Munich Personal RePEc Archive

PPML, Gravity, and Heterogeneous Trade Elasticities

Tyazhelnikov, Vladimir and Shi, Xuetao and Zhou, Xinbei

The University of Sydney, The University of Sydney, The University of Hong Kong

1 February 2026

Online at https://mpra.ub.uni-muenchen.de/128379/
MPRA Paper No. 128379, posted 12 Jun 2026 12:06 UTC

PPML, Gravity, and Heterogeneous Trade Elasticities*

Vladimir Tyazhelnikov† Xuetao Shi‡ Xinbei Zhou§

February 2026

Abstract

We revisit the gravity equation, the most widely used empirical tool in international trade, and show that common estimators recover distinct parameters when trade elasticities are heterogeneous: Poisson pseudo–maximum likelihood (PPML) estimates the aggregate elasticity, while ordinary least squares (OLS) and Gamma PML estimate the average elasticity. We demonstrate that, contrary to common belief, differences between OLS and PPML estimates are driven primarily by elasticity heterogeneity rather than heteroskedasticity of an error term or inclusion of zero trade flows. We develop a simple test for heterogeneity and, applying it to disaggregated trade data, reject homogeneity in at least 57% of industries. We then introduce a weighted PPML (WPPML) estimator to recover the aggregate effects of arbitrary, non-uniform shocks. Using WPPML, we show that trade flows respond less to realized tariff reductions from 2001–2016 than to uniform liberalization, reflecting systematic selection in where barriers fall, with substantial variation across industries.

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

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

*We thank Andrew Bernard, Peter Exterkate, Scott French, Nezih Guner, Vasiliy Korovkin, Ye Lu, Rachael Meager, Mark Melatos, Dennis Novy, Stephen Redding, John Romalis, Thomas Zylkin, 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

The gravity equation is the most widely used empirical tool in the field of International Trade. While there are multiple formulations of this empirical model (see Head and Mayer (2014)), they all broadly fall under the class of constant elasticity models, where the key parameter of interest is the trade elasticity—the sensitivity of trade volumes to changes in trade costs. In this paper, we examine the properties of several popular gravity estimators in the presence of unobserved coefficient heterogeneity. We demonstrate that these techniques generally estimate one of two distinct objects: the *average elasticity* or the *aggregate elasticity*.

We define the average elasticity as an unweighted arithmetic average of the underlying heterogeneous elasticities. While the average elasticity is a meaningful concept, it does not always align with the research objectives. In many cases, researchers are more interested in the aggregate elasticity, the effect of a one percent change in one variable on the percentage change of the average (or total) value of another variable.[1]

In this paper, we formally show that estimates obtained using Poisson pseudo maximum likelihood (PPML), the most common econometric technique used to estimate the gravity equation, can be interpreted as the aggregate elasticity.[2] By contrast, estimates, obtained with another popular estimator, the ordinary least squares (OLS) can be interpreted as the average elasticity.

The choice between estimating an average versus an aggregate elasticity should be guided by the research question. For example, in his seminal paper, Angrist (1990) uses the Vietnam draft lottery as an instrument to estimate the effect of education on wages. His estimates capture the impact of education on the average worker, giving similar weight to each individual regardless of income.[3] By contrast, in settings where the goal is to link micro-level responses to aggregate outcomes, such as firm-level shocks and total employment, an aggregate elasticity is more appropriate. For instance, Carlsson et al. (2021) studies how firm-specific shocks influence hiring, and from a policymaker's perspective, the relevant metric is the effect on overall employment, not on employment of an "average" firm.

The trade literature has long distinguished between micro- and macro-elasticities (for example, Imbs and Mejean (2015, 2017)). Feenstra et al. (2018) documents large discrep-

[1] Consider a simple numerical example with two individuals: one earns \$1,000,000 and pays a 10% tax rate, while the other earns \$1,000 and pays 1%. The unweighted average tax rate is 5.5%, yet total tax revenue is \$100,010 out of \$1,001,000, or 9.99%. The discrepancy arises because the simple average ignores income shares. Weighting each rate by income correctly recovers the aggregate or economy-wide rate.

[2] Ciani and Fisher (2019) make a related point and show that the interpretation of OLS and PPML dif-in-dif estimates of a multiplicative model differ.

[3] In principle, one could instead compute an aggregate elasticity: the effect of education on the total wage bill rather than on the average worker's earnings, but in the context of labor economics, the individual-level interpretation is more natural.

ancies between elasticities estimated at the industry level and those at the aggregate level, and Breinlich et al. (2022) shows that PPML has attractive aggregation properties when trade elasticities are heterogeneous across industries: PPML weighting ensures that estimates reflect the aggregate response to trade shocks. We extend this insight by demonstrating that PPML delivers similar aggregation properties within industries. In the context of international trade, aggregate elasticities are thus both meaningful and robustly estimated with PPML, leading us to recommend PPML as a natural default.

Much of the prior literature attributed the differences between PPML and OLS estimates to the fact that, unlike PPML, OLS estimates are biased when there is heteroskedasticity in an error term, a point that Santos Silva and Tenreyro (2006) (henceforth SST) raised. Using two more estimators: weighted least squares (WLS) and Gamma pseudo maximum likelihood (GPML), we challenge this notion, and perform a simple bias decomposition. We find that the share of bias in estimating the distance elasticity caused by heterogeneity is much larger than that caused by heteroskedasticity. In particular, the heteroskedasticity bias accounts for approximately 8-30% of the total difference between OLS and PPML.

The presence of elasticity heterogeneity is an empirical question and depends on the characteristics of a given data sample. It is therefore essential to determine whether a particular sample exhibits such heterogeneity. We develop a novel test for elasticity heterogeneity and provide a fast, straightforward implementation that can be executed in standard statistical software. Applying this test to 5,050 HS6 industries, we reject the null of elasticity homogeneity for 57% of industries at the 1% significance level, indicating that elasticity heterogeneity is a pervasive feature of empirical trade data.

We further contribute to the gravity literature and examine heterogeneous trade shocks in the context of heterogeneous trade elasticities. Expectedly, the effect of a given shock is contingent on which observations were treated and the respective intensity of such treatment. We show that a standard PPML estimator can be interpreted as the aggregate effect of a uniform decrease in the variable of interest. We show that a PPML estimator augmented with treatment intensity weights, we call weighted PPML (WPPML), estimates the aggregate effect of a given arbitrary shock.

It follows that the conventional logic that the larger number of observations and higher data coverage always improve the quality of the estimates, does not necessarily hold when the elasticities are heterogeneous and the counterfactual treatment is not uniform. Specifically, when a policymaker is interested in estimating the effects of a particular counterfactual scenario, it makes sense to restrict and re-weight sample in such a way that the resulting estimates would be relevant to the given counterfactual scenario.

We demonstrate the usefulness of our WPPML estimator by applying it to the recent

work of Fontagné et al. (2022), which estimates sector-level tariff elasticities. The elasticities from Fontagné et al. (2022) represent the effect of a uniform 1% decrease in tariff barriers worldwide. In contrast, we consider the scenario of *realized* changes in tariffs over the period from 2001 to 2016. Our findings indicate that the responsiveness of trade flows to actual changes in tariffs is generally lower than the response to hypothetical uniform trade liberalization, which suggests that trade liberalization predominantly occurs among country pairs with low trade elasticity. Furthermore, we show that the differences between WPPML and PPML estimates are highly heterogeneous between different sectors and is driven by both elasticity heterogeneity and the industry-specific profile of tariff changes between 2001 and 2016.

Our paper offers four clear and actionable recommendations for applied researchers. First, OLS and PPML estimate fundamentally different objects; the choice between them should be guided by the research question. Second, contrary to common belief, PPML does not effectively incorporate zero observations, as it places approximately zero weight on them. Third, when elasticity heterogeneity is a concern, we introduce a simple, practical statistical test to detect such heterogeneity. Fourth, interpreting estimated elasticities requires clarity about the implied counterfactual. When the relevant counterfactual differs from a uniform change in the variable of interest, we recommend using the weighted PPML estimator developed in this paper.

While our results are applicable in a broad context, we use the gravity equation in International Trade as an illustration of our results. Estimation of trade elasticities has been one of the central question in Trade as the correct estimates allow us 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.[4]

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 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

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

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.

The rest of the paper is organised as follows. In Section 2, we set up the theoretical background of the paper and formally show that PPML and OLS estimate aggregate and average elasticities correspondingly. In Section 3, we discuss the results from Section 2 in application to the context of international trade. In Section 4, we derive formal tests of heterogeneity and apply them to trade data. We extend the previous analysis into considering heterogeneous treatment and apply the novel WPPML to trade data in Section 5 and conclude with Section 6.

2 Average and Aggregate Elasticities

This section shows that PPML and OLS recover two distinct elasticity objects: an *aggregate elasticity* under PPML and an *average elasticity* under OLS.

To do this, we impose no parametric restrictions on heterogeneity: its functional form is left unspecified.[5] To keep the analysis tractable, we assume only that heterogeneity is driven by covariates other than the argument of the elasticity (tariffs in trade context). Allowing elasticity to vary with its own argument would violate the constant-elasticity framework and require a different specification or nonparametric estimation.

Formally, we allow trade elasticity to be a stochastic object, but assume its conditional expectation is a deterministic (yet unknown) function of observables. PPML and OLS estimation assume different functional forms of the conditional expectation. In homogeneous case, these two models correspond to the same definition of elasticity, while in the heterogeneous case they differ because of Jensen's inequality; we denote these elasticities as $h(\cdot)$ for PPML, and $\hbar(\cdot)$ for OLS and refer to the corresponding elasticities as the *h-elasticity* and *ħ-elasticity*, respectively.

We introduce the h-elasticity in Section 2.1 and formally discuss underlying assumptions. We then show that PPML approximately estimates the aggregate h-elasticity.[6] Section 2.2

[5] If heterogeneity were to be a known function of observable covariates, it could be addressed by interacting the variable of interest with those covariates.

[6] If the heterogeneity in the elasticity is detected from the data, completely ignoring it may result in misleading analysis. Econometricians have developed tools to non-parametrically estimate $h(\boldsymbol{X})$ (Ai and Chen (2003)). However, the procedures usually require assumptions on the point identification of $h(\boldsymbol{X})$, and the estimation result can be sensitive to the order of the sieve function. In the following, we show that using simple estimation procedures like PPML, we can estimate part of the information of $h(\boldsymbol{X})$ that can

introduces the $\hbar$-elasticity, and shows that OLS estimates are the unweighted average of the $\hbar$-elasticities. In Section 2.3, we discuss the link between the h- and $\hbar$-elasticity.

2.1 PPML and the Aggregate Elasticity

For any two vectors $\boldsymbol{u} = (u_1, \ldots, u_p)$ and $\boldsymbol{v} = (v_1, \ldots, v_p)$, we let $\boldsymbol{u}^{\boldsymbol{v}}$ denote $\prod_{j=1}^{p} u_j^{v_j}$. We use $\mathbb{R}_{\geq 0}$ to denote non-negative real numbers. We assume that all the random variables introduced in the paper have finite moments of any order. Consider the following model:

$$\text{Model 1: } Y = \boldsymbol{X}^{\beta^*} Z^{\xi} \epsilon \text{ with } \mathbb{E}\left[\epsilon \mid \boldsymbol{X}, Z, \xi\right] = 1,$$

where $\boldsymbol{X} \in \mathbb{R}_{\geq 0}^{p}$ and $Z \in \mathbb{R}_{\geq 0}$ are the covariates. We use the symbol $*$ to denote the true parameter value. Vector β^* is a coefficient vector of fixed values. ξ is a random variable. Regardless of the probability distribution of ξ, we can always define

$$h\left(\boldsymbol{X}, Z\right) \equiv \log_Z \left(\mathbb{E}\left[Z^{\xi} \mid \boldsymbol{X}, Z\right]\right). \tag{1}$$

As a result, we can write the conditional expectation of Y given $\boldsymbol{X}$ and Z as

$$\mathbb{E}\left[Y \mid \boldsymbol{X}, Z\right] = \boldsymbol{X}^{\beta^*} Z^{h(\boldsymbol{X}, Z)}. \tag{2}$$

To obtain a meaningful economic model from (2), it is essential to impose restrictions on $h\left(\cdot\right)$ (or equivalently on $\mathbb{E}\left[Y \mid \boldsymbol{X}, Z\right]$). Otherwise, the conditional expectation $\mathbb{E}\left[Y \mid \boldsymbol{X}, Z\right]$ can be of many different functional forms of $\boldsymbol{X}$ and Z, rendering the concept of elasticity difficult to define. To see this, assume that $h\left(\boldsymbol{X}, Z\right) = \log_Z \left(\frac{\beta^{*\top} \boldsymbol{X}}{\boldsymbol{X}^{\beta^*}}\right)$. Then, we will have $\mathbb{E}\left[Y \mid \boldsymbol{X}, Z\right] = \beta^{*\top} \boldsymbol{X}$, which is no longer a function of Z but a linear function of $\boldsymbol{X}$. We make the following assumption on $h\left(\cdot\right)$ to restore the tractability of the model.

Assumption 1. *$h\left(\boldsymbol{X}, Z\right)$ is constant with respect to Z, i.e., $h\left(\boldsymbol{X}, Z\right) = h\left(\boldsymbol{X}\right)$.*

Assumption 1 does not require ξ to be independent of Z, nor does it require ξ to be a deterministic function of $\boldsymbol{X}$. For instance, assume $\xi = h\left(\boldsymbol{X}\right) + \log_Z \left(\eta\right)$ such that $\mathbb{E}\left[\eta \mid \boldsymbol{X}, Z\right] = 1$. In this example, ξ is a function of Z and contains its own randomness represented by η. It can be verified that Assumption 1 is still satisfied. Additionally, ξ can also depend on ϵ if η correlates with ϵ.

Under Model 1 and Assumption 1, we obtain that $\mathbb{E}\left[Y \mid \boldsymbol{X}, Z\right] = \boldsymbol{X}^{\beta^*} Z^{h(\boldsymbol{X})}$. The Z-

be more useful than $h\left(\boldsymbol{X}\right)$ itself.

elasticity of $\mathbb{E}[Y \mid \boldsymbol{X}, Z]$ can be computed as:[7]

$$\frac{\partial \mathbb{E}[Y \mid \boldsymbol{X}, Z]}{\mathbb{E}[Y \mid \boldsymbol{X}, Z]} \frac{Z}{\partial Z} = h(\boldsymbol{X}).$$

In consequence, $h(\boldsymbol{X})$ measures the fractional response of the expectation of Y conditional on $(\boldsymbol{X}, Z)$ to a fraction change in Z. Since $h(\boldsymbol{X})$ is a function of $\boldsymbol{X}$, the elasticity can change with $\boldsymbol{X}$. From now on, we refer to $h(\boldsymbol{X})$ as the h-elasticity. Note that when ξ is constant, our model nests the classical model without underlying heterogeneity. In this case, the elasticity $h(\boldsymbol{X})$ is a constant.

