce incentives to impose strategic tariffs." *Economic Modelling* **61**: pp. 420–431.

Huchet-Bourdon, M. & E. Pishbahar (2009): "Armington elasticities and tariff regime: An application to European Union rice imports." *Journal of Agricultural Economics* **60(3)**: pp. 586–603.

ysis, June 15-17, 2006 in Addis Ababa, Ethiopia.

Hummels, D. (1999): "Toward a Geography of Trade Costs." *GTAP Working Paper 17/1999*, Purdue University, West Lafayette, Indiana.

IBRD (2009): "2009 Information and communications for development: Extending reach and increasing impact." *Technical report*, International Bank for Reconstruction and Development, The World Bank, Washington DC.

Imbs, J. M. & I. Mejean (2015): "Elasticity optimism." *American Economic Journal: Macroeconomics* **7(3)**: pp. 43–83.

Inglehart, R., C. Haerpfer, A. Moreno, C. Welzel, K. Kizilova, J. Diez-Medrano, M. Lagos, P. Norris, E. Ponarin, & B. Puranen (2014): "World Values Survey: All rounds - Country-pooled datafile at http://www.worldvaluessurvey.org/ WVSDocumentationWVL.jsp." *Technical report*, Madrid: JD Systems Institute.

Ioannidis, J. P., S. T. D., & D. H. (2017): "The Power of Bias in Economics Research." *Economic Journal* **127(605)**: pp. F236–F265.

Ivanova, N. (2005): "Estimation of own- and cross-price elasticities of disaggregated imported and domestic goods in Russia." *EERC Working Paper Series 13/2005*, Economics Education and Research Consortium Research Network, Moscow, Russia.

Kapuscinski, C. A. & P. G. Warr (1999): "Estimation of Armington Elasticities: An Application to the Philippines." *Departmental Working Papers 1996-08*, Australian National University, Economics RSPAS.

Kawashima, S. & D. A. P. Sari (2010): "Time-varying Armington elasticity and country-of-origin bias: from the dynamic perspective of the Japanese demand for beef imports." *Australian Journal of Agricultural and Resource Economics* **54(1)**: pp. 27–41.

Kehoe, T. J., P. S. Pujolas, & J. Rossbach (2017): "Quantitative trade models: Developments and challenges." *Staff Report 537*, Federal Reserve Bank of Minneapolis.

Kuiper, M. & F. van Tongeren (2006): "Using gravity to move Armington: An empirical approach to the small initial trade share problem in general equilibrium models." *Technical report*, Global Trade Analysis Project (GTAP). Paper prepared for the Ninth Annual Conference on Global Economic Analysis, June 15-17, 2006 in Addis Ababa, Ethiopia.

Lachler, U. (1984): "The elasticity of substitution between imported and domestically produced goods in Germany." *Kiel Working Papers 200/1984*, Kiel Institute of World Economics, Kiel.

Ley, E. & M. F. Steel (2009): "On the Effect of Prior Assumptions in Bayesian Model Averaging with Applications to Growth Regression." *Applied Econometrics* **24**: pp. 651–674.

Linder, S. B. (1961): *An essay on trade and transformation.* Almqvist & Wiksells boktryckeri: Uppsala.

Lopez, E. & E. Pagoulatos (2002): "Estimates and Determinants of Armington Elasticities for the U.S. Food Industry." *Journal of Industry, Competition and Trade* **2(3)**: pp. 247–258.

Lozano Karanauskas, C. (2004): "Elasticidades de sustitucion Armington para Colombia." *Archivos de Economia 271/2004*, Direccion de Estudios Economicos, Departamento Nacional de Planeacion, Republica de Colombia.

Lundmark, R. & S. Shahrammehr (2011a): "Forest biomass and Armington elasticities in Europe." *Biomass and Bioenergy* **35(1)**: pp. 415–420.

Lundmark, R. & S. Shahrammehr (2011b): "Sweden's import substitution possibilities for roundwood." *Scandinavian Journal of Forest Research* **26(2)**: pp. 146–153.

