PPML, Gravity, and Heterogeneous Trade Elasticities*

Vladimir Tyazhelnikov† Xuetao Shi‡ Xinbei Zhou§

March 2025

Abstract

The gravity equation is the most widely used empirical framework in international trade and is typically estimated using either Ordinary Least Squares (OLS) or Poisson Pseudo-Maximum Likelihood (PPML), with the latter offering robustness to heteroskedasticity in the error term. We show that when trade elasticities vary across country pairs, OLS and PPML estimates recover fundamentally different objects: OLS estimates the average elasticity, while PPML estimates the elasticity of the average. We develop a formal test for coefficient heterogeneity and demonstrate that the divergence between OLS and PPML estimates is largely attributable to this difference in interpretation rather than misspecification alone. In addition, we propose a novel weighted PPML (WPPML) estimator designed to recover the elasticity of the average in the presence of heterogeneous trade shocks.

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 and Australasian Trade Workshop for helpful suggestions.

†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 *elasticity of the average*.

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 elasticity of average, the effect of a one percent change in one variable on the percentage change of the average value of another variable.[1,2] This distinction between average elasticity and the elasticity of the average echoes an extensive discussion on the difference between microeconomic elasticities, obtained from firm-level or disaggregated sectoral data and corresponding macroeconomic elasticities. For instance, Imbs and Mejean (2015) show that a simple aggregation of micro-level elasticities leads to a biased macro-level elasticity (elasticity of the average in the context of our project). Our paper contributes to this literature by linking these conceptual distinctions to the aggregation properties of various econometric estimators.

In particular, we 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 elasticity of the average.[3] By contrast, estimates, obtained with

[1]Consider a simple numerical example: a country consists of two individuals, one rich, with an income of \$1,000,000 and facing a 10% tax rate and one poor with an income of \$1,000 and 1% tax rate. The unweighted average tax rate is calculated as 5.5%, yet total tax revenue is \$100, 000 + \$10 = \$100, 010. The percentage of total tax revenue in terms of total income is $\frac{100,010}{1,001,000} = 9.991\%$, which can be interpreted as the "tax rate on average". The latter is the true effect of tax rates on the whole economy. The unweighted average of tax rates differs from the true effect because it neglects the different weights that the two individuals have in the economy and includes both of them in the computation of average elasticity with equal weights. To accurately calculate the "tax rate on average", the heterogeneous tax rates for the two individuals should be assigned weights according to their share of the total income of the whole country: $10\% * \frac{1,000,000}{1,001,000} + 1\% * \frac{1,000}{1,001,000} = 9.991\%$. Similarly, in the context of international trade, countries differ in size and trade volumes, and failing to account for these differences biases elasticity estimates when using unweighted techniques such as OLS.

[2]Elasticity of the average can also be interpreted as the elasticity of the aggregate — the percentage change of the total value of the dependent variable, aggregated over all observations to a 1% shock in the covariate of interest.

[3]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.

another popular estimator, the ordinary least squares (OLS) can be interpreted as the average elasticity.

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. We provide evidence that heterogeneity in trade elasticities across country pairs—particularly in the effect of distance—also plays a significant role. We then provide a formal test, which detects the presence of coefficient heterogeneity in a given sample. Furthermore, we show, both theoretically and with Monte Carlo simulations, that the presence of such heterogeneity causes bias in the estimation of the elasticity of average by the OLS.[4]

To further investigate the link between heterogeneity and heteroskedasticity biases, we introduce a weighted least squares (WLS) estimator and show that both the PPML and WLS estimators are robust to heterogeneity. Still, unlike PPML and similar to OLS, WLS is not robust to the heteroskedasticity of an error term. The fourth estimation procedure we consider is Gamma pseudo maximum likelihood (GPML), which, similar to PPML, is consistent when the error term is heteroskedastic but is often overlooked because of its potentially lower efficiency compared to PPML. We show that, similar to OLS, the GPML estimates can be interpreted as the average elasticity, which explains why GPML and PPML estimates often differ in practice.

A comparison of OLS, PPML, WLS, and GPML estimates allows us to decompose the bias caused by using the OLS estimator with the log-linearized gravity equation into two different sources — heteroskedasticity in the error term and heterogeneity across country pairs — using PPML estimate as a benchmark. We document 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% of the total difference between OLS and PPML for the estimates of the average elasticity and 30% for the estimates of the elasticity of average. Our exercise is not a formal decomposition of heterogeneity and heteroskedasticity biases which itself is a challenging task. We argue that our results, while suggestive, are still meaningful and can be of use within and beyond the field of International Trade.

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 average effect of a

[4]In this paper, we focus on the estimation of the elasticity of the average; from this point of view, OLS estimates are biased. Alternatively, we can say that OLS and PPML estimates have different interpretations.

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 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 difference between WPPML and PPML estimates is highly heterogeneous between different sectors and is driven by both elasticity heterogeneity and the industry-specific profile of tariff changes between 2001 and 2016.

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.[5]

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

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

estimator approximates average marginal effects of the covariates on the true conditional expectation. Compared to Silva and Winkelmann (2024), this paper considers a special case of the misspecified exponential model where parameters might be random. As a result, we relax the normal distribution assumption and are able to connect the marginal effect of the covariates on the true conditional expectation with the heterogeneous trade elasticity.

The rest of the paper is organised as follows. In Section 2, we set up the theoretical background of the paper and compare the performances of different estimators of the gravity equation in the context of International Trade. Then we employ real trade data to provide motivational evidence for the existence of country-pair heterogeneous elasticities in Section 3. In Section 4, we present the results of Monte Carlo simulations and study the behaviour of the estimators in the presence of both heteroskedasticity and country-pair heterogeneity. We extend the previous analysis into considering heterogeneous treatment and propose the use of WPPML in Section 5 and conclude with Section 6.

2 Theoretical Background

In this section, we begin by outlining our assumptions regarding the nature of heterogeneity. Notably, we do not impose any assumptions on the functional form of heterogeneity. If heterogeneity arises from a known function of observable independent variables, the problem becomes straightforward as we can include an interaction term between the variable of interest and the variables that cause heterogeneity. To retain tractability, however, we make some minimal assumptions; in particular, we assume that heterogeneity is primarily driven by covariates other than trade costs. We assume that trade elasticity does not depend on trade costs because such a model cannot be interpreted as a constant elasticity model and would either require an alternative specification or estimation with non-parametric methods.

We then introduce a formal statistical test to detect the presence of heterogeneity. Crucially, this test maintains correct size: rejection of the null hypothesis of homogeneity provides credible evidence of heterogeneity. However, failure to reject the null does not necessarily imply homogeneity, as the test may lack power. The procedure is a standard J-test, available as a post-estimation option following generalized method of moments (GMM) estimation, and can be implemented using standard econometric software.

The section concludes with a formal proof that PPML estimates the elasticity of the average, followed by a comparison of various estimators and a discussion of the interpretation of their coefficients.

Model

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 $\prod_{j=1}^{p} u_j^{v_j}$. Consider the following model:

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

where $\boldsymbol{X} \in \mathbb{R}^p$ and $Z \in \mathbb{R}$ 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

$$h\left(\boldsymbol{X}, Z\right) \equiv \log_Z \left(\frac{\mathbb{E}\left[Y \mid \boldsymbol{X}, Z\right]}{\boldsymbol{X}^{\beta^*}} \right).$$

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 X and Z. 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 a linear function of $\boldsymbol{X}$ and a constant with respect to Z. We make the following assumption on $h\left(\cdot\right)$ to restore the tractability of the model.