Both ξ and $h(\boldsymbol{X})$ can be of interest; we, however, have seen that they can behave differently. It is possible for ξ to be random but $h(\boldsymbol{X})$ to be constant. Moreover, although the unobservable ξ might contain useful information about the relationship between Y and $(\boldsymbol{X}, Z)$ that $h(\boldsymbol{X})$ cannot capture, it is difficult to directly study ξ without imposing extra assumptions besides Assumption 1. On the other hand, $h(\boldsymbol{X})$ measures the elasticity, which is a key parameter that researchers are interested in. Thus, we focus on studying $h(\boldsymbol{X})$ instead of ξ in this paper.

Let $\dot{\boldsymbol{X}} \equiv (\log X_1, \ldots, \log X_p)^\top$ and $\dot{Z} \equiv \log Z$. For observations $\{(Y_i, \boldsymbol{X}_i, Z_i)\}_{i=1}^n$, let $\left(\widehat{\beta}, \widehat{\gamma}\right)$ be the solution to the following sample moment condition

$$\sum_{i=1}^{n} \left(Y_i - \boldsymbol{X}_i^{\widehat{\beta}} Z_i^{\widehat{\gamma}}\right) \begin{pmatrix} \dot{\boldsymbol{X}}_i \\ \dot{Z}_i \end{pmatrix} = \mathbf{0}.$$

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.

Lemma 1. *For Model 1, if Assumption 1 holds, then* $\widehat{\gamma} \xrightarrow{p} \gamma_0$ *as* $n \to \infty$, *where* $\gamma_0 \approx \mathbb{E}\left[Y \dot{Z}^2 h(\boldsymbol{X})\right] / \mathbb{E}\left[Y \dot{Z}^2\right]$.

Proof. See Appendix A.1 □

Lemma 1 shows that the PPML estimator $\widehat{\gamma}$ approximates the weighted average of the h-elasticity. The weight $\mathbb{E}\left[Y \dot{Z}^2\right]$ is the first-order derivative of the population moment

[7] Note that while this formulation allows us to estimate partial effect of Z on $\mathbb{E}[Y \mid \boldsymbol{X}, Z]$, the interpretation of the effect of other covariates $\boldsymbol{X}$ on the dependent variable β^* is problematic. The reason is that in this case, $\boldsymbol{X}$ affects Y both directly through β^* and indirectly through heterogeneity in ξ. This is a fundamental problem that arises because of how the model has been specified and can be addressed by considering a particular case where heterogeneity in ξ and conditional expectation of Y are governed by two non-overlapping sets of covariates. More specifically, we can assume $\mathbb{E}[Y \mid \boldsymbol{X}, Z] = \boldsymbol{X}_c^{\beta^*} Z^{h(\boldsymbol{X}_{nc})}$ such that $\boldsymbol{X} = (\boldsymbol{X}_c, \boldsymbol{X}_{nc})$ and $\boldsymbol{X}_c$ does not share common elements with $\boldsymbol{X}_{nc}$. In this case, the effect of $\boldsymbol{X}_c$ takes the original multiplicative form.

function with respect to γ.

Assumption 2. $\frac{Cov\left(Y,\dot{Z}^2\right)}{\mathbb{E}[Y]\mathbb{E}\left[\dot{Z}^2\right]} \approx 0$ *and* $\frac{Cov\left(Yh(\boldsymbol{X}),\dot{Z}^2\right)}{\mathbb{E}[Yh(\boldsymbol{X})]\mathbb{E}\left[\dot{Z}^2\right]} \approx 0$.

Assumption 2 requires that the squared regressor $\dot{Z}^2$ is nearly uncorrelated with either the outcome Y or the outcome being weighted as $Yh(\boldsymbol{X})$. A similar assumption is also imposed in Breinlich et al. (2022).

Proposition 1. *For Model 1, if Assumptions 1 and 2 hold, then* $\widehat{\gamma} \xrightarrow{p} \gamma_0$ *as* $n \to \infty$, *where* $\gamma_0 \approx \mathbb{E}[Yh(\boldsymbol{X})]/\mathbb{E}[Y]$.

Proof. See Appendix A.2 □

Compared to Lemma 1, Proposition A.2 shows that if Assumption 2 holds, then $\widehat{\gamma}$ approximates $\mathbb{E}[Yh(\boldsymbol{X})]/\mathbb{E}[Y]$, which we call the aggregate elasticity.

Unlike log–log OLS, which cannot include observations with a zero dependent variable because the logarithm is undefined, PPML estimation can mechanically incorporate them. Proposition 1, however, shows that PPML assigns weights approximately proportional to Y, so observations with $Y_i = 0$ effectively receive a weight of zero. It follows that while PPML formally allows zeros to enter the estimation, they have virtually no influence on the results, making their inclusion inconsequential.

2.2 OLS and the Average Elasticity

An alternative to Model 1 is the widely used log-log model:

$$\text{Model 2: } \log Y = \beta^* \dot{\boldsymbol{X}} + \xi \dot{Z} + \log \epsilon \text{ with } \mathbb{E}\left[\log \epsilon \mid \dot{\boldsymbol{X}}, \dot{Z}, \xi\right] = 0.$$

Since the log function is strictly monotone, we have $\mathbb{E}\left[\log \epsilon \mid \dot{\boldsymbol{X}}, \dot{Z}, \xi\right] = \mathbb{E}[\log \epsilon \mid \boldsymbol{X}, Z, \xi] = 0$. However, because of the Jensen's inequality, $\mathbb{E}[\log \epsilon \mid \boldsymbol{X}, Z, \xi] = 0$ is not implied by $\mathbb{E}[\epsilon \mid \boldsymbol{X}, Z, \xi] = 1$. As a result, Model 2 is not equivalent to Model 1. In this section, we provide a brief analysis of Model 2 that is analogous to Section 2.1.

Let $\hbar(\boldsymbol{X}, Z) \equiv \mathbb{E}[\xi \mid \boldsymbol{X}, Z]$. We obtain the conditional expectation of $\log Y$ given $\boldsymbol{X}$ and Z as

$$\mathbb{E}[\log Y \mid \boldsymbol{X}, Z] = \beta^* \dot{\boldsymbol{X}} + \hbar(\boldsymbol{X}, Z)\, \dot{Z}. \tag{3}$$

Similar to the discussion in Section 2.1, we impose the following assumption on $\hbar(\boldsymbol{X}, Z)$ to restore the meaning of the model while maintaining its flexibility.

Assumption 3. $\hbar(\boldsymbol{X}, Z)$ *is constant with respect to* Z, *i.e.,* $\hbar(\boldsymbol{X}, Z) = \hbar(\boldsymbol{X})$.

Under Model 2 and Assumption 3, of $\mathbb{E}[\log Y \mid \boldsymbol{X}, Z] = \beta^* \dot{\boldsymbol{X}} + \hbar(\boldsymbol{X})\dot{Z}$. Its partial derivative with respect to $\log Z$ can be easily computed as

$$\frac{\partial \mathbb{E}[\log Y \mid \boldsymbol{X}, Z]}{\partial \log Z} = \hbar(\boldsymbol{X}).$$

We refer to $\hbar(\boldsymbol{X})$ as the $\hbar$-elasticity.

Note that both $\hbar(\mathbf{X})$ and $h(\mathbf{X})$ are interpretable as elasticities. Although they represent distinct objects, we discuss the relationship between them in the next subsection.

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

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

Lemma 2. *For Model 2, if Assumption 3 holds, then* $\widetilde{\zeta} \xrightarrow{p} \zeta_0$ *as* $n \to \infty$, *where* $\zeta_0 \approx \mathbb{E}\left[\hbar(\boldsymbol{X})\dot{Z}^2\right] / \mathbb{E}\left[\dot{Z}^2\right]$.

Proof. See Appendix A.3 □

Lemma 2 shows that the OLS estimator $\widetilde{\zeta}$ approximates the weighted average of the $\hbar$-elasticity, where the weight $\mathbb{E}\left[\dot{Z}^2\right]$ is the first-order derivative of the population moment function with respect to γ.

To show that the $\widetilde{\zeta}$ approximates $\mathbb{E}[\hbar(\boldsymbol{X})]$, which we call the average elasticity, we rely on the following assumption, mirroring Assumption 2 from the previous section.

Assumption 4. $\frac{Cov\left(\hbar(\boldsymbol{X}), \dot{Z}^2\right)}{\mathbb{E}[\hbar(\boldsymbol{X})]\mathbb{E}\left[\dot{Z}^2\right]} \approx 0.$

Assumption 4 requires that $Cov\left(\hbar(\boldsymbol{X}), \dot{Z}^2\right)$, compared to $\mathbb{E}[\hbar(\boldsymbol{X})]\mathbb{E}\left[\dot{Z}^2\right]$, is relatively small. This can occur in many cases. Examples include $\boldsymbol{X}$ is (nearly) independent of $\dot{Z}$, the $\hbar$-elasticity has low variance, and $\hbar(\boldsymbol{X})$ and $\dot{Z}^2$ co-fluctuate and cancel out in expectation.

Proposition 2. *For Model 2, if Assumptions 3 and 4 hold, then* $\widetilde{\zeta} \xrightarrow{p} \zeta_0$ *as* $n \to \infty$, *where* $\zeta_0 \approx \mathbb{E}[\hbar(\boldsymbol{X})]$.

Proof. See Appendix A.4 □

2.3 Link Between Linear and Non-Linear Elasticities

As emphasized earlier, Models 1 and 2 are not equivalent. A researcher working with Model 1 cannot recover Model 2. In particular, the condition that $\mathbb{E}[\log \epsilon \mid \boldsymbol{X}, Z, \xi]$ is constant does not follow from Model 1. Consequently, the discussion in Section 2.2 does not apply,

and the $\hbar$-elasticity is not well-defined. Likewise, starting from Model 2 does not allow one to derive Model 1, and the h-elasticity is not defined. Thus, researchers must choose one of the two models and work with the elasticity appropriate for that specification.

However, both models can be made simultaneously meaningful under a slightly stronger condition on the error term.

Assumption 5. *ϵ is independent of $(\boldsymbol{X}, Z, \xi)$.*

Lemma 3. *Let Assumption 5 hold. (i) Under Model 2, it holds that*

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

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

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

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

Proof. See Appendix A.5 □

Lemma 3 establishes a link between the two specifications. Assumption 5 is sufficient but not necessary. The first part of the lemma shows that, even when the data generating process is described by Model 2, we still have $\mathbb{E}\left[Y \mid \boldsymbol{X}, Z, \xi\right] = c_1 \boldsymbol{X}^{\beta^*} Z^{\xi}$. Hence, $h\left(\boldsymbol{X}, Z\right) = \log_Z\left(\mathbb{E}\left[Z^{\xi} \mid \boldsymbol{X}, Z\right]\right)$ remains well defined, Assumption 1 can be imposed, and the h-elasticity being defined as $h\left(\boldsymbol{X}\right)$ is meaningful. The second part of the lemma shows that, even if the researcher adopts Model 1, the quantity $\hbar\left(\boldsymbol{X}\right) = \mathbb{E}\left[\xi \mid \boldsymbol{X}, Z\right]$ is still well defined under Assumption 3 and is the $\hbar$-elasticity. It is therefore meaningful to discuss the relationship between $h\left(\boldsymbol{X}\right)$ and $\hbar\left(\boldsymbol{X}\right)$.

Given Lemma 3, we now characterize the relationship between the h-elasticity defined in the multiplicative model and the $\hbar$-elasticity implied by the log-log specification. These elasticities coincide under homogeneity. Under heterogeneity, however, they capture slightly different objects, precisely why $h\left(\boldsymbol{X}\right)$ and $\hbar\left(\boldsymbol{X}\right)$ need to be interpreted with care. Although they are distinct, we show below that they remain closely related in a meaningful way.

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

Proposition 3. *Let Assumptions 1, 3, and 5 hold. (i) If ξ is constant, then both $h\left(\boldsymbol{X}\right)$ and $\hbar\left(\boldsymbol{X}\right)$ are constant and identical. (ii) If for any $\boldsymbol{x}_1 \leq \boldsymbol{x}_2$, $F_{\xi|\boldsymbol{X}=\boldsymbol{x}_1,Z=z}\left(t\right) \geq F_{\xi|\boldsymbol{X}=\boldsymbol{x}_2,Z=z}\left(t\right)$*

for any $t \in \mathbb{R}$, then both $h(\boldsymbol{X})$ and $\hbar(\boldsymbol{X})$ are monotonically increasing functions. (iii) If for any $\boldsymbol{x}_1 \leq \boldsymbol{x}_2$, $F_{\xi|\boldsymbol{X}=\boldsymbol{x}_1,Z=z}(t) \leq F_{\xi|\boldsymbol{X}=\boldsymbol{x}_2,Z=z}(t)$ for any $t \in \mathbb{R}$, then both $h(\boldsymbol{X})$ and $\hbar(\boldsymbol{X})$ are monotonically decreasing functions.

Proof. See Appendix A.6 □

When ξ is constant, neither the h-elasticity nor the $\hbar$-elasticity exhibits heterogeneity. This observation motivates the heterogeneity tests developed in Section 4: to test the null hypothesis that ξ is constant, it suffices to verify that either $h(\boldsymbol{X})$ or $\hbar(\boldsymbol{X})$ exhibit homogeneity.

The condition in Proposition 3 (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 $\boldsymbol{X}$ being larger. In this case, both $h(\boldsymbol{X})$ and $\hbar(\boldsymbol{X})$ increase with $\boldsymbol{X}$. The same idea applies to Proposition 3 (iii).

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

Assumption 6. *Assume that $\xi = g(\boldsymbol{X})$ for some unknown function $g(\cdot)$.*

Assumption 6 requires the unobserved heterogenous ξ to be a deterministic, although unrestricted, function of the observable $\boldsymbol{X}$. Since $\boldsymbol{X}$ is random, ξ remains random, but no unobserved variation remains beyond that induced by ξ.

Proposition 4. *Under Assumptions 5 and 6, it holds that $\hbar(\boldsymbol{x}) = h(\boldsymbol{x})$ for every $\boldsymbol{x}$ in the support of $\boldsymbol{X}$.*

Proof. See Appendix A.7 □

By Proposition 4, the two elasticities coincide pointwise. As a result, $\hbar(\boldsymbol{X})$ is constant if and only if $h(\boldsymbol{X})$ is constant. If Assumption 6 fails, it is possible for $\hbar(\boldsymbol{X})$ to be non-constant even when $h(\boldsymbol{X})$ remains constant. For example, let random variable η have equal probabilities to be 0.5 and 1.5 when $\boldsymbol{X} = \boldsymbol{x}_1$ and take the value 1 with probability 1 when $\boldsymbol{X} = \boldsymbol{x}_2$. If $\xi = 1 + \log_Z(\eta)$, then $h(\boldsymbol{X})$ is constant but not $\hbar(\boldsymbol{X})$. At the same time, if $\xi = \eta$, then $\hbar(\boldsymbol{X})$ is constant but not $h(\boldsymbol{X})$. These examples illustrate that Models 1 and 2 reflect different representations of the multiplicative relationship between Y and $(\boldsymbol{X}, Z)$. Without further structure they produce related but not identical elasticity objects.

