

**HOW DOES DATA AGGREGATION IMPACT  
ELASTICITY OF SUBSTITUTION ESTIMATION?  
EVIDENCE FROM A NEW ELASTICITY DATABASE**

Michael Schrammel

Samantha Schreiber

**ECONOMICS WORKING PAPER SERIES**  
Working Paper 2023–10–B

U.S. INTERNATIONAL TRADE COMMISSION  
500 E Street SW  
Washington, DC 20436

October 2023

The authors thank Saad Ahmad, David Riker and Peter Herman for helpful comments and suggestions. Office of Economics working papers are the result of ongoing professional research of USITC Staff and are solely meant to represent the opinions and professional research of individual authors. These papers are not meant to represent in any way the views of the U.S. International Trade Commission or any of its individual Commissioners.

# How does Data Aggregation Impact Elasticity of Substitution Estimation? Evidence from a New Elasticity Database

Michael Schrammel and Samantha Schreiber

Economics Working Paper 2023–10–B

October 2023

## **Abstract**

This paper uses a new elasticity of substitution database, estimated with variation in trade costs, to compare elasticities across different levels of data aggregation and regression types. We find that the OLS estimates in our database tend to be smaller than the PPML estimates on average. We also find that the magnitude of the elasticities estimated using aggregated data tend to be smaller than the elasticities estimated using disaggregated data. Additionally, we aggregate the disaggregated elasticity estimates with trade weights and compare them to the corresponding elasticities estimated with the aggregated data. The estimates aggregated by trade weight are on average larger than the estimates from the regression on the aggregated data. The estimates from the aggregate data may exhibit heterogeneity bias, implying that the trade-weighted estimates could be a better representation of the elasticities of substitution for aggregated product groups.

Michael Schrammel  
Office of Economics  
[michael.schrammel@usitc.gov](mailto:michael.schrammel@usitc.gov)

Samantha Schreiber  
Office of Economics  
[samantha.schreiber@usitc.gov](mailto:samantha.schreiber@usitc.gov)

# 1 Introduction

The elasticity of substitution (EOS), or Armington elasticity, is a key parameter used in economic models of trade policy changes. The EOS parameter describes the level of substitutability among domestic and imported varieties of a good, affecting the magnitude of changes in trade flows in response to relative price changes. Ahmad et al (2021) surveys many of the studies in the literature that estimate this parameter, including Hertel et al. (2007), Soderbery (2015, 2018), Broda and Weinstein (2006) and Ahmad and Riker (2019). The studies surveyed employ a range of econometric methods, use a variety of data sources, and estimate the elasticity at different levels of data aggregation. Ahmad et al (2021) shows that the range of estimates can vary widely by study, finding that there is little consensus in the literature on the best way to estimate the elasticity of substitution.

One topic discussed in the literature is how the level of data aggregation affects elasticity estimation. There have been several studies that have shown that elasticity of substitution estimates decrease with aggregation (for example, Imbs and Mejean 2015; Feenstra et al. 2018; McDaniel and Balistreri 2003; and Bajzik et al 2020). There are a few possible explanations for this pattern. First, the inclusion of a broader group of products at higher levels of aggregation would imply that substitutability should become smaller with more aggregate data. A line pipe of stainless steel with an outside diameter not exceeding 114.3 mm from China is likely more substitutable with the same product of same dimension produced in South Korea, compared to products within the more aggregated product grouping of tubes, pipes, and hollow profiles.

Second, as discussed in Imbs and Mejean (2015), aggregate data constrain heterogeneity present at the disaggregated tariff line level. For elasticities estimated using trade data, like this paper, changes in trade costs alter decisions about sourcing of imports at the disaggregated tariff line level. For a given U.S. Harmonized Tariff Schedule (HTS) 4-digit product,

for example, changes in trade costs over time are the average of trade cost changes in each of the 10-digit tariff line products within that 4-digit subheading. Heterogeneity bias can occur when estimating the elasticity using the aggregate data because heterogeneity from the disaggregated data is systematically pushed into the residual in the aggregate regressions.<sup>1</sup> This may result in aggregate elasticity estimates that are lower than the average of their disaggregated elasticity estimates.

In this paper we analyze the relationship between the EOS and the level of tariff line aggregation using a consistent econometric estimation method across all levels of aggregation. We employ the trade cost approach for estimating elasticities of substitution described in Riker (2020). We also build on the methods employed in Schreiber (2022) by comparing the Ordinary Least Squares (OLS) estimates, the method used in Riker (2020), to Pseudo-Poisson Maximum Likelihood (PPML) estimates. We estimate the EOS by product at the 2-digit, 4-digit, 6-digit, and 8-digit aggregation levels of the 2017 HTS and draw comparisons across levels of data aggregation. Finally, we compare estimates generated from the aggregate HTS codes with a weighted average of the disaggregated estimates, to understand potential heterogeneity bias present in the aggregate data.

In Section 2 of the paper, we outline the OLS and PPML equations and data used for the econometric estimation of the EOS estimates. In Section 3, we present our estimation results along with the key comparisons and statistics through a series of tables and graphs. Section 4 concludes the paper with a summary of the results and our thoughts on avenues for future research.

