Aggregation of Observation-Specific Elasticities*

Vladimir Tyazhelnikov† Xuetao Shi‡ Xinbei Zhou§

September 2026

Abstract

Constant-elasticity models are widely estimated using log–log regressions. In this paper, we show that when underlying elasticities are heterogeneous, standard estimators identify different economic objects. Ordinary least squares recovers the average elasticity, while Poisson pseudo–maximum likelihood (PPML) identifies the aggregate elasticity—the response of an aggregate outcome to a change in an explanatory variable. Estimated elasticities can therefore misstate aggregate responses. We introduce a novel test for elasticity heterogeneity and an estimator that accounts for heterogeneous treatment. We apply our approach to disaggregated trade data and reject homogeneity in 57 percent of industries and show that accounting for heterogeneous treatment lowers estimated responses to realized tariff changes relative to uniform liberalization, with substantial variation across sectors.

JEL Classification: F10, F14, C13, C21, C50

Keywords: Elasticity, heterogeneity, heteroskedasticity, gravity model, misspecification

*This paper has been previously circulated as 'PPML, Gravity, and Heterogeneous Trade Elasticities'. We are particularly grateful to Andrew Bernard and Thomas Zylkyn for generous and detailed comments that substantially improved the paper. We also thank Peter Exterkate, Scott French, Nezih Guner, Vasiliy Korovkin, Ye Lu, Rachael Meager, Mark Melatos, Dennis Novy, Stephen Redding, John Romalis, and participants of Melbourne Empirical Trade Conference, Australasian Trade Workshop, and European Trade Study Group for helpful suggestions. The authors have no relevant financial or other conflicts of interest to disclose.

†School of Economics, The University of Sydney, Sydney, Australia 2006. Corresponding author. (vladimir.tyazhelnikov@sydney.edu.au)

‡School of Economics, The University of Sydney, Sydney, Australia 2006 (xuetao.shi@sydney.edu.au)

§HKU Business School, The University of Hong Kong, Hong Kong SAR, China. (xbeizhou@connect.hku.hk).

1 Introduction

Empirical work across Economics routinely relies on constant-elasticity models, typically estimated using log–log regressions. These estimates are often interpreted as measuring economically meaningful elasticities that map directly into policy-relevant responses. This interpretation is correct under homogeneous elasticities. When elasticities are heterogeneous, however, an aggregate response depends on how observations are weighted and, for non-uniform changes, on which observations receive the largest shocks.

Consider first a uniform change in an explanatory variable. We show that Poisson pseudo–maximum likelihood (PPML) identifies the aggregate elasticity—the response of an aggregate outcome, such as total output, employment, or trade, to a 1% change in the explanatory variable. This is often the economically relevant object. For example, in Carlsson et al. (2021), firm-level shocks affect hiring, but the relevant response is that of total employment rather than that of the average firm.

Standard PPML therefore provides a natural benchmark for a uniform change. Many counterfactuals of interest, however, are non-uniform: the intensity of a policy change varies across observations. Under heterogeneous elasticities, the aggregate response then depends not only on the distribution of elasticities but also on the distribution of shocks, because observations receiving larger shocks contribute disproportionately to the outcome. We introduce a weighted PPML estimator (WPPML) that weights observations by treatment intensity and thereby recovers the aggregate response to arbitrary, non-uniform changes. WPPML aligns the estimated object with the counterfactual of interest.

Whether this adjustment is necessary depends on the presence of elasticity heterogeneity, which is typically unknown. We therefore develop a simple and easily implementable test for elasticity heterogeneity. The test can be applied in standard statistical software and allows researchers to determine whether a single elasticity adequately summarizes the responses in their setting.

Our framework also clarifies why estimates from standard constant-elasticity models differ. Ordinary least squares applied to the log-linear specification identifies the average elasticity, defined as the unweighted mean of heterogeneous responses. PPML instead identifies the aggregate elasticity associated with a uniform change. These objects coincide under homogeneity but diverge when elasticities vary across observations. The distinction is economically meaningful: in settings such as Angrist (1990), the object of interest is the effect of education on the earnings of the average individual, whereas many policy applications concern the response of an aggregate outcome.

Differences between OLS and PPML may therefore reflect differences in the economic

objects they identify rather than differences in statistical robustness. A large literature compares these estimators and attributes their divergence to heteroskedasticity or to the treatment of zero outcomes. We show instead that coefficient heterogeneity is the primary driver of these differences, while heteroskedasticity plays a secondary role. This reinterpretation shifts the focus from statistical corrections to economic identification.

We illustrate these mechanisms using international trade data, which provide a natural testing ground for three reasons. First, gravity equations are among the most widely used constant-elasticity models in Economics. Second, PPML is well established in this context following Santos Silva and Tenreyro (2006), making it a natural benchmark for empirical practice. Third, the trade literature has long distinguished between micro- and macro-elasticities, providing a clear conceptual link to the distinction between average and aggregate elasticities (Feenstra et al. (2018) and Imbs and Mejean (2015, 2017)).

Applying our framework to disaggregated trade data, we find that elasticity heterogeneity is pervasive. Using data on 5,050 HS6 industries, we reject the null of homogeneous elasticities in 57 percent of industries at the 1 percent level. We revisit the difference between OLS and PPML estimates using weighted least squares and Gamma PML and find that heteroskedasticity accounts for only 8–30 percent of the gap, while heterogeneity explains the majority of the difference. We then apply WPPML to tariff changes between 2001 and 2016. Standard PPML estimates correspond to the response to a hypothetical uniform reduction in tariffs, whereas WPPML captures the response to realized, uneven tariff changes. We find that trade flows respond less to actual tariff reductions than to uniform liberalization, implying that liberalization occurred disproportionately among country pairs with lower trade elasticities. The gap between PPML and WPPML estimates is economically large and varies substantially across sectors, reflecting both heterogeneity in elasticities and the non-random allocation of policy changes.

Taken together, our results clarify the interpretation of standard estimators in constant-elasticity models. They show that commonly used methods recover different economic objects under heterogeneity, provide tools to detect and address this heterogeneity, and demonstrate that failing to account for it can lead to systematic differences in estimated responses.

2 Literature Review

Elasticity estimation has been one of the central question in International Trade as the correct estimates allow to immediately calculate counterfactual outcomes of various policies. Elasticity estimation has attracted even more attention since the release of the seminal work of Arkolakis et al. (2012), in which they showed that for a large class of trade models, gains of

trade can be calculated using only observable import shares and the value of trade elasticity. A paper by Melitz and Redding (2015), however, challenged these results and noted that models considered in Arkolakis et al. (2012) start generating heterogeneous trade elasticities even with small and trivial deviations from the initial theoretical assumptions. The results of Melitz and Redding (2015) suggest that relying on an assumption of a homogeneous trade elasticity may be misleading and could lead to incorrect policy predictions.

The empirical estimation of heterogeneous country-pair trade elasticity has attracted limited attention, still a few empirical studies examine the distance elasticity across countries. Fratianni and Kang (2006) first use the log form of the gravity equation to estimate the distance elasticity of the full sample, which yields a significant estimate of -1.17. Then, they test whether the homogeneity assumption for different groups of countries holds with two tests: one for Organisation for Economic Co-operation and Development (OECD) countries and non-OECD countries and the other for Christian and Islamic countries; both tests reject the null hypothesis of distance homogeneity at the 1% level. They find that the distance elasticity is much smaller for OECD member countries in absolute values than for nonmember countries. They also find that the trading cost represented by the distance is the largest when trade occurs between an Islamic and Christian country, with a distance elasticity of -1.47. Magerman et al. (2016) summarize multiple studies on the distance and border effects in International Trade that perform sensitivity tests of the effects for various countries, regions, and periods with different methods; the authors confirm the presence of heterogeneity in distance and trade elasticities across different country pairs.

Chen and Novy (2022) provide an empirical assessment of heterogeneous trade elasticities closest to our paper. They show that trade elasticities exhibit a large degree of heterogeneity across country pairs; in particular, they find that trade elasticities tend to be especially strong when at least one of the trade partners is small. They extend their analysis and re-evaluate the effects of currency unions and trade agreements in the presence of heterogeneous trade elasticities. While their conclusion mirrors our findings regarding the application of PPML in the presence of heterogeneity, our paper is the first to provide a formal justification for these considerations and to consider differential treatment intensity.

A discussion of the mechanisms behind the heterogeneity of trade and distance elasticity is beyond the scope of this paper. Still, several studies provide microfoundations of the gravity equation with country-pair heterogeneity. Fieler (2011) first confirms that low-income countries trade less than rich countries, both with each other and with the rest of the world: in 2000, transactions to and from the 12 Western European countries accounted for 45% of global trade, while the 57 African countries accounted for only 4.2%. This fact suggests that pairs of large countries dominate global trade flows and, therefore, should

be given more weight in calculating the "elasticity of average". The author relaxes two of the traditional assumptions of trade models that generate the gravity equation by allowing for non-homothetic preferences and different quality of goods to be produced in high- and low-income countries. With this new model, the author can explain the prevalence of large trade flows among developed countries and small trade flows among developing countries. Novy (2013) shows that the gravity equation based on the translog demand system, unlike the standard CES gravity model, generates heterogeneous country-pairs trade elasticities. Bas et al. (2017) make a related point but focus on the supply-side mechanism of the heterogeneity. They introduce an extension of the Melitz (2003) model with the productivity distribution of heterogeneous firms following log-normal instead of the commonly used Pareto distribution and show that this assumption leads to bilateral-specific aggregate trade elasticity for each country pair.

Our paper fits into a large strand of literature devoted to the estimation of the gravity equation. A detailed review of this literature can be found in Head and Mayer (2014) and Yotov et al. (2016). In particular, we use Head and Mayer (2014) as a starting point; this paper considers the case of heterogeneous elasticities in Monte Carlo simulations and documents that OLS and GPML successfully estimate the average elasticity, while PPML fails to do so; this difference is being interpreted as the lack of robustness to misspecification and attribute it to the different weights that PPML and GPML assign to various observations. In our paper, we further develop this idea and thoroughly discuss the link between observation weights and the interpretation of various estimators. Mayer et al. (2019) make a similar point that PPML puts more weight on country pairs with larger trade flows; they, nevertheless, do not consider the presence of country-pair heterogeneous elasticity and do not discuss the difference in interpretation of the estimates.[1]

Finally, our paper is related to the strand of literature devoted to the issues associated with the aggregation of trade data. Costinot and Rodríguez-Clare (2014), Imbs and Mejean (2015), Kehoe et al. (2017), and French (2017) address issues caused by between-industry heterogeneity. Imbs and Mejean (2017) focus on the issues that arise when aggregating sectoral elasticities in case these elasticities exhibit heterogeneity on a country level. Coughlin and Novy (2016) analyze the consequences of the spatial aggregation of trade data. Larch et al. (2019) show that OLS and PPML estimates diverge when there is a large number of small countries in the data. Arvis and Shepherd (2013) find that the sum of predicted and actual trade flows are equal only for the PPML. Redding and Weinstein (2019) compare aggregate and sectoral gravity equations and derive an exact Jensen's inequality correction term, which

[1]Eaton et al. (2012) also discuss that the differences between PPML and OLS can be driven by different weights applied to trade flows.

drives the difference between estimates from those two equations. Most recently, Breinlich et al. (2022) show that in the presence of between-industry heterogeneity, a PPML estimator has a clear interpretation as a weighted average of sectoral elasticities. Our paper extends their result by considering a continuously changed heterogeneous elasticity. We also provide interpretations of other estimators obtained from OLS and WLS, and compare them with the PPML estimator. Moreover, we propose the weighted PPML estimator to accommodate heterogeneous trade shocks.

If the exponential model is misspecified, Silva and Winkelmann (2024) demonstrate that pseudo maximum likelihood estimators provide the optimal approximations to the conditional expectation in the sense of weighted mean squared error. Moreover, they show that if covariates are normally distributed, the PPML estimator approximates average marginal effects of the covariates on the true conditional expectation. Compared to Silva and Winkelmann (2024), this paper considers a special case of the misspecified exponential model where parameters might be random. As a result, we relax the normal distribution assumption and are able to connect the marginal effect of the covariates on the true conditional expectation with the heterogeneous trade elasticity.

3 Two Model Representations and Elasticity Objects

This section introduces the two conditional mean representations used throughout the paper. Model 1 is a multiplicative model for dependent variable in levels. Model 2 is a log linear model for positive dependent variables. The two models are alternative maintained specifications. They coincide in the standard homogeneous elasticity case, but they generally lead to different elasticity objects when elasticities are heterogeneous.

Let X_i denote the dependent variable in observation i, let τ_i denote the trade cost variable whose elasticity is of interest, and let $\boldsymbol{W}_i$ collect the remaining controls and fixed effects. For vectors $\boldsymbol{u} = (u_1, \ldots, u_p)$ and $\boldsymbol{v} = (v_1, \ldots, v_p)$, write $\boldsymbol{u}^{\boldsymbol{v}} = \prod_{j=1}^{p} u_j^{v_j}$. We use $\mathbb{R}_{\geq 0}$ to denote the nonnegative real numbers. Thus, for positive controls $\boldsymbol{W}_i$ and coefficient vector β^*, the term $\boldsymbol{W}_i^{\beta^*}$ is defined component by component. We also write $\dot{\tau}_i = \log \tau_i$ and use $\dot{\boldsymbol{W}}_i$ for log controls in the log linear specification. The random variable ξ_i denotes an observation specific trade cost elasticity. It may be correlated with $\boldsymbol{W}_i$ and τ_i. The assumptions below therefore do not impose random assignment of trade costs; they impose restrictions on how conditional means depend on trade costs after conditioning on the controls.

Model 1: multiplicative model. The multiplicative representation is

$$X_i = \boldsymbol{W}_i^{\beta^*} \tau_i^{\xi_i} \epsilon_i, \qquad \mathbb{E}\left[\epsilon_i \mid \boldsymbol{W}_i, \tau_i, \xi_i\right] = 1. \tag{M1}$$

The normalization is in levels. Conditional on $\boldsymbol{W}_i$ and τ_i, Model 1 averages $\tau_i^{\xi_i}$ in levels. Without an additional restriction, this level conditional mean need not have a constant elasticity with respect to τ_i. The following assumption imposes the form needed to summarize the conditional mean by one elasticity within a cell defined by the controls.

Assumption 1 (Multiplicative conditional elasticity)**.** *There exists a function $\theta^A(\boldsymbol{W}_i)$ such that*

$$\mathbb{E}\left[X_i \mid \boldsymbol{W}_i, \tau_i\right] = \boldsymbol{W}_i^{\beta^*} \tau_i^{\theta^A(\boldsymbol{W}_i)}$$

on the conditional support of τ_i.

assm:: h function says that, within a cell defined by the controls $\boldsymbol{W}_i$, expected trade in levels has a constant elasticity with respect to the trade cost variable. Equivalently, under the assumption,

$$\theta^A(\boldsymbol{W}_i) = \frac{\partial \log \mathbb{E}\left[X_i \mid \boldsymbol{W}_i, \tau_i\right]}{\partial \log \tau_i}.$$

The elasticity may vary across cells through $\boldsymbol{W}_i$, but it does not vary with τ_i once $\boldsymbol{W}_i$ is held fixed.

The assumption is not automatic in a random coefficient model. It rules out forms of sorting in which the conditional distribution of ξ_i changes with τ_i in a way that changes the elasticity of expected trade after conditioning on $\boldsymbol{W}_i$. A simple sufficient condition is that $\xi_i = \theta^A(\boldsymbol{W}_i)$ after conditioning on $\boldsymbol{W}_i$. More generally, write $\xi_i = \theta^A(\boldsymbol{W}_i) + \eta_i$. If $\mathbb{E}\left[\tau_i^{\eta_i} \mid \boldsymbol{W}_i, \tau_i\right]$ depends only on $\boldsymbol{W}_i$, then this term is a level shifter and can be absorbed into the conditional mean scale. Under this primitive condition, expected trade is proportional to $\tau_i^{\theta^A(\boldsymbol{W}_i)}$ within each cell. The condition is therefore weaker than homogeneous elasticities, but it is still a restriction on the conditional mean. This is why the paper later tests the homogeneous version of the restriction rather than assuming it is innocuous.

The object $\theta^A(\boldsymbol{W}_i)$ is a level based conditional elasticity. It is generally not equal to $\mathbb{E}\left[\xi_i \mid \boldsymbol{W}_i, \tau_i\right]$, because Model 1 averages $\tau_i^{\xi_i}$ before taking the elasticity of the conditional mean. For a uniform proportional change in trade costs, the corresponding aggregate elasticity is $\theta_A = \mathbb{E}\left[X_i \theta^A(\boldsymbol{W}_i)\right] / \mathbb{E}\left[X_i\right]$. A constant elasticity PPML regression is linked to this flow weighted object under the maintained multiplicative specification and the additional regularity conditions in Prop:: PPML h.

Model 2: log linear model. The log linear representation is

$$\log X_i = \beta^{*\top} \dot{\boldsymbol{W}}_i + \xi_i \dot{\tau}_i + \log \epsilon_i, \qquad \mathbb{E}\left[\log \epsilon_i \mid \dot{\boldsymbol{W}}_i, \dot{\tau}_i, \xi_i\right] = 0. \tag{M2}$$

This model is defined for positive values of dependent variable and imposes the normalization in logs rather than in levels. It is therefore not an algebraic transformation of Model 1. The level normalization in Model 1 does not imply the log normalization in Model 2, and the log normalization in Model 2 does not imply the level normalization in Model 1.

Taking conditional expectations in Model 2 gives $\mathbb{E}\left[\log X_i \mid \boldsymbol{W}_i, \tau_i\right] = \beta^{*\top} \dot{\boldsymbol{W}}_i + \mathbb{E}\left[\xi_i \mid \boldsymbol{W}_i, \tau_i\right] \dot{\tau}_i$. The log linear model delivers a stable conditional elasticity only if the conditional mean of ξ_i does not itself vary with τ_i after conditioning on $\boldsymbol{W}_i$. We impose the following parallel restriction.