3 Elasticities and Gravity

The results from the previous section apply to any constant elasticity or, more broadly, multiplicative model estimated in the presence of heterogeneity. We focus on the gravity equation in international trade to illustrate these insights, as PPML has become the dominant estimator in this literature and provides a natural setting to interpret our results in an applied context. We use two subscripts, i and j, when discussing the gravity equation; while a single subscript i is employed if the discussion is on the general model.

3.1 Comparison across Estimators

In the previous section, we showed that PPML estimates $\mathbb{E}\left[Yh\left(\boldsymbol{X}\right)\right]/\mathbb{E}\left[Y\right]$ and OLS estimates $\mathbb{E}\left[\hbar\left(\boldsymbol{X}\right)\right]$. To illustrate the practical implications of our findings, we examine the empirical moment conditions of these two estimators and compare them with alternative approaches.

The OLS estimator is defined as the solution to:

$$\sum_{i,j\in\Omega}\begin{pmatrix}\dot{\boldsymbol{X}}_{ij}\\ \dot{Z}_{ij}\end{pmatrix}\left(\log Y_{ij}-\log\widehat{Y}_{ij}\right)=\mathbf{0},$$

where Ω denotes the set of all country pairs i,j and we denote $\dot{\boldsymbol{X}}_{ij}\equiv\left(\log X_{1,ij},\ldots,\log X_{p,ij}\right)^{\top}$ and $\dot{Z}_{ij}\equiv\log Z_{ij}$, and $\widehat{Y}_{ij}=\boldsymbol{X}_{ij}^{\widehat{\beta}}Z_{ij}^{\widehat{\xi}}$.

Importantly, this moment condition includes log-deviations of the dependent variable Y, meaning that the loss function penalizes relative rather than absolute deviations from the observed values.

The empirical moment condition for PPML, is given by:

$$\sum_{i,j\in\Omega}\begin{pmatrix}\dot{\boldsymbol{X}}_{ij}\\ \dot{Z}_{ij}\end{pmatrix}\left(Y_{ij}-\widehat{Y}_{ij}\right)=\mathbf{0},$$

which includes absolute rather than relative deviations of predicted and observed values of the dependent variable, thus assigning larger weights to larger observations. (It essentially minimizes the sum of total deviations in trade volumes).

To recover estimates interpretable as an aggregate elasticity while retaining an OLS framework, one can instead employ weighted least squares (WLS) with appropriately chosen weights:

$$\sum_{i,j\in\Omega} \begin{pmatrix} \dot{\boldsymbol{X}}_{ij} \\ \dot{Z}_{ij} \end{pmatrix} \left(\log Y_{ij} - \log \widehat{Y}_{ij}\right) Y_{ij} = \mathbf{0}.$$

The share of the bilateral trade flows between countries i and j over the global trade flows, denoted by $Y_{ij}/\sum_{i,j} Y_{ij}$, can be used as the weights.[8] In the context of this study, trade elasticity can be interpreted as the effect of trade liberalization on the representative dollar of world trade flows. If every country pair is treated as a stratum, then the probability of a representative dollar falling in any stratum is $\frac{1}{M}$, where M is the number of country pairs. The share of the bilateral trade flows between a country pair is $\frac{Y_{ij}}{\sum_{i,j} Y_{ij}}$. The dollars from a country pair are underrepresented when $\frac{Y_{ij}}{\sum_{i,j} Y_{ij}} > \frac{1}{M}$. The relative probability of a particular dollar falling in a stratum is given by $\frac{1}{M}/\frac{Y_{ij}}{\sum_{i,j} Y_{ij}} = \frac{\sum_{i,j} Y_{ij}}{M Y_{ij}}$. Therefore, with normalization, the weight used in this study is $\frac{Y_{ij}}{\sum_{i,j} Y_{ij}}$.[9]

We follow Hausman and Wise (1981) by using weights that include an error term. To ensure that endogenous weights are not driving our results, we propose an alternative 2-stage procedure, which we refer as to feasible weighted least squares (FWLS). In the first stage, we run an OLS regression and generate predicted trade flows $\widetilde{Y}_{ij} = \boldsymbol{X}_{ij}^{\widetilde{\beta}} Z_{ij}^{\widetilde{\xi}}$. In the second stage, we run a WLS regression with the fitted values of trade flows from the previous stage as weights:[10]

$$\sum_{i,j\in\Omega} \begin{pmatrix} \dot{\boldsymbol{X}}_{ij} \\ \dot{Z}_{ij} \end{pmatrix} \left(\log Y_{ij} - \log \widehat{Y}_{ij}\right) \widetilde{Y}_{ij} = \mathbf{0}.$$

Alternatively, to estimate the average elasticity without log-linearizing the gravity equation, one can use the Gamma Poisson Maximum Likelihood (GPML) estimator, proposed by SST. While GPML estimates has the same interpretation as those from OLS, GPML is preferable because, as SST demonstrate, it is robust to heteroskedasticity. The GPML estimator solves:

$$\sum_{i,j\in\Omega} \begin{pmatrix} \dot{\boldsymbol{X}}_{ij} \\ \dot{Z}_{ij} \end{pmatrix} \left(\frac{Y_{ij}}{\widehat{Y}_{ij}} - 1\right) = \mathbf{0}.$$

One can see that moment conditions of GPML rely on relative (percentage, rather than log) deviations of Y_{ij} from its predicted value, so we can expect that the GPML and OLS

[8] One justification of this method follows the solutions to the endogenous stratified sampling issue proposed by Hausman and Wise (1981), where they use the Gary Income Maintenance Experiment as an example to show the extent of selection bias due to endogenous stratified sampling and demonstrate that the bias can be corrected by both the maximum likelihood estimator (MLE) and WLS estimator.

[9] Mayer et al. (2019) estimate the gravity equation and apply weights proportional to the value of trade costs, which is similar to our WLS estimator. They, however, do not discuss the interpretation of the estimates and the presence of country-pair heterogeneous elasticities.

[10] Chen and Novy (2022) make a similar point.

estimator to yield similar estimates.

Table I summarizes the empirical moment conditions for each of the discussed estimators. This table, once again, illustrates that OLS involves log deviations of Y_{ij} from its predicted value. Since percent deviations are approximately equal to log deviations, multiplying the log deviations by actual trade volumes leads to a result close to the level deviations, implied by the WLS first-order condition.

TABLE I
Empirical Moment Conditions

OLS	$\sum_{i,j\in\Omega}\begin{pmatrix}\dot{\boldsymbol{X}}_{ij}\\ \dot{Z}_{ij}\end{pmatrix}\left(\log Y_{ij}-\log\widehat{Y}_{ij}\right)=\mathbf{0}$	PPML	$\sum_{i,j\in\Omega}\begin{pmatrix}\dot{\boldsymbol{X}}_{ij}\\ \dot{Z}_{ij}\end{pmatrix}\left(Y_{ij}-\widehat{Y}_{ij}\right)=\mathbf{0}$
WLS	$\sum_{i,j\in\Omega}\begin{pmatrix}\dot{\boldsymbol{X}}_{ij}\\ \dot{Z}_{ij}\end{pmatrix}\left(\log Y_{ij}-\log\widehat{Y}_{ij}\right)Y_{ij}=\mathbf{0}$	GPML	$\sum_{i,j\in\Omega}\begin{pmatrix}\dot{\boldsymbol{X}}_{ij}\\ \dot{Z}_{ij}\end{pmatrix}\left(\frac{Y_{ij}}{\widehat{Y}_{ij}}-1\right)=\mathbf{0}$
FWLS	$\sum_{i,j\in\Omega}\begin{pmatrix}\dot{\boldsymbol{X}}_{ij}\\ \dot{Z}_{ij}\end{pmatrix}\left(\log Y_{ij}-\log\widehat{Y}_{ij}\right)\widetilde{Y}_{ij}=\mathbf{0}$		

Note: $\dot{\boldsymbol{X}}_{ij}\equiv(\log X_{1,ij},\ldots,\log X_{p,ij})^{\top}$ and $\dot{Z}_{ij}\equiv\log Z_{ij}$. $\widehat{Y}_{ij}=\boldsymbol{X}_{ij}^{\widehat{\beta}}Z_{ij}^{\widehat{\xi}}$ is the predicted bilateral trade volume. $\widetilde{Y}_{ij}=\boldsymbol{X}_{ij}^{\widetilde{\beta}}Z_{ij}^{\widetilde{\xi}}$ in the FWLS is the predicted trade flow from the first stage OLS regression.

3.2 Bias Decomposition

In this section, we apply estimation techniques discussed in the previous section to the gravity data. We document not only the existence of country-pair heterogeneity but also decompose the bias of OLS estimates by heterogeneity and heteroskedasticity channels.

To make our results comparable with those in SST, we employ the same data and follow their specification by including distance and indicator variables for remoteness, common language, colonial heritage, and preferential trade agreements in the model. We replicate the findings of SST's Table 5 of the gravity equation regression on OLS, PPML and GPML estimators, controlling for multilateral resistance terms with importer and exporter fixed effects, and complement these results with the estimated results of WLS and FWLS estimators. Table II reports the results of all five estimators and corresponding standard errors.

With PPML as the benchmark estimator, we can perform a decomposition to show that the proportion of bias caused by heterogeneity and heteroskedasticity. The total bias is calculated as the difference between the OLS estimates and PPML estimates. Following our discussion in Section 3.1, there are two alternative ways to decompose the total bias. With one method, we calculate the bias caused by heterogeneity as the difference between the OLS estimates and WLS estimates, and the bias caused by heteroskedasticity is the

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	0.174	0.395***	0.141**	0.369***	0.352***	-0.457**	0.326***
Border	(0.130)	(0.080)	(0.057)	(0.091)	(0.090)	(0.230)	(0.117)
Common	0.406***	0.419***	0.558***	0.383***	0.418***	0.681	0.416***
Language	(0.068)	(0.076)	(0.052)	(0.093)	(0.094)	(0.097)	(0.065)
Colonial	0.666***	-0.047	0.019	0.079	0.038	0.807***	0.510***
Ties	(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
N	9,613	9,613	9,613	18,360	9,613	18,360	9,613

Note: The OLS, WLS, FWLS, PPML, and GPML columns represent the estimates of the main regression estimated by the ordinary least squares, weighted least squares, fitted weighted least squares, Poisson pseudo-maximum likelihood, and gamma pseudo-maximum likelihood, respectively. Zero trade flows are excluded from the specifications of OLS, WLS, and FWLS. For PPML and GPML, the results from both cases of excluding and including zero trade flows in the main regression are reported. Standard errors are in parentheses. N stands for the number of observations. Details are in the main text.

difference between the WLS estimates and PPML estimates. With another method, we calculate the bias caused by heterogeneity as the difference between the GPML estimates and PPML estimates, and the bias caused by heteroskedasticity is the difference between the OLS estimates and GPML estimates. Table III summarizes the estimated $\widehat{\theta}$, the corresponding standard errors, and the bias decomposition results.

The difference between the two methods is that the former evaluates the size of the heteroskedasticity bias when estimating the average elasticity, while the latter evaluates it for the estimation of the aggregate elasticity. The size of biases obtained by these two methods naturally differ from each other, as the size and the sign of the heteroskedasticity bias depend on the pattern of heteroskedasticity and the degree of heterogeneity. We will illustrate this difference by Monte Carlo simulations in Appendix B.

Table III indicates that bias caused by heteroskedasticity accounts for a smaller share of the total bias, while heterogeneity is the primary driver of the differences between the

estimators.[11] These findings echo the conclusion of SST that the PML (Pseudo-Maximum Likelihood) estimators are the best option for estimating the trade elasticity and provide a fresh perspective on the interpretation of the estimates obtained with different methods.

TABLE III
Bias decomposition

(A) *Decomposition through WLS*

	OLS	*WLS*	*PPML*
Estimates	-1.347	-0.722	-0.770
	(0.031)	(0.049)	(0.042)
Heterogeneous Bias	-0.625	0	0
Percentage	108.32%		
Heteroskedastic Bias	0.048	0.048	0
Percentage	-8.32%	100%	
Total Bias	-0.577	0.048	0
Observations	9,613	9,613	9,613

(B) *Decomposition through GPML*

	OLS	*GPML*	*PPML*
Estimates	-1.347	-1.173	-0.770
	(0.031)	(0.029)	(0.042)
Heterogeneous Bias	-0.403	-0.403	0
Percentage	69.84%	100%	
Heteroskedastic Bias	-0.174	0	0
Percentage	30.16%		
Total Bias	-0.577	-0.403	0
Observations	9,613	9,613	9,613

Note: Total bias is computed as the difference between OLS and PPML estimates. Heteroskedastic bias is computed as the difference between WLS and PPML estimates or the difference between OLS and GPML estimates. Heterogeneous bias is computed as the difference between OLS and WLS estimates or the difference between GPML and PPML estimates. Zero trade flows were excluded from all estimation methods. Additional details are in the main text.

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 IV we repeat our decomposition exercise but include zero trade flows for GPML and PPML estimation. Not surprisingly, the PPML coefficient does not change much given that it assigns approximately zero weight to zero trade flows, illustrating the theoretical point we made in Section 2; at the same time, the GPML estimate changes dramatically because of higher weights on zero trade flows and higher trade elasticity corresponding to these observations. Our results suggest that the main driver of the difference between GMPL and OLS is not heterogeneity or heteroskedasticity but the capacity of the former estimator to include zero trade flows in the analysis.

While these results do not constitute a full formal decomposition, they provide a clear and informative illustration of the underlying mechanisms, even though a fully rigorous derivation would be extremely challenging, if not impossible. A key concern is the potential

[11] French (2017) performs an alternative bias decomposition 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 IV
Bias decomposition with Zero Trade Flows

(A) *Decomposition through WLS*

	OLS	*WLS*	*PPML*
Estimates	-1.347	-0.722	-0.750
	(0.031)	(0.049)	(0.041)
Heterogeneous Bias	-0.625	0	0
Percentage	104.69%		
Heteroskedastic Bias	0.028	0.028	0
Percentage	-4.69%	100%	
Total Bias	-0.597	0.028	0
Observations	9,613	9,613	18,360

(B) *Decomposition through GPML*

	OLS	*GPML*	*PPML*
Estimates	-1.347	-1.933	-0.750
	(0.031)	(0.055)	(0.041)
Heterogeneous Bias	-1.183	-1.183	0
Percentage	198.16%	100%	
Heteroskedastic Bias	0.586	0	0
Percentage	-98.16%		
Total Bias	-0.597	-1.183	0
Observations	9,613	18,360	18,360

Note: Total bias is computed as the difference between OLS and PPML estimates. Heteroskedastic bias is computed as the difference between WLS and PPML estimates or the difference between OLS and GPML estimates. Heterogeneous bias is computed as the difference between OLS and WLS estimates or the difference between GPML and PPML estimates. Zero trade flows were included for GPML and PPML estimation. Additional details are in the main text.

interaction between heterogeneity and heteroskedasticity. In Appendix B, we address this concern by performing Monte Carlo simulations based on the SST approach, under different assumptions on the variance of the error term. Across these exercises, we find no systematic or meaningful interaction between heteroskedasticity and heterogeneity, reinforcing confidence in the validity and interpretability of our decomposition.