---

<sup>1</sup>The heterogeneity bias referred to in this paper is defined in Pesaran and Smith (1995) and explored in detail in Imbs and Mejean (2015).

## 2 Methodology

The trade cost method from Riker (2020) is an econometric model that estimates the elasticity of substitution with variation in trade costs over time and a set of fixed effects that control for supply-side and demand-side factors. The method is based on the gravity model in the trade literature, simplified by estimating the model on a single importing country (the United States). One major benefit of using the trade cost approach is that the fixed effects limit the data requirements needed for estimation. In other words, a researcher does not need to separately identify the consumer price index, producer prices in the exporting country, total expenditures, etc, because the fixed effects control for these omitted variables.

Equation 1 represents demand for imports of product  $j$  from country  $c$  into customs district  $d$  by individual  $i$  in time  $t$ .<sup>2</sup> It is the landed-duty paid value (LDPV), assuming a non-nested constant elasticity of substitution demand functional form:

$$v_{jcdit} = k_{jct} E_{jit} P_{jit}^{\sigma_j - 1} (p_{jct} f_{jcdt})^{1 - \sigma_j} s_{jdit}^{-\sigma_j} \quad (1)$$

The international trade cost factor,  $f_{jcdt}$ , is calculated as the ratio of LDPV to customs value (CV) ( $f = \frac{LDPV}{CV}$ ). This measure is a value equal to one absent any trade costs, and greater than one for positive trade costs, capturing costs such as international shipping rates, tariffs, and insurance charges.<sup>3</sup>  $k_{jct}$  is a demand factor that represents the quality of imports of  $j$  from country  $c$ ,  $E_{jit}$  is individual  $i$ 's total expenditure on product  $j$ ,  $P_{jit}$  is the price index of individual  $i$  for product  $j$ ,  $\sigma_j$  is the elasticity of substitution for product  $j$ ,  $p_{jct}$  is the producer price of imports from country  $c$  for product  $j$ , and  $s_{jdit}$  captures the domestic shipping costs from district  $d$  to individual  $i$ .

---

<sup>2</sup>The customs district refers to U.S. district where the imported products clear customs.

<sup>3</sup>One limitation of calculating trade costs with the ratio of the landed duty-paid value to the customs value is that we cannot estimate trade costs of zero trade flows. Future research on this topic could use an alternative method, like in Fontagné et al. (2022), to incorporate zero trade flows in the analysis.

Equation 2 is the equivalent of equation 1 after summing across individual consumers:

$$v_{jcdt} = [f_{jcdt}^{1-\sigma_j}] [k_{jct} p_{jct}^{1-\sigma_j}] \left[ \sum_{i \in \omega_{jdt}} E_{jit} P_{jit}^{\sigma_j-1} s_{jdit}^{-\sigma_j} \right] \quad (2)$$

where  $\omega_{jdt}$  is the set of individuals who consume product  $j$ , imported into customs district  $d$  in year  $t$ . Taking the natural log of equation 2 produces the OLS log-linear estimating equation:

$$\ln v_{jcdt} = \beta_j \ln f_{jcdt} + \alpha_{jct} + \gamma_{jdt} + \epsilon_{jcdt} \quad (3)$$

where  $\epsilon_{jcdt}$  is the error term for the OLS model and  $\alpha_{jct}$  and  $\gamma_{jdt}$  represent the country-year and district-year fixed effects defined by equations 4 and 5.

$$\alpha_{jct} = \ln [k_{jct} p_{jct}^{1-\sigma_j}] \quad (4)$$

$$\gamma_{jdt} = \ln \left[ \sum_{i \in \omega_{jdt}} E_{jit} P_{jit}^{\sigma_j-1} s_{jdit}^{-\sigma_j} \right] \quad (5)$$

Equation 6 is the econometric specification used for the PPML regression derived by taking the exponential of equation 3 following Schreiber (2022):

$$v_{jcdt} = e^{(\beta_j \ln f_{jcdt} + \alpha_{jct} + \gamma_{jdt})} \times \delta_{jcdt} \quad (6)$$

where all variables are the same as their OLS counterparts and  $\delta_{jcdt}$  is the PPML model error term. The elasticity of substitution is calculated using equation 7:

$$\sigma_j = 1 - \beta_j \quad (7)$$

where  $\beta_j$  is the estimated coefficient on the natural log of the trade cost factor in the OLS

and PPML regressions.

We run the econometric model at four different levels of product aggregation (HTS-2, HTS-4, HTS-6, HTS-8) using panel data sets downloaded from the U.S. International Trade Commission’s DataWeb for the years 2018–2022. The data sets contain the LDPV and CV for each HTS product code disaggregated by country of origin, customs district of entry, and year. We generate a data set of elasticity of substitution estimates for each of the products that have significant and non-negative estimates.<sup>4</sup> In the section below, we summarize the estimates and compare across levels of aggregation.