Assumption 2 (Log linear conditional elasticity)**.** *There exists a function $\theta^a(\boldsymbol{W}_i)$ such that*

$$\mathbb{E}\left[\log X_i \mid \boldsymbol{W}_i, \tau_i\right] = \beta^{*\top} \dot{\boldsymbol{W}}_i + \theta^a(\boldsymbol{W}_i)\, \dot{\tau}_i$$

on the conditional support of τ_i.

Under assm:: h bar function, $\theta^a(\boldsymbol{W}_i)$ is the derivative of expected log trade with respect to log trade costs. Equivalently, under Model 2 the maintained specification sets $\theta^a(\boldsymbol{W}_i) = \mathbb{E}\left[\xi_i \mid \boldsymbol{W}_i, \tau_i\right]$. The associated average elasticity is $\theta_a = \mathbb{E}\left[\theta^a(\boldsymbol{W}_i)\right]$. A constant elasticity OLS regression is linked to this object under the maintained log linear specification and the regularity conditions in Prop:: OLS h bar.

Comparison of the two elasticity functions. The two assumptions impose the same type of stability restriction in two different conditional means. Model 1 requires expected trade in levels to have a stable elasticity with respect to trade costs. Model 2 requires expected log trade to have a stable slope with respect to log trade costs. When the elasticity is homogeneous, these two restrictions lead to the same elasticity. With heterogeneous elasticities, they generally define different objects. The level based elasticity $\theta^A(\boldsymbol{W}_i)$ reflects how heterogeneous elasticities enter expected trade in levels, while the log linear elasticity $\theta^a(\boldsymbol{W}_i)$ reflects the conditional mean of the random coefficient in the log equation.

The appendix records two useful implications of this comparison. First, under an independence condition on the multiplicative error, Model 1 and Model 2 can be embedded in the same random coefficient environment, up to constants in the error normalization; see lemma::Equivalence of two models. Second, if the conditional distribution of ξ_i shifts monotonically with the controls, then the level based elasticity and the log linear elasticity

move in the same direction; see Prop: h and h bar relation. These observations are useful for interpretation, but they are not required for the main empirical procedure. The paper treats Model 1 and Model 2 as alternative maintained specifications.

The distinction also matters for aggregation. The aggregate object θ_A is weighted by dependent variable and describes its aggregate expected response under Model 1. The average object θ_a weights observations equally and describes the average log linear elasticity under Model 2. Differences between PPML and OLS can therefore reflect both the level and log distinction and the different weights applied to heterogeneous elasticities. This motivates the heterogeneity tests in the next section and the WPPML estimator for nonuniform trade cost changes in sec:wppml.

4 Testing for Heterogeneous Elasticities

The homogeneous elasticity restrictions in Models 1 and 2 are testable implications of the maintained specification. Under Model 1, homogeneity requires $\theta^A(\boldsymbol{W})$ to be constant. Under Model 2, homogeneity requires $\theta^a(\boldsymbol{W})$ to be constant. Rejection of either restriction indicates that a single elasticity is not an adequate summary of the corresponding conditional mean.

The section proceeds in three steps. It first uses a standard trade dataset to illustrate why heterogeneity can affect common estimates. It then presents nonlinear and linear overidentification tests. Finally, it applies the linear implementation to HS 6 trade data and shows that elasticity heterogeneity is widespread across industries.

4.1 Illustrative Evidence

The standard gravity dataset of Santos Silva and Tenreyro (2006) provides a useful illustration. If trade elasticities are homogeneous, estimates from the same log linear gravity specification should be stable across subsamples that differ only in the size of the trade flows retained. If larger country pairs have systematically different elasticities, estimates from subsamples dominated by large flows should differ from those in the full sample.

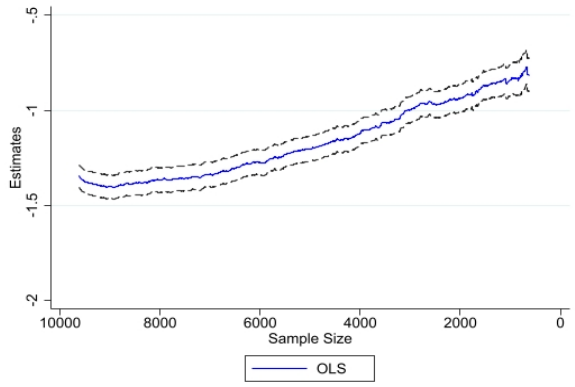
We estimate the benchmark log log OLS specification of Santos Silva and Tenreyro (2006). Starting from the full sample, we iteratively drop the ten smallest trade flows and reestimate the specification. Under homogeneity, the estimated distance elasticity should remain stable across iterations.

fig:olssorted instead shows a systematic relationship between the estimated elasticity $\widehat{\theta}$ and the average size of trade flows in the estimation sample. As small flows are removed,

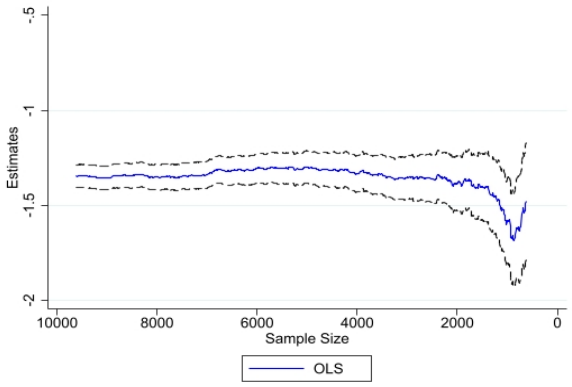
the estimated elasticity changes. By contrast, when ten observations are dropped at random in each iteration, fig:olsrandom shows that the estimates remain stable. Random deletion changes sample size but not the economic composition of the sample in the same systematic way. This pattern is consistent with elasticity heterogeneity that is correlated with trade flows.[2]

Figure 1. Evidence of Heterogeneity in Trade Elasticities

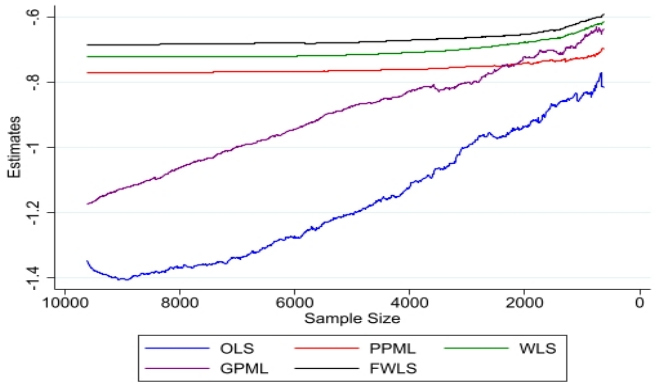
(a) Dropping the smallest trade flows



(b) Dropping random observations



(c) Comparison across estimators



Note: This figure illustrates how estimated trade elasticities vary when subsamples are constructed based on trade flows. Panel (a) sequentially drops the 10 smallest trade flows at each iteration, while Panel (b) drops 10 observations at random in each iteration. Panel (c) compares OLS, WLS, FWLS, GPML, and PPML estimates across subsamples ordered by trade flow. The full dataset contains 9,613 observations; the final iteration retains 623. Patterns in Panels (a) and (b) indicate that heterogeneity in country pair elasticities is correlated with trade flows rather than arising from sampling noise.

fig:comparison repeats the exercise for WLS, FWLS, GPML, and PPML. The estimates

[2]The specifications include exporter and importer fixed effects as well as standard bilateral controls. The patterns documented in fig:olssorted and fig:olsrandom are therefore not driven by the omission of these standard controls; they are consistent with unobserved heterogeneity of the form emphasized by Yotov (2012) and Larch et al. (2019).

follow the weighting interpretation developed below. OLS and GPML change more when the sample is restricted to larger trade flows. PPML, WLS, and FWLS already assign greater weight to larger trade flows, so removing small flow observations has less effect on their estimates.[3]

These patterns indicate that elasticity heterogeneity is empirically relevant in standard gravity data. The next subsections turn this observation into formal tests of the homogeneous elasticity restrictions.

4.2 Nonlinear Heterogeneity Test

We first test the null hypothesis that $\theta^A(\boldsymbol{W})$ is constant. Under the null hypothesis that $H_0 : \theta^A(\boldsymbol{W}) = \gamma^*$, where γ^* is constant, the conditional expectation becomes

$$\mathbb{E}[X \mid \boldsymbol{W}, \tau] = \boldsymbol{W}^{\beta^*} \tau^{\gamma^*}. \tag{1}$$

This conditional moment restriction implies

$$\mathbb{E}[g(\beta^*, \gamma^*)] = \mathbf{0}, \text{ where } g(\beta, \gamma) \equiv \left(X - \boldsymbol{W}^{\beta} \tau^{\gamma}\right) \begin{pmatrix} \dot{\boldsymbol{W}} \\ \dot{\tau} \\ \dot{\tau} \boldsymbol{W} \end{pmatrix}.$$

There are $2p+1$ moment conditions and $p+1$ unknown parameters. The resulting overidentifying restrictions motivate a Hansen J test.[4]

Given observations $\{(X_i, \boldsymbol{W}_i, \tau_i)\}_{i=1}^n$, let

$$\widehat{g}_n(\beta, \gamma) \equiv \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \boldsymbol{W}_i^{\beta} \tau_i^{\gamma}\right) \begin{pmatrix} \dot{\boldsymbol{W}}_i \\ \dot{\tau}_i \\ \dot{\tau}_i \boldsymbol{W}_i \end{pmatrix}$$

denote the sample average of $g(\beta, \gamma)$. The test statistic is

$$J_n \equiv n \left[\widehat{g}_n\left(\widetilde{\beta}, \widetilde{\gamma}\right)\right]^{\top} \widehat{\Omega}^{-1} \left[\widehat{g}_n\left(\widetilde{\beta}, \widetilde{\gamma}\right)\right],$$

[3] This statement is specific to the trade cost slope in this application. Zero and small trade flows can still matter for fixed effects, sample support, and general equilibrium adding up restrictions.

[4] An alternative approach is to first nonparametrically estimate $\theta^A(\boldsymbol{W})$ based on sieve estimation. Then because $\nabla \theta^A(\boldsymbol{W}) = \mathbf{0}$ under the null hypothesis, we compare the gradient of the estimator of $\theta^A(\boldsymbol{W})$ with the zero vector. However, such an approach requires nonparametric estimation of a heterogeneous elasticity function. The overidentifying restriction approach below is easier to implement in high dimensional gravity specifications.

where the estimator $\left(\widetilde{\beta},\widetilde{\gamma}\right)$ is computed from

$$\left(\widetilde{\beta},\widetilde{\gamma}\right)=\arg\min_{\beta,\gamma}\left[\widehat{g}_n\left(\beta,\gamma\right)\right]^{\top}\widehat{\Omega}^{-1}\left[\widehat{g}_n\left(\beta,\gamma\right)\right]$$

and $\widehat{\Omega}\xrightarrow{p}\Omega$, where Ω is the asymptotic variance of $\sqrt{n}\widehat{g}_n\left(\beta^*,\gamma^*\right)$. Define a first step estimator

$$\left(\widetilde{\beta}_I,\widetilde{\gamma}_I\right)=\arg\min_{\beta,\gamma}\left[\widehat{g}_n\left(\beta,\gamma\right)\right]^{\top}\left[\widehat{g}_n\left(\beta,\gamma\right)\right].$$

The common choice of $\widehat{\Omega}$ is

$$\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\boldsymbol{W}_i^{\widetilde{\beta}_I}\tau_i^{\widetilde{\gamma}_I}\right)^2\begin{pmatrix}\dot{\boldsymbol{W}}_i\\ \dot{\tau}_i\\ \dot{\tau}_i\boldsymbol{W}_i\end{pmatrix}\left(\dot{\boldsymbol{W}}_i^{\top},\dot{\tau}_i,\dot{\tau}_i\boldsymbol{W}_i^{\top}\right).$$

Theorem 1. *For Model 1, assume that the null hypothesis $H_0:\theta^A\left(\boldsymbol{W}\right)=\gamma^*$ is true. If the data are i.i.d. and the minimizations take place over a compact set that contains (β^*,γ^*), then $J_n\xrightarrow{d}\chi^2_p$ as $n\to\infty$.*

The proof follows the standard GMM argument in Newey and McFadden (1994). Thm:: J test implies that the test has the correct asymptotic size when the critical value is taken from the corresponding χ^2_p distribution.

The test is not a test against every possible departure from homogeneity. Some alternatives with nonconstant $\theta^A\left(\boldsymbol{W}\right)$ may satisfy the finite set of unconditional moments used to construct J_n. Rejection therefore has a clear interpretation, while failure to reject should not be interpreted as proof of homogeneity. A rejection is evidence against the homogeneous elasticity conditional mean restriction under the maintained gravity specification. Omitted interactions, measurement error, or endogeneity may also generate rejection and should be considered in empirical applications.

4.3 Linear Heterogeneity Test

The log linear test is the analogue of the previous test for Model 2. The null hypothesis is $H_0:\theta^a\left(\boldsymbol{W}\right)=\zeta^*$, where ζ^* is constant. Under the null,

$$\mathbb{E}\left[g^{\dagger}\left(\beta^*,\zeta^*\right)\right]=\mathbf{0},\text{ where }g^{\dagger}\left(\beta,\zeta\right)\equiv\left(\log X-\beta\dot{\boldsymbol{W}}-\zeta\dot{\tau}\right)\begin{pmatrix}\dot{\boldsymbol{W}}\\ \dot{\tau}\\ \dot{\tau}\boldsymbol{W}\end{pmatrix},$$

which gives $2p+1$ moment conditions for $p+1$ parameters. The testing procedure is the same as in subsec: nonlin-test, with $g(\cdot)$ replaced by $g^{\dagger}(\cdot)$.

Theorem 2. *For Model 2, assume that the null hypothesis $H_0 : \theta^a(\boldsymbol{W}) = \zeta^*$ is true. If the data are i.i.d. and the minimizations take place over a compact set that contains (β^*, ζ^*), then $J_n \xrightarrow{d} \chi^2_p$ as $n \to \infty$.*

Computation is simpler in the log linear case because $g^{\dagger}(\beta, \zeta)$ is linear in the parameters.

4.4 Empirical Application of the Heterogeneity Test

The empirical application uses the tariff elasticity setting of Fontagné et al. (2022). The setting is well suited for the test. It uses a standard gravity specification, reports product level trade elasticities, and covers 5,050 HS 6 product categories, 189 exporters, and 152 destinations over the period 2001 to 2016.

By Prop:: Equivalence, either the nonlinear or the linear test can be used to evaluate the homogeneous elasticity restriction under the maintained conditions. Both tests are based on Hansen's J statistic and require estimation of a covariance matrix for the moments. In empirical trade applications, this step can be computationally demanding because the specifications include many fixed effects. In the HS 6 application, the origin and destination industry time fixed effects can exceed 1,000 per industry.

The linear test is implemented through a pseudo IV formulation. The Hansen J test is routinely reported by IV software as a test of overidentifying restrictions. The GMM problem is recast as a linear IV model with the same moments as Model 2. Using the `ivreghdfe` command of Correia (2018), the model can be estimated quickly by industry and delivers the same J statistics as the corresponding linear GMM calculation.

Following Fontagné et al. (2022), we estimate the tariff elasticity using the following log linear specification:[5]

$$\ln X_{ijk,t} = \theta_k \ln(1 + \tau_{ijk,t}) + \zeta_k \boldsymbol{W}_{ij} + \eta_{ik,t} + \kappa_{jk,t} + \epsilon_{ijk,t},$$

where indices i, j, k, and t denote exporter, importer, HS 6 industry, and year, respectively. $\eta_{ik,t}$ and $\kappa_{jk,t}$ are origin and destination industry time fixed effects. θ_k is the industry specific tariff elasticity, the parameter of interest; $\tau_{ijk,t}$ is the *ad valorem* tariff imposed by importer j on good k in year t. $\boldsymbol{W}_{ij}$ captures bilateral distance, colonial history, contiguity, and common language.

[5] We discuss the baseline nonlinear specification in sec:wppml.

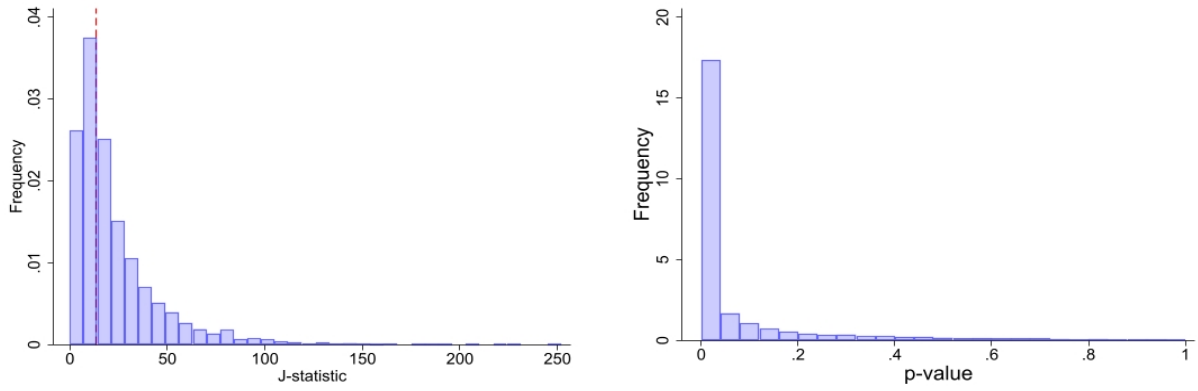
The corresponding moment conditions for industry k are:

$$\mathbb{E}\left[(\ln X_k - \theta_k \ln(1+\tau_k) - \zeta_k \boldsymbol{W}_k - \eta_k - \kappa_k)\begin{pmatrix} \ln(1+\tau_k) \\ \boldsymbol{W}_k \\ \eta_k \\ \kappa_k \\ \ln(1+\tau_k) \times \boldsymbol{W}_k \end{pmatrix}\right] = 0.$$

fig:Hansen summarizes the results. fig:jstathist shows the distribution of Hansen J statistics, with the critical value of 13.277 corresponding to rejection at the 1% significance level. fig:pstathist plots the corresponding p values.