4 Tests of Heterogeneity

The presence of elasticity heterogeneity is ultimately an empirical question: it may be present in some datasets and absent in others. It is therefore important to develop tools to detect and characterize such variation. We begin with a simple illustration, showing that a standard gravity dataset exhibits substantial elasticity heterogeneity.

We then formalize this analysis and introduce statistical tests for detecting heterogeneity. Because PPML and OLS impose different functional forms on the conditional mean, we propose two complementary approaches: a non-linear heterogeneity test, derived in Section 4.2, and a linear heterogeneity test, introduced in Section 4.3.

In Section 4.4, we apply one of these tests to 5,050 HS-6 industries from Fontagné et al. (2022) and find that the majority of industries in these data display significant heterogeneity.

4.1 Illustrative Evidence

We revisit the standard gravity dataset from SST to provide a simple illustration of the central distinction between OLS and PPML estimators in the presence of heterogeneity. If

trade elasticities are truly homogeneous, OLS estimates obtained from different subsamples should be similar. Conversely, if country pairs with larger trade volumes exhibit systematically different elasticities, OLS estimates from subsamples dominated by large flows will diverge, which is precisely what we find.

We begin by estimating the benchmark gravity regression using the log-log OLS specification of SST. Starting from the full sample, we iteratively drop the ten smallest trade flows and re-estimate the specification, recording the estimated elasticity at each step. Under homogeneity, these estimates would be stable across subsamples.

Figure 1a, however, reveals a clear negative relationship between the estimated elasticity $\widehat{\theta}$ and the average size of trade flows in the sample: as larger country pairs become dominant, the estimated elasticity increases. In contrast, when the same procedure is repeated but ten observations are dropped at random in each iteration (Figure 1b), the OLS estimates remain largely stable, even though heterogeneity is present. This pattern arises because such random deletion is orthogonal to the covariates, so the subsamples remain similar in composition to the full sample. This illustrates the mechanism discussed in Section 3: differences between OLS and PPML estimates only emerge when heterogeneity is systematically related to the covariates, such as when trade flows correlate with elasticity. Absent such correlation, heterogeneity acts like noise, and both estimators recover similar objects, though with different weights and confidence intervals.[12]

In line with the discussion in Section 3.2, we repeat this exercise for the WLS, FWLS, GPML, and PPML estimators. The results in Figure 1c are consistent with our theoretical findings: OLS and GPML estimators return different values of elasticities when estimated on the full sample and on the sample of only the largest countries. PPML, WLS, FWLS, on the other hand, already assign higher weight to larger trade flows, so the exclusion of country pairs with smaller trade volumes does not affect these estimates.[13]

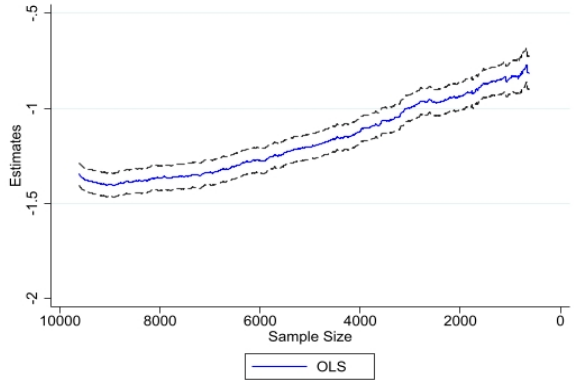
These patterns illustrate that heterogeneity is an empirically relevant feature of trade data and not a statistical artifact. We now turn to formal testing procedures that provide a rigorous framework for detecting elasticity heterogeneity.

[12]Note that these results are not driven by omitted variables, all specifications include exporter and importer fixed effects as well as standard bilateral controls. The patterns documented in Figures 1a and 1b are therefore attributable to unobserved heterogeneity, consistent with findings in Yotov (2012) and Larch et al. (2019).

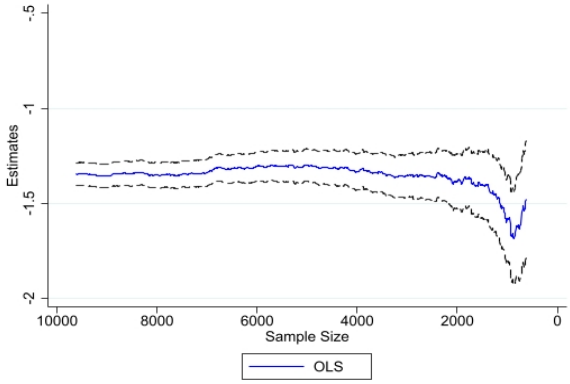
[13]These results, once again, illustrate our point made in Section 2 that the inclusion of zero or small trade flows do not have a significant effect on the PPML estimator.

Figure 1. Evidence of Heterogeneity in Trade Elasticities

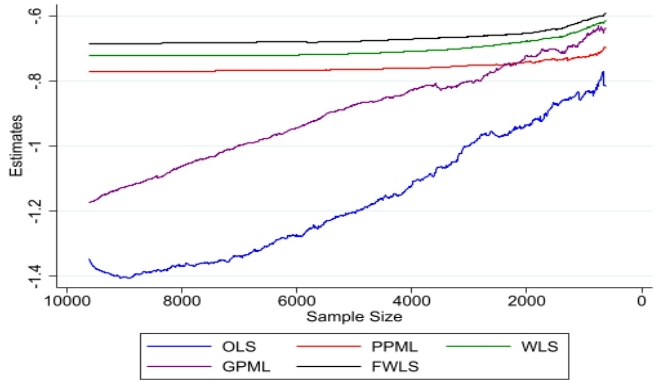
(a) Dropping the smallest trade flows



(b) Dropping random observations



(c) Comparison across estimators



Note: This figure illustrates how estimated trade elasticities vary when subsamples are constructed based on trade volumes. 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 volume. 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 volumes rather than arising from sampling noise.

4.2 Non-Linear Heterogeneity Test

We start with providing a statistical test for the null hypothesis that $h(\boldsymbol{X})$ is constant. Under the null hypothesis that $H_0 : h(\boldsymbol{X}) = \gamma^*$, where γ^* is constant, the conditional expectation becomes

$$\mathbb{E}[Y \mid \boldsymbol{X}, Z] = \boldsymbol{X}^{\beta^*} Z^{\gamma^*}. \tag{4}$$

This is a conditional moment equality condition. Under the null hypothesis, it holds that

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

These are $2p + 1$ number of moment conditions with $p + 1$ number of unknown parameters. This fact motivates us to apply idea of J-test.[14]

Given observations $\{(Y_i, \boldsymbol{X}_i, Z_i)\}_{i=1}^n$, let

$$\widehat{g}_n(\beta, \gamma) \equiv \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \boldsymbol{X}_i^{\beta} Z_i^{\gamma}\right) \begin{pmatrix} \dot{\boldsymbol{X}}_i \\ \dot{Z}_i \\ \dot{Z}_i \boldsymbol{X}_i \end{pmatrix}$$

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

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

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

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

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

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

[14] An alternative approach is to first non-parametrically estimate $h(\boldsymbol{X})$ based on sieve estimation. Then because $\nabla h(\boldsymbol{X}) = \mathbf{0}$ under the null hypothesis, we compare the gradient of the estimator of $h(\boldsymbol{X})$ with the zero vector. However, such a procedure is rather complicated and can be computationally costly. Our approach, on the other hand, is simple to run and often powerful enough to detect heterogeneity of the elasticity in the data.

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

$$\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \boldsymbol{X}_i^{\widetilde{\beta}_I} Z_i^{\widetilde{\gamma}_I}\right)^2 \begin{pmatrix} \dot{\boldsymbol{X}}_i \\ \dot{Z}_i \\ \dot{Z}_i \boldsymbol{X}_i \end{pmatrix} \left(\dot{\boldsymbol{X}}_i^{\top}, \dot{Z}_i, \dot{Z}_i \boldsymbol{X}_i^{\top}\right).$$

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

The proof of the theorem follows from the classical proof in Newey and McFadden (1994). Based on Theorem 1, we expect the distribution of J_n to be tight under the null hypothesis when n is large. The test has the correct asymptotic size when the critical value is obtained from the corresponding quantile value of χ^2_p distribution.

Although the test has the correct asymptotic size, it does not have the asymptotic power being 1 for all the alternative hypotheses that $H_1 : h(\boldsymbol{X})$ is not constant. As a result, it is possible that the distribution of J_n is tight while $h(\boldsymbol{X})$ is a proper function of $\boldsymbol{X}$. Intuitively, this is because we do not exploit all the possible moment equality conditions from the conditional moment equality condition in (4).

Because the test has the correct asymptotic size, the probability of making type I error is well controlled. In other words, if the observed data reject the null hypothesis that $h(\boldsymbol{X})$ is constant, we can be confident that the elasticity exhibits heterogeneity at the given significance level.

4.3 Linear Heterogeneity Test

Now, we propose a statistical test of $\hbar$-heterogeneity. If $\hbar(\boldsymbol{X})$ changes with $\boldsymbol{X}$, then the $\hbar$-elasticity corresponding to Model 2 exhibits heterogeneity. The test for testing the null hypothesis that $H_0 : \hbar(\boldsymbol{X}) = \zeta^*$, where ζ^* is constant, follows from the logic of the previous section. Under the null hypothesis, we have

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

which are $2p+1$ number of moment conditions with $p+1$ number of unknown parameters. The testing procedure is exactly the same as the one in Section 4.2 with $g(\cdot)$ replaced by $g^{\dagger}(\cdot)$.

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

The computation of the test statistic J_n is even simpler in this case because $g^\dagger(\beta, \zeta)$ is a linear function of β and ζ.

4.4 Empirical Application of the Heterogeneity Test

We now apply our proposed tests to trade data, drawing on the influential analysis of Fontagné et al. (2022). Their work is particularly well-suited for our purposes for two reasons. First, they provide widely cited and carefully constructed estimates of trade elasticities that align with the standard gravity framework. Second, their dataset offers exceptional coverage: their analysis spans 5,050 HS-6 product categories, 189 exporters, and 152 destinations over the period 2001–2016, making it one of the most comprehensive elasticity estimates available.

By the result of Proposition 4, researchers can apply either the non-linear or linear heterogeneity test for testing the null hypothesis that either h-elasticity or the $\hbar$-elasticity is constant. Both tests are based on the Hansen J-test, which requires inversion of a potentially high-dimensional covariance matrix. In practice, empirical trade specifications typically include a rich set of fixed effects, making this inversion computationally burdensome. In our specification, which includes origin- and destination-industry-time fixed effects, the number of fixed effects can exceed 1,000, leading to estimation times of several hours per industry and frequent convergence failures.

To address this, we propose a fast, robust, and straightforward approach that we call *Pseudo-IV estimation*. The Hansen J-test is widely implemented in standard IV estimation software as a diagnostic for instrument validity, which we exploit by recasting our GMM estimation problem as a linear IV model. We design the estimation procedure so that its moment conditions coincide with those of Model 2. Using this formulation, the `ivreghdfe` command introduced by Correia (2018) allows us to estimate the model in seconds per industry, reproducing identical J-statistics as linear GMM estimation while achieving a convergence rate of 99.9%.

Following Fontagné et al. (2022), we estimate the tariff elasticity using the following log-linear specification:[15]

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

[15] We discuss the baseline non-linear specification in Section 5.

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, our parameter of interest; $\tau_{ijk,t}$ is the *ad valorem* tariff imposed by importer j on good k in year t. $\boldsymbol{X}_{ij}$ captures bilateral distance, colonial history, contiguity, and common language.

The corresponding moment conditions for industry k are:

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

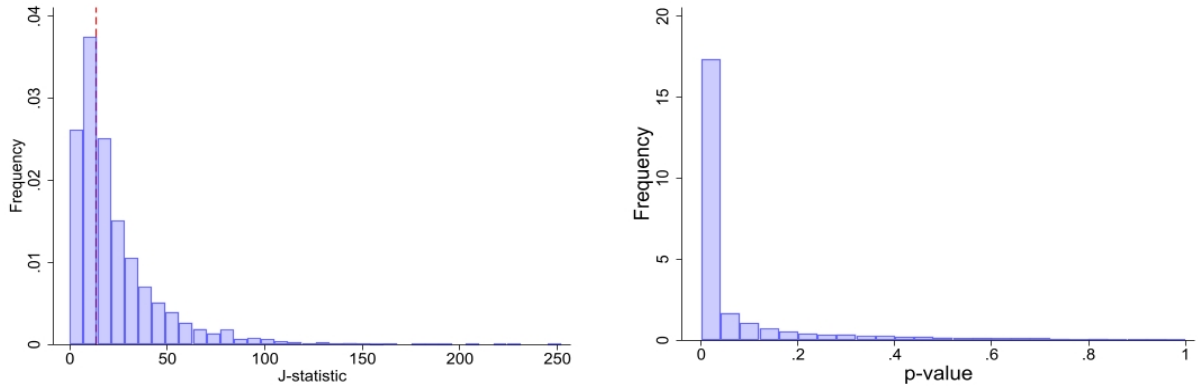
Figure 2 summarizes our estimation results. Panel 2a shows the distribution of Hansen J-statistics with the critical value of 13.277 corresponding to rejection of homogeneity at the 1% significance level. Panel 2b plots the distribution of p-values of the estimated J-statistics, representing the probability of incorrectly rejecting the null hypothesis.

Overall, the evidence indicates that most industries exhibit elasticity heterogeneity: the null of homogeneity is rejected for approximately 57.8%, 71.3%, and 78.8% of industries at the 1%, 5%, and 10% levels, respectively. As discussed in Section 4.2, while the size properties of the test are well defined, its power is unknown. Consequently, a failure to reject does not necessarily imply that trade elasticities are truly homogeneous.

The share of industries exhibiting heterogeneity varies widely across HS sections (Figure 3). For example, only around 40% of industries in "Pearls and Precious Stones" and "Arms and Ammunitions" reject the null at 1%, whereas nearly 80% of industries in "Works of Art" and "Footwear, Headgear, and Umbrellas" exhibit significant heterogeneity.

In this section, we demonstrated that heterogeneity in trade elasticities is not only a theoretical concern but a pervasive empirical feature of disaggregated trade data. By leveraging the Hansen J-test in a computationally efficient Pseudo-IV framework, we show that heterogeneity can be systematically detected at scale, even in specifications with hundreds of fixed effects. The substantial variation across HS sections further highlights that elasticities differ meaningfully within and across industries. These findings outline the limitations of a standard gravity approach with underlying assumption of homogeneous elasticity.