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

Under Assumption 2, we immediately obtain that $\mathbb{E}\left[Y \mid \boldsymbol{X}, Z\right] = \boldsymbol{X}^{\beta^*} Z^{h(\boldsymbol{X})}$. The Z-elasticity of $\mathbb{E}\left[Y \mid \boldsymbol{X}, Z\right]$ can be computed as:[6]

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

In consequence, $h\left(\boldsymbol{X}\right)$ measures the fractional response of the expectation of Y conditional on $\left(\boldsymbol{X}, Z\right)$ to a fraction change in Z. Since $h\left(\boldsymbol{X}\right)$ is a function of $\boldsymbol{X}$, the elasticity can

[6]Note that while this formulation allows us to estimate partial effect of Z on $\mathbb{E}\left[Y \mid \boldsymbol{X}, Z\right]$, 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 $\mathbb{E}\left[Y \mid \boldsymbol{X}, Z\right] = \boldsymbol{X}_c^{\beta^*} Z^{h(\boldsymbol{X}_{nc})}$ 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 $\boldsymbol{X}_c$ is the original multiplicative form. This formulation requires the effect of Z on $\mathbb{E}\left[Y \mid \boldsymbol{X}, Z\right]$ be a polynomial function with coefficient being $h\left(\boldsymbol{X}\right)$, which does not depend on Z.

change with $\boldsymbol{X}$. From now on, we refer to $h(\boldsymbol{X})$ as the Elasticity.

Although Assumption 2 imposes restriction on the unobserved heterogeneity, it allows for many interesting cases of heterogeneity. First of all, we can have ξ being a constant, which is the classical model where there is no unobserved heterogeneity. In this case, the elasticity $h(\boldsymbol{X})$ is a constant. In fact, as long as the random variable ξ satisfies that $\mathbb{E}\left[Z^{\xi-\mathbb{E}(\xi)} \mid \boldsymbol{X}, Z\right] = 1$, $h(\boldsymbol{X})$ is a constant. Second, Assumption 2 allows ξ to depend on Z and does not require ξ to be a deterministic function of $\boldsymbol{X}$. For instance, assume $\xi = h(\boldsymbol{X}) + \log_Z(\eta)$ such that $\mathbb{E}[\eta \mid \boldsymbol{X}, Z] = 1$. In this example, ξ is a function of Z and contains its own randomness represented by η. It can be verified that Assumption 2 is still satisfied. Additionally, ξ can also depend on ε if η correlates with ε.

Both ξ and $h(\boldsymbol{X})$ can be of interest for economists. However, we have seen that they can behave very differently with ξ being random but $h(\boldsymbol{X})$ being a 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 2. 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 study.

Test the Existence of Heterogeneity

Before studying estimators for $h(\boldsymbol{X})$, it can be important to test if the elasticity indeed varies with $\boldsymbol{X}$. If it does, then the elasticity exhibit heterogeneity. If not, then the elasticity is homogeneous. In other word, we aim to construct a test for testing whether the elasticity is heterogeneous.

Econometricians have developed tools for testing null hypotheses on functions. For the null hypothesis that $h(\boldsymbol{X})$ is a constant, one possible procedure 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. Instead, we propose a test that is simple to run and often powerful enough to detect heterogeneity of the elasticity in the data.

Under the null hypothesis that $H_0 : h(\boldsymbol{X}) = \gamma^*$, where γ^* is a constant, the conditional expectation becomes

$$\mathbb{E}[Y \mid \boldsymbol{X}, Z] = \boldsymbol{X}^{\beta^*} Z^{\gamma^*}. \tag{3}$$

This is a conditional moment equality condition. Let $\dot{\boldsymbol{X}} \equiv (\log X_1, \ldots, \log X_p)^\top$ and $\dot{Z} \equiv$

$\log Z$. Under the null hypothesis, it holds that

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

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.

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

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

denote the sample average of $g\left(\beta, \gamma\right)$. 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

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

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

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

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

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

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

The proof of the theorem follows from the classical proof in Newey and McFadden (1994).

Based on Theorem 2, we expect the distribution of J_n to be tight under the null hypothesis when n is sufficiently 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 hypothesis that $H_1 : h(\boldsymbol{X})$ is not a 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 (3).

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 sure that the elasticity exhibits heterogeneity.

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

PPML and the Elasticity of Average

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

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

We call $\left(\widehat{\beta}, \widehat{\gamma}\right)$ the PPML estimator. In this section, we study properties of the PPML estimator. In particular, we show that $\widehat{\gamma}$ approximates $\mathbb{E}[Yh(\boldsymbol{X})]/\mathbb{E}[Y]$, which we call the Elasticity of the Average.

To focus on $\widehat{\gamma}$, we replace $\widehat{\beta}$ with the true value β^* in the following discussion. Admittedly, this step is an approximation even when n is large because the probability limit of $\widehat{\beta}$ is not β^* in the presence of heterogenous elasticity. Assume that the variances of elements of $\boldsymbol{X}$, Z, ξ, and ε are finite. 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

$$\mathbb{E}\left[\left(Y-\boldsymbol{X}^{\beta^*}Z^{\gamma_0}\right)\dot{Z}\right]=0. \tag{4}$$

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 increasing K. As a result,

$$\mathbb{E}\left[\left(Y-\boldsymbol{X}^{\beta^*}Z^{\gamma_0}\right)\dot{Z}\right]=\sum_{k=1}^{K}\mathbb{E}\left[\left(Y-\boldsymbol{X}^{\beta^*}Z^{\gamma_0}\right)\dot{Z}\mid\boldsymbol{X}=\boldsymbol{x}_k,Z=z_k\right]f_{\boldsymbol{X},Z}(\boldsymbol{x}_k,z_k)\,\Delta_k=\mathbf{0} \tag{5}$$

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

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

Equation (6) implies that

$$\begin{aligned}\mathbb{E}\left[\left(Y-\boldsymbol{X}^{\beta^*}Z^{h(\boldsymbol{X})}\right)\dot{Z}\right]&=\int\mathbb{E}\left[\left(Y-\boldsymbol{X}^{\beta^*}Z^{h(\boldsymbol{X})}\right)\dot{Z}\mid\boldsymbol{X}=\boldsymbol{x},Z=z\right]f_{\boldsymbol{X},Z}(\boldsymbol{x},z)\,d\boldsymbol{x}dz\\&=\sum_{k=1}^{K}\mathbb{E}\left[\left(Y-\boldsymbol{X}^{\beta^*}Z^{h(\boldsymbol{X})}\right)\dot{Z}\mid\boldsymbol{X}=\boldsymbol{x}_k,Z=z_k\right]f_{\boldsymbol{X},Z}(\boldsymbol{x}_k,z_k)\,\Delta_k=\mathbf{0}.\end{aligned} \tag{7}$$

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

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

Together with Equations (5) and (7), it holds that $\sum_{k=1}^{K}\frac{\partial}{\partial\gamma}\phi_k(\gamma_k)(\gamma_0-\gamma_k)\approx 0$. Thus, we obtain that

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

Simple calculation would show that

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

By Equation (6), we have that

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

As a result, it holds that

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

This in turn shows that

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

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

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

Thus, we obtain that γ_0 can be approximated by $\mathbb{E}\left[Y\dot{Z}^2 h(\boldsymbol{X})\right]/\mathbb{E}\left[Y\dot{Z}^2\right]$. Since

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

$\mathbb{E}\left[Y\dot{Z}^2 h(\boldsymbol{X})\right]/\mathbb{E}\left[Y\dot{Z}^2\right]$ is approximately $\mathbb{E}\left[Yh(\boldsymbol{X})\right]/\mathbb{E}\left[Y\right]$ if $Cov\left(Yh(\boldsymbol{X}), \dot{Z}^2\right)$ is small relative to $\mathbb{E}\left[Yh(\boldsymbol{X})\right]\mathbb{E}\left[\dot{Z}^2\right]$ and $Cov\left(Y, \dot{Z}^2\right)$ is small relative to $\mathbb{E}\left[Y\right]\mathbb{E}\left[\dot{Z}^2\right]$, an assumption similar to Breinlich et al. (2022).