Lundmark, R. & S. Shahrammehr (2012): "Armington elasticities and induced industrial roundwood specialization in Europe." *Journal of Natural Resources Policy Research* **4(3)**: pp. 161–170.

Marquez, J. (2002): *Estimating Trade Elasticities.* Boston: Kluwer.

McCallum, J. (1995): "National borders matter: Canada–U.S. regional trade patterns." *American Economic Review* **85**: pp. 615–623.

McCloskey, D. N. & S. T. Ziliak (2019): "What Quantitative Methods Should We Teach to Graduate Students? A Comment on Swann's Is Precise Econometrics an Illusion?" *Journal of Economic Eduation* **(forthcoming)**.

McDaniel, C. A. & E. J. Balistreri (2003): "A discussion on Armington trade substitution elasticities." *Economie Internationale* **94-95(2)**: pp. 301–313.

Mohler, L. & M. Seitz (2012): "The gains from variety in the European Union." *Review of World Eco-*

nomics **148(3)**: pp. 475–500.

Nemeth, G., L. Szabo, & C. Juan-Carlos (2011): "Estimation of Armington elasticities in a CGE economy-energy-environment model for Europe." *Economic Modelling* **28(4)**: pp. 1993–1999.

Nganou, J.-P. (2005): "Estimates of Armington parameters for a landlocked economy." *MPRA Working Paper 31459/2005*, Munich Personal RePEc Archive, Munich.

Ogundeji, A. A., J. andre, & D. Uchezuba (2010): "Econometric estimation of Armington elasticities for selected agricultural products in South Africa." *South African Journal of Economic and Management Sciences* **13(2)**: pp. 123–134.

Olekseyuk, Z. & H. Schurenberg-Frosch (2016): "Are Armington elasticities different across countries and sectors? A European study." *Economic Modelling* **55**: pp. 328–342.

Parsley, D. C. & S.-J. Wei (2001): "Explaining the border effect: the role of exchange rate variability, shipping costs, and geography." *Journal of International Economics* **55(1)**: pp. 87–105.

Raftery, A. E., D. Madigan, & J. Hoeting (1997): "Bayesian model averaging for linear regression models." *Journal of the American Statistical Association* **92**: pp. 179–191.

Reinert, K. A. & D. W. Roland-Holst (1992): "Armington elasticities for United States manufacturing sectors." *Journal of Policy Modelling* **14(5)**: pp. 631–639.

Reinert, K. A. & C. R. Shiells (1991): "Estimated elasticities of substitution for analysis of a North American Free Trade Area." *Technical report*, Presented at the Conference on Modeling North American Economic Integration, March 1-2, 1991, Federal Reserve Bank of Minneapolis.

Rossi, P. E., G. M. Allenby, & R. McCulloch (2005): "Markov Chain Monte Carlo Methods." In V. Barnett, J. S. Hunter, & D. G. Kendall (editors), "Bayesian Statistics and Marketing," volume 1 of *Wiley Series in Probability and Statistics*, chapter 3, pp. 49–102. John Wiley & Sons, Ltd.

Saikkonen, L. (2015): "Estimation of substitution and transformation elasticities for South African trade." *UNU-WIDER Working Paper 104/2015*, United Nations University, World Institute for Development, Helsinki, Finland.

Saito, M. (2004): "Armington Elasticities in Intermediate Inputs Trade: A Problem in Using Multilateral Trade Data." *The Canadian Journal of Economics* **37(4)**: pp. 1097–1107.

Sauquet, A., F. Lecocq, P. Delacote, S. Caurla, A. Barkaoui, & S. Garcia (2011): "Estimating Armington elasticities for sawnwood and application to the French Forest Sector Model." *Resource and Energy Economics* **33(4)**: pp. 771–781.

Schurenberg-Frosch, H. (2015): "We Could Not Care Less About Armington Elasticities–But Should We? A Meta-Sensitivity Analysis of the Influence of Armington Elasticity Misspecification on Simulation Results." *Ruhr Economic Papers 594/2015*, Ruhr-Universität Bochum, Germany.