### 3 Analysis of Elasticity Estimates

#### Comparing OLS and PPML Estimates

First, we look at the number of significant non-negative EOS estimates at each level of aggregation and regression model—OLS and PPML. For the OLS estimates, 98 percent of the HTS 2-digit elasticities were significant and non-negative.<sup>5</sup> At the HTS 8-digit level, the number of significant and non-negative estimates drops to 35 percent. For the PPML estimates, 92 percent were significant and non-negative at the HTS 2-digit level and 43 percent at the HTS 8-digit level. The decrease in the number of significant elasticity estimates as we disaggregate product groups may be a result of the reduced number of observations for the narrowly-defined products at the HTS 8-digit level (e.g. a line pipe of stainless steel with an outside diameter not exceeding 114.3 mm).

Table 1 reports descriptive statistics of both OLS and PPML regressions at each level of data aggregation, after dropping all estimates that are negative or not statistically signifi-

---

<sup>4</sup>The elasticity data set is available on request.

<sup>5</sup>This paper assumes that the elasticity of substitution is greater than zero. In trade models with monopolistic competition, the EOS parameter is typically assumed to be greater than one. We did not constrain the regression, but instead dropped any negative elasticity estimates from the results.

cant. The median PPML and OLS elasticity estimates decrease as the level of aggregation increases. Consistent with the literature, this trend may in-part be caused by the removal of heterogeneity when aggregating and subsequent reduction in variation of trade costs. We also find that as the level of aggregation decreases the distribution of the EOS estimates becomes more disperse, with the distribution of PPML estimates being more disperse than that of OLS estimates. It is important to consider that the EOS has a zero lower-bound, which may explain the relatively big changes at the top end of the distribution and little to no change at the bottom end of the distribution.

Table 1: Elasticity of Substitution Descriptive Statistics

	median		min		5%		95%		max	
	OLS	PPML	OLS	PPML	OLS	PPML	OLS	PPML	OLS	PPML
HTS8										
sigma	6.00	10.92	0.55	0.07	2.71	3.77	19.90	55.90	2070.69	6274.00
standard error	(1.78)	(2.90)	(0.17)	(0.20)	(0.37)	(0.82)	(2.37)	(11.35)	(140.19)	(0.00)
HTS6										
sigma	5.76	10.45	0.13	0.07	2.65	3.75	16.23	47.67	2070.69	2208.97
standard error	(1.24)	(3.08)	(0.42)	(0.20)	(0.34)	(1.03)	(3.84)	(9.81)	(140.19)	(0.00)
HTS4										
sigma	5.22	9.53	1.67	1.69	2.45	3.52	13.03	34.18	182.99	409.32
standard error	(1.07)	(1.39)	(0.18)	(0.25)	(0.54)	(0.85)	(1.36)	(6.63)	(71.19)	(96.39)
HTS2										
sigma	4.47	9.59	0.65	1.49	2.24	3.29	10.22	24.06	11.99	63.83
standard error	(0.59)	(1.03)	(0.16)	(0.19)	(0.36)	(0.71)	(1.19)	(2.85)	(0.73)	(4.89)

Comparing the OLS estimates with PPML estimates, we see that the PPML estimates tend to be larger than their OLS counterparts. In general, there are two primary reasons why PPML estimates may differ from the OLS estimates: inclusion of zero trade flows, and the potential impact of heteroskedasticity in the data. The first reason, inclusion of zero trade flows, does not explain the difference in estimates because our PPML regression model does not include zero trade flows. As described above, this is due to a limitation of calculating trade costs using the LDPV/CV ratio. The second reason, the potential impact



of heteroskedasticity in the data, is the most relevant for the results in this paper. For the log-linear OLS method, if heteroskedasticity is present in the data, the transformed (log-linearized) error term in the regression model will be correlated with one of the explanatory variables, leading to bias (Santos Silva and Tenreyro, 2006). Exploring differences between OLS and PPML models, Borchert et al. (2020) estimated benchmark gravity estimates using an OLS model, PPML with zero trade flows model, and PPML without zero trade flows model. They find that the primary value of using PPML is to account for heteroskedasticity in the data and not inclusion of information contained in the zero trade flows. This suggests that our PPML estimates have the primary benefit of accounting for heteroskedasticity, and that inclusion of zero trade flows in the regression may not significantly change the estimates on average.

For the remainder of the paper, the comparisons of elasticity estimates at the sector level and by trade-weight use the OLS estimates. Although the PPML estimates may account for possible heteroskedasticity in the data, the OLS estimates more closely match the estimates found in the literature. This may be in part due to using data from only one importing country (the U.S.) for our estimation rather than using data that contains all bilateral trade flows. Future research could look into ways to improve the reasonableness of the PPML estimates.

### **Comparing Sector-level Estimates**

Next, we report 6-digit OLS estimates grouped by broad HTS section (table 2). Comparing across HTS sections, section 14 (natural or cultured pearls, precious stones, precious metals, imitation jewelry and coin) has the highest median and mean estimates (13.68 and 17.50, respectively). In contrast, HTS section 13 (articles of stone, plaster, cement, ceramic, and glass) has the smallest median and mean estimates at 3.45 and 3.98, respectively. Agricultural products, like HTS sections 1-3, tend to have median and mean values above the

database averages, whereas sections with more differentiated products (like apparel, footwear and miscellaneous goods such as furniture and toys) tend to have medians and means below the database averages. Within these sections, the standard deviation column shows that there is significant heterogeneity across 6-digit products.