The results indicate substantial heterogeneity. The null of homogeneity is rejected for 57.8%, 71.3%, and 78.8% of industries at the 1%, 5%, and 10% levels, respectively. As discussed in subsec: nonlin-test, the test has correct asymptotic size but is not uniformly powerful. A failure to reject therefore should not be interpreted as evidence that an industry has a homogeneous trade elasticity.

Figure 2. Results of Pseudo IV Estimation



(a) Histogram

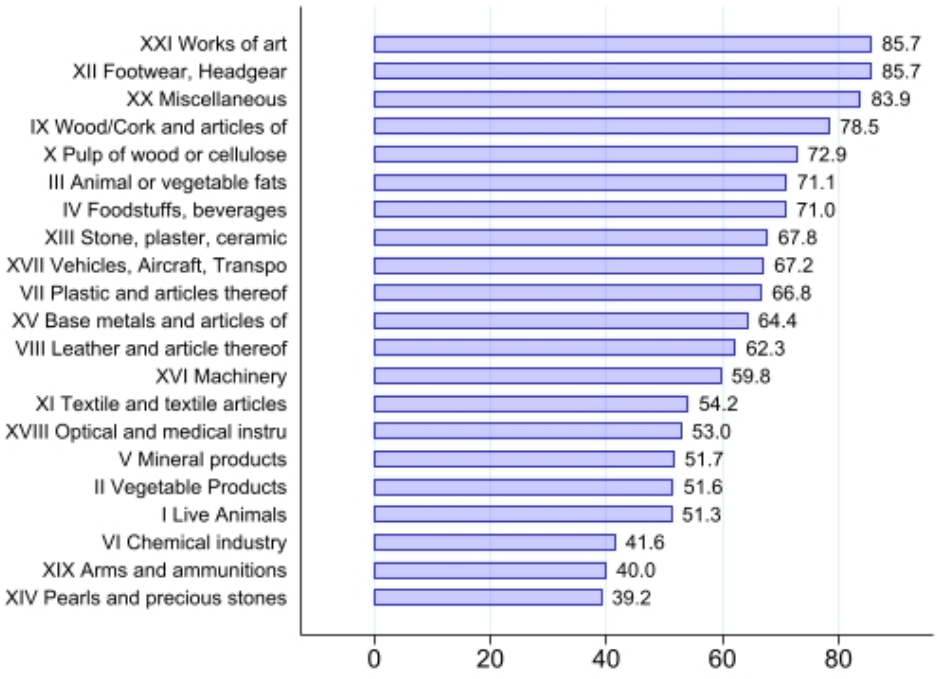
(b) Kernel Density

Note: This figure shows the results of the heterogeneity test applied to 5,050 HS 6 industries from Fontagné et al. (2022). Panel (a) reports the distribution of Hansen J statistics. The red dashed line indicates the critical value at the 1% level, 13.277. Panel (b) shows the corresponding distribution of p values; lower p values indicate stronger evidence of heterogeneity.

The share of industries exhibiting heterogeneity varies widely across HS sections (fig:By$_i ndustry$).$Forex$

The evidence indicates that heterogeneity is a pervasive feature of disaggregated trade data. The linear test can be implemented at scale, including in specifications with many fixed effects, and the rejection rates vary substantially across HS sections. These findings motivate the counterfactual specific estimator in the next section.

Figure 3. Heterogeneity Test by HS Section



Note: This figure shows the share of industries within each HS section for which homogeneity is rejected at the 1% significance level. Section titles have been shortened for presentation.

5 WPPML for Nonuniform Trade Cost Changes

The preceding section shows that the homogeneous elasticity restriction is rejected in many product level gravity specifications. When elasticities are heterogeneous, the relevant elasticity depends on the trade cost change being considered. A uniform proportional change weights all trade flows in one way. A tariff reform that affects some products or country pairs more than others weights them differently. This section defines the elasticity for a nonuniform reform and then introduces WPPML.

5.1 Treatment Intensities and Target Elasticity

Let Ω index the trade flows affected by a policy reform. For each observation $i \in \Omega$, let X_i be the initial trade flow and let X_i' be the counterfactual trade flow after the trade cost variable changes from t_i to t_i'. Define the proportional trade cost change by $\Delta_i = (t_i' - t_i)/t_i$. Let $\bar{\Delta} = |\Omega|^{-1} \sum_{i\in\Omega} \Delta_i$, with $\bar{\Delta} \neq 0$, and define

$$s_i = \frac{\Delta_i}{\bar{\Delta}}, \qquad \frac{1}{|\Omega|} \sum_{i\in\Omega} s_i = 1.$$

The vector $\boldsymbol{s} = \{s_i\}_{i\in\Omega}$ is the treatment profile of the reform. It records the exposure of each trade flow relative to the average proportional trade cost change. The special case $s_i = 1$ for every i is a uniform proportional change.

For observations with $\Delta_i \neq 0$, let $\theta_i = ((X_i' - X_i)/X_i)/\Delta_i$ denote the finite change elasticity. The aggregate response to the reform, normalized by the average proportional trade cost change, is

$$\theta_{\boldsymbol{s}} = \frac{\left(\sum_{i\in\Omega} X_i' - \sum_{i\in\Omega} X_i\right) / \sum_{i\in\Omega} X_i}{\bar{\Delta}} = \frac{\sum_{i\in\Omega} X_i s_i \theta_i}{\sum_{i\in\Omega} X_i}.$$

This expression combines trade flow weights and treatment weights. Trade flow weights appear because the outcome is aggregate trade in levels. Treatment weights appear because the reform changes trade costs by different amounts across observations. Observations with no exposure have $s_i = 0$ and do not enter the reform specific response.

For small trade cost changes, θ_i is approximated by the Model 1 elasticity $\theta^A(\boldsymbol{W}_i)$. Define

$$\bar{\theta}_{\boldsymbol{s}} = \frac{\sum_{i\in\Omega} X_i s_i \theta^A(\boldsymbol{W}_i)}{\sum_{i\in\Omega} X_i s_i}, \qquad e_{\boldsymbol{s}} = \frac{\sum_{i\in\Omega} X_i s_i}{\sum_{i\in\Omega} X_i}.$$

Then $\theta_{\boldsymbol{s}} \approx e_{\boldsymbol{s}} \bar{\theta}_{\boldsymbol{s}}$. The object $\bar{\theta}_{\boldsymbol{s}}$ is the treatment weighted elasticity. It averages heterogeneous elasticities over the trade value exposed to the reform. The scalar $e_{\boldsymbol{s}}$ is the trade weighted

exposure factor. If the reform is uniform, then $s_i = 1$, $e_{\boldsymbol{s}} = 1$, and $\bar{\theta}_{\boldsymbol{s}} = \theta_A$. If elasticities are homogeneous, $\bar{\theta}_{\boldsymbol{s}}$ equals the common elasticity for every treatment profile.

5.2 WPPML Estimator

WPPML estimates $\bar{\theta}_{\boldsymbol{s}}$ by weighting the PPML moment condition by the treatment profile. Let ω_i be weights proportional to s_i. Multiplying all weights by the same positive constant does not change the estimator. The WPPML estimator $\left(\widehat{\beta}_{\boldsymbol{s}}, \widehat{\theta}^W_{\boldsymbol{s}}\right)$ solves

$$\sum_{i \in \Omega} \omega_i \begin{pmatrix} \dot{\boldsymbol{W}}_i \\ \dot{\tau}_i \end{pmatrix} \left(X_i - \boldsymbol{W}_i^{\widehat{\beta}_{\boldsymbol{s}}} \tau_i^{\widehat{\theta}^W_{\boldsymbol{s}}} \right) = \mathbf{0}.$$

If all nonzero treatment intensities have the same sign, WPPML is an ordinary weighted PPML estimator after a rescaling that makes the weights nonnegative. If a policy combines trade cost increases and decreases, the researcher can either interpret the equation as a signed moment condition or focus on one component of the reform, such as tariff liberalization.

WPPML estimates $\bar{\theta}_{\boldsymbol{s}}$, not $\theta_{\boldsymbol{s}}$ directly. To report the aggregate trade flow response per unit of the average trade cost change, the WPPML coefficient is multiplied by $e_{\boldsymbol{s}}$. Standard PPML is the special case in which $s_i = 1$ for all observations.

Assumption 3. *The reform is local:* $\max_{i \in \Omega} |\Delta_i|$ *is small.*

assm:: WPPML allows the finite change elasticity θ_i to be approximated by the Model 1 elasticity $\theta^A(\boldsymbol{W}_i)$. The following proposition gives the corresponding large sample interpretation of WPPML. The weak correlation condition is used in its treatment weighted form, meaning that the additional local leverage from $\dot{\tau}_i^2$ does not materially change the treatment weighted composition of elasticities.

Proposition 1. *Under assm:: h function, assm:: WPPML, and the treatment weighted version of assup:: PPML weak correlation, the WPPML coefficient satisfies* $\left|\widehat{\theta}^W_{\boldsymbol{s}} - \bar{\theta}_{\boldsymbol{s}}\right| \xrightarrow{p} \delta$ *as* $N \to \infty$*, where* $\delta \approx 0$*. The corresponding aggregate response is approximated by* $e_{\boldsymbol{s}} \widehat{\theta}^W_{\boldsymbol{s}}$*.*

Proof. See Proof: WPPML. □

5.3 Application to Tariff Liberalization

WPPML is applied to the tariff setting of Fontagné et al. (2022). Their baseline specification is a product level structural gravity equation estimated separately by HS 6 product:

$$X_{ijk,t} = \exp\left(\theta_k \ln\left(1 + \tau_{ijk,t}\right) + \gamma_k \ln\left(d_{ij}\right) + \zeta_k \boldsymbol{W}_{ij} + \eta_{ik,t} + \kappa_{jk,t}\right) \epsilon_{ijk,t},$$

where i, j, k, and t denote exporter, importer, HS 6 product, and year. The terms $\eta_{ik,t}$ and $\kappa_{jk,t}$ are origin industry time and destination industry time fixed effects. The coefficient θ_k is the product level tariff elasticity, $\tau_{ijk,t}$ is the ad valorem tariff imposed by importer j on product k, d_{ij} is bilateral distance, and $\boldsymbol{W}_{ij}$ contains other bilateral controls.

The counterfactual is the realized tariff liberalization from 2001 to 2016. For each product k, define the liberalization intensity by $r_{ijk} = \max\left\{-\left(\tau_{ijk,2016} - \tau_{ijk,2001}\right)/\tau_{ijk,2001}, 0\right\}$. The WPPML treatment weight is

$$s_{ijk} = \frac{r_{ijk}}{\left[N\left(N-1\right)\right]^{-1}\sum_{i,j} r_{ijk}}.$$

Observations with no tariff reduction receive zero weight. The exercise therefore studies the liberalization component of realized tariff changes, rather than the net effect of tariff reductions and tariff increases.[6]

fig:wppmlhist and fig:wppmldens compare the distribution of PPML elasticities with the distribution of WPPML elasticities.[7] For elasticities that are large in absolute value, the two distributions are similar. For smaller elasticities, the WPPML distribution is shifted to the right. This pattern indicates that tariff reductions were concentrated among trade flows with smaller elasticities in absolute value. The implied partial equilibrium trade flow response is therefore smaller than the response to a uniform liberalization with the same average treatment. Translating this trade flow response into welfare requires an explicit structural welfare formula.

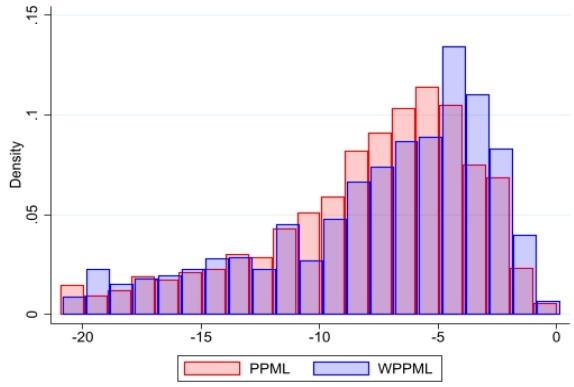
table:WPPML reports PPML and WPPML estimates by HS section. The PPML column replicates Fontagné et al. (2022); the WPPML column reports the treatment weighted elasticity based on realized tariff reductions. The difference column reports WPPML minus PPML. The differences vary across HS sections and reflect both within industry elasticity heterogeneity and the profile of tariff liberalization.

[6]Although tariff changes are used to construct the weights, WPPML is not a three way fixed effects panel estimator. The three way PPML specification uses within pair industry changes over time for identification. WPPML can be applied in a single cross section because the treatment profile enters as estimation weights.

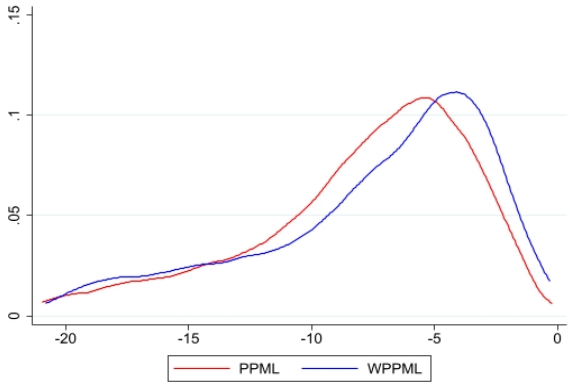
[7]Following Fontagné et al. (2022), we replace insignificant estimates of trade elasticities with 1.

Figure 4. Empirical Distribution of PPML and WPPML Trade Elasticities

(a) Histogram



(b) Kernel Density



Note: This figure presents the empirical distribution of PPML and WPPML trade elasticities calculated for HS 6 products with $\epsilon_k \in (-22, 0)$.

TABLE I

PPML and WPPML Trade Elasticities by HS Section

HS	Section description	HS 6	PPML mean (s.d.)	WPPML mean (s.d.)	Diff. WPPML minus PPML	Nonmissing PPML/WPPML
I	Live Animals and Animal Products	228	-7.51 (9.06)	-8.35 (12.08)	-0.84	223/213
II	Vegetable Products	256	-5.95 (4.06)	-5.66 (6.61)	0.29	252/250
III	Animal or vegetable fats and oils	45	-8.53 (8.69)	-11.16 (15.64)	-2.63	43/40
IV	Prepared foodstuffs, beverages and tobacco	193	-6.17 (4.50)	-5.28 (5.09)	0.89	193/187
V	Mineral products	148	-18.53 (17.62)	-21.48 (34.71)	-2.95	143/137
VI	Products of chemical industries	789	-11.84 (38.16)	-10.75 (12.06)	1.09	758/748
VII	Plastic and articles thereof	211	-8.39 (7.19)	-5.72 (6.99)	2.67	211/207
VIII	Raw hides and skins, leather and articles thereof	69	-5.62 (4.63)	-6.19 (7.07)	-0.57	68/66
IX	Wood/Cork and articles of Wood/Cork	93	-8.47 (8.11)	-8.43 (8.19)	0.04	93/86
X	Pulp of wood or other cellulosic materials	144	-9.92 (7.42)	-9.71 (31.51)	0.21	142/143
XI	Textile and textile articles	801	-7.13 (6.86)	-5.51 (8.07)	1.62	793/777
XII	Footwear, headgear, umbrellas and prepared feathers	49	-3.61 (2.77)	-2.37 (1.80)	1.24	46/46
XIII	Articles of stone, plaster, ceramic and glass	143	-6.62 (4.19)	-4.33 (3.07)	2.29	142/136
XIV	Natural cultured pearls and precious stones and metals	51	-13.59 (13.52)	-14.52 (16.58)	-0.93	50/49
XV	Base metals and articles of base metals	568	-9.59 (9.75)	-9.94 (14.62)	-0.35	557/530
XVI	Machinery, mechanical appliances, and electrical machinery	769	-6.09 (5.56)	-7.13 (19.66)	-1.04	756/698
XVII	Vehicles, aircraft, and transport equipment	131	-10.46 (8.53)	-10.22 (18.95)	0.24	129/121
XVIII	Optical, photographic, precision, and medical instruments	217	-5.65 (5.60)	-5.80 (8.30)	-0.15	210/191
XIX	Arms and ammunitions	20	-6.52 (5.14)	-13.55 (11.35)	-7.03	20/20
XX	Miscellaneous	118	-4.85 (3.42)	-2.73 (2.59)	2.12	117/101
XXI	Works of art	7	-5.96 (4.37)	-7.70 (6.21)	-1.74	7/7

Note: The table reports HS section means of HS 6 level trade elasticities; standard deviations are in parentheses. "Diff." is WPPML minus PPML. Nonmissing reports PPML/WPPML HS 6 counts. PPML uses unweighted estimates, and WPPML uses tariff liberalization weights. Products with $\beta_k < -1$ are included.

6 Gravity Estimates and Estimator Comparisons

This section considers the standard gravity setting that motivates the paper. PPML is widely used in applied trade because it accommodates zero trade flows and addresses the heteroskedasticity concern emphasized by Santos Silva and Tenreyro (2006). The results above show that, with heterogeneous elasticities, estimator choice also changes the weights placed on different trade flows. We use the SST data to document how much of the difference across gravity estimators is associated with reweighting and how much is associated with the residual scale. In this section, subscripts i and j denote exporter and importer; the single subscript notation is reserved for the general model.

6.1 Interpreting Fixed Effects and Controls under Heterogeneity

The analysis concerns the coefficient on the trade cost variable τ, not the coefficients on the controls collected in $\boldsymbol{W}$. This distinction is important in gravity applications because $\boldsymbol{W}$ often includes exporter, importer, exporter year, importer year, industry time, or origin and destination industry time fixed effects. These controls condition the mean of trade and determine the residual variation used to estimate the trade cost elasticity. Their coefficients are not the primary estimands in this paper.