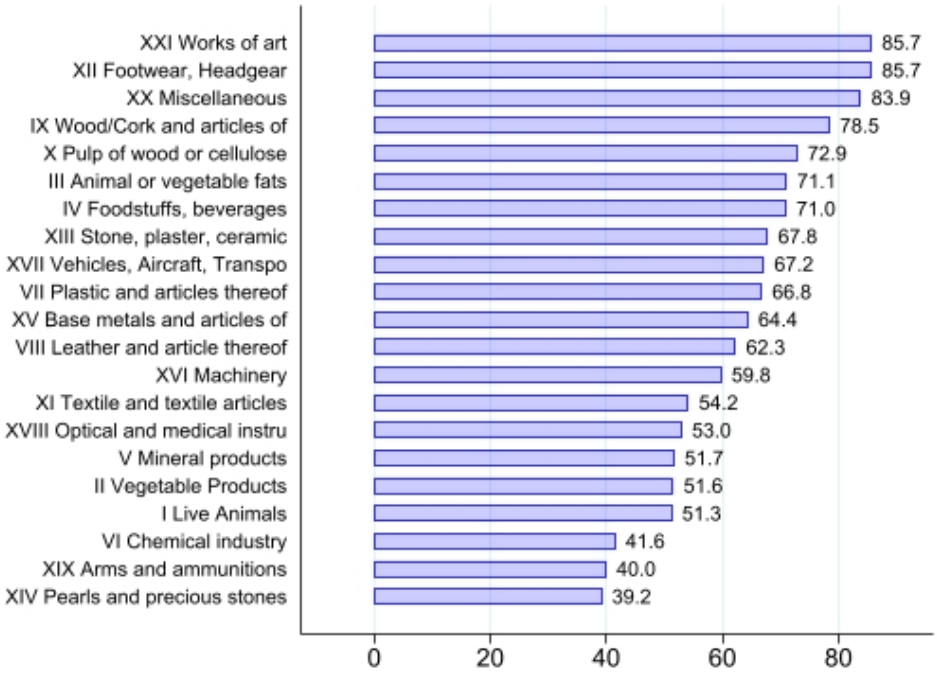
Figure 2. Results of Pseudo-IV Estimation



(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 1% level, 13.277. Panel (b) shows the corresponding distribution of p-values; lower p-values indicate stronger evidence of heterogeneity.

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 Heterogeneous Treatment

In previous sections, we defined the aggregate elasticity as the average percentage response of trade flows to a uniform change in trade costs. Although such a shock represents a particular case of a generic trade shock, our emphasis is placed on it due to its explicit interpretability and its immediate association with a PPML estimator. However, in the presence of heterogeneous effects, non-uniform treatment would have a different effect on the elasticity of average depending on which observations are treated and with what intensity.[16] Therefore, we further derive an expression for the aggregate elasticity in case of an arbitrary treatment and show that it can be estimated by a weighted PPML (WPPML) estimator with weights proportionate to treatment intensity.

5.1 WPPML Estimator

To illustrate the concept of heterogeneous treatment, let us write down the empirical counterpart of aggregate elasticity in case changes in country-pair trade costs t_{ij} are not uniform. We denote the counterfactual trade volume between countries i and j when the trade cost t_{ij} changes to t'_{ij} as Y'_{ij}. In this case, aggregate elasticity $\theta_{\boldsymbol{s}}$ can be defined as the average percentage change in Y_{ij} in response to an average change of 1% in t:

$$\begin{aligned}
\theta_{\boldsymbol{s}} &= \frac{\frac{\sum_{i,j} Y'_{ij} - \sum_{i,j} Y_{ij}}{\sum_{i,j} Y_{ij}}}{\sum_{i,j} \frac{t'_{ij}-t_{ij}}{t_{ij}}/N^2} = \frac{\frac{\sum_{i,j}\left(Y'_{ij}-Y_{ij}\right)}{\sum_{i,j} Y_{ij}}}{\sum_{i,j} \frac{t'_{ij}-t_{ij}}{t_{ij}}/N^2} = \sum_{i,j} \frac{\frac{Y'_{ij}-Y_{ij}}{Y_{ij}} \frac{Y_{ij}}{\sum_{i,j} Y_{ij}}}{\sum_{i,j} \frac{t'_{ij}-t_{ij}}{t_{ij}}/N^2} \\
&= \sum_{i,j} \left(\frac{\frac{Y'_{ij}-Y_{ij}}{Y_{ij}}}{\frac{t'_{ij}-t_{ij}}{t_{ij}}} \frac{Y_{ij}}{\sum_{i,j} Y_{ij}} \frac{\frac{t'_{ij}-t_{ij}}{t_{ij}}}{\sum_{i,j} \frac{t'_{ij}-t_{ij}}{t_{ij}}/N^2} \right) \\
&= \sum_{i,j} \left(\theta_{ij} \frac{Y_{ij}}{\sum_{i,j} Y_{ij}} s_{ij} \right) = \frac{\sum_{i,j} \theta_{ij} Y_{ij} s_{ij}}{N^2} / \frac{\sum_{i,j} Y_{ij}}{N^2},
\end{aligned} \tag{5}$$

where $\theta_{i,j} \equiv \frac{Y'_{ij}-Y_{ij}}{Y_{ij}} / \frac{t'_{ij}-t_{ij}}{t_{ij}}$, and $s_{ij} \equiv \frac{t'_{ij}-t_{ij}}{t_{ij}} / \left(\sum_{i,j} \frac{t'_{ij}-t_{ij}}{t_{ij}}/N^2 \right)$ and $\boldsymbol{s} \equiv \{s_{ij}\}_{i,j\in\Omega}$ are the weights and the vector of treatment intensity. When all the weights are equal, ($s_{ij} = s_{kl}$ for all i, j, k, l), it corresponds to the case of a uniform decrease in all the trade barriers $\{t_{ij}\}_{i,j\in\Omega}$. When $\boldsymbol{s} = \{1\}_{i,j\in\Omega}$, the right-hand-side of Equation (5) is the empirical counterpart of $\mathbb{E}\left[Yh\left(\boldsymbol{X}\right)\right] / \mathbb{E}\left[Y\right]$, corresponding to the approximate probability limit of the regular (unweighted) PPML estimator.

[16] It is true even for linear models with heterogenous coefficients.

Note that $\boldsymbol{s}$ is not random and instead represents the counterfactual of interest. Naturally, in the presence of heterogeneity, the aggregate elasticity (average responsiveness of the dependent variable to treatment) will depend on the distribution of treatment. For example, if the counterfactual of interest is the decrease in trade barriers between countries 1 and 2 with all other trade barriers remaining unchanged ($t'_{ij} \neq t_{ij}$ if $i = 1, j = 2$ and $t'_{ij} = t_{ij}$ if $i \neq 1, j \neq 2$), the true elasticity of such a treatment will be equal to θ_{12}.

Now, for a given treatment vector of interest $\boldsymbol{s}$, a policymaker can properly estimate the aggregate elasticity by assigning corresponding weights to the relevant observations. In particular, if only a subset of observations are being treated, untreated observations should be omitted from the estimation by setting the corresponding weights to zero. Otherwise, values of θ_{ij} irrelevant to a given treatment would contaminate the true aggregate elasticity.[17] More generally, if some observations receive larger treatment than others, the weight of their elasticity in the calculation of the total elasticity should be higher. We allow s_{ij} to be positive or negative to accommodate cases where researchers are interested in knowing the change of the global trade volume if trade barriers between some countries decrease while increase between others.

Note that unlike size weights $Y_{ij}/\sum_{i,j} Y_{ij}$, which appear because we deal with a non-linear model, treatment weights would appear in any model, including linear, with heterogeneous coefficients. Now, the regular PPML estimator considers the baseline case of a uniform treatment $\boldsymbol{s} = \{1\}_{i,j\in\Omega}$. If a policymaker is interested in the outcomes of any other treatment vector, they can apply the WPPML estimator $\left(\widehat{\beta}, \widehat{\theta}\right)$, which is defined as

$$\sum_{i,j\in\Omega} w_{ij} \begin{pmatrix} \dot{\boldsymbol{X}}_{ij} \\ \dot{Z}_{ij} \end{pmatrix} \left(Y_{ij} - \widehat{Y}_{ij}\right) = \mathbf{0},$$

where $\widehat{Y}_{ij} = \boldsymbol{X}_{ij}^{\widehat{\beta}} Z_{ij}^{\widehat{\theta}}$ and $w_{ij} = \frac{t'_{ij} - t_{ij}}{t_{ij}}$. Finally, there exists such a treatment vector such that the average elasticity and the aggregate elasticity are identical; it happens when larger observations (country pairs) receive proportionally lower treatment and thus two weights (size and treatment) cancel out.

Interestingly, the average elasticity does not depend on the treatment vector, because it is a simple arithmetic average of heterogenous elasticities θ_{ij} and does not account the the relative size and importance of different observations or whether these observations were treated or not (which is another advantage of the aggregate elasticity).

[17] Consider, for example, the case when there are two subpopulations with different true values of trade elasticity θ_{g_1} and θ_{g_2}. Then, if only the first group was treated, the responsiveness of the group to this treatment will correspond to θ_{g_1}, while values of θ_{g_2} are completely irrelevant.

It is important to note that in the absence of heterogeneity researchers could estimate the elasticity first and then calculate $\theta_{\boldsymbol{s}}$. To see this, assume $\theta_{ij} = \theta$ to be the same for all country pairs i, j. Then it holds that $\theta_{\boldsymbol{s}} = \theta \frac{\sum_{i,j} Y_{ij} s_{ij}}{\sum_{i,j} Y_{ij}}$. For any vector of treatment intensity $\boldsymbol{s}$, we can first estimate θ and then compute $\theta_{\boldsymbol{s}}$ by multiplying the estimated θ with a factor $\frac{\sum_{i,j} Y_{ij} s_{ij}}{\sum_{i,j} Y_{ij}}$.[18] Such an exercise, however, is not feasible when θ_{ij} is heterogeneous across countries. It is impossible to start with the estimates of either aggregate or average elasticities and obtain $\theta_{\boldsymbol{s}}$ by adjusting OLS or PPML estimates with a factor. The introduced WPPML is essential if researchers are interested in the effect of a non-uniform treatment in the presence of heterogeneity.

Assumption 7. $\Delta t_{ij} \equiv t'_{ij} - t_{ij} \approx 0$ *for all* i *and* j.

Assumption 7 requires that the change of the trade cost be small enough for θ_{ij} to approximate the h-elasticity. Under this assumption, we show that the WPPML estimator approximates $\theta_{\boldsymbol{s}}$ when N is large.

Proposition 5. *Under Assumptions 1, 2, and 7, it holds that* $\left|\hat{\theta} - \theta_{\boldsymbol{s}}\right| \xrightarrow{p} \delta$ *as* $N \to \infty$, *where* $\delta \approx 0$.

Proof. See Appendix A.8 □

5.2 Empirical Application of WPPML

As demonstrated in the previous section, the PPML estimate can be interpreted as an average elasticity corresponding to a 1% uniform decrease in trade barriers. While this may serve as a meaningful counterfactual, depending on the degree of heterogeneity, the relevant average elasticity associated with a specific counterfactual scenario might differ significantly.

We illustrate this difference by comparing baseline PPML elasticity estimates from Fontagné et al. (2022) with our WPPML estimates. To estimate the tariff elasticity, their baseline specification relies on the following standard structural gravity formulation:

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

where i, j, k, t denote countries of origin and destination, industry, and year. $\kappa_{ik,t}$ and $\eta_{jk,t}$ are origin-industry-time and destination-industry-time fixed effects. θ_k is the industry-specific tariff elasticity, our coefficient of interest; and $\tau_{ijk,t}$ is the *ad valorem* tariff imposed by country

[18] Even without heterogeneity, this normalization is useful for assessing the impact of a non-uniform treatment on the average change in the dependent variable. It should be noted that in this scenario, this normalization and the WPPML method would produce identical results.

j on good k in year t. d_{ij} and $\boldsymbol{X}_{ij}$ are distance and other relevant covariates controlling for bilateral-specific geographic-related trade costs, respectively.

The WPPML estimator can be used to estimate the average elasticity corresponding to any counterfactual scenario, so any case of tariff liberalization that happened between 2001-2016 can be considered. We, however, choose to estimate the average elasticity of *realized* changes in tariffs over the whole period. In particular, we use the normalized percentage changes in tariffs between 2001 and 2016 as weights:[19,20]

$$s_{ijk} = -\frac{\tau_{ijk,2016} - \tau_{ijk,2001}}{\tau_{ijk,2001}} \Big/ \left(\sum_{i,j} \frac{\tau_{ijk,2016} - \tau_{ijk,2001}}{\tau_{ijk,2001}} / N\left(N-1\right) \right)$$

We illustrate the difference between baseline elasticities estimated by Fontagné et al. (2022) and *realized* elasticities in Figures 4a and 4b.[21] One can see that when elasticities are large (in absolute terms), the two estimators generate comparable results. However, for elasticities lower than 15, the distribution of realized elasticities is significantly skewed to the right, suggesting that tariff liberalization was concentrated among country-pairs with lower corresponding trade elasticities and is thus corresponding to lower welfare gains than would have been predicted by Fontagné et al. (2022).

Table V presents the estimated trade elasticities separately for each HS section using both PPML and WPPML estimators. The column of PPML elasticities replicates the results by Fontagné et al. (2022). It is clear that after considering the heterogeneous treatment intensities, measured by the percentage changes in tariffs over the whole period, the WPPML trade elasticities substantially differ from the PPML elasticities for certain HS sections. Moreover, the magnitude and direction of the differences between the baseline PPML and WPPML estimates exhibit substantial heterogeneity across industries. Notably, there are large positive differences for industries such as Animal or Vegetable Fats and Oils, and Mineral Products. In contrast, negative differences are observed for industries like Plastics and Articles Thereof, and Textiles and Textile Articles. For other industries, such as Wood/Cork and Articles of Wood/Cork, and Base Metals and Articles of Base Metals, the estimates remain generally similar. The size and sign of these differences is unknown a priori and depends on both within-industry heterogeneity of elasticities and a given profile of tariff changes.

[19] Negative weights are replaced with zeros in the estimation process.

[20] Note that even though we use changes in tariffs as weights, our estimation is conceptually different from the three-way fixed effects estimation, which relies on country-pair-industry fixed effects and uses variation in tariffs over time. While WPPML also excludes observations without changes in tariffs, our approach relies on the cross-sectional variation in tariffs. The difference is evident from the fact that the three-way PPML requires at least two time periods for identification, while WPPML can be applied using data for one period only.