Hence, the PPML estimator $\widehat{\gamma}$ converges in probability to γ_0, which is approximately $\mathbb{E}\left[Yh(\boldsymbol{X})\right]/\mathbb{E}\left[Y\right]$.

Gravity Equations: International Trade Context

The results in this paper are applicable to the estimation of any multiplicative model in the presence of heterogeneity. To illustrate our point, however, we consider a particular case of such a model in the context of International Trade. We use the 'naive form' of the gravity equation from Head and Mayer (2014) to be consistent with the constant-elasticity models in SST and constrain the trade elasticity to be distance elasticity:[7]

$$Y_{ij} = \frac{\text{GDP}_i^{\beta_1^*} \times \text{GDP}_j^{\beta_2^*}}{\text{Distance}_{ij}^{-\xi^*}}\varepsilon_{ij}, \tag{8}$$

where Y_{ij} is the trade flow between country i and j, Distance_{ij} stands for bilateral distance, and ε_{ij} is the error term. Note that in this naive model, ξ^* is a constant rather than a random variable.

Taking logs on both sides of the equation yields the following:

$$\log Y_{ij} = \beta_1^* \log \text{GDP}_i + \beta_2^* \log \text{GDP}_j + \xi^* \log \text{Distance}_{ij} + \log \varepsilon_{ij}, \tag{9}$$

where $\xi^* < 0$ is the value of trade elasticity.

Denote $\boldsymbol{X}_{ij} \equiv (\text{GDP}_i, \text{GDP}_j)$ and $Z_{ij} \equiv \text{Distance}_{ij}$. To obtain consistent estimates of the coefficients in Equation (8) using the log-linearized form of Equation (9) by OLS, it is necessary for $\mathbb{E}\left(\log \varepsilon \mid \boldsymbol{X}, Z, \xi^*\right)$ to be constant. According to SST, however, ε is generally heteroskedastic. Running a simple OLS regression of $\log Y_{ij}$ on $\dot{\boldsymbol{X}}_{ij}$ and $\dot{Z}_{ij}$ leads to inconsistent and biased estimates of ξ^*. This is because even though the exogeneity

[7] In this paper, we follow SST as closely as possible, to model heterogeneous trade elasticities, but we have to deviate and use a more complex data generating process.

assumption $\mathbb{E}(\varepsilon \mid \boldsymbol{X}, \boldsymbol{Z}, \xi^*) = 0$ is satisfied in Equation (8), with heteroskedasticity in ε, $\mathbb{E}(\log \varepsilon \mid \boldsymbol{X}, \boldsymbol{Z}, \xi^*)$ does not necessarily equal a constant.

We further incorporate heterogeneous country-pair trade elasticities into Equation (9) by allowing ξ^* to change across country-pair:

$$\log Y_{ij} = \beta_1^* \log \text{GDP}_i + \beta_2^* \log \text{GDP}_j + \xi_{ij}^* \log \text{Distance}_{ij} + \log \varepsilon_{ij}.$$

Our objective is to estimate the (aggregate) trade elasticity that reflects the effect of a uniform change in the trade costs on bilateral trade volumes, i.e., the elasticity of average $\frac{\mathbb{E}(Y\xi^*)}{\mathbb{E}(Y)}$.

Empirically, let d_{ij} denote the distance between countries i and j, and Y'_{ij} be the counterfactual trade volume between countries i and j when the distance d_{ij} changes to d'_{ij}. The average elasticity is calculated as

$$\delta^{AE} = \frac{1}{N^2} \sum_{i,j} \left(\frac{Y'_{ij} - Y_{ij}}{Y_{ij}} / \frac{d'_{ij} - d_{ij}}{d_{ij}} \right) = \frac{1}{N^2} \sum_{i,j} \delta_{ij}, \tag{10}$$

which is the average ratio of the percentage change in trade volumes and the percentage change in distance. $\delta_{ij} \equiv \frac{Y'_{ij} - Y_{ij}}{Y_{ij}} / \frac{d'_{ij} - d_{ij}}{d_{ij}}$ is the pair-specific elasticity.

The elasticity of average is defined as the ratio of the total percentage change in trade flows and the average percentage change in distances:

$$\begin{aligned}
\delta^{EA} &= \frac{\frac{\sum_{i,j} Y'_{ij} - \sum_{i,j} Y_{ij}}{\sum_{i,j} Y_{ij}}}{\sum_{i,j} \frac{d'_{ij} - d_{ij}}{d_{ij}} / N^2} = \frac{\frac{\sum_{i,j} (Y'_{ij} - Y_{ij})}{\sum_{i,j} Y_{ij}}}{\sum_{i,j} \frac{d'_{ij} - d_{ij}}{d_{ij}} / N^2} \\
&= \frac{1}{\sum_{i,j} Y_{ij} \sum_{i,j} \frac{d'_{ij} - d_{ij}}{d_{ij}} / N^2} \sum_{i,j} \frac{Y'_{ij} - Y_{ij}}{Y_{ij}} \frac{d_{ij}}{d'_{ij} - d_{ij}} \frac{d'_{ij} - d_{ij}}{d_{ij}} Y_{ij} \\
&= \frac{1}{\sum_{i,j} Y_{ij} \sum_{i,j} \frac{d'_{ij} - d_{ij}}{d_{ij}} / N^2} \sum_{i,j} \delta_{ij} \frac{d'_{ij} - d_{ij}}{d_{ij}} Y_{ij} \\
&= \frac{\sum_{i,j} \delta_{ij} Y_{ij}}{\sum_{i,j} Y_{ij}} = \frac{\sum_{i,j} \delta_{ij} Y_{ij}}{N^2} / \frac{\sum_{i,j} Y_{ij}}{N^2},
\end{aligned} \tag{11}$$

where the second to last equality holds because $\frac{d'_{ij} - d_{ij}}{d_{ij}}$ is assumed to be constant across country pairs, representing the uniform percentage change in trade costs.

Comparison across Estimators

In this subsection, we discuss the behaviour of some commonly used estimators for the trade elasticity. We use Ω to denote the set of all country pairs i, j. Note that we denote $\dot{\boldsymbol{X}}_{ij} \equiv (\log X_{1,ij}, \ldots, \log X_{p,ij})^\top$ and $\dot{Z}_{ij} \equiv \log Z_{ij}$. Let $\widehat{Y}_{ij} = \boldsymbol{X}_{ij}^{\widehat{\beta}} Z_{ij}^{\widehat{\xi}}$.

The OLS estimator is defined as the solution to

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

In the case that the error term is homoskedastic and that the heterogeneity ξ^* is independent of GDP and Distance, we show in Section **??** that the OLS estimator $\widehat{\xi}$ is a consistent estimator for APE: $\mathbb{E}(\xi^*)$. This estimator was defined previously as the "average trade elasticity". However, even if both assumptions hold, using the OLS to estimate the "elasticity of average" leads to bias with heterogeneity because $\mathbb{E}(\xi^*)$ assigns equal weight to every individual ξ^*_{ij}; while different trading country pairs have a differential impact on the world's aggregate trade flows.

A natural solution candidate for the weighting problem is the WLS estimator, which is defined as the solution to

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

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]

[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.

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]

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

We also use PPML and GPML estimators in this study following the method discussed by SST. The PPML estimator is defined as the solution to

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

The moment condition of PPML indicates that it also involves the level deviations of Y_{ij} from its predicted value. As a result, similar to WLS, it addresses the heterogeneity issue. We have shown in Section 2 that the probability limit of the PPML estimator $\widehat{\xi}$ is approximately $\frac{\mathbb{E}(Y\xi^*)}{\mathbb{E}(Y)}$.