A given observable dimension can affect trade through two channels. It can shift the level of trade through $\boldsymbol{W}_{ij}^{\beta^*}$, and it can also be associated with the elasticity functions $\theta^A(\boldsymbol{W}_{ij})$ or $\theta^a(\boldsymbol{W}_{ij})$. For example, if large countries or large product cells have different distance or tariff elasticities, fixed effects absorb part of the associated level difference. They do not, by themselves, eliminate slope heterogeneity in τ. Unless the specification includes interactions between τ and the relevant dimension, the coefficient on τ remains a weighted summary of heterogeneous slopes.

This point can be stated formally using the notation of subsec:PPML$_e$*stimator.Supposethecontrolscanb* $(\boldsymbol{W}_c, \boldsymbol{W}_h)$, and the conditional mean satisfies

$$\mathbb{E}[X \mid \boldsymbol{W}, \tau] = \boldsymbol{W}_c^{\beta^*} \tau^{\theta^A(\boldsymbol{W}_h)}.$$

In this special case, β^* has the usual multiplicative level interpretation for variables in $\boldsymbol{W}_c$, because those variables do not also determine the elasticity with respect to τ. If the same variable affects both the level term and θ^A, its marginal effect combines a level component and an elasticity heterogeneity component. If θ^A is differentiable in W_ℓ, then

$$\frac{\partial \log \mathbb{E}[X \mid \boldsymbol{W}, \tau]}{\partial \log W_\ell} = \beta_\ell^* + \log \tau \cdot \frac{\partial \theta^A(\boldsymbol{W})}{\partial \log W_\ell}.$$

Thus β_ℓ^* is not generally the total elasticity of the conditional mean with respect to W_ℓ. In gravity applications this overlap is plausible because exporter, importer, pair, product, or cell characteristics can both shift trade levels and sort observations into different trade cost elasticities.

Thus fixed effects and controls are essential for conditioning, but their coefficients are not the main elasticity objects studied here. Conditional on the chosen controls and fixed effects, PPML, OLS, and related estimators target different weighted summaries of the heterogeneous trade cost elasticity. If the coefficients on controls or fixed effects are themselves the objects of interest, the model requires additional structure, such as nonoverlapping covariate sets or explicit interactions between τ and the relevant fixed effect groups.

The moment condition comparison used in this empirical exercise is collected in subsec:Country-pair-heterogeneity, after the formal discussion of PPML and OLS. That appendix shows how OLS, PPML, WLS, FWLS, and GPML differ in the residual scale and in the weights attached to observations. The empirical exercise below uses this comparison to separate changes in the gravity coefficient associated with reweighting observations from changes associated with the treatment of heteroskedasticity.

6.2 Interpreting Differences Across Gravity Estimators

The same gravity equation is estimated using OLS, WLS, FWLS, PPML, and GPML. The moment condition comparison in subsec:Country-pair-heterogeneity is used to interpret differences across coefficients. The exercise separates changes in the distance coefficient associated with reweighting observations from changes associated with residual scaling and heteroskedasticity.

To make the results comparable with those in SST, the analysis uses the same data and follows their specification by including distance and indicator variables for remoteness, common language, colonial heritage, and preferential trade agreements. The analysis replicates the findings in Table 5 of SST for OLS, PPML, and GPML estimators, controlling for multilateral resistance terms with importer and exporter fixed effects, and complements these results with WLS and FWLS estimates. table:data reports the results of all five estimators and corresponding standard errors.

Taking PPML as the benchmark, the analysis asks how much of the OLS minus PPML gap is associated with heterogeneity and how much is associated with heteroskedasticity. The total gap is the difference between the OLS and PPML estimates. Following the estimator comparison in subsec:Country-pair-heterogeneity, we compute two versions of the comparison. In the first, the heterogeneity component is the difference between OLS and WLS, and

TABLE II
Gravity Equation

	OLS	WLS	FWLS	PPML (with ZTF)	PPML (no ZTF)	GPML (with ZTF)	GPML (no ZTF)
Distance	-1.347***	-0.722***	-0.685***	-0.750***	-0.770***	-1.933***	-1.173***
	(0.031)	(0.049)	(0.038)	(0.041)	(0.049)	(0.055)	(0.029)
Common border	0.174	0.395***	0.141**	0.369***	0.352***	-0.457**	0.326***
	(0.130)	(0.080)	(0.057)	(0.091)	(0.090)	(0.230)	(0.117)
Common language	0.406***	0.419***	0.558***	0.383***	0.418***	0.681	0.416***
	(0.068)	(0.076)	(0.052)	(0.093)	(0.094)	(0.097)	(0.065)
Colonial ties	0.666***	-0.047	0.019	0.079	0.038	0.807***	0.510***
	(0.070)	(0.103)	(0.098)	(0.134)	(0.134)	(0.100)	(0.067)
FTA	0.310***	0.424***	0.585***	0.376***	0.374***	1.472***	0.579***
	(0.098)	(0.080)	(0.085)	(0.077)	(0.077)	(0.257)	(0.085)
Exporter FE	Y	Y	Y	Y	Y	Y	Y
Importer FE	Y	Y	Y	Y	Y	Y	Y
Observations	9,613	9,613	9,613	18,360	9,613	18,360	9,613

Note: Standard errors are in parentheses. OLS, WLS, FWLS, PPML, and GPML denote ordinary least squares, weighted least squares, fitted weighted least squares, Poisson pseudo maximum likelihood, and gamma pseudo maximum likelihood. Zero trade flows are excluded from OLS, WLS, and FWLS. For PPML and GPML, we report estimates both including and excluding zero trade flows (ZTF).

the heteroskedasticity component is the difference between WLS and PPML. In the second, the heterogeneity component is the difference between GPML and PPML, and the heteroskedasticity component is the difference between OLS and GPML. table:decomposition reports the estimates, standard errors, and the two comparison terms.

The two calculations answer slightly different questions. The WLS comparison evaluates the heteroskedasticity component while moving from the unweighted log linear object toward a trade weighted log linear object. The GPML comparison evaluates the same issue relative to a relative level deviation estimator. The size and sign of the components need not coincide because they depend on the pattern of heteroskedasticity and the distribution of heterogeneous elasticities. The Monte Carlo exercises in app:heteroskedasticity examine this interaction.

table:decomposition indicates that the component associated with heteroskedasticity accounts for a smaller share of the total gap, while heterogeneity is the primary driver of the differences between the estimators.[8] These findings are consistent with the recommendation in Santos Silva and Tenreyro (2006) to use PML estimators in multiplicative gravity models. They also add an interpretation of why alternative estimators differ: much of the difference

[8]French (2017) performs an alternative estimator comparison between OLS and PPML, focusing on the aggregation properties of both estimators. In this paper, we are agnostic about the source of the elasticity heterogeneity, but if it is industrial heterogeneity, our findings are close to those in French (2017).

TABLE III
Estimator Comparisons

(A) Decomposition through WLS

	OLS	*WLS*	*PPML*
Estimate	-1.347	-0.722	-0.770
Standard error	(0.031)	(0.049)	(0.042)
Heterogeneity component	-0.625	0	0
Percentage	108.32%	0	0
Heteroskedasticity component	0.048	0.048	0
Percentage	-8.32%	100%	0
Total gap	-0.577	0.048	0
Observations	9,613	9,613	9,613

(B) Decomposition through GPML

	OLS	*GPML*	*PPML*
Estimate	-1.347	-1.173	-0.770
Standard error	(0.031)	(0.029)	(0.042)
Heterogeneity component	-0.403	-0.403	0
Percentage	69.84%	100%	0
Heteroskedasticity component	-0.174	0	0
Percentage	30.16%	0	0
Total gap	-0.577	-0.403	0
Observations	9,613	9,613	9,613

Note: The total gap is the difference between the OLS and PPML estimates. The heteroskedasticity component is the difference between WLS and PPML in Panel A and the difference between OLS and GPML in Panel B. The heterogeneity component is the difference between OLS and WLS in Panel A and the difference between GPML and PPML in Panel B. Zero trade flows are excluded from all estimators.

is associated with the weights placed on heterogeneous trade elasticities.

Note that OLS and WLS do not allow for using zero trade flows due to the log linearization of the dependent variable, while GPML and PPML allow for zero values of the dependent variable. To make the results comparable, in our baseline analysis we excluded zero trade for all estimators. In practice, however, zero trade flows tend to be included in the PML estimation, which may be another driver of the difference in the estimates between OLS and PPML.

In table:decomposition$_z tf we repeat our decomposition exercise but include zero trade flows for GPML and

The exercise is descriptive rather than a structural decomposition. The main limitation is that heterogeneity and heteroskedasticity can interact. app:heteroskedasticity evaluates this

TABLE IV
Estimator Comparisons with Zero Trade Flows

(A) Decomposition through WLS

	OLS	*WLS*	*PPML*
Estimate	-1.347	-0.722	-0.750
Standard error	(0.031)	(0.049)	(0.041)
Heterogeneity component	-0.625	0	0
Percentage	104.69%	0	0
Heteroskedasticity component	0.028	0.028	0
Percentage	-4.69%	100%	0
Total gap	-0.597	0.028	0
Observations	9,613	9,613	18,360

(B) Decomposition through GPML

	OLS	*GPML*	*PPML*
Estimate	-1.347	-1.933	-0.750
Standard error	(0.031)	(0.055)	(0.041)
Heterogeneity component	-1.183	-1.183	0
Percentage	198.16%	100%	0
Heteroskedasticity component	0.586	0	0
Percentage	-98.16%	0	0
Total gap	-0.597	-1.183	0
Observations	9,613	18,360	18,360

Note: The total gap is the difference between the OLS and PPML estimates. The heteroskedasticity component is the difference between WLS and PPML in Panel A and the difference between OLS and GPML in Panel B. The heterogeneity component is the difference between OLS and WLS in Panel A and the difference between GPML and PPML in Panel B. Zero trade flows are included for GPML and PPML.

concern using Monte Carlo simulations based on the SST design and alternative assumptions on the error variance. Across these exercises, we find no systematic interaction that overturns the estimator interpretation used in the comparison exercise.

Consistent with app:estimator-foundations, the simulations are interpreted as evidence on when the approximation is quantitatively informative, not as a proof that the approximation is exact.

7 Conclusion

Gravity estimates are often summarized by a single trade elasticity. This paper shows that, when elasticities are heterogeneous, the economic meaning of that coefficient depends on the maintained conditional mean model and on the counterfactual. Under the multiplicative gravity representation, PPML is linked to an aggregate elasticity: the trade flow weighted response of total trade to a uniform proportional change in trade costs. Under the log linear representation, OLS is linked to an average elasticity across observations.

The empirical evidence indicates that the distinction matters. The heterogeneity test rejects the homogeneous elasticity restriction in a majority of HS 6 industries. In the distance application, much of the OLS and PPML gap is associated with reweighting heterogeneous elasticities rather than only with heteroskedasticity. In the tariff application, WPPML implies that realized tariff reductions from 2001 to 2016 generated smaller trade flow responses than a uniform liberalization with the same average tariff change.

The implication for applied gravity work is direct. The researcher should state the counterfactual, define the relevant elasticity object, and choose the estimator and weights that correspond to that object. If the object is the response of aggregate trade to a uniform change in trade costs, the relevant object is an aggregate elasticity. If the object is the average response across observations, the relevant object is an average elasticity. If the object is a particular nonuniform reform, the relevant object is a treatment weighted elasticity. Standard gravity methods remain useful, but their coefficients need to be interpreted through the weights implied by the estimator and the policy experiment.

References

Angrist, J. D. (1990). 'Lifetime earnings and the vietnam era draft lottery: Evidence from social security administrative records', *The american economic review*, pp. 313–336.

Arkolakis, C., Costinot, A.and Rodrıguez-Clare, A. (2012). 'New trade models, same old gains?', *American Economic Review*, Vol. 102(1), pp. 94–130.

Arvis, J. F.and Shepherd, B. (2013). 'The poisson quasi-maximum likelihood estimator: A solution to the 'adding up' problem in gravity models', *Applied Economics Letters*, Vol. 20, pp. 515–519.

Bas, M., Mayer, T.and Thoenig, M. (2017). 'From micro to macro: Demand, supply, and heterogeneity in the trade elasticity', *Journal of International Economics*, Vol. 108, pp. 1–19.

Breinlich, H., Novy, D.and Santos Silva, J. (2022). 'Trade, gravity and aggregation', *Review of Economics and Statistics*, pp. 1–29.

Carlsson, M., Messina, J.and Nordström Skans, O. (2021). 'Firm-level shocks and labour flows', *The Economic Journal*, Vol. 131(634), pp. 598–623.

Chen, N.and Novy, D. (2022). 'Gravity and heterogeneous trade cost elasticities', *The Economic Journal*, Vol. 132(644), pp. 1349–1377.

Correia, S. (2018). 'Ivreghdfe: Stata module for extended instrumental variable regressions with multiple levels of fixed effects',

Costinot, A.and Rodríguez-Clare, A. (2014). 'Trade theory with numbers: Quantifying the consequences of globalization', *Handbook of international economics*, Vol. 4, pp. 197–261.

Coughlin, C. C.and Novy, D. (2016). 'Estimating border effects: The impact of spatial aggregation', *FRB St. Louis Working Paper*.

Eaton, J., Kortum, S. S.and Sotelo, S. (2012). *International trade: Linking micro and macro* (tech. rep.). National bureau of economic research.

Fally, T. (2015). 'Structural gravity and fixed effects', *Journal of International Economics*, Vol. 97, pp. 76–85.

Feenstra, R. C., Luck, P., Obstfeld, M.and Russ, K. N. (2018). 'In search of the armington elasticity', *Review of Economics and Statistics*, Vol. 100(1), pp. 135–150.

Fieler, A. C. (2011). 'Nonhomotheticity and bilateral trade: Evidence and a quantitative explanation', *Econometrica*, Vol. 79, pp. 1069–1101.

Fontagné, L., Guimbard, H.and Orefice, G. (2022). 'Tariff-based product-level trade elasticities', *Journal of International Economics*, Vol. 137, pp. 103593.

Fratianni, M.and Kang, H. (2006). 'Heterogeneous distance–elasticities in trade gravity models', *Economics Letters*, Vol. 90, pp. 68–71.

French, S. (2017). 'Comparative advantage and biased gravity', *UNSW Business School Research Paper*.

Head, K.and Mayer, T. (2014). 'Gravity equations: Workhorse, toolkit, and cookbook', *Handbook of International Economics*, Vol. 4, pp. 131–195.

Imbs, J.and Mejean, I. (2015). 'Elasticity optimism', *American Economic Journal: Macroeconomics*, Vol. 7, pp. 43–83.

Imbs, J.and Mejean, I. (2017). 'Trade elasticities', *Review of International Economics*, Vol. 25, pp. 383–402.

Kehoe, T. J., Pujolas, P. S.and Rossbach, J. (2017). 'Quantitative trade models: Developments and challenges', *Annual Review of Economics*, Vol. 9, pp. 295–325.

Larch, M., Wanner, J., Yotov, Y. V.and Zylkin, T. (2019). 'Currency unions and trade: A ppml re-assessment with high-dimensional fixed effects', *Oxford Bulletin of Economics and Statistics*, Vol. 81, pp. 487–510.

Magerman, G., Studnicka, Z.and Van Hove, J. (2016). 'Distance and border effects in international trade: A comparison of estimation methods', *Economics: The Open-Access, Open-Assessment E-Journal*, Vol. 10, pp. 1–31.

Mayer, T., Vicard, V.and Zignago, S. (2019). 'The cost of non-europe, revisited', *Economic Policy*, Vol. 34(98), pp. 145–199.

Melitz, M. J. (2003). 'The impact of trade on intra-industry reallocations and aggregate industry productivity', *Econometrica*, Vol. 71, pp. 1695–1725.

Melitz, M. J.and Redding, S. J. (2015). 'New trade models, new welfare implications', *American Economic Review*, Vol. 105(3), pp. 1105–1146.

Newey, W. K.and McFadden, D. (1994). 'Large sample estimation and hypothesis testing', *Handbook of econometrics*, Vol. 4, pp. 2111–2245.

Novy, D. (2013). 'International trade without ces: Estimating translog gravity', *Journal of International Economics*, Vol. 89, pp. 271–282.

Redding, S. J.and Weinstein, D. E. (2019). 'Aggregation and the gravity equation', *AEA Papers and Proceedings*, Vol. 109, pp. 450–55.

Santos Silva, J. M. C.and Tenreyro, S. (2006). 'The log of gravity', *The Review of Economics and statistics*, Vol. 88, pp. 641–658.

Silva, J. S.and Winkelmann, R. (2024). 'Misspecified exponential regressions: Estimation, interpretation, and average marginal effects', *Review of Economics and Statistics*, pp. 1–25.

Weidner, M.and Zylkin, T. (2021). 'Bias and consistency in three-way gravity models', *Journal of International Economics*, Vol. 132, pp. 103513.

Wooldridge, J. M. (2005). 'Unobserved heterogeneity and estimation of average partial effects', *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg*, pp. 27–55.

Yotov, Y. V. (2012). 'A simple solution to the distance puzzle in international trade', *Economics Letters*, Vol. 117, pp. 794–798.

Yotov, Y. V., Piermartini, R., Monteiro, J.-A.and Larch, M. (2016). *An advanced guide to trade policy analysis: The structural gravity model*. World Trade Organization Geneva.

A Notation and Definitions

This appendix records notation used in more than one part of the paper. Notation that is specific to a proof, a simulation design, or a single table is defined at the point where it is used.

Core notation

Symbol	Definition
X_i	Trade flow outcome for observation i. In the tariff application, $X_{ijk,t}$ denotes exports from origin i to destination j in HS 6 product k and year t.
τ_i	Trade cost argument whose elasticity is of interest, such as distance or one plus a tariff.
$\dot{\tau}_i$	Log trade cost argument, $\dot{\tau}_i \equiv \log \tau_i$.
$\boldsymbol{W}_i$	Controls and fixed effects other than τ_i.
$\dot{\boldsymbol{W}}_i$	Log controls used in the log linear representation.
ξ_i	Observation specific trade cost elasticity in the random coefficient gravity representation.
ϵ_i	Error term. Model 1 normalizes it in levels; Model 2 normalizes it in logs.
β^*	Coefficient vector on the controls in the maintained conditional mean representation.
$\boldsymbol{u}^{\boldsymbol{v}}$	Componentwise vector power, $\boldsymbol{u}^{\boldsymbol{v}} = \prod_{m=1}^{p} u_m^{v_m}$.
$\mathbb{R}_{\geq 0}$	Set of nonnegative real numbers.