Table 2: Elasticity Estimates by HTS Section

Section	Description	Num	Median	Mean	Std Dev
1	Animal Products	54	9.38	13.05	9.86
2	Vegetable Products	84	6.63	9.92	14.08
3	Animal and Vegetable Fats	14	7.60	9.7	6.89
4	Prepared Foodstuffs	84	5.44	6.43	5.33
5	Mineral Products	20	7.90	9.28	6.29
6	Chemical Products	339	7.27	16.05	112.36
7	Plastic	158	3.67	3.94	1.45
8	Raw hides, skins, leather	32	4.34	5.47	4.12
9	Wood/Cork	42	4.01	4.46	1.83
10	Pulp of wood	49	3.85	5.2	5.73
11	Textile and articles	372	4.49	5.03	2.16
12	Footwear, headgear	38	4.08	4.36	1.71
13	Stone, plaster, ceramic	71	3.45	3.98	2.08
14	Precious stones and metals	39	13.68	17.5	12.08
15	Base metals	268	5.14	7.46	16.48
16	Machinery	642	6.93	7.36	3.06
17	Vehicles and aircraft	103	5.93	9.48	14.86
18	Medical instruments	163	7.97	9.01	4.72
19	Arms and ammunition	12	8.99	10.13	4.94
20	Misc	109	4.12	4.63	2.02
21	Works of art	16	5.50	6.21	2.72
<b>Total</b>		<b>2709</b>	<b>5.76</b>	<b>8.11</b>	<b>40.49</b>

Our estimates can be compared to the estimates in Fontagné et al. (2022), a recent study that provides a database of 6-digit elasticity estimates using a similar trade cost approach as in this paper. One important difference between our approach and the approach in Fontagné et al. (2022) is that we only use U.S. imports data to estimate the EOS, whereas the Fontagné paper uses all bilateral trade flows. Surprisingly, the mean of the elasticity database at the

6-digit level is 8.11 for both this paper and the Fontagné paper. However, our database has a significantly higher standard deviation (40.49 compared to 8.50). In addition, the approach used by Fontagné results in more significant estimates at the 6-digit level (4,135 significant estimates compared to 2,709 estimates in our database). At the sector level, the Fontagné paper also find larger mean elasticity estimates for section 14 (precious stones), section 5 (mineral products), section 17 (vehicles and aircraft), and section 6 (chemical products), but there are differences in the ranking and size of mean estimates across other sections.

### Comparing Aggregate and Trade-weighted Estimates

Next, we aggregate the EOS estimates from the more disaggregated HTS tariff lines by trade weight and compare the results to the estimates from the regressions on the corresponding aggregated product group. This allows us to explore the accuracy of traded weighted aggregation and the aggregation bias of the EOS estimates. To calculate the trade weights, we drop the negative and statistically insignificant elasticity estimates from the aggregated and disaggregated data sets. Then, we sum the trade values for each product code by domestic district and country of origin. Following this step, we calculate the five-year—2018–2022—average trade value for each product code and use these results as our trade weights. To accurately compare estimates, we drop the elasticities from the disaggregated set for products that are the sole good found within the more aggregated grouping.<sup>6</sup> If these estimates remained in the analysis, they would skew the results because they would make it appear as if the trade weighted elasticity was exactly equal to the estimated elasticity for the aggregated product group.

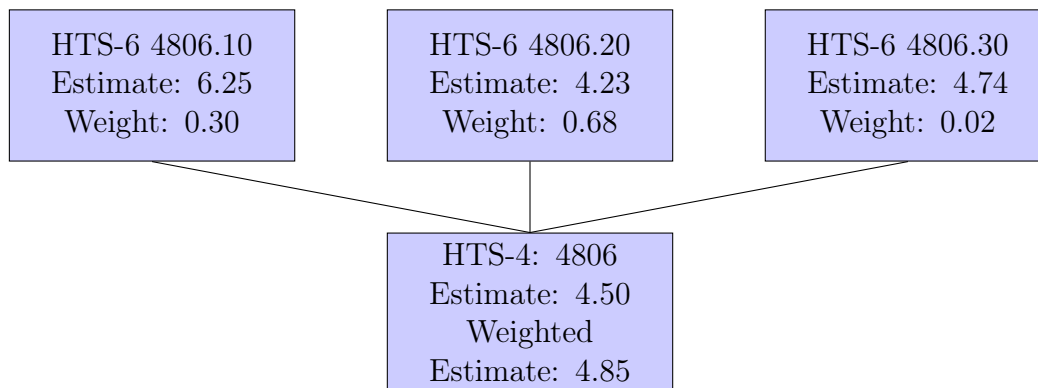
Figure 1 shows an example comparison between HTS-4 and HTS-6. The 4-digit estimate for HTS-4 (parchment paper) is 4.50. The 6-digit estimates within HTS 4806 are 6.25 (HTS 4806.10 vegetable parchment), 4.23 (HTS 4806.20 greaseproof papers), and 4.74 (HTS

---

<sup>6</sup>An example of this is the 6-digit HTS code, 0101.21, contains only one HTS 8-digit code, 0101.21.00.

4806.30 tracing papers). Trade-weighting the 6-digit estimates produces a 4-digit estimate of 4.85 which is slightly larger than the aggregate estimate of 4.50.