The GPML estimator is the solution to

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

Conversely, the moment condition of GPML relies on the percent deviations of Y_{ij} from its predicted value, so we can expect that the GPML and OLS estimator will yield similar estimates.

Table 1 below summarises the empirical moment conditions for each of the estimators discussed in this study.

From these conditions, one can see 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.

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

TABLE 1
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 Empirical Evidence

In this section, we apply different estimation techniques 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 provide the estimated results of WLS and FWLS estimators. Table 2 reports the results of all five estimators and corresponding standard errors.

First, we present the evidence obtained from the gravity dataset that country-pair elasticities are not homogeneous. Naturally, we cannot establish a link between the unobserved variable and distance elasticity; trade volumes, conversely, are observed. If the OLS estimates of the distance elasticity do not vary much depending on how large trade flows are, then there will not be a significant difference between the average elasticity and the elasticity of average. The reason is that trade shares serve as weights while computing the elasticity of average.

After sorting the bilateral trade flows into ascending order, we run OLS regression multiple times, dropping the 10 smallest trade flows in the sample at every iteration. The estimated $\widehat{\theta}$ and corresponding 95% confidence interval are saved and plotted in Figure 1a. This is equivalent to running regressions on different subsamples containing various levels of trade flows. As the number of iterations increases, the weight of large trade volumes increases as well. The graph demonstrates an unambiguously positive relationship between estimated

TABLE 2
Gravity equation

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

$\widehat{\theta}$ and larger average trade flows in the subsamples. Compared to the results generated from the same procedure but with 10 trade flows being randomly dropped at every iteration on Figure 1b, it is clear that a positive relationship between the estimated $\widehat{\theta}$ and larger trade volumes in the subsamples provides evidence of country-pair heterogeneity.[11] Note that we included exporter and importer fixed effects and standard controls for country pairs, so the results in Figures 1a and 1b are driven by unobserved heterogeneity.[12]

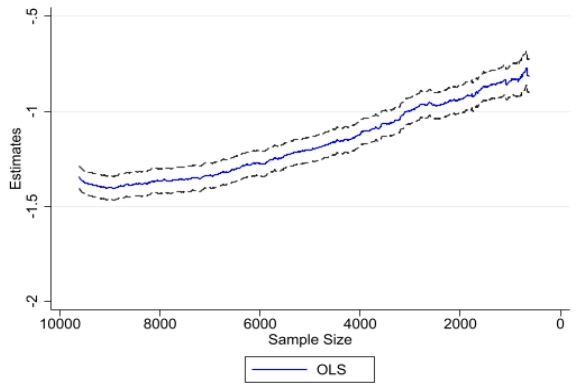
Note: This figure presents evidence of unobserved heterogeneity in country-pair elasticity. In panel (a), at each iteration, we drop 10 observations with the smallest corresponding volume of trade, while in panel (b), we drop 10 random observations. In panel (c), we sort the data into ascending order and show the OLS, WLS, FWLS, GPML, and PPML estimates on different subsamples. There are 9,613 observations in the full sample. The last iteration

[11]The standard errors for the estimates with smaller subsamples in Figure 1b are much larger because in small subsamples the number of countries increases relative to the number of observations, and consequently more fixed effects are included leading to higher standard errors of $\hat{\theta}$.

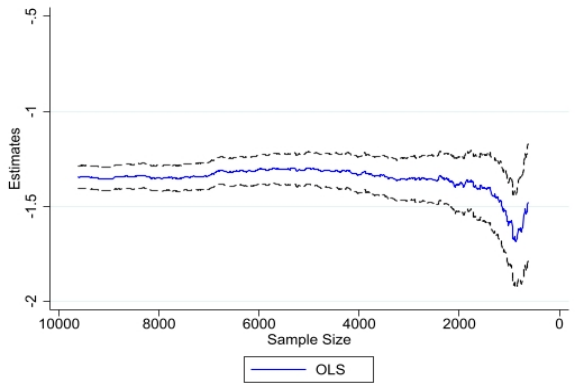
[12]Our findings are consistent with the study Yotov (2012) and Larch et al. (2019).

Figure 1. Evidence of heterogeneity

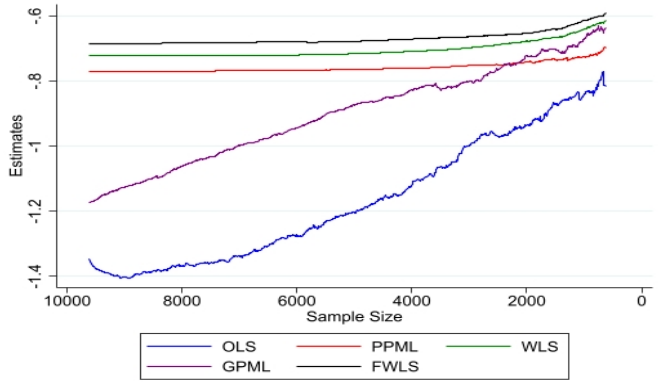
(a) Increasing trade volumes in subsamples - OLS



(b) Random subsamples - OLS



(c) Comparison across all estimators



Note: This figure presents evidence of unobserved heterogeneity in country-pair elasticity. In panel (a), at each iteration, we drop 10 observations with the smallest corresponding volume of trade, while in panel (b), we drop 10 random observations. In panel (c), we sort the data into ascending order and show the OLS, WLS, FWLS, GPML, and PPML estimates on different subsamples. There are 9,613 observations in the full sample. The last iteration has 623 observations.

has 623 observations.

In Section 4, we use Monte Carlo simulations to illustrate that the WLS, FWLS, and PPML estimators are robust to the presence of country-pair heterogeneity when estimating the "elasticity of average" in the simulated economy, and the GPML estimator yields similar results as the OLS estimator, which estimates the average elasticity. Now we show that these methods lead to similar results with the same gravity dataset. We still drop the 10 smallest trade values on each iteration but now apply the WLS, FWLS, PPML, and GPML estimators to the regression equation. For WLS, the weights are calculated as the share of each observation's bilateral trade flows on the global total trade flows. For FWLS, the weights are calculated as the share of each observation's predicted bilateral trade flows on

the sum of total predicted trade flows.

Compared to Figure 1a, the estimators from WLS, FWLS, and PPML illustrated in Figure 1c do not depend on the choice of subsamples except for the case when the sample size becomes very small. The overall constant estimates show that these estimators are robust to country-pair heterogeneity. The three estimators exhibit a similar pattern, indicating that all three of them are capable of addressing the heterogeneity in the country-pair elasticity problem, consistent with our findings in Section 2. Moreover, the estimator from GPML exhibits an increasing trend as the average trade flows in the subsamples grow larger. This result is also consistent with our findings in the previous section that GPML and OLS would yield similar results.

With PPML as the benchmark estimator, we can perform a decomposition to show the proportion of bias caused by heterogeneity and heteroskedasticity. The total bias is calculated as the difference between the OLS estimates and PPML estimates. According to our previous results, however, 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 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 3 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 elasticity of average. 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 Section 4.

Table 3 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.[13] 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.

Note that OLS and WLS do not allow for using zero trade flows due to the log-linearization

[13]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 heterogeneity elasticity, but if it is industrial heterogeneity, our findings are close to those in French (2017).

TABLE 3
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.

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 4 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; 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.

4 Monte Carlo Simulations

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 elasticity of average) and that the application of the PPML and WLS estimators corrects this bias.

TABLE 4
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.