Elasticity and treatment objects

Symbol	Definition
$\theta^A(\boldsymbol{W}_i)$	Level based conditional elasticity under Model 1.
θ_A	Aggregate elasticity under Model 1, $\theta_A = \mathbb{E}\left[X_i \theta^A(\boldsymbol{W}_i)\right] / \mathbb{E}\left[X_i\right]$.
$\theta^a(\boldsymbol{W}_i)$	Log linear conditional elasticity under Model 2.
θ_a	Average elasticity under Model 2, $\theta_a = \mathbb{E}\left[\theta^a(\boldsymbol{W}_i)\right]$.
Δ_i	Proportional trade cost change for observation i in a nonuniform reform.
s_i	Normalized treatment intensity for observation i.
$\boldsymbol{s}$	Treatment profile, $\boldsymbol{s} = \{s_i\}_{i \in \Omega}$.
$\bar{\theta}_{\boldsymbol{s}}$	Treatment weighted elasticity estimated by WPPML.
$\theta_{\boldsymbol{s}}$	Aggregate response to the treatment profile, normalized by the average trade cost change.

$e_{\boldsymbol{s}}$	Trade weighted exposure factor linking $\bar{\theta}_{\boldsymbol{s}}$ to $\theta_{\boldsymbol{s}}$.

Estimator and empirical notation

Symbol or object	Definition
PPML	Poisson pseudo maximum likelihood estimator for the multiplicative gravity equation.
OLS	Log linear least squares estimator.
WPPML	Weighted PPML estimator with weights proportional to treatment intensity.
J_n	Hansen J statistic used in the overidentifying restriction tests of homogeneity.
d_{ij}	Bilateral distance between origin i and destination j.
$\tau_{ijk,t}$	Ad valorem tariff imposed by importer j on product k from origin i in year t.
$\eta_{ik,t}$, $\kappa_{jk,t}$	Origin product year and destination product year fixed effects in the tariff application.

B Estimator Foundations: Average and Aggregate Elasticities

This appendix gives the detailed estimator foundations summarized in sec:elasticity-objects. Assumptions and formal results stated in this appendix are numbered with the prefix B, in their order of appearance. It characterizes the local weighted elasticity objects selected by PPML and OLS when trade elasticities are heterogeneous. After the PPML and OLS discussion, subsec:Country-pair-heterogeneity compares the moment conditions of the estimators used in the gravity application. The main text uses these results to define the objects targeted by the heterogeneity test, WPPML, and the estimator comparisons.

We use the notation introduced in sec:elasticity-objects. The outcome X is a trade flow, τ is the trade cost argument, and $\boldsymbol{W}$ collects the remaining controls and fixed effects. The multiplicative elasticity is denoted by $\theta^A(\boldsymbol{W})$ and the log linear elasticity by $\theta^a(\boldsymbol{W})$. These objects are functions of observables because the underlying trade elasticity may vary across country pairs, products, or cells.

To keep a constant elasticity gravity interpretation, the formal model imposes a restriction on the conditional mean: after conditioning on $\boldsymbol{W}$, the trade cost argument enters with a constant elasticity. This does not rule out unobserved random coefficient heterogeneity. It only rules out a conditional mean in which the elasticity changes directly with the level of

the same trade cost argument. If the elasticity depends directly on the tariff level or on distance itself, a different specification or a nonparametric elasticity would be needed.

The results in this appendix are local interpretations of population moment conditions. They are stated in the order in which they are used. lemma:: PPML h gives the local weighted object targeted by PPML. Prop:: PPML h states when this object reduces to the aggregate elasticity introduced in sec:elasticity-objects. lemma:: h bar gives the corresponding local object targeted by OLS, and Prop:: OLS h bar states when it reduces to the average elasticity. The asymptotic testing result is in sec:tests; the Monte Carlo exercises in app:heteroskedasticity and the empirical exercises show when the interpretation is quantitatively useful in trade applications.

The formulas are written in a compact one regressor notation. With controls or fixed effects, the exact local weights are based on the remaining variation in $\dot{\tau}$ after those controls are accounted for. To keep the notation readable, the derivations below use the simple $\dot{\tau}^2$ expressions; the economic interpretation is unchanged.

B.1 PPML and the Aggregate Elasticity

This subsection gives the formal PPML interpretation summarized in sec:elasticity-objects. We take Model 1 and assm:: h function as maintained. Thus the relevant conditional mean is $\mathbb{E}[X \mid \boldsymbol{W}, \tau] = \boldsymbol{W}^{\beta^*} \tau^{\theta^A(\boldsymbol{W})}$. The objective is to relate the constant elasticity PPML coefficient on $\log \tau$ to the heterogeneous elasticity function $\theta^A(\boldsymbol{W})$.

Let $\dot{\boldsymbol{W}} \equiv (\log W_1, \ldots, \log W_p)^\top$ and $\dot{\tau} \equiv \log \tau$. For observations $\{(X_i, \boldsymbol{W}_i, \tau_i)\}_{i=1}^n$, let $\left(\widehat{\beta}, \widehat{\gamma}\right)$ be the solution to the following sample moment condition

$$\sum_{i=1}^{n} \left(X_i - \boldsymbol{W}_i^{\widehat{\beta}} \tau_i^{\widehat{\gamma}}\right) \begin{pmatrix} \dot{\boldsymbol{W}}_i \\ \dot{\tau}_i \end{pmatrix} = \mathbf{0}.$$

The estimator $\left(\widehat{\beta}, \widehat{\gamma}\right)$ is the PPML estimator. In the remaining part of this section, we study properties of the PPML estimator.

For Model 1, if assm:: h function holds, then $\widehat{\gamma} \xrightarrow{p} \gamma_0$ as $n \to \infty$, where $\gamma_0 \approx \mathbb{E}\left[X \dot{\tau}^2 \theta^A(\boldsymbol{W})\right] / \mathbb{E}[X \dot{\tau}^2]$.

Proof. See Proof of lemma: PPML h □

lemma:: PPML h shows that the PPML estimator $\widehat{\gamma}$ approximates a weighted average of the elasticity θ^A. The local observation weight is proportional to $X\dot{\tau}^2$, which comes from the derivative of the population moment condition with respect to the trade cost coefficient.

The approximation sign in lemma:: PPML h is intentional: it records the local linearization of the PPML moment condition around the pseudo true coefficient, not an exact finite sample or population identity.

The extra leverage from $\dot{\tau}^2$ does not materially change the trade flow weighted composition of elasticities:

$$\frac{\mathbb{E}\left[X\dot{\tau}^2\theta^A\left(\boldsymbol{W}\right)\right]}{\mathbb{E}\left[X\dot{\tau}^2\right]} \approx \frac{\mathbb{E}\left[X\theta^A\left(\boldsymbol{W}\right)\right]}{\mathbb{E}\left[X\right]}.$$

assup:: PPML weak correlation is not an exogeneity condition. It does not require trade flows to be unrelated to distance, tariffs, or other trade costs. In gravity applications, trade flows and trade costs are related by design; this is why the aggregate elasticity uses trade flow weights. The assumption only says that the additional local factor $\dot{\tau}^2$ in lemma:: PPML h does not change the weighted mix of elasticities very much. It is more plausible when observations with high and low $\dot{\tau}^2$ have similar elasticities after trade flows have already been used as weights. It is more likely to fail when, for example, very distant pairs or high tariff pairs have systematically different trade elasticities. In that case, the PPML coefficient should be interpreted as the local object weighted by $X\dot{\tau}^2$ in lemma:: PPML h, rather than as the simpler flow weighted aggregate elasticity in Prop:: PPML h. With many controls or fixed effects, the same idea applies to the remaining variation in $\dot{\tau}$ after those controls are accounted for.

For Model 1, if assm:: h function and assup:: PPML weak correlation hold, then $\widehat{\gamma} \xrightarrow{p} \gamma_0$ as $n \to \infty$, where $\gamma_0 \approx \mathbb{E}\left[X\theta^A\left(\boldsymbol{W}\right)\right]/\mathbb{E}\left[X\right]$.

Proof. See Proof: PPML h □

Compared to lemma:: PPML h, Prop:: PPML h shows that if assup:: PPML weak correlation holds, then $\widehat{\gamma}$ approximates $\mathbb{E}\left[X\theta^A\left(\boldsymbol{W}\right)\right]/\mathbb{E}\left[X\right]$, which we call the aggregate elasticity.

Thus Prop:: PPML h is an interpretation device: it explains when the PPML coefficient can be read as an aggregate elasticity, while the formal testing contribution below does not require this approximation itself to be exact.

Unlike log log OLS, which cannot include observations with a zero dependent variable because the logarithm is undefined, PPML estimation can mechanically incorporate them. This should not be read as saying that PPML discards zero trade flows. Consistent with Santos Silva and Tenreyro (2006), PPML is useful precisely because the score is defined when $X_i = 0$. If $\widehat{\mu}_i = \boldsymbol{W}_i^{\widehat{\beta}}\tau_i^{\widehat{\gamma}}$ denotes the fitted conditional mean, then a zero observation enters the trade cost score as $-\widehat{\mu}_i\dot{\tau}_i$, while the local slope derivative is proportional to $-\widehat{\mu}_i\dot{\tau}_i^2$. Thus the relevant leverage is governed by fitted trade, not by an indicator for whether the realized

flow is zero. Prop:: PPML h therefore implies a narrower statement: zero flow observations whose fitted trade is small typically move the trade cost slope little relative to high fitted value trade flows.

Zeros are not inconsequential for the PPML problem as a whole. They can affect the fitted fixed effects and the adding up or general equilibrium content of structural gravity; this is the sense emphasized by Fally (2015) and by work on high dimensional fixed effects such as Weidner and Zylkin (2021). Our claim is a clarification of the weighting interpretation for the trade cost slope, not a correction to SST's argument for estimating gravity in levels with PPML.

B.2 OLS and the Average Elasticity

This subsection gives the log linear counterpart to the PPML result. We take Model 2 and assm:: h bar function as maintained. Thus $\mathbb{E}\left[\log X \mid \boldsymbol{W}, \tau\right] = \beta^{*\top}\dot{\boldsymbol{W}} + \theta^a\left(\boldsymbol{W}\right)\dot{\tau}$, and the objective is to relate the constant elasticity OLS coefficient on $\log\tau$ to the heterogeneous elasticity function $\theta^a\left(\boldsymbol{W}\right)$.

For observations $\{(X_i, \boldsymbol{W}_i, \tau_i)\}_{i=1}^n$, the OLS estimators for $\widetilde{\beta}$ and $\widetilde{\zeta}$ are defined as the solution to

$$\sum_{i=1}^{n}\begin{pmatrix}\dot{\boldsymbol{W}}_i\\ \dot{\tau}_i\end{pmatrix}\left[\log X_i - \left(\widetilde{\beta}, \widetilde{\zeta}\right)\begin{pmatrix}\dot{\boldsymbol{W}}_i\\ \dot{\tau}_i\end{pmatrix}\right] = 0.$$

For Model 2, if assm:: h bar function holds, then $\widetilde{\zeta} \xrightarrow{p} \zeta_0$ as $n \to \infty$, where $\zeta_0 \approx \mathbb{E}\left[\theta^a\left(\boldsymbol{W}\right)\dot{\tau}^2\right]/\mathbb{E}\left[\dot{\tau}^2\right]$.

Proof. See Proof of lemma: OLS h bar □

lemma:: h bar shows that the OLS estimator $\widetilde{\zeta}$ approximates the weighted average of the elasticity θ^a, where the local observation weight is proportional to $\dot{\tau}^2$, which comes from the derivative of the population moment condition with respect to the log trade cost coefficient.

To show that the $\widetilde{\zeta}$ approximates $\mathbb{E}\left[\theta^a\left(\boldsymbol{W}\right)\right]$, which we call the average elasticity, we rely on the following assumption, mirroring assup:: PPML weak correlation from the previous section.

$\frac{Cov\left(\theta^a(\boldsymbol{W}),\dot{\tau}^2\right)}{\mathbb{E}[\theta^a(\boldsymbol{W})]\mathbb{E}[\dot{\tau}^2]} \approx 0.$

assm:: OLS weak correlation requires that $Cov\left(\theta^a\left(\boldsymbol{W}\right), \dot{\tau}^2\right)$, compared to $\mathbb{E}\left[\theta^a\left(\boldsymbol{W}\right)\right]\mathbb{E}\left[\dot{\tau}^2\right]$, is relatively small. This can occur in several cases: $\boldsymbol{W}$ may be nearly independent of $\dot{\tau}$, the elasticity θ^a may have low variance, or $\theta^a\left(\boldsymbol{W}\right)$ and $\dot{\tau}^2$ may move together in a way that cancels out in expectation.

This is the log linear analogue of assup:: PPML weak correlation: it asks that the mechanical leverage of $\dot{\tau}^2$ not be systematically aligned with the heterogeneous log linear elasticity. When that alignment is strong, OLS should be interpreted as the weighted by $\dot{\tau}^2$ object in lemma:: h bar, rather than as the simple average in Prop:: OLS h bar.

For Model 2, if assm:: h bar function and assm:: OLS weak correlation hold, then $\widetilde{\zeta} \xrightarrow{p} \zeta_0$ as $n \to \infty$, where $\zeta_0 \approx \mathbb{E}\left[\theta^a\left(\boldsymbol{W}\right)\right]$.

Proof. See Proof: OLS h bar □

B.3 Comparison across Estimators

The previous subsections characterize the elasticity objects associated with PPML and OLS. We now collect the corresponding empirical moment conditions and compare them with WLS, FWLS, and GPML. This comparison is useful for the gravity application in sec: Gravity, where differences across estimators are used to account for the gap between OLS and PPML distance elasticities.

Let Ω denote the set of country pairs and define $Z_{ij} \equiv \left(\dot{\boldsymbol{W}}_{ij}^{\top}, \dot{\tau}_{ij}\right)^{\top}$, where $\dot{\boldsymbol{W}}_{ij}$ collects the log controls and fixed effects and $\dot{\tau}_{ij}$ is the log trade cost argument. Let $\widehat{X}_{ij} = \boldsymbol{W}_{ij}^{\widehat{\beta}} \tau_{ij}^{\widehat{\xi}}$ denote the fitted trade flow. OLS solves $\sum_{i,j\in\Omega} Z_{ij}\left(\log X_{ij} - \log \widehat{X}_{ij}\right) = 0$. Thus OLS penalizes log deviations, or relative deviations, in trade flows. PPML solves $\sum_{i,j\in\Omega} Z_{ij}\left(X_{ij} - \widehat{X}_{ij}\right) = 0$. It therefore penalizes level deviations. In trade applications, this gives more leverage to country pairs with larger dollar trade flows.

Weighted least squares provides a simple way to separate the role of weights from the role of the residual scale. With weights proportional to X_{ij}, WLS solves $\sum_{i,j\in\Omega} Z_{ij}\left(\log X_{ij} - \log \widehat{X}_{ij}\right) X_{ij} = 0$. It keeps the log linear residual used by OLS but gives more weight to larger trade flows, as PPML does in levels. We use WLS as an accounting device rather than as a preferred estimator. Because the weight contains the dependent variable, we also report FWLS. FWLS first obtains fitted trade flows $\widetilde{X}_{ij}$ from an OLS regression and then solves $\sum_{i,j\in\Omega} Z_{ij}\left(\log X_{ij} - \log \widehat{X}_{ij}\right) \widetilde{X}_{ij} = 0$.

Finally, GPML solves $\sum_{i,j\in\Omega} Z_{ij}\left(X_{ij}/\widehat{X}_{ij} - 1\right) = 0$. Its residual is a relative deviation rather than a level deviation. For this reason, GPML is closer in interpretation to OLS than to PPML, although it is estimated in levels and is robust to the form of heteroskedasticity emphasized by Santos Silva and Tenreyro (2006). table:EMC summarizes these moment conditions.

This comparison clarifies the estimator comparisons used in the main text. The OLS and WLS difference isolates the effect of replacing equal observation weights with trade flow

TABLE VIII
Empirical Moment Conditions

Estimator	Empirical moment condition
OLS	$\sum_{i,j\in\Omega} Z_{ij}\left(\log X_{ij} - \log \widehat{X}_{ij}\right) = 0$
PPML	$\sum_{i,j\in\Omega} Z_{ij}\left(X_{ij} - \widehat{X}_{ij}\right) = 0$
WLS	$\sum_{i,j\in\Omega} Z_{ij}\left(\log X_{ij} - \log \widehat{X}_{ij}\right) X_{ij} = 0$
FWLS	$\sum_{i,j\in\Omega} Z_{ij}\left(\log X_{ij} - \log \widehat{X}_{ij}\right) \widetilde{X}_{ij} = 0$
GPML	$\sum_{i,j\in\Omega} Z_{ij}\left(X_{ij}/\widehat{X}_{ij} - 1\right) = 0$

Note: $Z_{ij} \equiv \left(\dot{\boldsymbol{W}}_{ij}^{\top}, \dot{\tau}_{ij}\right)^{\top}$. The fitted trade flow is $\widehat{X}_{ij} = \boldsymbol{W}_{ij}^{\widehat{\beta}} \tau_{ij}^{\widehat{\xi}}$. In FWLS, $\widetilde{X}_{ij}$ is the fitted trade flow from the first stage OLS regression.

weights while keeping a log linear residual. The WLS and PPML difference captures the remaining change associated with moving from log deviations to level deviations. Similarly, the OLS and GPML difference reflects the effect of using a level estimator with a relative deviation residual, while the GPML and PPML difference reflects the move from relative deviations to level deviations. These decompositions are descriptive; they are not structural welfare decompositions.