Figure 1: Example comparison of HTS-4 estimation with trade-weighted HTS-6 estimation

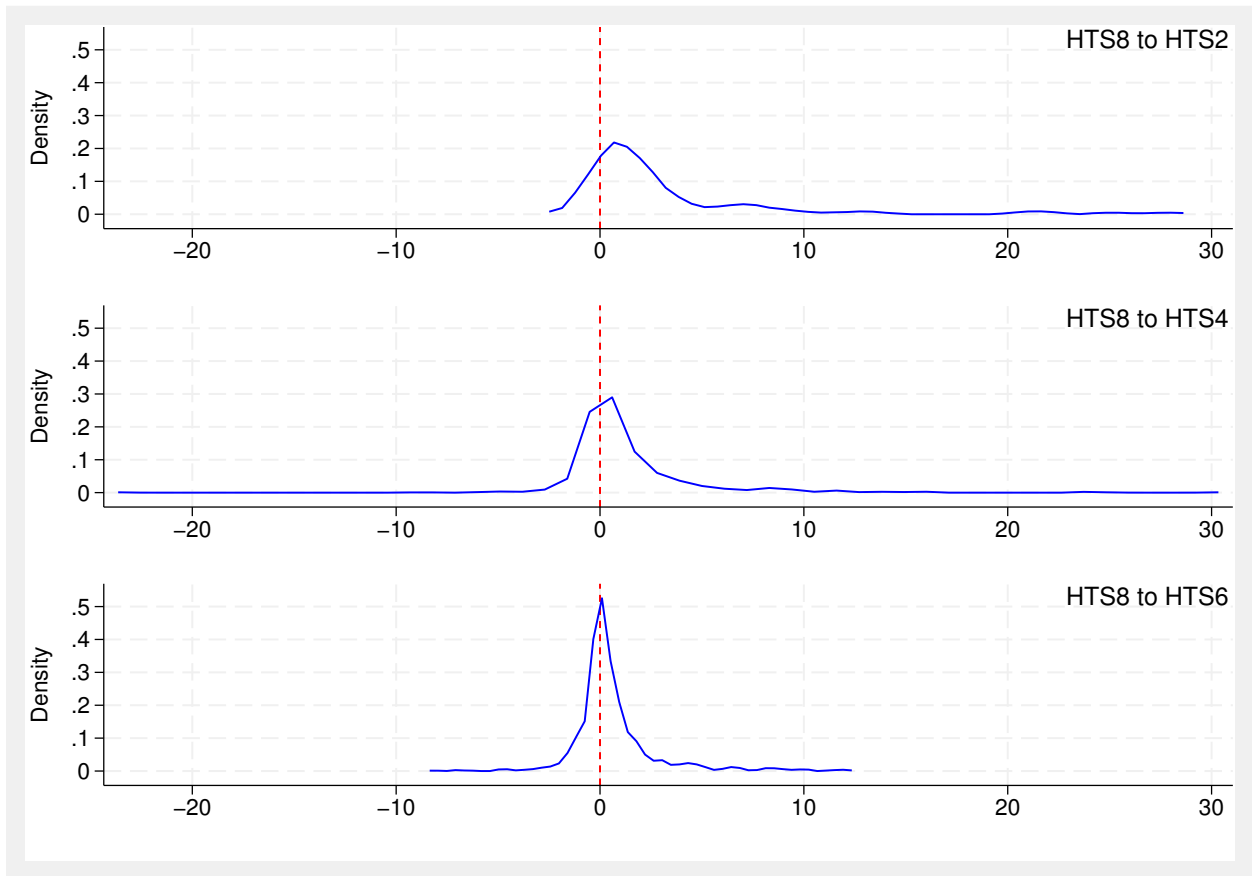


To compare the trade-weighted estimates to the regression estimates of the corresponding aggregated grouping, we take the difference between the two ( $\sigma_{trade\ weighted} - \sigma_{estimated}$ ). Figure 2 provides the kernel density graphs of the differences between the HTS 8-digit estimates, aggregated by trade weight, and the corresponding econometrically estimated elasticities for the HTS 6-digit, 4-digit, and 2-digit product groups.<sup>7</sup>

As the aggregation distance—the number of aggregation levels—decreases, the estimates aggregated by trade weight become closer to corresponding EOS estimate calculated for the aggregate product groups. The more concentrated the density of differences is around zero, the more representative the trade-weighted EOS estimates are of the estimates calculated for the aggregated groups. We find that the elasticity estimates when looked at collectively tend to have a slight negative aggregation bias that is reduced as the aggregation distance decreases. The peak of the kernel density curve falls further right of zero when aggregating from the HTS 8-digit to 2-digit level than when aggregating from the HTS 8-digit to 6-digit level. It is important to note that the direction of the bias is not always negative; there

<sup>7</sup>For the purpose of readability we dropped the observations found in the 99th percentile from the graphs. They were significantly larger than the other estimates due to a combination of large trade weights and the insignificant and negative elasticities being dropped.

