CGPA WORKING PAPER
2025-12

A Generalized Poisson-Pseudo Maximum Likelihood Estimator

Ohyun Kwon
Jangsu Yoon
Yoto V. Yotov

DREXEL UNIVERSITY
Center for
Global Policy Analysis
LeBow College of Business

A Generalized Poisson-Pseudo Maximum Likelihood Estimator*

Ohyun Kwon   Jangsu Yoon   Yoto V. Yotov

Abstract

We examine the Constant Variance to Mean Ratio (CVMR) assumption – a key condition to make PPML an efficient estimator – and propose Generalized Poisson-Pseudo Maximum Likelihood (G-PPML) as a complementary estimator. We estimate the conditional variance of the dependent variable using an iterated GMM, thereby providing a specification test for the CVMR assumption. The proposed G-PPML estimator, which capitalizes on conditional variance estimates, is more efficient than existing PML estimators. After establishing the asymptotic properties of the G-PPML estimator, we verify that it performs well under fairly general assumptions about the conditional variance. Our empirical application to trade flows data demonstrates that the CVMR assumption is satisfied in most but not all cases. The standard errors of G-PPML are approximately 20% smaller than those of PPML, demonstrating its improved estimation efficiency.

JEL classification: *C13, C50, F10*

Keywords: *Poisson-Pseudo Maximum Likelihood, Iterated GMM, Gravity Models*

*Acknowledgements will be added later. Authors have no conflict of interest to declare. Ohyun Kwon (Corresponding Author): School of Economics, Drexel University, ok85@drexel.edu.; Jangsu Yoon: Department of Economics, University of Wisconsin-Milwaukee, yoon22@uwm.edu.; Yoto V. Yotov: School of Economics, Drexel University; ifo Institute; CESifo, yotov@drexel.edu.

1 Introduction

Owing to the seminal work of Santos Silva and Tenreyro (2006), the Poisson Pseudo Maximum Likelihood (PPML) estimator has established itself as the leading estimator for trade gravity regressions. In addition to handling the issue of heteroskedasticity, the multiplicative form of PPML makes it appropriate for utilizing the information that is contained in zero trade flows, which often take up a significant fraction of the data (both at the aggregate and, especially, at the product level). Santos Silva and Tenreyro also clarify that the PPML estimator does not require the data to follow a Poisson distribution, and that although it is a count data estimator, PPML is appropriate for regressions with continuous data even in the presence of a mass point at zero.[1]

Since its introduction to trade, PPML has attracted significant attention, and a number of papers have added to the list of its attractive properties. For example, Fernández-Val and Weidner (2016) show that PPML estimations with two-way fixed effects do not suffer from the incidental parameter problem (IPP). Weidner and Zylkin (2021) show that PPML with three-way fixed effects is consistent albeit with asymptotic bias, for which they provide a correction. Correia et al. (2020) introduce the `ppmlhdfe` STATA command, which simultaneously addresses the issues of computational speed and convergence with PPML. Fally (2015) demonstrates that the PPML is perfectly consistent with a wide class of structural gravity models. Capitalizing on this property, Anderson et al. (2018) demonstrate how PPML can be used to obtain not only partial equilibrium gravity estimates but also benchmark general equilibrium effects.

Despite all the attractive properties of PPML, some researchers are skeptical about its validity and use as the "workhorse" gravity estimator. As discussed in Head and Mayer (2014), the main argument against the use of PPML is that the relationship between the variance of bilateral exports and their expected value may not be consistent with the assumption of the Poisson distribution –

[1] We refer the reader to Santos Silva and Tenreyro (2022) and to the dedicated PPML web site https://personal.lse.ac.uk/tenreyro/lgw.html for many helpful tips and relevant information about PPML.

an idea later rejected by Weidner and Zylkin (2021).[2] Based on Monte Carlo simulations, Head and Mayer (2014) challenge the PPML's Constant Variance to Mean Ratio (CVMR) assumption and recommend that researchers also estimate gravity with other estimators, including OLS in logs and Gamma-PML, for robustness checks.[3]