Data Generating Process

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

$$Y_{ij} = \exp\left(\beta_1^* \log \text{GDP}_i + \beta_2^* \log \text{GDP}_j + \theta_{ij}^* \log \text{Distance}_{ij}\right) * \varepsilon_{ij} \tag{12}$$

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

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

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

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

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.[15]

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 elasticity of average δ^{EA} is calculated from Equation (11) 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 (12). This corresponds to the uniform treatment. The average elasticity δ^{AE} is calculated from Equation (10).

Estimates

In this section, we investigate the behaviour of the two distinct parameters of interest, the elasticity of average 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 elasticity of average as follows:

$$bias_k=1-\frac{\widehat{\theta}_k}{\delta_k^{EA}}, \tag{13}$$

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

[15] 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.

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 2 displays the behaviour of all estimators of the trade elasticity. As a reference, we include the elasticity of average of each heterogeneity case, δ_k^{EA}, and the average elasticity, δ_k^{AE},[16] in the figures of estimates.

The stark difference between the values of the elasticity of average and the average elasticity is straightforward from the left panel graphs of Figure 2. 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 2. 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 2b and 2d). 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}.

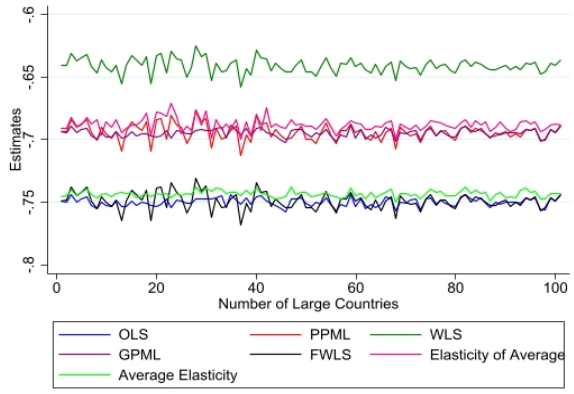
As discussed by SST, Case 3 is a special case that the log linear model corrects the heteroskedasticity in the data and that the GPML is the optimal PML estimator in this case. Therefore, in Figure 2f, the GPML estimator still has a negligible bias, while the bias of the PPML estimator, in this case, is much larger.

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 (elasticity of average) 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 (13), which are the percentage difference between the elasticity of average and the estimated coefficient for different estimators.

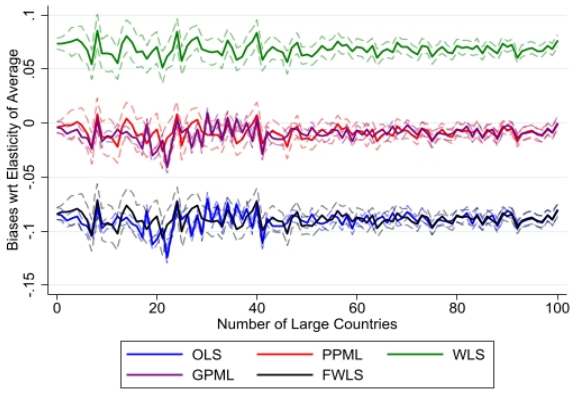
[16] For each k, the value of δ_k^{AE} is calculated as the average of 80 simulations

Figure 2. Estimates and biases of OLS, WSL, GPML, and PPML (Independent Heterogeneity)

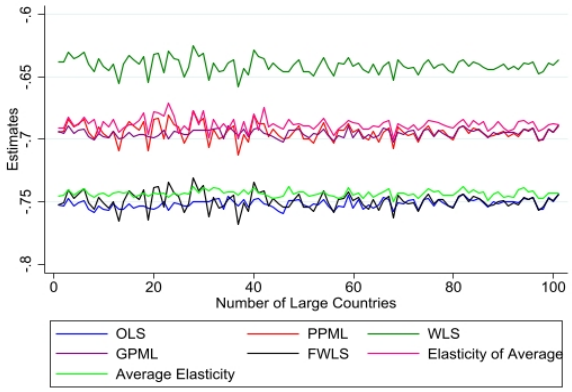
(a) Estimates Case 1



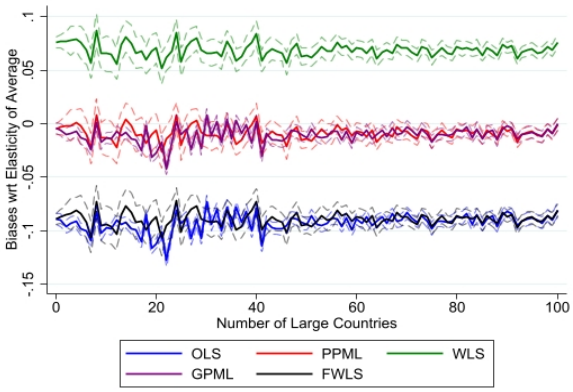
(b) Biases Case 1



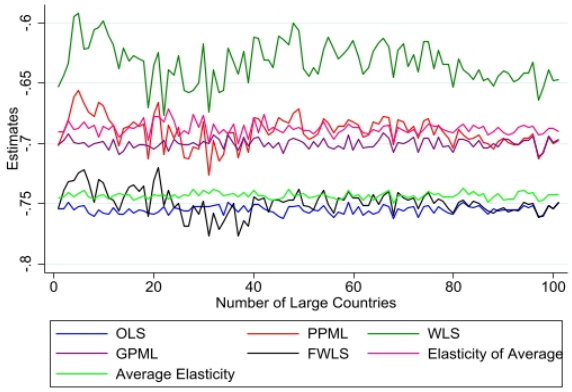
(c) Estimates Case 2



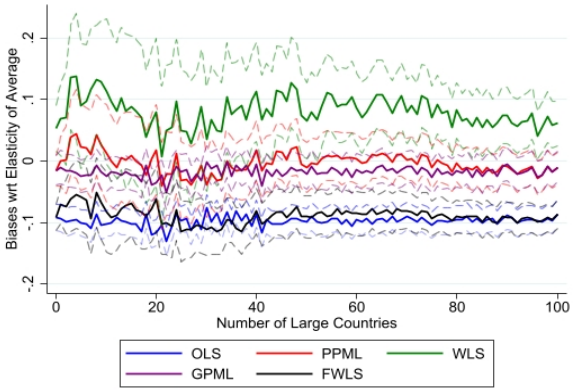
(d) Biases Case 2



(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 (elasticity of average) 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 (13), which are the percentage difference between the elasticity of average and the estimated coefficient for different estimators. The 95% confidence intervals of the estimated biases are included as dashed lines.

The 95% confidence intervals of the estimated biases are included as dashed lines.

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 5) 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 5, 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 5, the OLS estimates are drastically 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 elasticity of average.

Figure 3 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 elasticity of average of each heterogeneity case, δ_k^{EA}, and the average elasticity, δ_k^{AE}, in the figures of estimates.