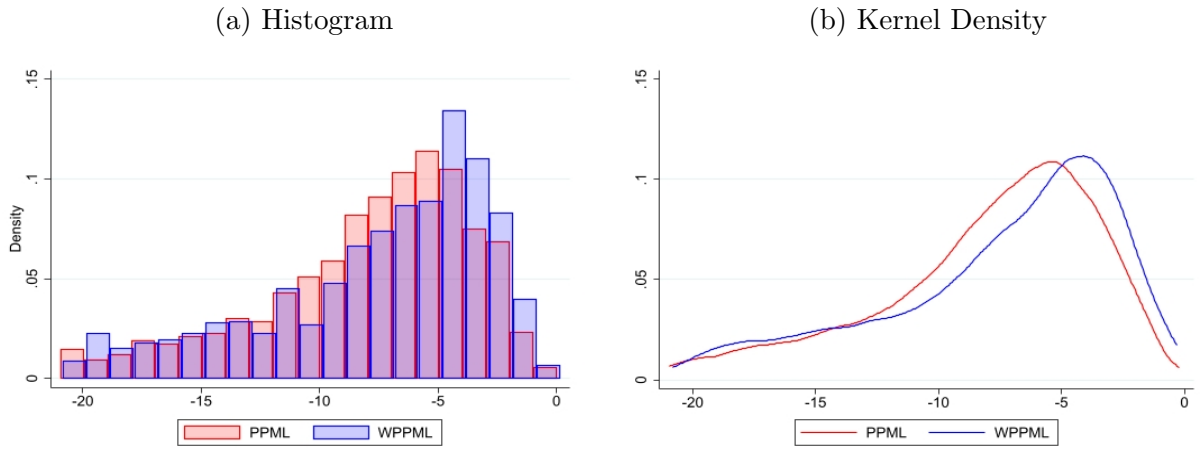
[21] Following Fontagné et al. (2022), we replace insignificant estimates of trade elasticities with 1.

Table V
PPML and WPPML Trade Elasticities by HS Section

Section	Description	No. of HS6	PPML	No of HS6 non-missing	WPPML	No of HS6 non-missing
I	Live Animals and Animal Products	228	-7.51 (9.06)	223	-8.35 (12.08)	213
II	Vegetable Products	256	-5.95 (4.06)	252	-5.66 (6.61)	250
III	Animal or vegetable fats and oils	45	-8.53 (8.69)	43	-11.16 (15.64)	40
IV	Prepared foodstuffs, beverages and tobacco	193	-6.17 (4.50)	193	-5.28 (5.09)	187
V	Mineral products	148	-18.53 (17.62)	143	-21.48 (34.71)	137
VI	Products of chemical industries	789	-11.84 (38.16)	758	-10.75 (12.06)	748
VII	Plastic and articles thereof	211	-8.39 (7.19)	211	-5.72 (6.99)	207
VIII	Raw hides and skins, leather and article thereof	69	-5.62 (4.63)	68	-6.19 (7.07)	66
IX	Wood/Cork and articles of Wood/Cork	93	-8.47 (8.11)	93	-8.43 (8.19)	86
X	Pulp of wood or other cellulosic materials	144	-9.92 (7.42)	142	-9.71 (31.51)	143
XI	Textile and textile articles	801	-7.13 (6.86)	793	-5.51 (8.07)	777
XII	Footwear, Headgear, Umbrellas and prepared feathers	49	-3.61 (2.77)	46	-2.37 (1.80)	46
XIII	Articles of stone, plaster, ceramic and glass	143	-6.62 (4.19)	142	-4.33 (3.07)	136
XIV	Natural cultured pearls and precious stones and metals	51	-13.59 (13.52)	50	-14.52 (16.58)	49
XV	Base metals and articles of base metals	568	-9.59 (9.75)	557	-9.94 (14.62)	530
XVI	Machinery and mechanical appliances and electrical machinery	769	-6.09 (5.56)	756	-7.13 (19.66)	698
XVII	Vehicles, Aircraft and transport equipment	131	-10.46 (8.53)	129	-10.22 (18.95)	121
XVIII	Optical, photographic, precision and medical instruments	217	-5.65 (5.60)	210	-5.80 (8.30)	191
XIX	Arms and ammunitions	20	-6.52 (5.14)	20	-13.55 (11.35)	20
XX	Miscellaneous	118	-4.85 (3.42)	117	-2.73 (2.59)	101
XXI	Works of art	7	-5.96 (4.37)	7	-7.70 (6.21)	7

Note: This table lists the descriptive statistics (mean, standard deviation (in brackets), and non-missing values) for the ϵ_k parameter estimated using both PPML and WPPML for each HS section. Only products with tariff elasticity $\beta_k < -1$ are considered.

Figure 4. Empirical Distribution of PPML and WPPML Trade Elasticities



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

6 Conclusion

This paper revisits the interpretation of the most widely used estimators of the gravity equation and demonstrates that they capture fundamentally different objects. We show that Poisson pseudo-maximum likelihood (PPML) estimates the aggregate elasticity: the effect of a uniform percentage change in trade costs on aggregate trade, while ordinary least squares (OLS) estimates the average elasticity across country pairs. The choice between these estimators should therefore be guided by the research question rather than econometric convention.

We provide new theoretical results and Monte Carlo simulations to clarify how heterogeneity in trade elasticities across country pairs drives systematic differences between OLS, GPML, WLS, and PPML estimates. While heteroskedasticity bias has been the dominant explanation in the literature, we find that heterogeneity bias is far more consequential, accounting for the bulk of the divergence between OLS and PPML. We also introduce a simple, practical test for detecting elasticity heterogeneity and show that heterogeneity is pervasive in global trade data: 57% of HS6 industries reject the null hypothesis of homogeneity at the 1% level.

Finally, we extend the interpretation of PPML to the case of heterogeneous treatment intensities and propose a weighted PPML (WPPML) estimator, which recovers the aggregate effect of a given non-uniform shock. Our findings suggest that the relevance of estimated elasticities depends critically on the counterfactual of interest: uniform liberalization, targeted tariff changes, or industry-specific trade cost shocks may all warrant different weighting

schemes and corresponding trade elasticities.

Our results carry two overarching lessons. First, estimator choice is not merely a question of statistical efficiency but of economic interpretation: applied researchers must be explicit about whether they seek an "average" or "aggregate" elasticity and about the counterfactual treatment scenario implicit in their estimation. Second, heterogeneity is not a nuisance but a central feature of international trade data. Properly addressing it through estimator choice, weighting, and hypothesis testing leads to more robust inference and policy-relevant conclusions. Although our analysis focuses on the gravity equation, the insights generalize to a wide class of multiplicative models estimated with PPML and related techniques, making these contributions relevant beyond the field of international trade.

References

Ai, C.and Chen, X. (2003). 'Efficient estimation of models with conditional moment restrictions containing unknown functions', *Econometrica*, Vol. 71(6), pp. 1795–1843.

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.

Ciani, E.and Fisher, P. (2019). 'Dif-in-dif estimators of multiplicative treatment effects', *Journal of Econometric Methods*, Vol. 8(1).

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.

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*.

Hausman, J. A.and Wise, D. A. (1981). 'Stratification on endogenous variables and estimation: The gary income maintenance experiment', *Structural analysis of discrete data with econometric applications*, pp. 365–391.

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.

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 Technical Proofs

A.1 Proof of Lemma 1

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: $\sum_{i=1}^{n}\left(Y_i - \boldsymbol{X}_i^{\beta^*} Z_i^{\widehat{\gamma}}\right)\dot{Z}_i = 0$. Let Γ be the compact parameter space for γ. The uniform law of large numbers implies that

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

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

$$\mathbb{E}\left[\left(Y - \boldsymbol{X}^{\beta^*} Z^{\gamma_0}\right)\dot{Z}\right] = 0. \tag{6}$$

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

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

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

We discretize the above integral by partitioning the support of $(\boldsymbol{X}, Z)$ 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 $\mathbb{E}\left[\left(Y - \boldsymbol{X}^{\beta^*} Z^{\gamma_0}\right)\dot{Z}\right]$ as

$$\sum_{k=1}^{K}\mathbb{E}\left[\left(Y - \boldsymbol{X}^{\beta^*} Z^{\gamma_0}\right)\dot{Z} \mid \boldsymbol{X} = \boldsymbol{x}_k, Z = z_k\right] f_{\boldsymbol{X},Z}(\boldsymbol{x}_k, z_k)\,\Delta_k = \mathbf{0}. \tag{7}$$

By Model (2) and Assumption 1, we have that for any $(\boldsymbol{x}, z)$,

$$\mathbb{E}\left[\left(Y - \boldsymbol{X}^{\beta^*} Z^{h(\boldsymbol{X})}\right)\dot{Z} \mid \boldsymbol{X} = \boldsymbol{x}, Z = z\right] = 0. \tag{8}$$

Equation (8) implies that

$$\begin{aligned}\mathbb{E}\left[\left(Y - \boldsymbol{X}^{\beta^*} Z^{h(\boldsymbol{X})}\right)\dot{Z}\right] &= \int \mathbb{E}\left[\left(Y - \boldsymbol{X}^{\beta^*} Z^{h(\boldsymbol{X})}\right)\dot{Z} \mid \boldsymbol{X} = \boldsymbol{x}, Z = z\right] f_{\boldsymbol{X},Z}(\boldsymbol{x}, z)\, d\boldsymbol{x}dz \\ &= \sum_{k=1}^{K}\mathbb{E}\left[\left(Y - \boldsymbol{X}^{\beta^*} Z^{h(\boldsymbol{X})}\right)\dot{Z} \mid \boldsymbol{X} = \boldsymbol{x}_k, Z = z_k\right] f_{\boldsymbol{X},Z}(\boldsymbol{x}_k, z_k)\,\Delta_k = \mathbf{0}.\end{aligned} \tag{9}$$

Denote $\gamma_k \equiv h(\boldsymbol{x}_k)$ and

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

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

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

Together with Equations (7) and (9), it holds that $\sum_{k=1}^{K} \phi_k'(\gamma_k)(\gamma_0 - \gamma_k) \approx 0$. Thus, we obtain that

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

Taking the first-order derivative provides us that

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

By Equation (8), we have that

$$\mathbb{E}\left[\boldsymbol{X}^{\beta^*} Z^{\gamma_k} \mid \boldsymbol{X} = \boldsymbol{x}_k, Z = z_k\right] = \mathbb{E}\left[\boldsymbol{X}^{\beta^*} Z^{h(\boldsymbol{x}_k)} \mid \boldsymbol{X} = \boldsymbol{x}_k, Z = z_k\right] = \mathbb{E}\left[Y \mid \boldsymbol{X} = \boldsymbol{x}_k, Z = z_k\right].$$

This shows that

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

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

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

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

A.2 Proof of Proposition 1

Proof. Based on the result of Lemma 1, by the definition of covariance, it holds that

$$\frac{\mathbb{E}\left[Y\dot{Z}^2 h\left(\boldsymbol{X}\right)\right]}{\mathbb{E}\left[Y\dot{Z}^2\right]} = \frac{\mathbb{E}\left[Yh\left(\boldsymbol{X}\right)\right]\mathbb{E}\left[\dot{Z}^2\right] + Cov\left(Yh\left(\boldsymbol{X}\right), \dot{Z}^2\right)}{\mathbb{E}\left[Y\right]\mathbb{E}\left[\dot{Z}^2\right] + Cov\left(Y, \dot{Z}^2\right)}.$$

Applying Assumption 2, we can obtain that $\mathbb{E}\left[Y\dot{Z}^2 h\left(\boldsymbol{X}\right)\right]/\mathbb{E}\left[Y\dot{Z}^2\right]$ can be approximated by $\mathbb{E}\left[Yh\left(\boldsymbol{X}\right)\right]/\mathbb{E}\left[Y\right]$. Hence, the PPML estimator $\widehat{\gamma}$ converges in probability to γ_0, which is approximately $\mathbb{E}\left[Yh\left(\boldsymbol{X}\right)\right]/\mathbb{E}\left[Y\right]$. □

A.3 Proof of Lemma 2

Proof. Following the proof of Proposition 1, 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 $\sum_{i=1}^{n}\left(\log Y_i - \beta^*\dot{\boldsymbol{X}}_i^{\top} - \widetilde{\zeta}\dot{Z}_i\right)\dot{Z}_i = 0$. Because of the linearity in $\widetilde{\zeta}$, we can derive the closed-form expression of $\widetilde{\zeta}$ as

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

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

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

as $n \to \infty$. By the assumption of Model 2, $\mathbb{E}\left[\log\epsilon\dot{Z}\right] = 0$. Assumption 3 implies that

$$\mathbb{E}\left[\xi\dot{Z}^2\right] = \mathbb{E}\left[\mathbb{E}\left[\xi\dot{Z}^2 \mid \boldsymbol{X}, Z\right]\right] = \mathbb{E}\left[\hbar\left(\boldsymbol{X}\right)\dot{Z}^2\right].$$

Therefore, $\widetilde{\zeta} \xrightarrow{p} \mathbb{E}\left[\hbar\left(\boldsymbol{X}\right)\dot{Z}^2\right]/\mathbb{E}\left[\dot{Z}^2\right]$. □

A.4 Proof of Proposition 2

Proof. By Lemma 2, we have that $\widetilde{\zeta} \xrightarrow{p} \mathbb{E}\left[\hbar\left(\boldsymbol{X}\right)\dot{Z}^2\right]/\mathbb{E}\left[\dot{Z}^2\right]$. Under Assumption 4, $\mathbb{E}\left[\hbar\left(\boldsymbol{X}\right)\dot{Z}^2\right]$ is approximately $\mathbb{E}\left[\hbar\left(\boldsymbol{X}\right)\right]\mathbb{E}\left[\dot{Z}^2\right]$. As a result, $\widetilde{\zeta} \xrightarrow{p} \zeta_0$, where $\zeta_0 \approx \mathbb{E}\left[\hbar\left(\boldsymbol{X}\right)\right]$. □

A.5 Proof of Lemma 3

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

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

A.6 Proof of Proposition 3

Proof. Part (i) is straightforward. If ξ is constant, then $h(\boldsymbol{X}) = \hbar(\boldsymbol{X}) = \xi$. For part (ii), if $F_{\xi|\boldsymbol{X}=\boldsymbol{x}_1, Z=z}(t) \geq F_{\xi|\boldsymbol{X}=\boldsymbol{x}_2, Z=z}(t)$ for any $t \in \mathbb{R}$, then we have

$$\int \mu(t, z)\, dF_{\xi|\boldsymbol{X}=\boldsymbol{x}_1, Z=z}(t) \leq \int \mu(t, z)\, dF_{\xi|\boldsymbol{X}=\boldsymbol{x}_2, Z=z}(t), \tag{10}$$

for any function $\mu : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ that is non-deceasing in its first argument given the second argument fixed. Let $\mu(t, z) = z^t$, which is increasing in t given any $z > 0$. Inequality (10) implies that

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

By Assumption 1, the left-hand-side of the inequality equals $h(\boldsymbol{x}_1)$ and the right-hand-side of the inequality equals $h(\boldsymbol{x}_2)$. Because $\boldsymbol{x}_1 \leq \boldsymbol{x}_2$, we have that $h(\cdot)$ is monotonically increasing. By letting $\mu(t, z) = t$, we can obtain that $\hbar(\cdot)$ is also monotonically increasing. Part (iii) of the proposition follows from the similar argument. □