Against this backdrop, we make two contributions. First, we propose a method to examine the CVMR assumption by estimating the conditional variance of the model using an iterated Generalized Method of Moments (iGMM).[4] The suggested iGMM estimator provides a valid specification test for the underlying assumptions on the conditional variance of broad class of PML estimators, and we complement it with a Stata command - `cvmrtest`. As demonstrated by Hansen and Lee (2021), iGMM is robust to mild misspecification and provides stable estimates regardless of the initial guess on parameters, making it ideal for this purpose.[5] Our Monte Carlo analysis shows that the proposed iGMM method is consistent and more accurate than all existing methods to recover the conditional variance across a wide range of parameter values.

Second, we propose the *Generalized* PPML (G-PPML) estimator based on the consistent estimates of the conditional variance. The G-PPML estimator, optimally weighted by conditional variance estimates, is generally more efficient than existing PML estimators.[6] We verify the asymptotic properties of the G-PPML estimator and confirm that G-PPML is also immune to the

[2]Weidner and Zylkin (2021) show that PPML is a consistent estimator for all scenarios considered in Head and Mayer (2014).

[3]Throughout the paper, OLS refers to ordinary least squares estimation in logarithmic form, while the PML estimators are applied directly to the dependent variable without requiring a logarithmic transformation.

[4]The iterative nature of iGMM lies in its repeated updating of parameter estimates and the weighting matrix until convergence.

[5]Hansen and Lee (2021) demonstrates that iGMM estimates a pseudo-true parameter, which, as our Monte Carlo analysis shows, serves as a valuable benchmark for evaluating PPML efficiency. Furthermore, this pseudo-true parameter facilitates G-PPML, which offers improved estimation efficiency over PPML when the CVMR assumption is violated. We therefore argue that iGMM remains robust to mild misspecification in this context.

[6]We labeled our estimator G-PPML because, under the assumption of pseudo maximum likelihood, we find the relationship between G-PPML and PPML to resemble the relationship between the Generalized Least Squares (GLS) estimator and OLS. Alternatively, G-PPML can be viewed as a "weighted" PPML estimator. We thank João Santos Silva and Seojeong Lee for helpful discussions and suggestions on the name of the estimator.

IPP under the correct specification of the conditional variance.[7] Under a mild assumption on the class of PML estimators, G-PPML is consistent, free from the IPP issue, exhibits improved estimation efficiency (as reflected in lower standard errors), and is not subject to concerns regarding potential misspecification of the heteroskedastic structure (Head and Mayer, 2014). Thus, G-PPML complements PPML when the CVMR assumption is rejected from our iGMM-based specification test. We complement G-PPML with a Stata command - `gppmlhdfe`.

We conduct a series of Monte Carlo experiments to demonstrate the performance of iGMM and G-PPML. First, we show that the iGMM estimator can estimate the true conditional variance parameters consistently across a wide range of parameter values. Second, we use the estimates of the conditional variance parameter to estimate the gravity equation with the simulated data. The Monte Carlo results confirm that: (i) G-PPML is more efficient when the true conditional variance parameter deviates from the distributional assumption of other PML estimators; (ii) G-PPML encompasses other PML estimators as special cases when the DGP conforms to the distributional assumption embedded in each PML; (iii) the efficiency gain from using G-PPML is more pronounced when the level of noise in the data is greater; (iv) the OLS estimator, which can be unbiased in a knife-edge case, is considerably biased in a general parametric setting; and (v) G-PPML can account for potential misspecification in the error term structure.