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

TABLE 5
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)

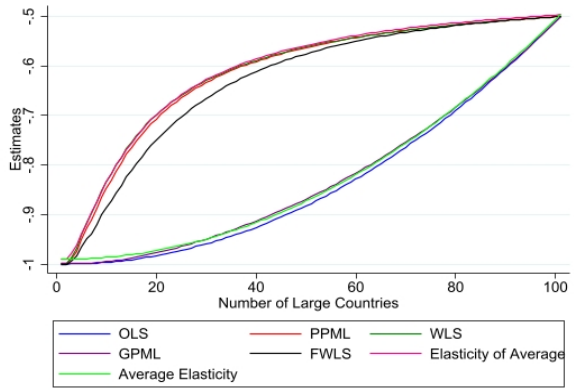
$N_L = 50$

	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)

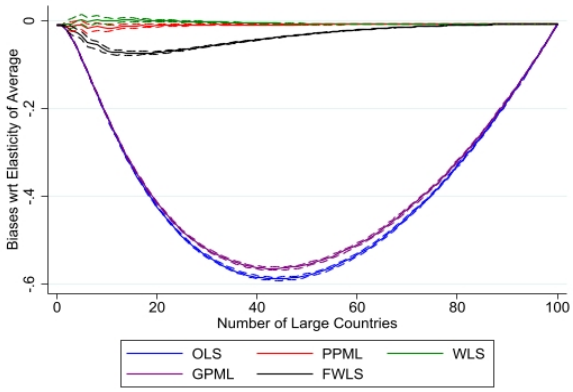
Note: This table contains both the case with no country-pair heterogeneity and the heterogeneous case with the number of large and small countries both equal to 50. N_L is the number of large countries in the economy. Case 1, 2, and 3 heteroskedasticity follow the corresponding specifications in Section 3.1. Biases are calculated using equation 13. The estimated 95% confidence intervals are included in brackets.

Figure 3. Estimates and biases of OLS, WSL, GPML and PPML

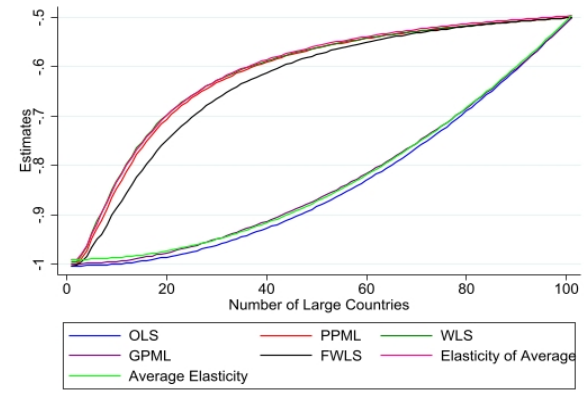
(a) Estimates Case 1



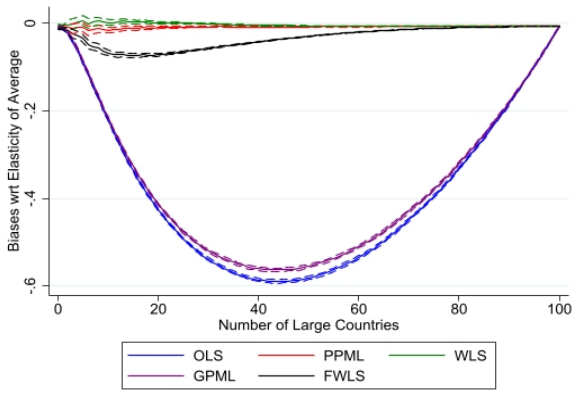
(b) Biases Case 1



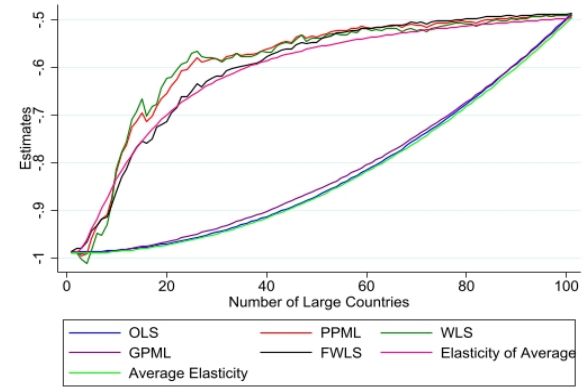
(c) Estimates Case 2



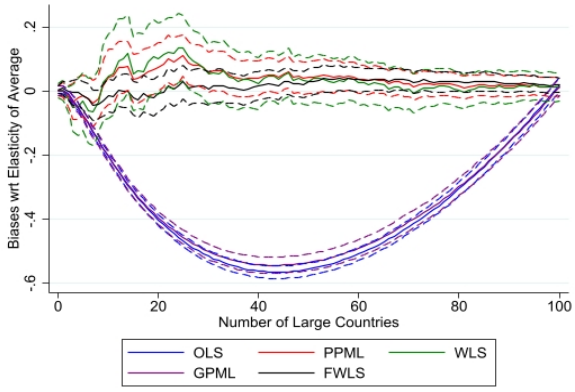
(d) Biases Case 2



(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 (elasticity of average) 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 (13), which are the percentage difference between the elasticity of average and the estimated coefficient for different estimators. The 95% confidence intervals of the estimated biases are included as dashed lines.

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 3b, 3d, and 3f 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 2, it is crucial to assign proper weights to observations if we want to estimate the elasticity of average, δ_k^{EA}, correctly. [17]

The results in Table 5 are also presented in these graphs. In Figure 3a, 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 elasticity of average δ_k^{EA}. Conversely, the OLS estimator performs best without heterogeneity in Case 3 (Figure 3e).

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

5 Heterogeneous Treatment

In previous discussions, we define the elasticity of average as the average percentage response of trade flows to a uniform change in trade costs (distances). 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.[18] Therefore, we further derive an expression for the elasticity of average 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.

WPPML Estimator

Similar to Section 2, we denote the distance between countries i and j as d_{ij}, and the counterfactual trade volume between countries i and j when the distance d_{ij} changes to d'_{ij} as Y'_{ij}. The elasticity of average is defined in the same way as Equation (11) except that the

[17] 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.

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

change in trade costs is not uniform:

$$\begin{aligned}\delta_{\boldsymbol{s}}^{EA} &= \frac{\frac{\sum_{i,j} Y'_{ij} - \sum_{i,j} Y_{ij}}{\sum_{i,j} Y_{ij}}}{\sum_{i,j} \frac{d'_{ij}-d_{ij}}{d_{ij}}/N^2} = \frac{\frac{\sum_{i,j}\left(Y'_{ij}-Y_{ij}\right)}{\sum_{i,j} Y_{ij}}}{\sum_{i,j} \frac{d'_{ij}-d_{ij}}{d_{ij}}/N^2} = \sum_{i,j} \frac{\frac{Y'_{ij}-Y_{ij}}{Y_{ij}} \frac{Y_{ij}}{\sum_{i,j} Y_{ij}}}{\sum_{i,j} \frac{d'_{ij}-d_{ij}}{d_{ij}}/N^2} \\ &= \sum_{i,j} \left(\frac{\frac{Y'_{ij}-Y_{ij}}{Y_{ij}}}{\frac{d'_{ij}-d_{ij}}{d_{ij}}} \frac{Y_{ij}}{\sum_{i,j} Y_{ij}} \frac{\frac{d'_{ij}-d_{ij}}{d_{ij}}}{\sum_{i,j} \frac{d'_{ij}-d_{ij}}{d_{ij}}/N^2} \right) \\ &= \sum_{i,j} \left(\delta_{ij} \frac{Y_{ij}}{\sum_{i,j} Y_{ij}} s_{ij} \right) = \frac{\sum_{i,j} \delta_{ij} Y_{ij} s_{ij}}{N^2} / \frac{\sum_{i,j} Y_{ij}}{N^2}, \end{aligned} \tag{14}$$

where $s_{ij} \equiv \frac{d'_{ij}-d_{ij}}{d_{ij}} / \left(\sum_{i,j} \frac{d'_{ij}-d_{ij}}{d_{ij}}/N^2 \right)$ 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 $\{d_{ij}\}_{i,j\in\Omega}$. The right-hand-side of Equation (14) is the empirical counterpart of $\frac{\frac{1}{N^2}\sum_{i,j}\mathbb{E}\left(s_{ij}\theta^*_{ij}Y\right)}{\mathbb{E}(Y)}$. When $\boldsymbol{s} = \{1\}_{i,j\in\Omega}$, this expression becomes $\frac{\mathbb{E}[Y\theta^*]}{\mathbb{E}[Y]}$, corresponding to the approximate probability limit of the regular (unweighted) PPML estimator.