B.4 Link Between Linear and Nonlinear Elasticities

sec:elasticity-objects treats Model 1 and Model 2 as alternative maintained specifications. This subsection records the formal link between them. The link is useful for comparing $\theta^{A}(\boldsymbol{W})$ and $\theta^{a}(\boldsymbol{W})$, but it is not required for the heterogeneity test or the WPPML estimator.

ϵ is independent of $(\boldsymbol{W}, \tau, \xi)$.

Let assm:: Equivalence of two models hold. (i) Under Model 2, it holds that

$$\mathbb{E}\left[X \mid \boldsymbol{W}, \tau, \xi\right] = c_1 \boldsymbol{W}^{\beta^*} \tau^{\xi},$$

where $c_1 = \mathbb{E}\left[\epsilon \mid \boldsymbol{W}, \tau, \xi\right] = \mathbb{E}\left[\epsilon\right]$ is constant. (ii) Under Model 1, it holds that

$$\mathbb{E}\left[\log X \mid \dot{\boldsymbol{W}}, \dot{\tau}, \xi\right] = \beta^* \dot{\boldsymbol{W}} + \xi \dot{\tau} + c_2,$$

where $c_2 = \mathbb{E}\left[\log \epsilon \mid \boldsymbol{W}, \tau, \xi\right] = \mathbb{E}\left[\log \epsilon\right]$ is constant.

Proof. See Proof of lemma h and h bar □

lemma::Equivalence of two models establishes a sufficient condition under which both level and log conditional means have the random coefficient gravity form, up to constants in the error normalization. Under this condition, it is informative to compare the elasticity θ^A from Model 1 with the elasticity θ^a from Model 2. The next result gives simple cases in which the two objects coincide or inherit the same monotonic relationship with the controls.

For any $(\boldsymbol{w}, z)$, denote $F_{\xi|\boldsymbol{W}=\boldsymbol{w},\tau=z}(\cdot)$ as the conditional cumulative distribution function of ξ given $(\boldsymbol{W}, \tau) = (\boldsymbol{w}, z)$. For any two vectors $\boldsymbol{w}_1, \boldsymbol{w}_2 \in \mathbb{R}^p$, we say $\boldsymbol{w}_1 \leq \boldsymbol{w}_2$ if each element of $\boldsymbol{w}_1$ is less or equal to the corresponding element of $\boldsymbol{w}_2$.

Let assm:: h function, assm:: h bar function, and assm:: Equivalence of two models hold. For the monotonicity parts, also take the relevant support of τ to satisfy $\tau > 1$. (i) If ξ is constant, then both $\theta^A(\boldsymbol{W})$ and $\theta^a(\boldsymbol{W})$ are constant and identical. (ii) If for any $\boldsymbol{w}_1 \leq \boldsymbol{w}_2$, $F_{\xi|\boldsymbol{W}=\boldsymbol{w}_1,\tau=z}(t) \geq F_{\xi|\boldsymbol{W}=\boldsymbol{w}_2,\tau=z}(t)$ for any $t \in \mathbb{R}$, then both $\theta^A(\boldsymbol{W})$ and $\theta^a(\boldsymbol{W})$ are monotonically increasing functions. (iii) If for any $\boldsymbol{w}_1 \leq \boldsymbol{w}_2$, $F_{\xi|\boldsymbol{W}=\boldsymbol{w}_1,\tau=z}(t) \leq F_{\xi|\boldsymbol{W}=\boldsymbol{w}_2,\tau=z}(t)$ for any $t \in \mathbb{R}$, then both $\theta^A(\boldsymbol{W})$ and $\theta^a(\boldsymbol{W})$ are monotonically decreasing functions.

Proof. See Proof: h and h bar □

When ξ is constant, neither the elasticity θ^A nor the elasticity θ^a exhibits heterogeneity. This observation motivates the heterogeneity tests developed in sec:tests: to test the null hypothesis that ξ is constant, it suffices to verify that either $\theta^A(\boldsymbol{W})$ or $\theta^a(\boldsymbol{W})$ exhibit homogeneity.

The condition in Prop: h and h bar relation (ii) relates to the concept of first order stochastic dominance. Intuitively, it means that ξ has a smaller probability of being small (thus larger probability of being large) conditioning on $\boldsymbol{W}$ being larger. In this case, both $\theta^A(\boldsymbol{W})$ and $\theta^a(\boldsymbol{W})$ increase with $\boldsymbol{W}$. The same idea applies to Prop: h and h bar relation (iii).

It is worth mentioning that without additional assumption, the link between $\theta^A(\boldsymbol{W})$ and $\theta^a(\boldsymbol{W})$ remains subtle. For instance, neither $\theta^a(\boldsymbol{W})$ being constant implies $\theta^A(\boldsymbol{W})$ being constant, nor does $\theta^A(\boldsymbol{W})$ being constant imply $\theta^a(\boldsymbol{W})$ being constant. To obtain equivalence, one can impose stronger structure on the source of heterogeneity.

Assume that $\xi = g(\boldsymbol{W})$ for some unknown function $g(\cdot)$.

assm:: Equivalence requires the unobserved heterogeneous ξ to be a deterministic, although unrestricted, function of the observable $\boldsymbol{W}$. Since $\boldsymbol{W}$ is random, ξ remains random, but no unobserved variation remains beyond that induced by ξ.

Under assm:: Equivalence of two models and assm:: Equivalence, it holds that $\theta^a(\boldsymbol{w}) = \theta^A(\boldsymbol{w})$ for every $\boldsymbol{w}$ in the support of $\boldsymbol{W}$.

Proof. See Proof: Equivalence □

By Prop:: Equivalence, the two elasticities coincide pointwise. As a result, $\theta^a(\boldsymbol{W})$ is constant if and only if $\theta^A(\boldsymbol{W})$ is constant. If assm:: Equivalence fails, it is possible for $\theta^a(\boldsymbol{W})$ to be nonconstant even when $\theta^A(\boldsymbol{W})$ remains constant. For example, let random variable η have equal probabilities to be 0.5 and 1.5 when $\boldsymbol{W} = \boldsymbol{w}_1$ and take the value 1 with probability 1 when $\boldsymbol{W} = \boldsymbol{w}_2$. If $\xi = 1 + \log_\tau(\eta)$, then $\theta^A(\boldsymbol{W})$ is constant but not $\theta^a(\boldsymbol{W})$. At the same time, if $\xi = \eta$, then $\theta^a(\boldsymbol{W})$ is constant but not $\theta^A(\boldsymbol{W})$. These examples illustrate that Models 1 and 2 reflect different representations of the multiplicative relationship between X and $(\boldsymbol{W}, \tau)$. Without further structure they produce related but not identical elasticity objects.

C Technical Proofs

This appendix contains the proofs of the formal results. Proof headings cross reference the result being proved; results stated in Appendix B retain their B numbering.

C.1 Proof of lemma:: PPML h

Proof. To focus on $\widehat{\gamma}$, we replace $\widehat{\beta}$ with the true value β^* in the following discussion. This step is an approximation even when n is large, because the probability limit of $\widehat{\beta}$ is not β^* in the presence of heterogeneous elasticity. The PPML estimator $\widehat{\gamma}$ satisfies the first order condition: $\sum_{i=1}^n \left(X_i - \boldsymbol{W}_i^{\beta^*} \tau_i^{\widehat{\gamma}}\right) \dot{\tau}_i = 0$. Let Γ be the compact parameter space for γ. The uniform law of large numbers implies that

$$\sup_{\gamma \in \Gamma} \left\| \frac{1}{n} \sum_{i=1}^n \left(X_i - \boldsymbol{W}_i^{\beta^*} \tau_i^{\gamma}\right) \dot{\tau}_i - \mathbb{E}\left[\left(X - \boldsymbol{W}^{\beta^*} \tau^{\gamma}\right) \dot{\tau}\right] \right\| \xrightarrow{p} 0.$$

Therefore, we have that $\widehat{\gamma} \xrightarrow{p} \gamma_0$, where γ_0 satisfies the moment condition

$$\mathbb{E}\left[\left(X - \boldsymbol{W}^{\beta^*} \tau^{\gamma_0}\right) \dot{\tau}\right] = 0. \tag{2}$$

Note that γ_0 is a fixed constant that does not exhibit any randomness or change with $\boldsymbol{W}$. Our objective is to study how γ_0 relates to $\theta^A(\boldsymbol{W})$.

Let $f_{\boldsymbol{W},\tau}(\cdot)$ be the probability density functions of $(\boldsymbol{W}, \tau)$. By the law of iterated expectations, we have that

$$\mathbb{E}\left[\left(X - \boldsymbol{W}^{\beta^*} \tau^{\gamma_0}\right) \dot{\tau}\right] = \int \mathbb{E}\left[\left(X - \boldsymbol{W}^{\beta^*} \tau^{\gamma_0}\right) \dot{\tau} \mid \boldsymbol{W} = \boldsymbol{w}, \tau = z\right] f_{\boldsymbol{W},\tau}(\boldsymbol{w}, z)\, d\boldsymbol{w} dz.$$

We discretize the above integral by partitioning the support of $(\boldsymbol{W}, \tau)$ as $\{\Delta_k\}_{k=1}^K$. Such an approximation can be made as precise as possible by refining the partition and increasing K. As a result, we can express $\mathbb{E}\left[\left(X - \boldsymbol{W}^{\beta^*}\tau^{\gamma_0}\right)\dot{\tau}\right]$ as

$$\sum_{k=1}^{K} \mathbb{E}\left[\left(X - \boldsymbol{W}^{\beta^*}\tau^{\gamma_0}\right)\dot{\tau} \mid \boldsymbol{W} = \boldsymbol{w}_k, \tau = z_k\right] f_{\boldsymbol{W},\tau}\left(\boldsymbol{w}_k, z_k\right)\Delta_k = \mathbf{0}. \tag{3}$$

By Model 1 and assm:: h function, we have that for any $(\boldsymbol{w}, z)$,

$$\mathbb{E}\left[\left(X - \boldsymbol{W}^{\beta^*}\tau^{\theta^A(\boldsymbol{W})}\right)\dot{\tau} \mid \boldsymbol{W} = \boldsymbol{w}, \tau = z\right] = 0. \tag{4}$$

eq: conditional expectation being zero implies that

$$\begin{aligned}\mathbb{E}\left[\left(X - \boldsymbol{W}^{\beta^*}\tau^{\theta^A(\boldsymbol{W})}\right)\dot{\tau}\right] &= \int \mathbb{E}\left[\left(X - \boldsymbol{W}^{\beta^*}\tau^{\theta^A(\boldsymbol{W})}\right)\dot{\tau} \mid \boldsymbol{W} = \boldsymbol{w}, \tau = z\right] f_{\boldsymbol{W},\tau}\left(\boldsymbol{w}, z\right) d\boldsymbol{w}dz \\ &= \sum_{k=1}^{K} \mathbb{E}\left[\left(X - \boldsymbol{W}^{\beta^*}\tau^{\theta^A(\boldsymbol{W})}\right)\dot{\tau} \mid \boldsymbol{W} = \boldsymbol{w}_k, \tau = z_k\right] f_{\boldsymbol{W},\tau}\left(\boldsymbol{w}_k, z_k\right)\Delta_k = \mathbf{0}.\end{aligned} \tag{5}$$

Denote $\gamma_k \equiv \theta^A\left(\boldsymbol{w}_k\right)$ and

$$\phi_k\left(\gamma\right) \equiv \mathbb{E}\left[\left(X - \boldsymbol{W}^{\beta^*}\tau^{\gamma}\right)\dot{\tau} \mid \boldsymbol{W} = \boldsymbol{w}_k, \tau = z_k\right] f_{\boldsymbol{W},\tau}\left(\boldsymbol{w}_k, z_k\right)\Delta_k.$$

We linearly approximate $\sum_{k=1}^K \phi_k\left(\gamma_0\right)$ using the first order Taylor expansion at $(\gamma_1, \ldots, \gamma_K)$:

$$\sum_{k=1}^{K} \phi_k\left(\gamma_0\right) \approx \sum_{k=1}^{K} \phi_k\left(\gamma_k\right) + \sum_{k=1}^{K} \phi_k'\left(\gamma_k\right)\left(\gamma_0 - \gamma_k\right).$$

Together with eq:1 and eq:2, it holds that $\sum_{k=1}^K \phi_k'\left(\gamma_k\right)\left(\gamma_0 - \gamma_k\right) \approx 0$. Thus, we obtain that

$$\gamma_0 \approx \left[\sum_{k=1}^{K} \phi_k'\left(\gamma_k\right)\right]^{-1}\left[\sum_{k=1}^{K} \phi_k'\left(\gamma_k\right)\theta^A\left(\boldsymbol{w}_k\right)\right].$$

Taking the first order derivative provides us that

$$\phi_k'\left(\gamma_k\right) = -\mathbb{E}\left[\boldsymbol{W}^{\beta^*}\tau^{\gamma_k}\dot{\tau}^2 \mid \boldsymbol{W} = \boldsymbol{w}_k, \tau = z_k\right] f_{\boldsymbol{W},\tau}\left(\boldsymbol{w}_k, z_k\right)\Delta_k.$$

By eq: conditional expectation being zero, we have that

$$\mathbb{E}\left[\boldsymbol{W}^{\beta^*}\tau^{\gamma_k} \mid \boldsymbol{W}=\boldsymbol{w}_k, \tau=z_k\right]=\mathbb{E}\left[\boldsymbol{W}^{\beta^*}\tau^{\theta^A(\boldsymbol{w}_k)} \mid \boldsymbol{W}=\boldsymbol{w}_k, \tau=z_k\right]=\mathbb{E}\left[X \mid \boldsymbol{W}=\boldsymbol{w}_k, \tau=z_k\right].$$

This shows that

$$\begin{aligned}\sum_{k=1}^{K}\phi_k'\left(\gamma_k\right) &= -\sum_{k=1}^{K}\mathbb{E}\left[X\dot{\tau}^2 \mid \boldsymbol{W}=\boldsymbol{w}_k, \tau=z_k\right] f_{\boldsymbol{W},\tau}\left(\boldsymbol{w}_k, z_k\right)\Delta_k \\ &= -\int \mathbb{E}\left[X\dot{\tau}^2 \mid \boldsymbol{W}=\boldsymbol{w}, \tau=z\right] f_{\boldsymbol{W},\tau}\left(\boldsymbol{w}, z\right) d\boldsymbol{w}dz = -\mathbb{E}\left[X\dot{\tau}^2\right],\end{aligned}$$

where the last equality follows from the law of iterated expectations. Following a similar argument, we can obtain that

$$\begin{aligned}\sum_{k=1}^{K}\phi_k'\left(\gamma_k\right)\theta^A\left(\boldsymbol{w}_k\right) &= -\sum_{k=1}^{K}\mathbb{E}\left[X\dot{\tau}^2\theta^A\left(\boldsymbol{w}_k\right) \mid \boldsymbol{W}=\boldsymbol{w}_k, \tau=z_k\right] f_{\boldsymbol{W},\tau}\left(\boldsymbol{w}_k, z_k\right)\Delta_k \\ &= -\int \mathbb{E}\left[X\dot{\tau}^2\theta^A\left(\boldsymbol{W}\right) \mid \boldsymbol{W}, \tau\right] f_{\boldsymbol{W},\tau}\left(\boldsymbol{w}, z\right) d\boldsymbol{w}dz = -\mathbb{E}\left[X\dot{\tau}^2\theta^A\left(\boldsymbol{W}\right)\right].\end{aligned}$$

Thus, we obtain that the PPML estimator $\widehat{\gamma}$ converges in probability to γ_0, which can be approximated by $\mathbb{E}\left[X\dot{\tau}^2\theta^A\left(\boldsymbol{W}\right)\right]/\mathbb{E}\left[X\dot{\tau}^2\right]$. The theorem then follows. □

C.2 Proof of Prop:: PPML h

Proof. Based on lemma:: PPML h, the local object targeted by PPML is

$$\frac{\mathbb{E}\left[X\dot{\tau}^2\theta^A\left(\boldsymbol{W}\right)\right]}{\mathbb{E}\left[X\dot{\tau}^2\right]}.$$

assup:: PPML weak correlation states directly that this object is close to $\mathbb{E}\left[X\theta^A\left(\boldsymbol{W}\right)\right]/\mathbb{E}\left[X\right]$.
Hence, the PPML estimator $\widehat{\gamma}$ converges in probability to γ_0, which is approximately $\mathbb{E}\left[X\theta^A\left(\boldsymbol{W}\right)\right]/\mathbb{E}\left[X\right]$.
□

C.3 Proof of lemma:: h bar

Proof. Following the proof of Prop:: PPML h, we focus on $\widetilde{\zeta}$ and replace $\widetilde{\beta}$ with its true value β^*. We then have that $\widetilde{\zeta}$ satisfies the first order condition $\sum_{i=1}^{n}\left(\log X_i - \beta^*\dot{\boldsymbol{W}}_i^{\top} - \widetilde{\zeta}\dot{\tau}_i\right)\dot{\tau}_i =$

0. Because of the linearity in $\widetilde{\zeta}$, we can derive the closed form expression of $\widetilde{\zeta}$ as

$$\widetilde{\zeta} = \left(\sum_{i=1}^{n} \dot{\tau}_i^2 \right)^{-1} \sum_{i=1}^{n} \left(\log X_i - \beta^* \dot{\boldsymbol{W}}_i^\top \right) \dot{\tau}_i.$$

By the law of large numbers, it holds that $\frac{1}{n} \sum_{i=1}^{n} \dot{\tau}_i^2 \xrightarrow{p} \mathbb{E}\left[\dot{\tau}^2\right]$ as $n \to \infty$. Replacing $\log X_i$ with the right hand side of the equality in Model 2, we obtain that

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^{n} \left(\log X_i - \beta^* \dot{\boldsymbol{W}}_i^\top \right) \dot{\tau}_i &= \frac{1}{n} \sum_{i=1}^{n} \left(\xi_i \dot{\tau}_i + \log \epsilon_i \right) \dot{\tau}_i \\ &= \frac{1}{n} \sum_{i=1}^{n} \xi_i \dot{\tau}_i^2 + \frac{1}{n} \sum_{i=1}^{n} \log \epsilon_i \dot{\tau}_i \\ &\xrightarrow{p} \mathbb{E}\left[\xi \dot{\tau}^2\right] + \mathbb{E}\left[\log \epsilon \dot{\tau}\right] \end{aligned}$$

as $n \to \infty$. By the assumption of Model 2, $\mathbb{E}\left[\log \epsilon \dot{\tau}\right] = 0$. assm:: h bar function implies that

$$\mathbb{E}\left[\xi \dot{\tau}^2\right] = \mathbb{E}\left[\mathbb{E}\left[\xi \dot{\tau}^2 \mid \boldsymbol{W}, \tau\right]\right] = \mathbb{E}\left[\theta^a\left(\boldsymbol{W}\right) \dot{\tau}^2\right].$$

Therefore, $\widetilde{\zeta} \xrightarrow{p} \mathbb{E}\left[\theta^a\left(\boldsymbol{W}\right) \dot{\tau}^2\right] / \mathbb{E}\left[\dot{\tau}^2\right]$. □

C.4 Proof of Prop:: OLS h bar

Proof. By lemma:: h bar, we have that $\widetilde{\zeta} \xrightarrow{p} \mathbb{E}\left[\theta^a\left(\boldsymbol{W}\right) \dot{\tau}^2\right] / \mathbb{E}\left[\dot{\tau}^2\right]$. Under assm:: OLS weak correlation, $\mathbb{E}\left[\theta^a\left(\boldsymbol{W}\right) \dot{\tau}^2\right]$ is approximately $\mathbb{E}\left[\theta^a\left(\boldsymbol{W}\right)\right] \mathbb{E}\left[\dot{\tau}^2\right]$. As a result, $\widetilde{\zeta} \xrightarrow{p} \zeta_0$, where $\zeta_0 \approx \mathbb{E}\left[\theta^a\left(\boldsymbol{W}\right)\right]$. □

C.5 Proof of lemma::Equivalence of two models

Proof. To prove the first part of the theorem, we take the exponential on both sides of the equality in Model 2. It holds that $X = \boldsymbol{W}^{\beta^*} \tau^{\xi} \epsilon$. Because ϵ is independent of $(\boldsymbol{W}, \tau, \xi)$, we have $\mathbb{E}\left[X \mid \boldsymbol{W}, \tau, \xi\right] = c_1 \boldsymbol{W}^{\beta^*} \tau^{\xi}$, where c_1 is defined in the theorem.