Shiells, C. R. & K. A. Reinert (1993): "Armington models and the terms-of-trade effects: Some econometric evidence for North America." *The Canadian Journal of Economics* **26(2)**: pp. 299–316.

Shiells, C. R., R. M. Stern, & A. V. Deardorff (1986): "Estimates of the elasticities of substitution between imports and home goods for the United States." *Weltwirtschaftliches Archiv* **122(3)**: pp. 497–519.

Stanley, T. & H. Doucouliagos (2010): "Picture This: A Simple Graph That Reveals Much Ado About Research." *Journal of Economic Surveys* **24(1)**: pp. 170–191.

Stanley, T. D. (2001): "Wheat from Chaff: Meta-analysis as Quantitative Literature Review." *Journal of Economic Perspectives* **15(3)**: pp. 131–150.

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

Stanley, T. D., J. S. B., & D. H. (2010): "Could It Be Better to Discard 90Paradox." *The American Statistician* **64**: pp. 70–77.

Steel, M. F. J. (2019): "Model Averaging and its Use in Economics." *Journal of Economic Literature* **(forthcoming)**.

Tourinho, O., A. Fontes, H. Kume, & A. C. de Souza Pedroso (2003): "Armington Elasticities for Brazil – 1986-2002: New Estimates." *IPEA Discussion Papers 974/2003*, Institute for Applied Economic Research, Rio de Janeiro, Brazil.

Tourinho, O., A. Fontes, H. Kume, & A. C. de Souza Pedroso (2010): "Armington Elasticities for Brazil." In C. de Miguel, J. D. Lima, P. Giordano, J. Guzmán, A. Schuschny, & M. Watanuki (editors), "Modeling Public Policies in Latin America and the Caribbean," chapter 9, pp. 285–316. United Nations Publication, Santiago, Chile.

Trefler, D. (1995): "The case of the missing trade and other mysteries." *American Economic Review* **85(5)**: pp. 1029–46.

Turner, K., S. Y. Ha, G. Hewings, P. McGregor, & K. Swales (2012): "Econometric estimation of Armington import elasticities for a regional CGE model of the Illinois economy." *Economic Systems Research* **24(1)**: pp. 1–19.

Warr, P. G. & E. A. Lapiz (1994): "Estimation of Elasticities of Substitution/Transformation between Domestic and Imported/Exported Commodities." *Technical report*, mimeo.

WB (2017): "World Development Indicators (WDI): Public database." *Database*, The World Bank: Washington, DC. Online at https://data.worldbank.org/ [Accessed Sept-5, 2017].

Welsch, H. (2006): "Armington elasticities and induced intra-industry specialization: The case of France, 1970-1997." *Economic Modelling* **23(3)**: pp. 556–567.

Welsch, H. (2008): "Armington elasticities for energy policy modeling: Evidence from four European countries." *Energy Economics* **30(5)**: pp. 2252–2264.

Wolf, H. (2000): "Intranational home bias in trade." *Review of Economics and Statistics* **82**: p. 555–563.

Zeugner, S. & M. Feldkircher (2015): "Bayesian model averaging employing fixed and flexible priors: The BMS package for R." *Journal of Statistical Software* **68(4)**: pp. 1–37.

Zhang, X.-g. & G. Verikios (2006): "Armington Parameter Estimation for a Computable General Equilibrium Model: A Database Consistent Approach." *Economics Discussion/Working Papers 10/2006*, The University of Western Australia, Department of Economics.