Figure 2: Kernel Density Graphs of the Differences Between Trade-Weighted Estimates and Regression Estimates for Aggregated Groups



are some HTS codes for which their aggregate estimate is larger than their trade-weighted estimate. The direction of the bias depends on the correlation between heterogeneity in the residuals and the regressor. Furthermore, we find when aggregating up one product level (e.g. HTS 8-digit to 6-digit or HTS 6-digit to 4-digit), the more disaggregated the starting level, the closer the estimates aggregated by trade weight will be to the estimates calculated using the aggregated data. An estimate aggregated by trade weight from the HTS 8-digit to 6-digit level will tend to be closer to the regression estimate using the 6-digit data than a trade-weighted estimate aggregated from the HTS 6-digit to the 4-digit level will be to the

regression estimate using the 4-digit data.<sup>8</sup>

A similar trend in aggregation bias is also found in Imbs and Mejean (2015) when calculating the price elasticity of imports using pooled microeconomic data. They attribute the aggregation bias to the existence of heterogeneity bias in the models using aggregate data. Intuitively, with well-behaved residuals, the elasticity estimates from the regressions using aggregated data should be equal to the trade-weighted aggregation of the disaggregated elasticities. In reality, we know this not to be the case. Imbs and Mejean (2015) attribute this to heterogeneity from the disaggregated data being systematically pushed into the residual when running a regression with the more aggregated data. This causes the residual to be systematically correlated with the regressor resulting in heterogeneity bias. The authors suggest that for a one-sector trade model, it is better to use a weighted average of the disaggregated elasticities, than an elasticity estimated with aggregated data due to the presence of heterogeneity bias. Another paper that explores differences by level of aggregation is Redding and Weinstein (2019), who discuss that a log-linearized model cannot simultaneously hold for more than one level of aggregation due to Jensen’s inequality. If we assume that the model holds at the most disaggregated level used in this paper (HTS 8-digit) then the EOS estimates from the more aggregate data are at best a log-linear approximation.

An additional reason that the trade-weighted estimates may differ from the aggregate estimates in figure 2 is that there are some disaggregated HTS codes that were not statistically significant. As mentioned above, only 35 percent of HTS 8-digit products were significant and non-negative for the OLS model. When the disaggregated elasticities are aggregated by trade-weight, those products with non-significant or negative estimates were dropped. The regression to produce the aggregated elasticity estimates would inherently include those dropped trade values when arriving at an estimate. However, these non-significant disaggre-

---

<sup>8</sup>The graphs comparing the differences between the HTS 6-digit and 4-digit estimates, aggregated by trade weight, and corresponding econometrically estimated elasticities for HTS 4-digit and 2-digit product groups can be found in the Appendix (figure 3).

gated estimates are not the only driver of differences between trade-weighted and aggregate estimates; there are product groupings in the database where all 8-digit estimates are significant and non-negative and have differences between trade-weighted and aggregate estimates.

To illustrate aggregation bias by sector, we report the five largest and smallest differences between the econometrically estimated and trade-weighted elasticities when aggregating from the HTS 8-digit to the 2-digit level (table 3). There are a few reasons for differences in the size of aggregation bias by sector: removal of negative and statistically insignificant estimates in the trade-weighted aggregation, and differences in the heterogeneity of elasticity within sectors. The first reason may have an impact on the magnitude of the difference, although the direction of the impact is not clear. There are HTS 2-digit sectors with a larger number of significant and non-negative elasticity estimates at the 8-digit level that have a large aggregation bias and some that have small aggregation bias. The same is true for sectors with a small number of significant and non-negative estimates at the HTS 8-digit level.<sup>9</sup> The direction of the impact on the magnitude of the aggregation bias is clearer for the second reason. Sectors that contain heterogeneous elasticity estimates are likely to have a higher aggregation bias than those that contain more homogeneous estimates. The elasticity estimates for HTS 8801.00 and HTS 8806.22 are 4.50 and 20.90, while the estimates for HTS 6302.51 and HTS 6302.93 are 3.45 and 3.78.<sup>10</sup> We observe that high variation in elasticity estimates within product groups tends to lead to higher differences between the trade-weighted and aggregate estimates.

---

<sup>9</sup>We ran a simple regression of the difference between the aggregated and trade-weighted elasticities on the percent of significant non-negative elasticity estimates for the HTS 8-digit products within each sector. The percent of significant and non-negative HTS 8-digit elasticity estimates within each 2-digit sector explained little of the variation in differences.

<sup>10</sup>The corresponding product category descriptions are: HTS 8801.00 (balloons and dirigibles; gliders, hang gliders and other non-powered aircraft), HTS 8806.22 (unmanned aircraft for remote-controlled flight only, with maximum take-off weight more than 250g but not more than 7kg), HTS 6302.51 (table linen of cotton, not knitted or crocheted), HTS 6302.93 (toilet and kitchen linen of man-made fibers).