A.7 Proof of Proposition 4

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

$$\hbar(\boldsymbol{X}, Z) = \mathbb{E}[g(\boldsymbol{X}) \mid \boldsymbol{X}, Z] = g(\boldsymbol{X}) = \hbar(\boldsymbol{X}).$$

At the same time, we have that

$$\begin{aligned}h(\boldsymbol{X}, Z) &= \log_Z\left(\frac{\mathbb{E}[Y \mid \boldsymbol{X}, Z]}{\boldsymbol{X}^{\beta^*}}\right) = \log_Z\left(\frac{\mathbb{E}\left[\boldsymbol{X}^{\beta^*} Z^{\xi} \varepsilon \mid \boldsymbol{X}, Z\right]}{\boldsymbol{X}^{\beta^*}}\right) \\ &= \log_Z\left(\mathbb{E}\left[Z^{\xi} \mid \boldsymbol{X}, Z\right]\right) = \log_Z\left(\mathbb{E}\left[Z^{g(\boldsymbol{X})} \mid \boldsymbol{X}, Z\right]\right) \\ &= \log_Z\left(Z^{g(\boldsymbol{X})}\right) = g(\boldsymbol{X}) = h(\boldsymbol{X}).\end{aligned}$$

As a result, it holds that $h(\boldsymbol{X}) = \hbar(\boldsymbol{X}) = g(\boldsymbol{X})$. □

A.8 Proof of Proposition 5

Proof. The WPPML estimator can be written as

$$\sum_{i,j} w_{ij} \left(Y_{ij} - \boldsymbol{X}_{ij}^{\widehat{\beta}} Z_{ij}^{\widehat{\theta}} \right) \left(\dot{\boldsymbol{X}}_{ij}^{\top}, \dot{Z}_{ij} \right)^{\top} = \mathbf{0}.$$

Similar to the proof of Proposition 1, we replace $\widehat{\beta}$ with β^* to focus on the study of $\widehat{\theta}$. The estimator $\widehat{\theta}$ satisfies the sample moment condition: $\sum_{i,j} w_{ij} \left(Y_{ij} - \boldsymbol{X}_{ij}^{\beta^*} Z_{ij}^{\widehat{\theta}} \right) \dot{Z}_{ij} = 0$. By Assumption 7, we have that $\frac{1}{N^2} \sum_{i,j} w_{ij}^2 \to \eta$ for some finite constant $\eta > 0$ because each w_{ij}^2 is small. For any $\theta \in \Theta$, the weak law of large numbers for triangular arrays provides that

$$\frac{1}{N^2} \sum_{i,j} w_{ij} \left(Y_{ij} - \boldsymbol{X}_{ij}^{\beta^*} Z_{ij}^{\widehat{\theta}} \right) \dot{Z}_{ij} - \frac{1}{N^2} \sum_{i,j} \mathbb{E} \left[w_{ij} \left(Y_{ij} - \boldsymbol{X}_{ij}^{\beta^*} Z_{ij}^{\theta} \right) \dot{Z}_{ij} \right] \xrightarrow{p} 0,$$

where the expectation is taken with respect to Y_{ij}, $\boldsymbol{X}_{ij}$ and Z_{ij}. For a compact parameter space Θ, the uniform law of large numbers provides that

$$\sup_{\theta \in \Theta} \left| \frac{1}{N^2} \sum_{i,j} w_{ij} \left(Y_{ij} - \boldsymbol{X}_{ij}^{\beta^*} Z_{ij}^{\theta} \right) \dot{Z}_{ij} - \frac{1}{N^2} \sum_{i,j} \mathbb{E} \left[w_{ij} \left(Y_{ij} - \boldsymbol{X}_{ij}^{\beta^*} Z_{ij}^{\theta} \right) \dot{Z}_{ij} \right] \right| \xrightarrow{a.s.} 0.$$

Since $\widehat{\theta}$ satisfies that $\sum_{i,j} w_{ij} \left(Y_{ij} - \boldsymbol{X}_{ij}^{\beta^*} Z_{ij}^{\widehat{\theta}} \right) \dot{Z}_{ij} = 0$, we expect $\left| \widehat{\theta} - \theta_0 \right| \xrightarrow{p} 0$, where θ_0 satisfies that $\sum_{i,j} \mathbb{E} \left[w_{ij} \left(Y_{ij} - \boldsymbol{X}_{ij}^{\beta^*} Z_{ij}^{\theta_0} \right) \dot{Z}_{ij} \right] = 0$. Similar to the proof of Proposition 1, we partition the support of $(\boldsymbol{X}_{ij}, Z_{ij})$ as $\{\Delta_k\}_{k=1}^K$. Note that the partition is not indexed by i or j. We use the same partition $\{\Delta_k\}_{k=1}^K$ across i and j to simplify the notation. It holds that

$$\begin{aligned}
0 =& \sum_{i,j} \mathbb{E} \left[w_{ij} \left(Y_{ij} - \boldsymbol{X}_{ij}^{\beta^*} Z_{ij}^{\theta_0} \right) \dot{Z}_{ij} \right] \\
=& \sum_{i,j} \sum_{k=1}^{K} \mathbb{E} \left[w_{ij} \left(Y_{ij} - \boldsymbol{X}_{ij}^{\beta^*} Z_{ij}^{\theta_0} \right) \dot{Z}_{ij} \mid \boldsymbol{X}_{ij} = \boldsymbol{x}_k, Z_{ij} = z_k \right] f_{\boldsymbol{X},Z} \left(\boldsymbol{x}_k, z_k \right) \Delta_k \\
=& \sum_{i,j} \sum_{k=1}^{K} \mathbb{E} \left[w_{ij} \left(\boldsymbol{X}_{ij}^{\beta^*} Z_{ij}^{\xi_{ij}} + \epsilon_{ij} - \boldsymbol{X}_{ij}^{\beta^*} Z_{ij}^{\theta_0} \right) \dot{Z}_{ij} \mid \boldsymbol{X}_{ij} = \boldsymbol{x}_k, Z_{ij} = z_k \right] f_{\boldsymbol{X},Z} \left(\boldsymbol{x}_k, z_k \right) \Delta_k \\
=& \sum_{i,j} \sum_{k=1}^{K} \mathbb{E} \left[w_{ij} \left(\boldsymbol{X}_{ij}^{\beta^*} Z_{ij}^{\xi_{ij}} - \boldsymbol{X}_{ij}^{\beta^*} Z_{ij}^{\theta_0} \right) \dot{Z}_{ij} \mid \boldsymbol{X}_{ij} = \boldsymbol{x}_k, Z_{ij} = z_k \right] f_{\boldsymbol{X},Z} \left(\boldsymbol{x}_k, z_k \right) \Delta_k. \qquad (11)
\end{aligned}$$

By Model 1 and Assumption 1, we have that

$$\sum_{k=1}^{K} \mathbb{E} \left[\left(\boldsymbol{X}_{ij}^{\beta^*} Z_{ij}^{\xi_{ij}} - \boldsymbol{X}_{ij}^{\beta^*} Z_{ij}^{h(\boldsymbol{X}_{ij})} \right) \dot{Z}_{ij} \mid \boldsymbol{X}_{ij}, Z_{ij} \right] = 0 \qquad (12)$$

holds for any i and j. Define

$$\phi_{ij,k}(\theta) \equiv \mathbb{E}\left[w_{ij}\left(\boldsymbol{X}_{ij}^{\beta^*} Z_{ij}^{\xi_{ij}} - \boldsymbol{X}_{ij}^{\beta^*} Z_{ij}^{\theta}\right)\dot{Z}_{ij} \mid \boldsymbol{X}_{ij} = \boldsymbol{x}_k, Z_{ij} = z_k\right] f_{\boldsymbol{X},Z}(\boldsymbol{x}_k, z_k)\,\Delta_k.$$

Analogous to the proof of Proposition 1, we linearly approximate $\sum_{i,j}\sum_{k=1}^{K}\phi_{ij,k}(\theta_0)$ using the first-order Taylor expansion at $h(\boldsymbol{x}_k)$:

$$\sum_{i,j}\sum_{k=1}^{K}\phi_{ij,k}(\theta_0) \approx \sum_{i,j}\sum_{k=1}^{K}\phi_{ij,k}(h(\boldsymbol{x}_k)) + \sum_{i,j}\sum_{k=1}^{K}\phi'_{ij,k}(h(\boldsymbol{x}_k))(\theta_0 - h(\boldsymbol{x}_k)), \tag{13}$$

where $\phi'_{ij,k}(\cdot)$ denotes the first-order derivative of $\phi_{ij,k}(\cdot)$. Equations (11) and (12) imply that

$$\sum_{i,j}\sum_{k=1}^{K}\phi_{ij,k}(\theta_0) = \sum_{i,j}\sum_{k=1}^{K}\phi_{ij,k}(h(\boldsymbol{x}_k)) = 0.$$

As a result, by Equation (13), it holds that

$$\theta_0 \approx \left[\sum_{i,j}\sum_{k=1}^{K}\phi'_{ij,k}(h(\boldsymbol{x}_k))\right]^{-1}\left[\sum_{i,j}\sum_{k=1}^{K}\phi'_{ij,k}(h(\boldsymbol{x}_k))\,h(\boldsymbol{x}_k)\right].$$

We first compute $\sum_{k=1}^{K}\phi'_{ij,k}(h(\boldsymbol{x}_k))\,h(\boldsymbol{x}_k)$ as

$$\begin{aligned}
&-\sum_{k=1}^{K}\mathbb{E}\left[w_{ij}\boldsymbol{X}_{ij}^{\beta^*} Z_{ij}^{h(\boldsymbol{x}_k)}\dot{Z}_{ij}^2 h(\boldsymbol{x}_k) \mid \boldsymbol{X}_{ij} = \boldsymbol{x}_k, Z_{ij} = z_k\right] f_{\boldsymbol{X},Z}(\boldsymbol{x}_k, z_k)\,\Delta_k \\
&= -w_{ij}\sum_{k=1}^{K}\mathbb{E}\left[Y_{ij}\dot{Z}_{ij}^2 h(\boldsymbol{x}_k) \mid \boldsymbol{X}_{ij} = \boldsymbol{x}_k, Z_{ij} = z_k\right] f_{\boldsymbol{X},Z}(\boldsymbol{x}_k, z_k)\,\Delta_k,
\end{aligned}$$

where the equality follows because w_{ij} is non-random given $\boldsymbol{X}_{ij}$ and Z_{ij}, and

$$\mathbb{E}[Y_{ij} \mid \boldsymbol{X}_{ij} = \boldsymbol{x}_k, Z_{ij} = z_k] = \mathbb{E}\left[\boldsymbol{X}_{ij}^{\beta^*} Z_{ij}^{h(\boldsymbol{x}_k)} \mid \boldsymbol{X}_{ij} = \boldsymbol{x}_k, Z_{ij} = z_k\right]$$

by Assumption 1. By letting K be sufficiently large, we have that

$$\begin{aligned}
\sum_{k=1}^{K}\phi'_{ij,k}(h(\boldsymbol{x}_k))\,h(\boldsymbol{x}_k) &= -w_{ij}\int\mathbb{E}\left[Y_{ij}\dot{Z}_{ij}^2 h(\boldsymbol{X}) \mid \boldsymbol{X}_{ij} = \boldsymbol{x}, Z_{ij} = z\right] f_{\boldsymbol{X}_{ij},Z_{ij}}(\boldsymbol{x}, z)\,d\boldsymbol{x}dz \\
&= -w_{ij}\mathbb{E}\left[Y_{ij}\dot{Z}_{ij}^2 h(\boldsymbol{X}_{ij})\right],
\end{aligned}$$

where the last equality follows from the law of iterated expectations. Based on a similar argument, we can obtain that $\sum_{k=1}^{K} \phi'_{ij,k}\left(h\left(\boldsymbol{x}_k\right)\right)$ equals

$$-w_{ij}\sum_{k=1}^{K}\mathbb{E}\left[Y_{ij}\dot{Z}_{ij}^2 \mid \boldsymbol{X}_{ij}=\boldsymbol{x}_k, Z_{ij}=z_k\right] f_{\boldsymbol{X},Z}\left(\boldsymbol{x}_k, z_k\right)\Delta_k = -w_{ij}\mathbb{E}\left[Y_{ij}\dot{Z}_{ij}^2\right].$$

Because the unconditional distribution of (Y_{ij}, Z_{ij}) is the same across i and j, it holds that $\mathbb{E}\left[Y_{ij}\dot{Z}_{ij}^2\right] = \mathbb{E}\left[Y\dot{Z}^2\right]$. By the definition of w_{ij}, we have that $s_{ij} = N^2 w_{ij}/\sum_{ij} w_{ij}$. Therefore, we can approximate θ_0 by $\frac{1}{\mathbb{E}\left[Y\dot{Z}^2\right]}\sum_{ij} s_{ij}\mathbb{E}\left[Y\dot{Z}^2 h\left(\boldsymbol{X}_{ij}\right)\right]$. By Assumption 2, we can further obtain that

$$\theta_0 \approx \frac{\sum_{ij} s_{ij}\mathbb{E}\left[Y\dot{Z}^2 h\left(\boldsymbol{X}_{ij}\right)\right]}{\mathbb{E}\left[Y\dot{Z}^2\right]} = \frac{\mathbb{E}\left[\dot{Z}^2\right]\sum_{ij} s_{ij}\mathbb{E}\left[Y h\left(\boldsymbol{X}_{ij}\right)\right]}{\mathbb{E}\left[\dot{Z}^2\right]\mathbb{E}\left[Y\right]} = \frac{\sum_{ij} s_{ij}\mathbb{E}\left[Y h\left(\boldsymbol{X}_{ij}\right)\right]}{\mathbb{E}\left[Y\right]}.$$

Thus, we have shown that $\left|\widehat{\theta}-\theta_0\right| \xrightarrow{p} 0$, where $\theta_0 \approx \frac{\sum_{ij} s_{ij}\mathbb{E}[Y h(\boldsymbol{X}_{ij})]}{\mathbb{E}[Y]}$. At the same time, for Δt_{ij} being sufficiently small by Assumption 7, θ_{ij} coincides with the h-elasticity. Therefore, $\theta_{\boldsymbol{s}}$ is close to $\frac{\sum_{i,j} s_{ij}\mathbb{E}[Y h(\boldsymbol{X}_{ij})]}{\mathbb{E}[Y]}$, which is approximately θ_0, for large N by the weak law of large numbers for triangular arrays. Hence, we can conclude that $\left|\widehat{\theta}-\theta_{\boldsymbol{s}}\right| \xrightarrow{p} \delta$, where $\delta \approx 0$. □

B Heterogeneity and Heteroskedasticity

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):

$$Y_{ij} = \exp\left(\beta_1^* \log \mathrm{GDP}_i + \beta_2^* \log \mathrm{GDP}_j + \theta_{ij}^* \log \mathrm{Distance}_{ij}\right) * \epsilon_{ij} \tag{14}$$

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.[22] $\boldsymbol{X}_{ij} \equiv (\mathrm{GDP}_i, \mathrm{GDP}_j)$ and $Z_{ij} \equiv \mathrm{Distance}_{ij}$.