For the second part, we take the logarithm on both sides of the equality in Model 1. We have that $\log X = \beta^* \dot{\boldsymbol{W}} + \xi \dot{\tau} + \log \epsilon$. Because ϵ is independent of $(\boldsymbol{W}, \tau, \xi)$, any function of ϵ is also independent of $(\boldsymbol{W}, \tau, \xi)$. As a result, $\mathbb{E}\left[\log X \mid \boldsymbol{W}, \tau, \xi\right] = \beta^* \dot{\boldsymbol{W}} + \xi \dot{\tau} + c_2$ for c_2 being defined in the theorem. Hence, the theorem holds. □

C.6 Proof of Prop: h and h bar relation

Proof. Part (i) is straightforward. If ξ is constant, then $\theta^A(\boldsymbol{W}) = \theta^a(\boldsymbol{W}) = \xi$. For part (ii), if $F_{\xi|\boldsymbol{W}=\boldsymbol{w}_1,\tau=z}(t) \geq F_{\xi|\boldsymbol{W}=\boldsymbol{w}_2,\tau=z}(t)$ for any $t \in \mathbb{R}$, then we have

$$\int \mu(t,z)\, dF_{\xi|\boldsymbol{W}=\boldsymbol{w}_1,\tau=z}(t) \leq \int \mu(t,z)\, dF_{\xi|\boldsymbol{W}=\boldsymbol{w}_2,\tau=z}(t), \tag{6}$$

for any function $\mu : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ that is nondecreasing in its first argument given the second argument fixed. Let $\mu(t,z) = z^t$, which is increasing in t when $z > 1$. Inequality eq: inequality expectation implies that

$$\begin{aligned}\mathbb{E}\left[\tau^{\xi} \mid \boldsymbol{W} = \boldsymbol{w}_1, \tau = z\right] &= \int z^t dF_{\xi|\boldsymbol{W}=\boldsymbol{w}_1,\tau=z}(t) \\ &\leq \int z^t dF_{\xi|\boldsymbol{W}=\boldsymbol{w}_2,\tau=z}(t) = \mathbb{E}\left[\tau^{\xi} \mid \boldsymbol{W} = \boldsymbol{w}_2, \tau = z\right].\end{aligned}$$

By assm:: h function, the left hand side is $z^{\theta^A(\boldsymbol{w}_1)}$ and the right hand side is $z^{\theta^A(\boldsymbol{w}_2)}$. Since $z > 1$, applying $\log_z(\cdot)$ preserves the ordering. Because $\boldsymbol{w}_1 \leq \boldsymbol{w}_2$, we have that $\theta^A(\cdot)$ is monotonically increasing. By letting $\mu(t,z) = t$, we can obtain that $\theta^a(\cdot)$ is also monotonically increasing. Part (iii) of the theorem follows from the similar argument. □

C.7 Proof of Prop:: Equivalence

Proof. If $\xi = g(\boldsymbol{W})$, then

$$\theta^a(\boldsymbol{W}, \tau) = \mathbb{E}\left[g(\boldsymbol{W}) \mid \boldsymbol{W}, \tau\right] = g(\boldsymbol{W}) = \theta^a(\boldsymbol{W}).$$

At the same time, we have that

$$\begin{aligned}\theta^A(\boldsymbol{W}, \tau) &= \log_\tau\left(\frac{\mathbb{E}\left[X \mid \boldsymbol{W}, \tau\right]}{\boldsymbol{W}^{\beta^*}}\right) = \log_\tau\left(\frac{\mathbb{E}\left[\boldsymbol{W}^{\beta^*}\tau^{\xi}\varepsilon \mid \boldsymbol{W}, \tau\right]}{\boldsymbol{W}^{\beta^*}}\right) \\ &= \log_\tau\left(\mathbb{E}\left[\tau^{\xi} \mid \boldsymbol{W}, \tau\right]\right) = \log_\tau\left(\mathbb{E}\left[\tau^{g(\boldsymbol{W})} \mid \boldsymbol{W}, \tau\right]\right) \\ &= \log_\tau\left(\tau^{g(\boldsymbol{W})}\right) = g(\boldsymbol{W}) = \theta^A(\boldsymbol{W}).\end{aligned}$$

As a result, it holds that $\theta^A(\boldsymbol{W}) = \theta^a(\boldsymbol{W}) = g(\boldsymbol{W})$. □

C.8 Proof of Prop: WPPML

Proof. The proof uses the same population moment argument as the PPML result in Prop:: PPML h, but applies it to the weighted sample defined by the treatment profile. We use the single index notation of sec:wppml. Let ω_i be proportional to s_i, and let θ_0 denote the probability limit of the WPPML coefficient when β is evaluated at β^*. The sample moment for the trade cost coefficient is

$$\sum_{i\in\Omega} \omega_i \dot{\tau}_i \left(X_i - \boldsymbol{W}_i^{\beta^*} \tau_i^{\widehat{\theta}_{\boldsymbol{s}}^W} \right) = 0.$$

Under the standard uniform law of large numbers for the weighted PPML moment, θ_0 solves the corresponding population moment,

$$\mathbb{E}\left[\omega_i \dot{\tau}_i \left(X_i - \boldsymbol{W}_i^{\beta^*} \tau_i^{\theta_0} \right) \right] = 0,$$

where the expectation is taken over observations in the reform sample.

By Model 1 and assm:: h function, $\mathbb{E}\left[X_i \mid \boldsymbol{W}_i, \tau_i\right] = \boldsymbol{W}_i^{\beta^*} \tau_i^{\theta^A(\boldsymbol{W}_i)}$. Therefore the population moment can be written as

$$\mathbb{E}\left[\omega_i \dot{\tau}_i \boldsymbol{W}_i^{\beta^*} \left(\tau_i^{\theta^A(\boldsymbol{W}_i)} - \tau_i^{\theta_0} \right) \right] = 0.$$

A first order expansion of $\tau_i^{\theta_0}$ around $\theta^A\left(\boldsymbol{W}_i\right)$ gives

$$\tau_i^{\theta_0} \approx \tau_i^{\theta^A(\boldsymbol{W}_i)} \left[1 + \dot{\tau}_i \left(\theta_0 - \theta^A\left(\boldsymbol{W}_i\right) \right) \right].$$

Substituting this expansion into the population moment yields

$$\mathbb{E}\left[\omega_i X_i \dot{\tau}_i^2 \left(\theta_0 - \theta^A\left(\boldsymbol{W}_i\right) \right) \right] \approx 0.$$

Thus the local WPPML coefficient satisfies

$$\theta_0 \approx \frac{\mathbb{E}\left[\omega_i X_i \dot{\tau}_i^2 \theta^A\left(\boldsymbol{W}_i\right) \right]}{\mathbb{E}\left[\omega_i X_i \dot{\tau}_i^2 \right]}.$$

The treatment weighted version of assup:: PPML weak correlation removes the additional local leverage term $\dot{\tau}_i^2$ from this ratio, so that

$$\theta_0 \approx \frac{\mathbb{E}\left[\omega_i X_i \theta^A\left(\boldsymbol{W}_i\right) \right]}{\mathbb{E}\left[\omega_i X_i \right]}.$$

Because ω_i is proportional to s_i, this object is the population counterpart of

$$\bar{\theta}_{\boldsymbol{s}} = \frac{\sum_{i\in\Omega} s_i X_i \theta^A(\boldsymbol{W}_i)}{\sum_{i\in\Omega} s_i X_i}.$$

It follows that $\left|\widehat{\theta}^W_{\boldsymbol{s}} - \bar{\theta}_{\boldsymbol{s}}\right| \xrightarrow{p} \delta$, with $\delta \approx 0$. Finally, assm:: WPPML makes the reform local, so the finite change elasticity is approximated by $\theta^A(\boldsymbol{W}_i)$. The aggregate response to the reform is therefore approximated by

$$\theta_{\boldsymbol{s}} \approx \bar{\theta}_{\boldsymbol{s}} \frac{\sum_{i\in\Omega} s_i X_i}{\sum_{i\in\Omega} X_i} = e_{\boldsymbol{s}} \bar{\theta}_{\boldsymbol{s}},$$

and the plug in estimator is $e_{\boldsymbol{s}}\widehat{\theta}^W_{\boldsymbol{s}}$. □

D Heterogeneity and Heteroskedasticity

For reference, the Monte Carlo below uses the following average and aggregate elasticity definitions:

$$\delta^{AE} = \frac{1}{N^2}\sum_{i,j}\left(\frac{X'_{ij} - X_{ij}}{X_{ij}} / \frac{d'_{ij} - d_{ij}}{d_{ij}}\right) = \frac{1}{N^2}\sum_{i,j}\delta_{ij}, \tag{7}$$

$$\delta^{EA} = \frac{\sum_{i,j}\delta_{ij}X_{ij}}{\sum_{i,j}X_{ij}} = \frac{\sum_{i,j}\delta_{ij}X_{ij}}{N^2} / \frac{\sum_{i,j}X_{ij}}{N^2}, \tag{8}$$

In this section, we simulate bilateral trade flows under the assumptions of two types of countries and heterogeneous trade elasticities across country pairs. The simulation results indicate that using the OLS estimator under such conditions yields biased estimates for the parameter of interest (the aggregate elasticity) and that the application of the PPML and WLS estimators corrects this bias.

Data Generating Process

We incorporate the heterogeneous elasticities into SST's specifications of heteroskedastic errors in the data generating process (DGP):

$$X_{ij} = \exp\left(\beta_1^* \log \mathrm{GDP}_i + \beta_2^* \log \mathrm{GDP}_j + \theta_{ij}^* \log \mathrm{Distance}_{ij}\right) * \epsilon_{ij} \tag{9}$$

where ϵ_{ij} is a log normal random variable with mean 1 and variance σ^2.

Following SST, we consider the following three specifications of σ^2 to assess the performance of different estimators under different patterns of heteroskedasticity.[9] $\boldsymbol{W}_{ij} \equiv (\text{GDP}_i, \text{GDP}_j)$ and $\tau_{ij} \equiv \text{Distance}_{ij}$.

Case 1: $\sigma^2 = \exp(\log X_{ij})^{-2}$; $Var\left(X_{ij} \mid \boldsymbol{W}_{ij}, \tau_{ij}, \theta^*_{ij}\right) = 1$.

Case 2: $\sigma^2 = \exp(\log X_{ij})^{-1}$; $Var\left(X_{ij} \mid \boldsymbol{W}_{ij}, \tau_{ij}, \theta^*_{ij}\right) = \exp(\log X_{ij})$.

Case 3: $\sigma^2 = 1$; $Var\left(X_{ij} \mid \boldsymbol{W}_{ij}, \tau_{ij}, \theta^*_{ij}\right) = \exp(\log X_{ij})^2$.

We consider a simple economy in which there are two types of countries, large and small, with $\text{GDP}_L = 100$ and $\text{GDP}_S = 10$, respectively. The total number of countries in the economy is $N = 100$. The effect of GDP on bilateral trade flows is assumed to be 1, i.e., $\beta^*_1 = \beta^*_2 = 1$. The distance between a country pair is randomly generated from the uniform distribution on $[2, 3]$, and a counterfactual case with a 1% increase in each pairwise distance is also generated to reflect the changes in trade costs.

In terms of heterogeneity, the value of the trade elasticity between two countries θ^*_{ij} can be either -1 or -0.5. We first examine the scenario where $\Pr\left(\theta^*_{ij} = -1\right) = \Pr\left(\theta^*_{ij} = -0.5\right) = 0.5$. This satisfies the exogeneity with heterogeneity assumption in Wooldridge (2005) that the unobserved heterogeneity is independent of the structural covariates. Then we rely on the findings of Fieler (2011) and assume that the trade elasticity between two large countries is $\theta^*_{LL} = -0.5$ and that the elasticity for the pairs of two small or one small and one large country is $\theta^*_{SS} = \theta^*_{SL} = -1$. Our results, however, do not depend on the source of heterogeneity, and in general, we are agnostic about it.[10]

This design varies one source of sorting behind the PPML minus OLS gap: the association between trade flow size and heterogeneous elasticities. The diagnostic implied by assup:: PPML weak correlation is more specific. Prop:: PPML h becomes less accurate when the leverage term $\dot{\tau}^2$ is systematically sorted toward pairs with different values of $\theta^A(\boldsymbol{W})$ after flow weighting. Thus, a direct stress test of the assumption would vary the strength of this leverage sorting; the discussion following assup:: PPML weak correlation shows the direction in which the approximation should deteriorate.

We start by applying the DGP when all countries are small, i.e., $N_L = 0$, and increase the number of large countries in the economy by one after each iteration until $N_L = 100$. For each of the $K = 101$ cases, we run 80 simulations with new draws of errors and bilateral distances. For each simulation, the aggregate elasticity δ^{EA} is calculated from eq:delta EA

[9] Due to the differences between our DGP and that in SST, the fourth case of heteroskedasticity is omitted.

[10] While GDPs are observed and controlled for by fixed effects, its effect on distance elasticity is of bilateral nature and is unobserved. We chose the distance elasticity to depend on GDP to generate a link between the size of trade flows and the elasticity heterogeneity. In the absence of such a link, Jensen's inequality becomes equality, and the average elasticity becomes equal to the elasticity of the average.

with d'_{ij} being the counterfactual distance with a 1% increase from d_{ij} and X'_{ij} being the corresponding trade flows generated from eq:dgp. This corresponds to the uniform treatment. The average elasticity δ^{AE} is calculated from eq:delta AE.

Estimates

In this section, we investigate the behavior of the two distinct parameters of interest, the aggregate elasticity and the average elasticity, and the performance of different estimators for models with heterogeneous country pairs in the simulated economy and the above mentioned heteroskedastic errors. The WLS estimator is applied with the share of every country pair in total world trade $X_{ij}/\sum_{i,j} X_{ij}$ as weights. For the FWLS estimator, the share of predicted trade flows $\widetilde{X}_{ij}/\sum_{i,j} X_{ij}$ is used as weights. The GPML and PPML estimators use X_{ij} as the dependent variable and the log of GDPs and log of bilateral distance as independent variables, while all other estimators use $\log X_{ij}$ as the dependent variable.

To provide a clearer picture of the differences between estimators, We express the bias of using different estimators as the percentage difference between the estimates and the aggregate elasticity as follows:

$$bias_k = 1 - \frac{\widehat{\theta}_k}{\delta_k^{EA}}, \tag{10}$$

where $k \in \{0, 1, \ldots, 100\}$ is the number of large countries in the economy and $\widehat{\theta}_k$ is the estimated result when there are k large countries in the economy. The value of δ_k^{EA} in the k-th case is calculated as the average of 80 simulations.

In the case where unobserved heterogeneity is assumed to be independent of the country sizes, each country pair will have an equal opportunity of having a trade elasticity of -1 or -0.5. fig:indp displays the behavior of all estimators of the trade elasticity. As a reference, we include the aggregate elasticity of each heterogeneity case, δ_k^{EA}, and the average elasticity, δ_k^{AE},[11] in the figures of estimates.

The stark difference between the values of the aggregate elasticity and the average elasticity is straightforward from the left panel graphs of fig:indp. Consequently, the behavior of the selected estimators differentiates from each other and centers around the two distinctive concepts. Firstly, PPML, WLS, and FWLS have similar patterns as δ_k^{EA}, while GPML and OLS have similar patterns as δ_k^{AE}. This is consistent with our discussion of the empirical moment conditions summarized in subsec:Country-pair-heterogeneity. In terms of estimated values, PPML and GPML estimates are close to δ_k^{EA}, while OLS and FWLS estimates are close to δ_k^{AE}, and WLS estimates are close to neither.