Table 3: Magnitude of Elasticity Estimate Aggregation Bias

	HTS 2-digit	Description	Difference	Percent of Significant 8-digit HTS Codes
Largest	88	Aircraft, Spacecraft...	28.51	50.00
	07	Edible Vegetables...	27.87	9.90
	08	Edible Fruit And Nuts; Peel...	25.15	17.27
	28	Inorganic Chemicals; Organic...	21.79	20.15
	80	Tin and Articles Thereof	20.98	30.00
	:	:	:	:
Smallest	34	Soap etc.; Lubricating...	0.10	45.28
	57	Carpets and Other Textile...	-0.07 <sup>a</sup>	46.30
	39	Plastics and Articles Thereof	-0.06 <sup>a</sup>	69.33
	18	Cocoa and Cocoa Preparations	0.03	19.12
	63	Made-up Textile Articles...	0.01	54.00

<sup>a</sup> The differences should be looked at in absolute value. Aggregation bias is measured by the distance of the differences from zero.

## 4 Conclusion

This paper describes how the Armington elasticity, or elasticity of substitution, changes across different levels of data aggregation. The EOS estimates are calculated using the trade cost approach specified in Riker (2020) and Schreiber (2022) at the 8-digit, 6-digit, 4-digit, and 2-digit product levels of the 2017 U.S. Harmonized Tariff Schedule. As the level of aggregation increases, the magnitude and variation of the elasticity estimates decreases. The PPML elasticity estimates tend to be larger and the variation increases with aggregation at a faster rate than the OLS estimates. We also explore the potential for heterogeneity bias in elasticities estimated on aggregate data by comparing them to a trade-weighted average of the disaggregated elasticities. On average, the aggregate estimates were smaller than the trade-weighted disaggregated estimates, suggesting a slight negative aggregation bias. These findings are similar to the results in Imbs and Mejean (2015).



Further research in this area can explore whether or not these trends hold when using alternative methods for estimating the elasticity of substitution. Also, this paper finds a large number of the EOS estimates are not statistically significant when using the trade cost approach at the very disaggregated product levels. Due to the lack of significant values we did not measure the elasticities at the most disaggregated HTS 10-digit level, where trade cost changes occur. A more robust interpretation of how elasticities change across levels of data aggregation might be found using an approach that results in more significant estimates at the disaggregated level.

## References

- Ahmad, S., Montgomery, C. and Schreiber, S. (2021). A Comparison of Sectoral Armington Elasticity Estimates in the Trade Literature, *Journal of International Commerce and Economics* .
- Ahmad, S. and Riker, D. (2019). A Method for Estimating the Elasticity of Substitution and Import Sensitivity by Industry, *USITC Economics Working Paper Series, 2019-05-B* .
- Bajzik, J., Havranek, T., Irsova, Z. and Schwarz, J. (2020). Estimating the Armington Elasticity: The Importance of Study Design and Publication Bias, *Journal of International Economics* **127**: 1–23.
- Blundell, R., Pashardes, P. and Weber, G. (1993). What do we Learn about Consumer Demand Patterns from Micro Data?, *The American Economic Review* **83**(3): 570–97.
- Blundell, R. and Stoker, T. M. (2005). Heterogeneity and Aggregation, *Journal of Economic Literature* **43**(2): 347–91.
- Borchert, I., Larch, M., Shikher, S. and Yotov, Y. (2021). Disaggregated Gravity: Benchmark Estimates and Stylized Facts from a New Database, *Review of International Economics* **30**(1): 113–36.
- Broda, C. and Weinstein, D. (2006). Globalization and the Gains from Variety, *Quarterly Journal of Economics* **121**(2): 541–585.
- Feenstra, R. C., Luck, P., Obstfeld, M. and Russ, K. N. (2018). In Search of the Armington Elasticity, *The Review of Economics and Statistics* **100**(1): 135–150.
- Fontagné, L., Guimbard, H. and Orefice, G. (2022). Tariff-based Product-level Trade Elasticities, *Journal of International Economics* **137**.

- Halvorsen, B. and Larsen, B. M. (2012). How Serious is the Aggregation Problem? An Empirical Illustration, *Applied Economics* **45**(26): 3786–94.
- Hertel, T., Hummels, D., Ivanic, M. and Keeney, R. (2007). How Confident can we be of CGE-based Assessments of Free Trade Agreements?, *Economic Modelling* **24**(4): 611–635.
- Hillberry, R. and Hummels, D. (2013). Chapter 18–Trade Elasticity Parameters for a Computable General Equilibrium Model, *Handbook of Computable General Equilibrium Modeling* **1**: 1213–1269.
- Horridge, M. (2018). Aggregating CES Elasticities for CGE Models, *21st Annual Conference on Global Economic Analysis, Cartagena, Colombia* .
- Imbs, J. and Mejean, I. (2015). Elasticity Optimism, *American Economic Journal: Macroeconomics* **7**(3): 43–83.
- McDaniel, C. A. and Balistreri, E. J. (2003). A Discussion on Armington Trade Substitution Elasticities, *Economie Internationale* **94-95**: 301–313.
- Miller, M. and Alberini, A. (2016). Sensitivity of Price Elasticity of Demand to Aggregation, Unobserved Heterogeneity, Price Trends, and Price Endogeneity: Evidence from U.S. Data, *Energy Policy* **97**: 235–49.
- Pesaran, M. and Smith, R. (1995). Estimating Long-Run Relationships from Dynamic Heterogeneous Panels, *Journal of Econometrics* **68**(1): 79–113.
- Powers, W. and Schreiber, S. (2023). Measuring the Impact of Trade Cost Volatility on Supply Chains: An Application to Plastics and Light Truck Manufacturing, *USITC Economics Working Paper Series, 2023-06-A* .