Case 1: $\sigma^2 = \exp\left(\log Y_{ij}\right)^{-2}$; $Var\left(Y_{ij} \mid \boldsymbol{X}_{ij}, Z_{ij}, \theta_{ij}^*\right) = 1$.

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

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

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

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 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.[23]

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 Equation (**??**) with d'_{ij} being the counterfactual distance with a 1% increase from d_{ij} and Y'_{ij} being the corresponding trade flows generated from Equation (14). This corresponds to the uniform treatment. The average elasticity δ^{AE} is calculated from Equation (**??**).

Estimates

In this section, we investigate the behaviour 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 $Y_{ij}/\sum_{i,j} Y_{ij}$ as weights. For the FWLS estimator, the share of predicted trade flows $\widetilde{Y}_{ij}/\sum_{i,j} Y_{ij}$ is used as weights. The GPML and PPML estimators use Y_{ij} as the dependent variable and the log of GDPs and log of bilateral distance as independent variables, while all other estimators use $\log Y_{ij}$ 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:

$$bias_k = 1 - \frac{\widehat{\theta}_k}{\delta^{EA}_k}, \tag{15}$$

where $k \in \{0, 1, \ldots, 100\}$ is the number of large countries in the economy and $\widehat{\theta}_k$ is the

[23] 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.

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. Figure 5 displays the behaviour 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},[24] 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 Figure 5. Consequently, the behaviour of the selected estimators differentiates from each other and centres 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 estimators' empirical moment conditions in Section 3.1. 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.

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 (Figure 5b and 5d). On the other hand, OLS, WLS, and FWLS estimators all suffer from having heteroskedastic errors. While WLS and FWLS estimators also attempt to address the country-pair heterogeneity issue by assigning different weights to observations, they all result in notable biases from the parameter of interest, δ_k^{EA}.

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 Figure 5f, the GPML estimator still has a negligible bias, while the bias of the PPML estimator, in this case, is much larger.

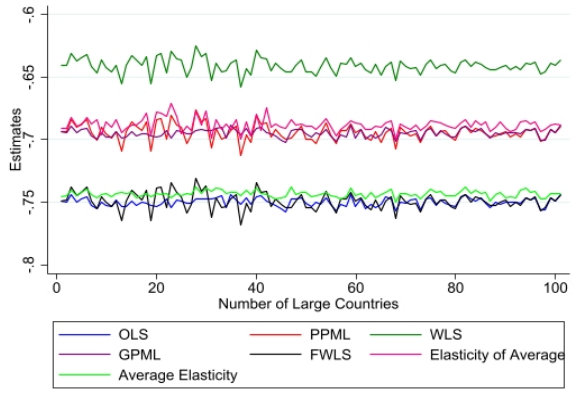
Next, we follow Fieler (2011) and introduce the dependence of heterogeneity on country size, where trade elasticities between a pair of large countries will be larger in absolute value than those between country pairs involving a small country. Before analyzing the main estimated results, we present a summary table below (Table VI) for the benchmark results where no country-pair heterogeneity is considered, and all countries are small. Recall that we have $\theta_{LL}^* = -0.5$ and $\theta_{SS}^* = \theta_{SL}^* = -1$. Under these conditions, according to SST, we could expect good performance from the PPML estimator across all three specifications of heteroskedasticity, while the OLS estimator only performs well under Case 3. As shown in Table VI, under Case 1 and 2, PPML gives an estimate closer to -1, the universal trade elasticity, than the OLS estimator with smaller standard errors. The reverse is true under Case 3, consistent with SST's findings under their DGP.

Then we introduce country-pair heterogeneity into the benchmark by making half of the countries large. As shown in the second half of Table VI, the OLS estimates are drastically

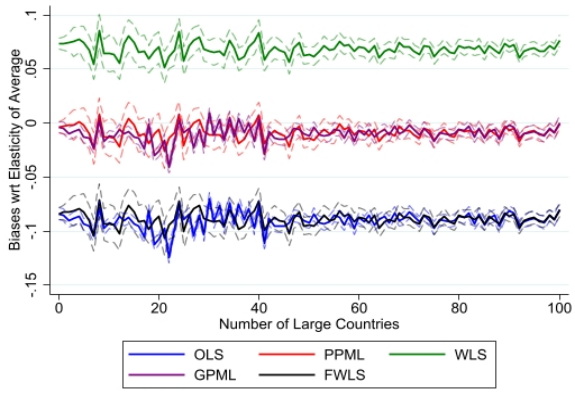
[24] For each k, the value of δ_k^{AE} is calculated as the average of 80 simulations

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

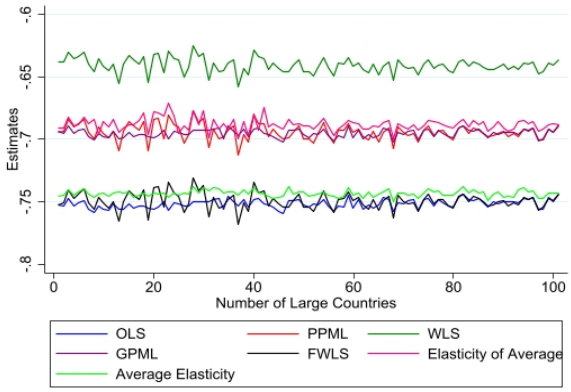
(a) Estimates Case 1



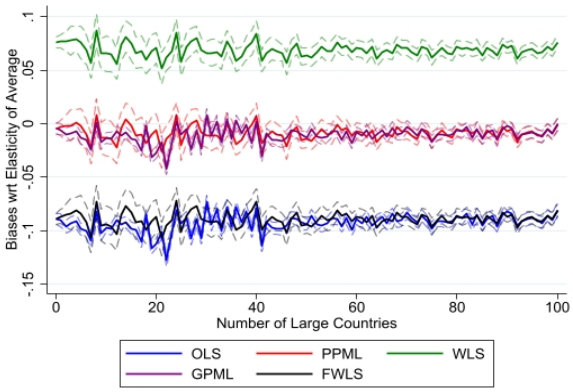
(b) Biases Case 1



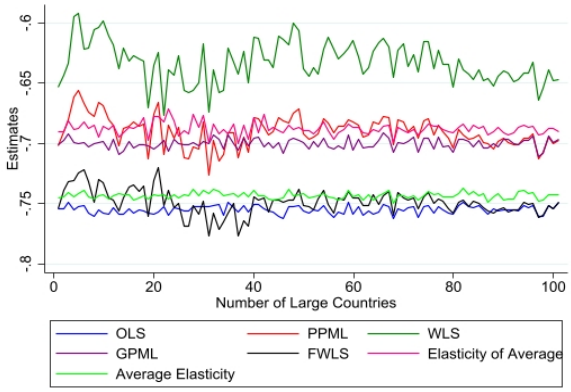
(c) Estimates Case 2



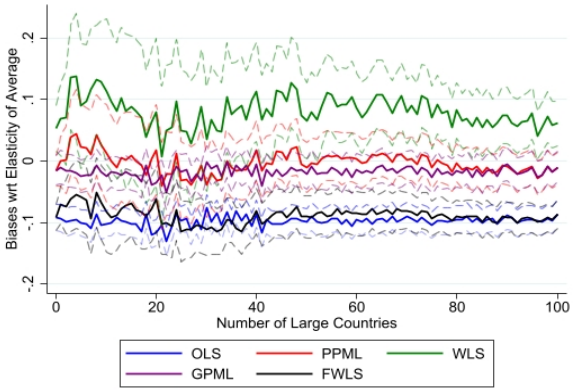
(d) Biases Case 2



(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 volume 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 Equation (15), 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 VI
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)

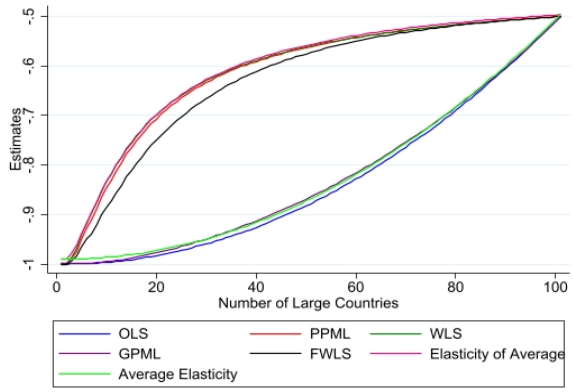
Note: This table contains both the case with no country-pair heterogeneity and the heterogeneous case with the number of large and small countries both equal to 50. N_L is the number of large countries in the economy. Case 1, 2, and 3 heteroskedasticity follow the corresponding specifications in Section 3.1. Biases are calculated using equation 15. The estimated 95% confidence intervals are included in brackets.

different from the PPML ones. We also include WLS estimates to show the similarity between the estimated results using WLS and PPML, which is consistent with the discussion in Section 3. The estimated results from FWLS are reasonably close to those of WLS, indicating that endogenous weights do not significantly impact our results. Furthermore, while GPML performs well under both specifications of heteroskedasticity when there are no large countries in the economy, it fails to address the issue of heterogeneity, resulting in similar results as OLS and much larger biases than PPML in estimating the aggregate elasticity.

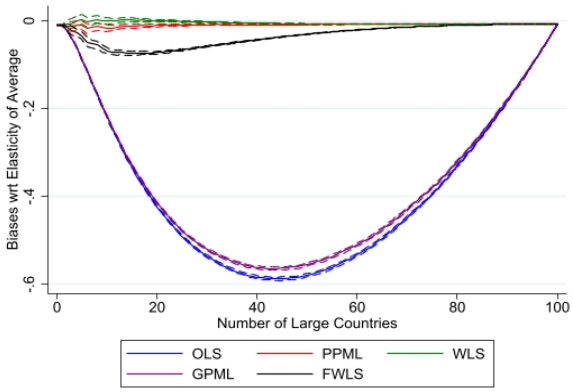
Figure 6 displays the behaviour of all estimators of the trade elasticity under this circumstance together with the aforementioned heteroskedasticity. With the increasing number of large countries in the economy, we should expect an increase in the average of estimated elasticities. As before, we also include the aggregate elasticity of each heterogeneity case, δ_k^{EA}, and the average elasticity, δ_k^{AE}, in the figures of estimates.

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

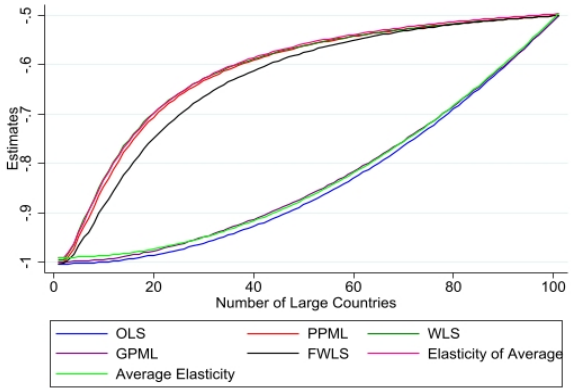
(a) Estimates Case 1



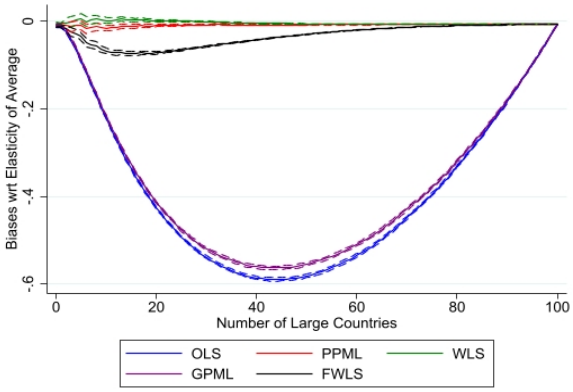
(b) Biases Case 1



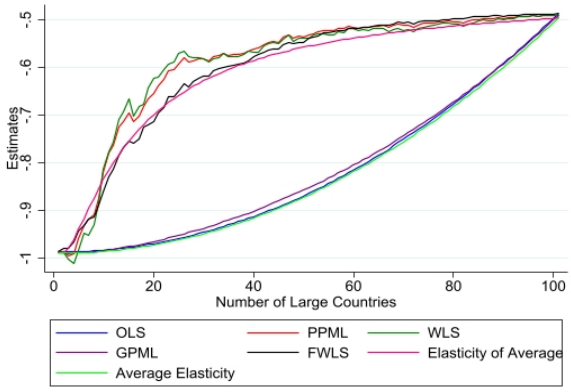
(c) Estimates Case 2



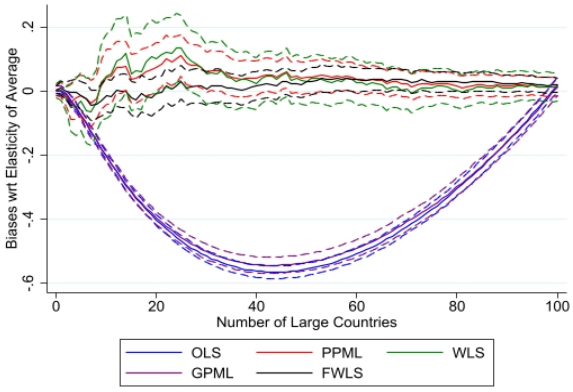
(d) Biases Case 2



(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 volume 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 Equation (15), 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.

It is, again, clear from the left panel graphs that OLS and GPML estimators behave differently than other estimators and the elasticity of the average, but are in close alignment with the average elasticity. In general, the PPML, WLS, and FWLS estimators outperform OLS and GPML, with both smaller deviations from δ_k^{EA} and similar behaviour as the number of large countries in the economy k increases. The differences between the two categories of estimators are significant when there is more heterogeneity in the economy, i.e., the number of large and small countries are close to each other. In contrast, in the case of little heterogeneity in the economy, the estimates are not drastically different. These two points indicate that in our simulations, country-pair heterogeneity dependent on country size generates a larger share of bias than heteroskedastic errors, which confirms our point that the PPML, WLS, and FWLS estimators are capable of addressing heterogeneity-related issues better than OLS and GPML.

Figures 6b, 6d, and 6f represent biases for Cases 1, 2, and 3, respectively. We can see that despite the distinct patterns of heteroskedasticity in the error term, the effect of this distinction is trivial on both the shape of the graph and the scale of the bias. Here, the bias caused by country-pair heterogeneity dependent on country size outweighs the bias caused by heteroskedastic errors, so unlike the results in Figure 5, it is crucial to assign proper weights to observations if we want to estimate the aggregate elasticity, δ_k^{EA}, correctly. [25]

The results in Table VI are also presented in these graphs. In Figure 6a, 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 (Figure 6e).

The dashed lines show the narrow 95% confidence interval of the biases under each case, indicating that the differences across estimators are statistically significant.

[25]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.