Note that $\boldsymbol{s}$ is not random and instead represents the counterfactual of interest. Naturally, in the presence of heterogeneity, the elasticity of average (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 ($d'_{ij} \neq d_{ij}$ if $i = 1, j = 2$ and $d'_{ij} = d_{ij}$ if $i \neq 1, j \neq 2$), the true elasticity of such a treatment will be equal to δ_{12}.

Now, for a given treatment vector of interest $\boldsymbol{s}$, a policymaker can properly estimate the elasticity of average 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 δ_{ij} irrelevant to a given treatment would contaminate the true elasticity of average.[19] More generally, if some observations receive larger treatment than others,[20] 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 others increase.

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

[20] A simple example would be if the trade costs between Germany and Italy decreased by 99% while trade costs between France and Poland decreased by only 0.01%.

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 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 $w_{ij} = \frac{d'_{ij} - d_{ij}}{d_{ij}}$. Finally, there exists such a treatment vector such that the average elasticity and the elasticity of average are identical; it happens when larger observations (country pairs) receive proportionally lower treatment and thus two weights (size and treatment) cancel out.

Interestingly, average elasticity δ^{AE} does not depend on the treatment vector, because it is a simple arithmetic average of heterogenous elasticities δ_{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 elasticity of average).

It is important to note that in the absence of heterogeneity researchers could estimate the elasticity first and then calculate $\delta^{EA}_{\boldsymbol{s}}$. To see this, assume $\delta_{ij} = \delta$ to be the same for all country pairs i, j. Then it holds that $\delta^{EA}_{\boldsymbol{s}} = \delta \frac{\sum_{i,j} Y_{ij} s_{ij}}{\sum_{i,j} Y_{ij}}$. For any vector of treatment intensity $\boldsymbol{s}$, we can first estimate δ and then compute $\delta^{EA}_{\boldsymbol{s}}$ by multiplying the estimated δ with a factor $\frac{\sum_{i,j} Y_{ij} s_{ij}}{\sum_{i,j} Y_{ij}}$.[21] Such an exercise, however, is not feasible when δ_{ij} is heterogeneous across countries. It is impossible to start with either δ^{EA} (estimable using PPML) or δ^{AE} (estimable using OLS) and obtain $\delta^{EA}_{\boldsymbol{s}}$ by adjusting δ^{EA} or δ^{AE} 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.

Next, we formally show that the WPPML estimator approximates $\frac{\frac{1}{N^2}\sum_{i,j}\mathbb{E}\left(s_{ij}\theta^*_{ij}Y\right)}{\mathbb{E}(Y)}$ when N is large. The WPPML estimator can be written as

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

Following the discussion for Model (1), because we can find consistent estimators for the fixed model parameter vector β^*, we replace $\widehat{\beta}$ with β^* to simplify the discussion. The estimator

[21] 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.

$\widehat{\theta}$ satisfies the following sample moment condition:

$$\sum_{i,j} w_{ij} \left[Y_{ij} - \boldsymbol{X}_{ij}^{\beta^*} Z_{ij}^{\widehat{\theta}} \right] \dot{Z}_{ij} = 0.$$

For any $\theta \in \Theta$, the Kronecker's lemma with the Kolmogorov convergence criteria implies that

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

where $\overset{a.s.}{\to}$ denotes convergence almost surely. To make the above convergence non-trivial, we need the weight vector to satisfy that $\frac{1}{N^2} \sum_{i,j} w_{ij}^2 \to \eta$ for some constant $\eta > 0$. For compact parameter space Θ, the uniform law of large numbers provides that

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

Thus, it holds that $\widehat{\theta} \overset{p}{\to} \theta_0$, where θ_0 satisfies that

$$\sum_{i,j} w_{ij} \mathbb{E} \left(\left[Y_{ij} - \boldsymbol{X}_{ij}^{\beta^*} Z_{ij}^{\theta_0} \right] \dot{Z}_{ij} \right) = 0.$$

Let θ_{ij}^* denote θ^* for individual i, j. We have that

$$\begin{aligned}
0 =& \sum_{i,j} w_{ij} \mathbb{E} \left(\left[Y_{ij} - \boldsymbol{X}_{ij}^{\beta^*} Z_{ij}^{\theta_0} \right] \dot{Z}_{ij} \right) \\
=& \sum_{i,j} w_{ij} \mathbb{E} \left(\left[\boldsymbol{X}_{ij}^{\beta^*} Z_{ij}^{\theta_{ij}^*} + \varepsilon_{ij} - \boldsymbol{X}_{ij}^{\beta^*} Z_{ij}^{\theta_0} \right] \dot{Z}_{ij} \right) \\
=& \sum_{i,j} w_{ij} \left[\mathbb{E} \left(\left[\boldsymbol{X}_{ij}^{\beta^*} Z_{ij}^{\theta_{ij}^*} - \boldsymbol{X}_{ij}^{\beta^*} Z_{ij}^{\theta_0} \right] \dot{Z}_{ij} \right) + \mathbb{E} \left(\varepsilon_{ij} \dot{Z}_{ij} \right) \right] \\
=& \sum_{i,j} w_{ij} \mathbb{E} \left(\left[\boldsymbol{X}_{ij}^{\beta^*} Z_{ij}^{\theta_{ij}^*} - \boldsymbol{X}_{ij}^{\beta^*} Z_{ij}^{\theta_0} \right] \dot{Z}_{ij} \right).
\end{aligned}$$

By Model (1), we have that for any θ_{ij}^*,

$$\sum_{i,j} w_{ij} \mathbb{E} \left(\left[\boldsymbol{X}_{ij}^{\beta^*} Z_{ij}^{\theta_{ij}^*} - \boldsymbol{X}_{ij}^{\beta^*} Z_{ij}^{\theta_{ij}^*} \right] Z_{ij} \right) = 0.$$

Define

$$\phi_{ij}(\theta) \equiv w_{ij} \mathbb{E} \left(\left[\boldsymbol{X}_{ij}^{\beta^*} Z_{ij}^{\theta_{ij}^*} - \boldsymbol{X}_{ij}^{\beta^*} Z_{ij}^{\theta} \right] Z_{ij} \right).$$

Analogous to the discussion in Section 2 for the PPML estimator, we have that

$$\begin{aligned} 0 = \sum_{i,j} \phi_{ij}\left(\theta_0\right) &= \sum_{i,j} \phi_{ij}\left(\theta_{ij}^*\right) + \sum_{i,j} \frac{d}{d\theta}\phi_{ij}\left(\theta_{ij}^\star\right)\left(\theta_0 - \theta_{ij}^*\right) \\ &= \sum_{i,j} \frac{d}{d\theta}\phi_{ij}\left(\theta_{ij}^\star\right)\left(\theta_0 - \theta_{ij}^*\right), \end{aligned}$$

which implies that

$$\theta_0 = \frac{\sum_{i,j} \frac{d}{d\theta}\phi_{ij}\left(\theta_{ij}^\star\right)\theta_{ij}^*}{\sum_{i,j} \frac{d}{d\theta}\phi_{ij}\left(\theta_{ij}^\star\right)}.$$