- Redding, S. and Weinstein, D. (2019). Aggregation and the Gravity Equation, *AEA Papers and Proceedings* **109**: 450–455.
- Riker, D. (2020). A Trade Cost Approach to Estimating the Elasticity of Substitution, *USITC Economics Working Paper Series, 2020-07-D* .
- Santos Silva, J. M. C. and Tenreyro, S. (2006). The Log of Gravity, *The Review of Economics and Statistics* **88**(4): 641–658.
- Schreiber, S. (2022). A Practical Approach to Estimating Sector-Level Substitution Elasticities with PPML, *USITC Economics Working Paper Series, 2022-03-A* .
- Soderbery, A. (2015). Estimating Import Supply and Demand Elasticities: Analysis and Implications, *Journal of International Economics* **96**: 1–17.
- Soderbery, A. (2018). Trade Elasticities, Heterogeneity, and Optimal Tariffs, *Journal of International Economics* **114**: 44–62.

## 5 Appendix

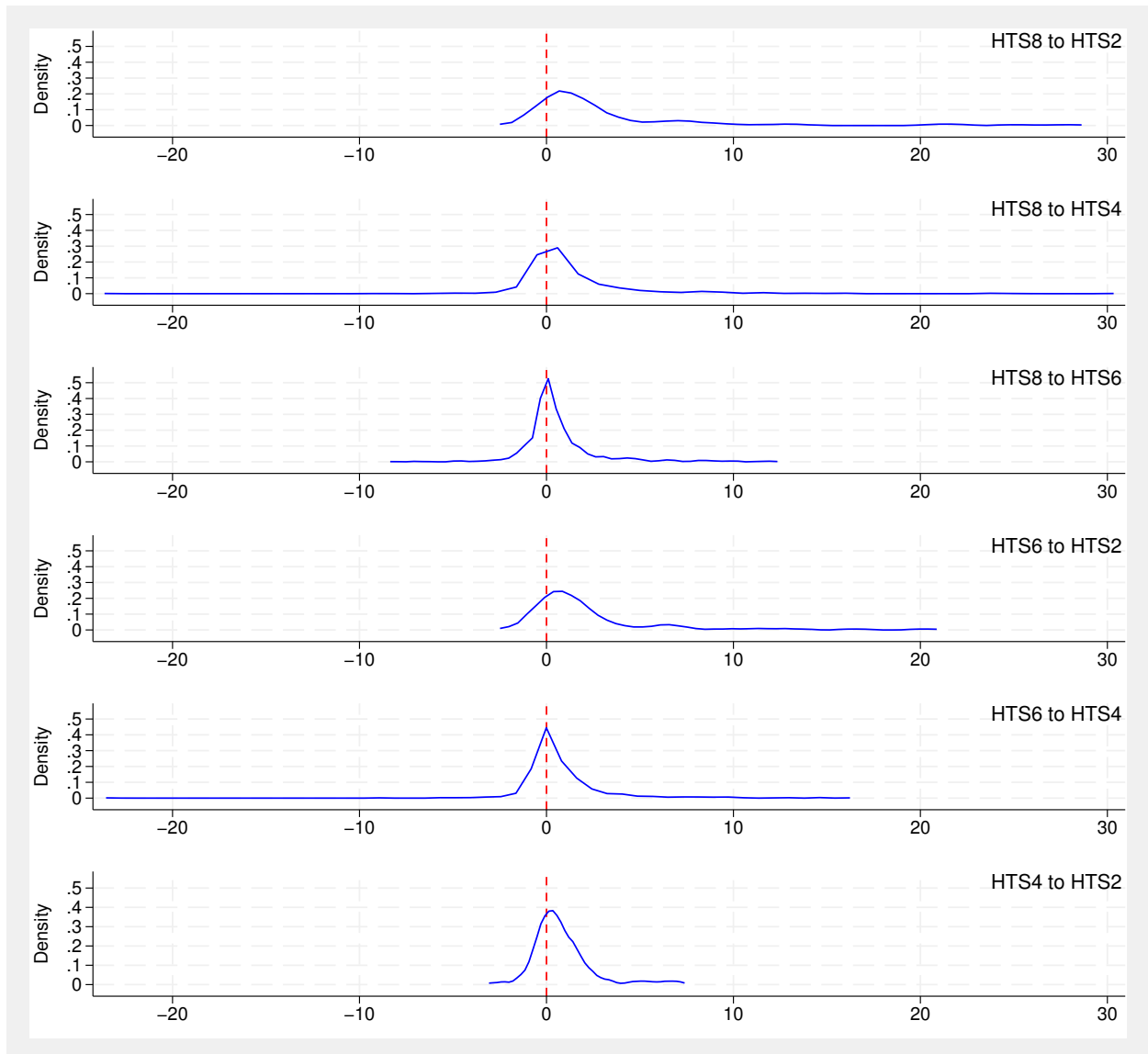


Figure 3: Kernel Density Graphs of the Differences Between Trade-Weighted Estimates and Regression Estimates for Aggregated Groups