[11] For each k, the value of δ_k^{AE} is calculated as the average of 80 simulations

The distinction between δ_k^{EA} and δ_k^{AE} is due to the existence of country pair heterogeneity and heteroskedasticity in the error term. SST shows that both PPML and GPML estimators address the issue of heteroskedastic errors, so the difference between these two estimators results from the fact that PPML assigns different weights to country pairs in the sample while GPML does not. When country pair heterogeneity is randomly determined, the weighting does not seem to be crucial, as the biases from these two estimators are almost negligible when Case 1 and 2 types of heteroskedastic error are present (fig:biases1$_i ndp and fig : biases2_i ndp). On the other hand, OLS, WLS, and FWLS estimators all suffer from$

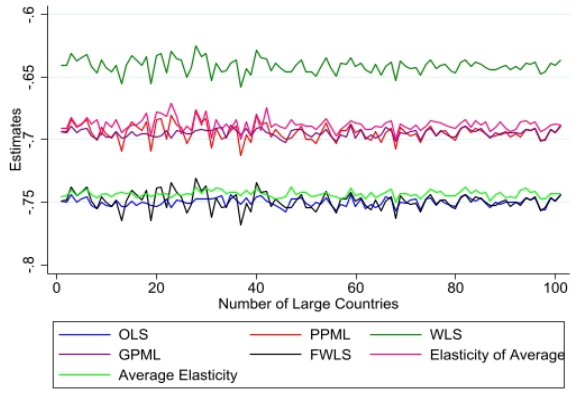
As discussed by SST, Case 3 is a special case that the log linear model corrects the heteroskedasticity in the data and that the GPML is the optimal PML estimator in this case. Therefore, in fig:biases3$_i ndp, the GPML estimator still has a negligible bias, while the bias of the PPML estimator, i$

Next, we follow Fieler (2011) and introduce the dependence of heterogeneity on country size, where trade elasticities between a pair of large countries will be larger in absolute value than those between country pairs involving a small country. Before turning to the heterogeneous simulations, table:benchmark first reports the benchmark case with no country pair heterogeneity, in which all countries are small. Recall that $\theta_{LL}^* = -0.5$ and $\theta_{SS}^* = \theta_{SL}^* = -1$. Under these conditions, the SST framework implies a common trade elasticity of -1 for the relevant country pairs. In Cases 1 and 2, PPML is closer to this value than OLS and has smaller standard errors. In Case 3, OLS is closer to the benchmark elasticity, consistent with the SST simulation design.

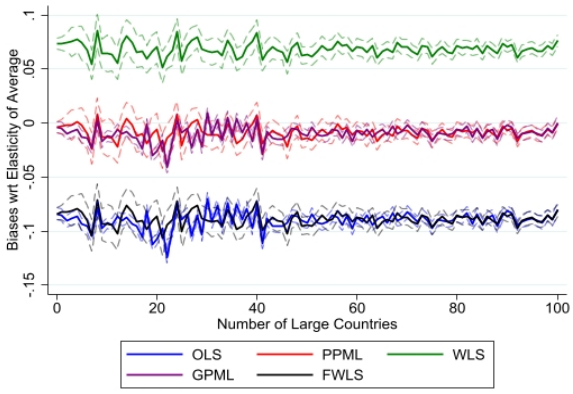
Then we introduce country pair heterogeneity into the benchmark by making half of the countries large. As shown in the second half of table:benchmark, introducing country pair heterogeneity widens the gap between the log linear and level estimators. The WLS estimates are close to the PPML estimates, consistent with the weighting interpretation in Section 3. The FWLS estimates are also close to WLS, suggesting that replacing true trade weights with fitted trade weights has little effect in this design. GPML is close to OLS once heterogeneity is introduced. In these simulations, therefore, OLS and GPML are closer to average elasticity objects, whereas PPML, WLS, and FWLS are closer to the aggregate elasticity.

Figure 5. Estimates and biases of OLS, WLS, GPML, and PPML (Independent Heterogeneity)

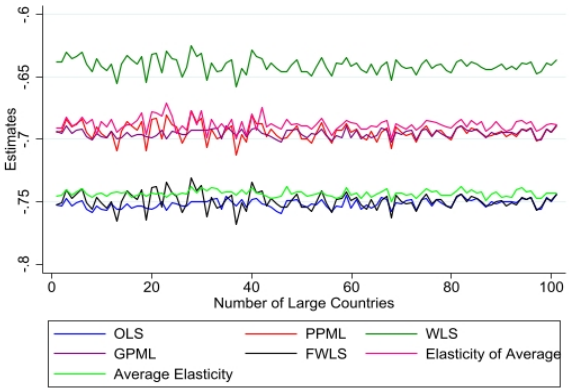
(a) Estimates Case 1



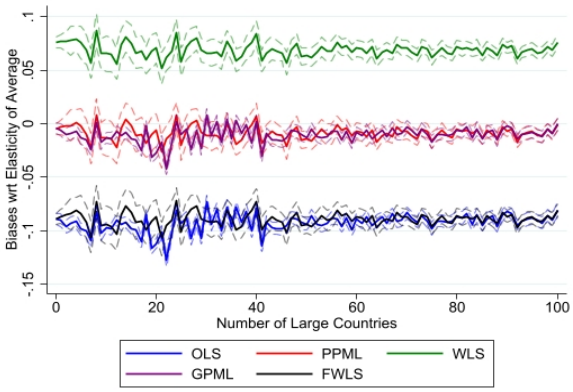
(b) Biases Case 1



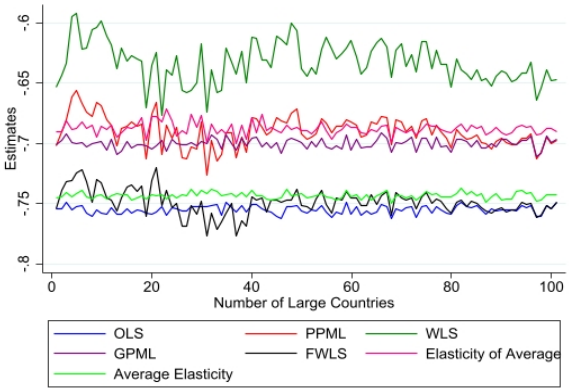
(c) Estimates Case 2



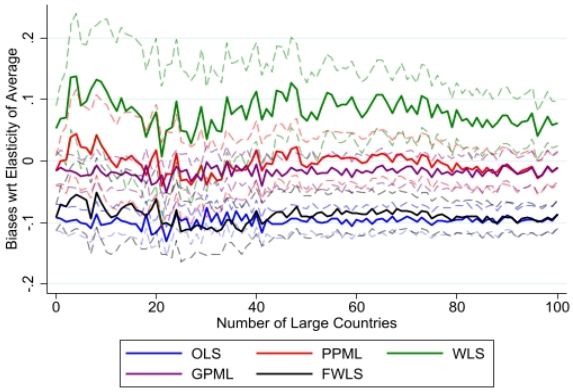
(d) Biases Case 2



(e) Estimates Case 3



(f) Biases Case 3



Note: The horizontal axis shows the number of large countries in the economy, and the vertical axis is the estimated result. Cases 1, 2, and 3 heteroskedastic errors follow the same specification as in Section 3.1. Trade elasticity (aggregate elasticity) is calculated as the change in level trade flow with a 1% increase in bilateral distance and is shown in red. Average elasticity is calculated as the unweighted average of the true θ's in the DGP and is shown in orange. Biases are calculated using eq:bias, which are the percentage difference between the aggregate elasticity and the estimated coefficient for different estimators. The 95% confidence intervals of the estimated biases are included as dashed lines.

Table IX
Sample estimates and biases

$N_L = 0$	*Case 1*		*Case 2*		*Case 3*	
	Estimate	Bias	Estimate	Bias	Estimate	Bias
OLS	-0.99997	-0.000067	-1.00340	-0.003456	-0.99043	0.009467
	(-1.00010, -0.99980)	(-0.000230, 0.000097)	(-1.00500, -1.00170)	(-0.005137, -0.001775)	(-1.00500, -0.97584)	(-0.005128, 0.024062)
GPML	-0.99988	0.000016	-0.99883	0.001069	-0.98833	0.011568
	(-1.00000, -0.99972)	(-0.000148, 0.000181)	(-1.00050, -0.99715)	(-0.000610, 0.002479)	(-1.00570, -0.97094)	(-0.005824, 0.028959)
PPML	-0.99989	0.000012	-0.99886	0.001038	-0.98788	0.012024
	(-1.00000, -0.99973)	(-0.000145, 0.000169)	(-1.00050, -0.99724)	(-0.000586, 0.002661)	(1.00470, -0.97110)	(-0.004754, 0.028802)
WLS	-0.99981	0.000095	-0.99436	0.005541	-0.98379	0.016107
	(-0.99996, -0.99965)	(-0.000062, 0.000252)	(-0.99599, -0.99273)	(0.003913, 0.007170)	(-1.01020, -0.95735)	(-0.010338, 0.042553)
FWLS	-0.99997	-0.000071	-1.00340	-0.003466	-0.99047	0.009429
	(-1.00010, -0.99981)	(-0.000227, 0.000086)	(-1.00500, -1.00170)	(-0.005093, -0.001839)	(-1.00480, -0.97617)	(-0.0004876, 0.023735)
$N_L = 50$	*Case 1*		*Case 2*		*Case 3*	
	Estimate	Bias	Estimate	Bias	Estimate	Bias
OLS	-0.88112	-0.567010	-0.88177	-0.568160	-0.87162	-0.549710
	(-0.88294, -0.87931)	(-0.570290, -0.563730)	(-0.88369, -0.87985)	(-0.571610, -0.564710)	(-0.88598, -0.85726)	(-0.575180, -0.524240)
GPML	-0.86736	-0.542540	-0.86653	-0.541060	-0.85408	-0.518510
	(-0.86918, -0.86554)	(-0.545820, -0.539250)	(-0.86852, -0.86455)	(-0.544630, -0.537490)	(-0.87129, -0.83686)	(-0.549030, -0.487980)
PPML	-0.56238	-0.000141	-0.56220	0.000176	-0.53943	0.040914
	(-0.56264, -0.56212)	(-0.000616, 0.000333)	(-0.56254, -0.56186)	(-0.000433, 0.000784)	(-0.56987, -0.50889)	(-0.013219, 0.095047)
WLS	-0.56130	0.001769	-0.56101	0.002284	-0.53531	-0.032658
	(-0.56157, -0.56104)	(0.001284,0.002253)	(-0.56135, -0.56067)	(0.001670, 0.002897)	(-0.58083, -0.48978)	(-0.032658,0.129260)
FWLS	-0.57515	-0.022858	-0.57494	-0.002483	-0.55603	0.011342
	(-0.57545, -0.57485)	(-0.023411, -0.22305)	(-0.57531, -0.57457)	(-0.023143, -0.021824)	(-0.58076, -0.53130)	(-0.032656, -0.055340)

Note: This table contains both the case with no country pair heterogeneity and the heterogeneous case with the number of large and small countries both equal to 50. N_L is the number of large countries in the economy. Case 1, 2, and 3 heteroskedasticity follow the corresponding specifications in Section 3.1. Biases are calculated using eq:bias. The estimated 95% confidence intervals are included in brackets.

fig:dep displays the behavior of all estimators of the trade elasticity as the share of large countries varies. The figure also reports the aggregate elasticity, δ_k^{EA}, and the average elasticity, δ_k^{AE}, for each value of k. As the number of large countries increases, the average estimated elasticity moves with the changing composition of country pairs.

The left panels show that OLS and GPML track the average elasticity more closely than the aggregate elasticity. PPML, WLS, and FWLS are closer to δ_k^{EA} and display similar movements as k changes. The gap between these two groups of estimators is largest when the economy contains similar numbers of large and small countries, because country pair heterogeneity is then strongest. When heterogeneity is limited, the estimates are closer to one another. These patterns indicate that, in this design, the weighting induced by country pair heterogeneity is quantitatively more important than the particular form of heteroskedasticity.

fig:biases1, fig:biases2, and fig:biases3 report biases for Cases 1, 2, and 3, respectively. The bias profiles are similar across the three heteroskedasticity specifications. In this design, the bias associated with country pair heterogeneity is larger than the bias associated with the error variance specification. This reinforces the central aggregation point: to estimate the aggregate elasticity, δ_k^{EA}, the estimator must place the appropriate weights on observations. [12]

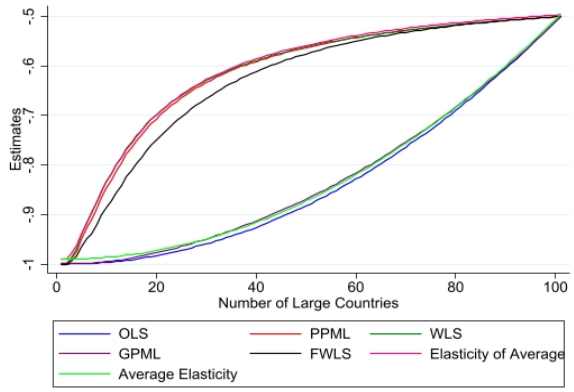
The results in table:benchmark are also presented in these graphs. In fig:estimates1, when the number of large countries $k = 0$ (no large countries) and $k = 100$ (no small countries), i.e., no heterogeneity, the PPML estimates are the closest to the aggregate elasticity δ_k^{EA}. Conversely, the OLS estimator performs best without heterogeneity in Case 3 (fig:estimates3).

The dashed lines report 95% confidence intervals for the bias estimates. The intervals are narrow, reflecting the precision of the Monte Carlo comparisons.

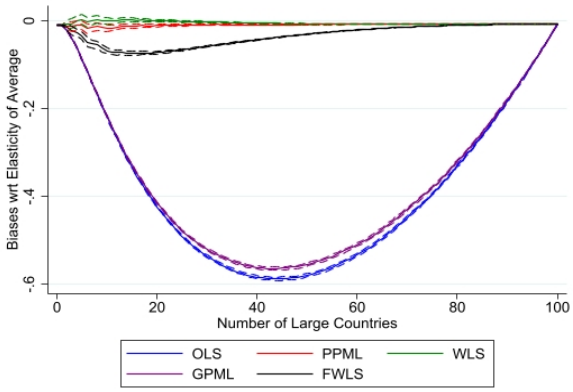
[12] Note that the estimates and biases are similar for the cases of $N_L = 0$ and $N_L = 1$ because in both cases, there are no country pairs where both countries are large, hence in both cases the country pair elasticities are uniform.

Figure 6. Estimates and biases of OLS, WLS, GPML, and PPML

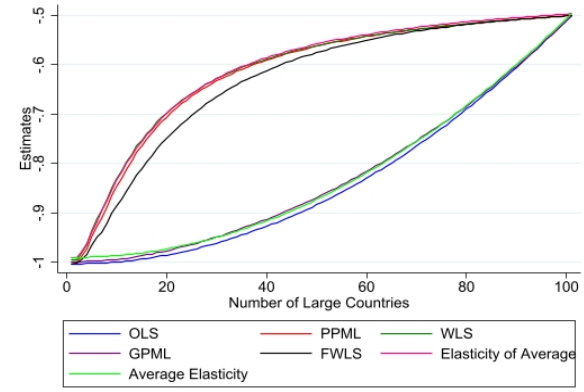
(a) Estimates Case 1



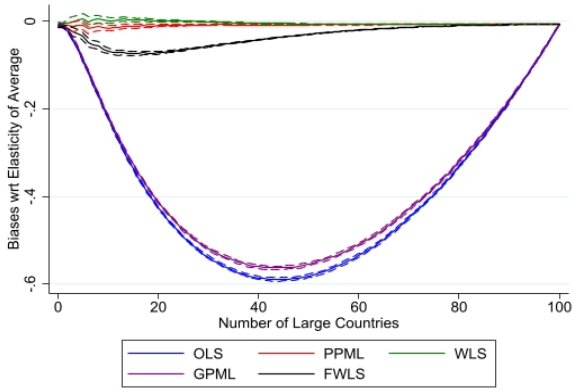
(b) Biases Case 1



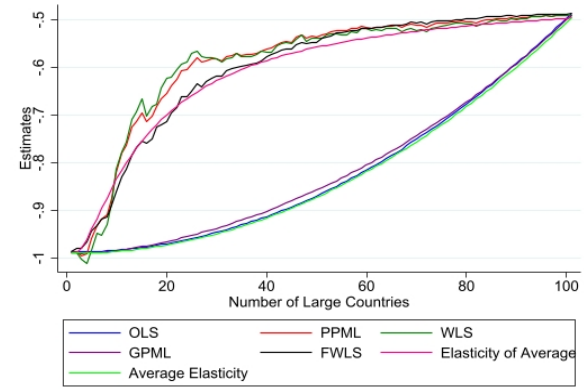
(c) Estimates Case 2



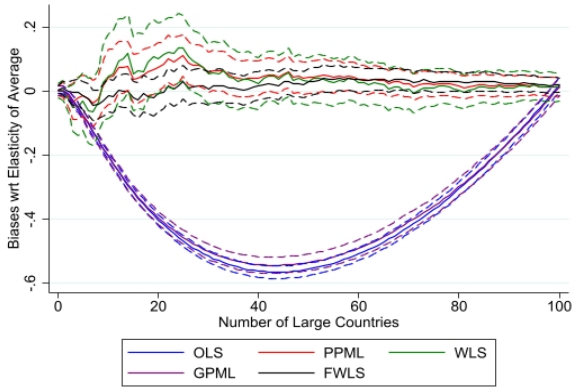
(d) Biases Case 2



(e) Estimates Case 3



(f) Biases Case 3



Note: The horizontal axis shows the number of large countries in the economy, and the vertical axis is the estimated result. Cases 1, 2, and 3 heteroskedastic errors follow the same specification as in Section 3.1. Trade elasticity (aggregate elasticity) is calculated as the change in level trade flow with a 1% increase in bilateral distance and is shown in red. Average elasticity is calculated as the unweighted average of the true θ^*'s in the DGP and is shown in orange. Biases are calculated using eq:bias, which are the percentage difference between the aggregate elasticity and the estimated coefficient for different estimators. The 95% confidence intervals of the estimated biases are included as dashed lines.