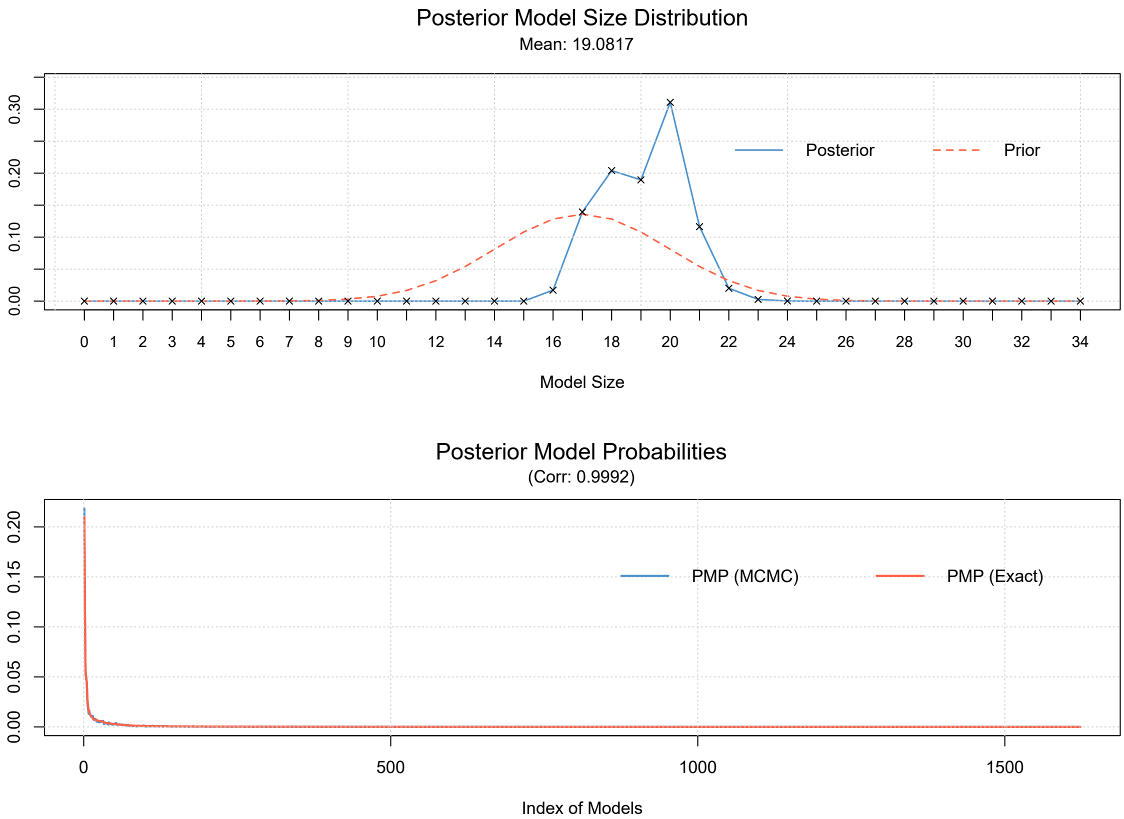
Appendix: Diagnostics of BMA and Robustness Checks

Table A1: Diagnostics of the baseline BMA estimation

Mean no. regressors	*Draws*	*Burn-ins*	*Time*	*No. models visited*
19.0817	$3 \cdot 10^5$	$1 \cdot 10^5$	30.26075 secs	28,730
Modelspace	*Visited*	*Topmodels*	*Corr PMP*	*No. obs.*
$1.7 \cdot 10^{10}$	0.00017%	100%	0.9992	3,524
Model prior	*g-prior*	*Shrinkage-stats*		
Uniform	UIP	Av = 0.9997		

Notes: We employ the priors suggested by Eicher *et al.* (2011), who recommend using 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 in the data). The results of this BMA exercise are reported in Table 5.

Figure A1: Model size and convergence of the baseline BMA estimation



Notes: The figure depicts the posterior model size distribution and the posterior model probabilities of the BMA exercise reported in Table 5.

Table A2: Why elasticities vary (alternative priors and weights)