A similar calculation would show that

$$\begin{aligned} \frac{d}{d\theta}\phi_{ij}\left(\theta_{ij}^\star\right) &= -\sum_{i,j} w_{ij}\mathbb{E}\left(\boldsymbol{X}_{ij}^{\beta^*} Z_{ij}^{\theta_{ij}^\star}\dot{Z}_{ij}^2\right) \sim -\sum_{i,j} w_{ij}\mathbb{E}\left(\boldsymbol{X}_{ij}^{\beta^*} Z_{ij}^{\theta_{ij}^*}\dot{Z}_{ij}^2\right) \\ &= -\sum_{i,j} w_{ij}\mathbb{E}\left(\left[\boldsymbol{X}_{ij}^{\beta^*} Z_{ij}^{\theta_{ij}^*} + \varepsilon_{ij}\right]\dot{Z}_{ij}^2\right) = -\sum_{i,j} w_{ij}\mathbb{E}\left(Y_{ij}\dot{Z}_{ij}^2\right) \\ &= -\sum_{i,j} w_{ij}\mathbb{E}\left(Y\dot{Z}^2\right) = -\mathbb{E}\left(Y\dot{Z}^2\right)\sum_{i,j} w_{ij}, \end{aligned}$$

where the third equality follows from $\mathbb{E}\left(\varepsilon \mid \boldsymbol{X}, Z, \theta^*\right) = 0$ and the second to last equality follows from that the unconditional distribution of Y_{ij} is the same across i, j. Similarly, we can obtain that

$$\begin{aligned} \sum_{i,j} \frac{d}{d\theta}\phi_{ij}\left(\theta_{ij}^\star\right)\theta_{ij}^* &\sim -\sum_{i,j} w_{ij}\theta_{ij}^*\mathbb{E}\left(\boldsymbol{X}_{ij}^{\beta^*} Z_{ij}^{\theta_{ij}^*}\dot{Z}_{ij}^2\right) \\ &= -\sum_{i,j} w_{ij}\mathbb{E}\left(\theta_{ij}^* Y\dot{Z}^2\right). \end{aligned}$$

After ignoring the variability of $\dot{Z}$ because of its small magnitude compared to Y, we obtain the following expression:

$$\theta_0 \sim \frac{\sum_{i,j} w_{ij}\mathbb{E}\left(\theta_{ij}^* Y\right)}{\mathbb{E}\left(Y\right)\sum_{i,j} w_{ij}} = \frac{\sum_{i,j} \frac{w_{ij}}{\sum_{i,j} w_{ij}}\mathbb{E}\left(\theta_{ij}^* Y\right)}{\mathbb{E}\left(Y\right)} = \frac{\frac{1}{N^2}\sum_{i,j}\mathbb{E}\left(s_{ij}\theta_{ij}^* Y\right)}{\mathbb{E}\left(Y\right)},$$

where the first equality holds by the definitions of w_{ij} and s_{ij}. Hence, the approximate probability limit of the WPPML estimator $\widehat{\theta}$ is the population analogue of $\delta_{\boldsymbol{s}}^{EA}$ in Equation (14).

Empirical Application

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 a recent paper by Fontagné et al. (2022) with our WPPML estimates. Fontagné et al. (2022) utilizes highly disaggregated data for 5,050 HS-6 product categories, covering 189 exporters and 152 destinations over the period from 2001 to 2016. To estimate the tariff elasticity, they rely on the following standard structural gravity formulation:

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

where i, j, k, t denote countries of origin and destination, industry, and year. $\theta_{ik,t}$ and $\theta_{jk,t}$ are origin-industry-time and destination-industry-time fixed effects. β_k is the industry-specific tariff elasticity, our coefficient of interest; and $\tau_{ijk,t}$ is the *ad valorem* tariff imposed by country 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 industry-level trade elasticity can be recovered accordingly as $\varepsilon_k = 1 + \beta_k$.

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:[22,23]

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

We illustrate the difference between baseline elasticities estimated by Fontagné et al. (2022) and *realized* elasticities in Figures 4a and 4b [24]. One can see that when elasticities are large (in absolute terms), the two estimators generate comparable results. However, for

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

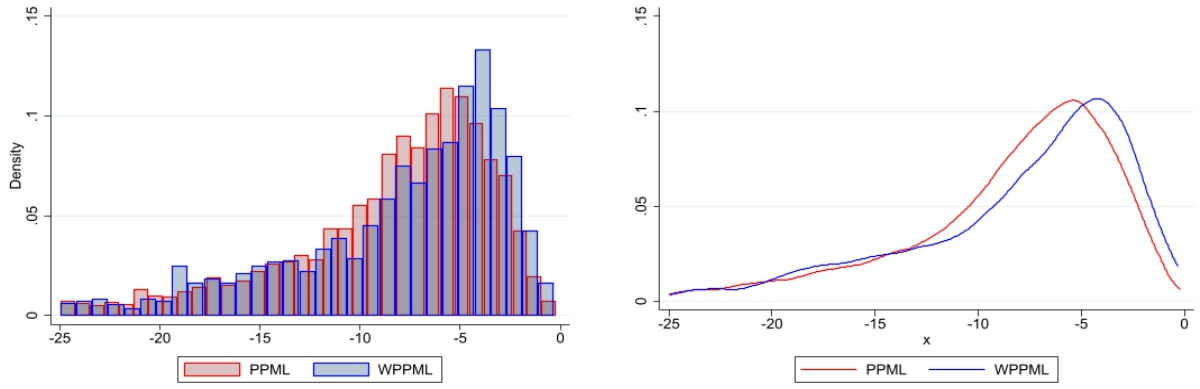
[23] Note that even though we use changes in tariffs as weights, our estimation is conceptually different from the three-way fixed effects estimation, which relies on country-pair-industry fixed effects $\theta_{ij,k}$ and uses variation in tariffs over time. While WPPML also excludes observations without 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.

[24] The insignificant estimated ε_k's are replaced with 1.

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).

Figure 4. Empirical Distribution of PPML and WPPML Trade Elasticities

(a) Histogram (b) Kernel Density



Note: This figure presents the empirical distribution of PPML and WPPML trade elasticities calculated for HS-6 products with $\varepsilon_k \in (-25, 0)$.

Table 6 presents the estimated trade elasticities separately for each HS chapter 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 show 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.

Table 6
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.

6 Conclusion

In this paper, we argue that the interpretation of PPML regression, the most popular method for estimating the gravity equation, is different from the results generated by the previously dominant OLS regression. We show that the former can be interpreted as the elasticity of average and the latter as the average elasticity. When choosing between the OLS and PPML estimates, we recommend making a decision based on the desired interpretation of the results or providing the results obtained with both methods.

We find evidence of the existence of heterogeneity across different country pairs in trade data. Furthermore, we decompose the differences between the OLS and PPML estimates into two channels by comparing the OLS, WLS, GPML, and PPML estimators: one is the previously extensively studied heteroskedasticity bias, and the other is the different interpretations of the two estimators, which we define as the bias in the estimation of the elasticity of average. Our results show that while the bias associated with heteroskedasticity exists, it is 7 to 12 times smaller than that associated with country-pair heterogeneity and has the opposite sign compared to the previous findings. As the presence of heteroskedasticity bias may still be a problem, to obtain the estimates of the average elasticity, we recommend applying GPML.

We use Monte Carlo simulations to show that when distance elasticity is systematically heterogeneous between country pairs, OLS and GPML cannot be used to estimate the elasticity of average, while WLS and PPML are the appropriate methods. We further show that the elasticity of average also depends on the observation-specific treatment intensity and the (ordinary) PPML elasticity estimates can be interpreted as the elasticity of average in the case of a uniform treatment. For the cases with heterogeneous treatment intensities, we introduce a new WPPML estimator, which estimates the corresponding elasticity of average for a given vector of treatment intensities. The weights are proportionate to the treatment intensities.

While our focus in this paper is on the properties of OLS, GPML, and PPML in estimating the gravity equation, our findings are much more general. Any standard log-log regression estimated by PPML will have a different interpretation, which, depending on the research question, may be preferable to the interpretation of a standard OLS regression.

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.

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.

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).

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.

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.