Response variable:	Alternative BMA prior			Study-weighted BMA			Precision-weighted BMA		
Armington elasticity	Post. mean	Post. SD	PIP	Post. mean	Post. SD	PIP	Post. mean	Post. SD	PIP
Constant	-1.40	NA	1.00	-0.02	NA	1.00	1.05	NA	1.00
SE * Long-run effect	0.73	0.02	1.00	0.60	0.02	1.00	0.56	0.04	1.00
Long-run effect	0.00	0.03	0.05	0.04	0.10	0.21	0.02	0.16	0.03
Data characteristics									
Data disaggregation	0.19	0.04	0.99	0.36	0.03	1.00	-0.11	0.01	1.00
Results disaggregation	-0.23	0.04	1.00	-0.43	0.03	1.00	0.00	0.00	0.03
Monthly data	-0.57	0.13	0.99	-0.44	0.11	0.99	-0.37	0.06	1.00
Annual data	-1.15	0.12	1.00	-0.98	0.08	1.00	-0.43	0.11	0.99
Time series	0.56	0.13	0.99	0.00	0.04	0.03	0.01	0.02	0.09
Cross-section	2.23	0.19	1.00	1.64	0.12	1.00	0.45	0.19	0.95
Data period	0.03	0.00	1.00	0.03	0.00	1.00	0.02	0.00	1.00
Data size	0.32	0.02	1.00	0.41	0.02	1.00	0.04	0.01	0.98
Midyear	0.00	0.00	0.06	-0.02	0.00	0.99	0.04	0.00	1.00
Structural variation									
Secondary sector	0.00	0.01	0.02	-0.03	0.07	0.22	0.19	0.03	1.00
Tertiary sector	0.00	0.04	0.03	-0.54	0.14	0.99	0.06	0.06	0.61
Developed countries	0.00	0.03	0.03	-0.45	0.11	0.99	-0.65	0.09	1.00
Market size	-0.12	0.02	1.00	-0.15	0.02	1.00	0.00	0.00	0.02
Tariffs	0.03	0.01	1.00	0.03	0.01	0.99	0.00	0.00	0.04
Non-tariff barriers	0.25	0.19	0.70	0.08	0.12	0.37	0.00	0.02	0.04
FX volatility	0.30	0.07	0.99	0.16	0.04	0.99	0.09	0.10	0.52
National pride	0.00	0.03	0.02	0.00	0.02	0.02	0.01	0.04	0.04
Internet usage	0.00	0.00	0.07	0.00	0.00	0.02	-0.01	0.01	0.69
Estimation technique									
Static model	0.00	0.02	0.03	-0.27	0.07	1.00	-0.68	0.06	1.00
Distributed lag and trend model	0.00	0.01	0.02	-0.03	0.09	0.15	-0.79	0.10	1.00
Partial adjustment model	-0.01	0.04	0.08	0.00	0.02	0.02	-0.01	0.03	0.06
First-difference model	0.00	0.03	0.04	-0.19	0.19	0.56	0.00	0.01	0.02
Nonlinear model	-0.02	0.08	0.07	0.53	0.22	0.92	0.00	0.02	0.02
OLS	-0.18	0.19	0.60	0.00	0.01	0.02	0.48	0.06	1.00
CORC	-0.06	0.15	0.19	0.00	0.03	0.04	0.25	0.06	0.99
TSLS	0.38	0.20	0.85	0.00	0.02	0.02	0.00	0.03	0.03
GMM	-0.04	0.13	0.13	-0.92	0.13	1.00	0.48	0.08	1.00
Import constraint	0.50	0.17	0.96	0.65	0.11	1.00	0.00	0.03	0.03
Seasonality	-0.55	0.11	1.00	0.00	0.01	0.02	0.71	0.09	1.00
Publication characteristics									
Impact factor	0.29	0.22	0.70	0.00	0.03	0.028	0.71	0.13	1.000
Citations	0.54	0.05	1.00	0.47	0.04	1.000	-0.10	0.03	0.995
Published	0.58	0.09	1.00	0.88	0.08	1.000	0.27	0.06	0.997
Studies	42			42			42		
Observations	3,524			3,524			3,524		

Notes: SD = standard deviation. SE = standard error. PIP = posterior inclusion probability. All three panels represent results of Bayesian model averaging (BMA). The first panel employs an alternative model prior, the beta-binomial prior advocated by Ley & Steel (2009); Zellner's g prior is set according to Fernandez et al. (2001). The two weighted specifications employ priors suggested by Eicher *et al.* (2011), who recommend using the uniform model prior and the unit information prior. The study-weighted specification represents BMA applied to data weighted by the inverse of the number of observations reported per study, precision-weighted specification represents BMA applied to data weighted by the inverse of the standard error. All variables are described in Table 4.