To demonstrate the practical importance of our methods, we estimate the gravity equation on a standard set of covariates (e.g., distance, trade agreements, etc.) using trade data for 105 sectors. Our estimates of the conditional variance parameter reveal four salient patterns: (i) they are all strictly positive; (ii) they are clustered around one, suggesting that, in many cases, the PPML estimator should perform quite well; (iii) all estimates of the conditional variance parameter are less than two, implying that the Gamma-PML estimator may not be appropriate

[7] In related work, Jochmans (2017) and Yang and Zhang (2023) propose estimators without the IPP using an alternative GMM-based approach. This approach does not require any assumptions on conditional variances and enables a more computationally efficient estimator, as its moment condition is free from high-dimensional fixed effects. Our paper differs from this work by following the traditional PML approach and providing a formal test for the CVMR assumption.

for gravity estimations; and (iv) most important for our purposes, we also observe deviations of the conditional variance parameter from the CVMR assumption (e.g., $\lambda = 1$), implying that there can be an estimation efficiency gain associated with G-PPML.[8]

Comparisons between the PPML and G-PPML gravity estimates reveal the following. We find that many of the PPML and G-PPML gravity estimates are very similar to each other owing to the fact that most λ estimates are clustered around 1. However, we also observe that a significant fraction of the G-PPML estimates are different from the corresponding PPML estimates. Moreover, consistent with our theory, the further away the estimates of the conditional variance parameter are from 1, the larger the difference is between the PPML and G-PPML estimates. Finally, comparisons of standard errors and z-statistics reveal that, in general, G-PPML estimation is more efficient and enables more efficient hypothesis testing. These findings suggest that G-PPML can be useful for researchers in the gravity literature when our CVMR test rejects the null hypothesis that PPML is the most efficient estimator.

The rest of the paper is organized as follows. Section 2 offers a formal description of the main challenge to the PPML estimator, introduces our G-PPML estimator, and establishes its consistency and asymptotic properties. Section 3 implements Monte Carlo simulations that showcase the main properties of our estimator. Section 4 demonstrates the validity and importance of our methods with an application to real sectoral trade data. Finally, Section 5 offers concluding remarks and points to possible directions for further research.

2 A Generalized PPML Estimator

The objective of this section is to develop an estimator that inherits the key desired properties of PPML while being more conservative about the conditional variance structure of the trade volume. To this end, we synthesize the key insights of PPML and propose an intuitive modifi-

[8] In the application, the standard errors of G-PPML are roughly 20% smaller than those of PPML, and the point estimate differences between PPML and G-PPML can reach 40% when the CVMR assumption is not satisified.

cation to improve the efficiency of the PPML estimator. Capitalizing on a recent development in the econometrics literature (Hansen and Lee, 2021), we implement an iGMM estimator that estimates the conditional variance of the dependent variable in the gravity equation. The next stage of estimation solves for the coefficients with the weighted moment conditions, which are informed by the conditional variance estimates, to achieve an improvement in estimation efficiency. Given the prominence of PPML for estimating trade gravity equations, we specify the following econometric model as our departure point:

$$y_{ijt} = \exp\left(x'_{ijt}\beta + \gamma_{it} + \eta_{jt}\right) + \epsilon_{ijt}, \tag{1}$$

where $x_{ijt} \in \mathbb{R}^k$ is a vector of regressors that capture trade costs, and γ_{it} and η_{jt} are exporter- and importer-level fixed effects, which capture the importers and exporters' observed and unobserved characteristics that may vary across different periods indexed by t. We conventionally assume that the conditional mean of y_{ijt} follows $\mathbb{E}\left[y_{ijt}|x_{ijt}, \gamma_{it}, \eta_{jt}\right] = \mu_{ijt} \equiv \exp\left(x'_{ijt}\beta + \gamma_{it} + \eta_{jt}\right)$. The measurement error ϵ_{ijt} satisfies $\mathbb{E}\left[\epsilon_{ijt}|x_{ijt}, \gamma_{it}, \eta_{jt}\right] = 0$.[9] The model is consistent with the broad class of models considered in Head and Mayer (2014) since the error term ϵ_{ijt} is potentially heteroskedastic. The moment conditions of the PPML method are constructed based on the assumption of a Poisson distribution: the conditional variance of y_{ijt} is proportional to the conditional mean of y_{ijt} (i.e., CVMR). Specifically, given a general form of the conditional variance of y_{ijt},

$$\text{Var}\left(y_{ijt}|x_{ijt}, \gamma_{it}, \eta_{jt}\right) = \text{Var}\left(\epsilon_{ijt}|x_{ijt}, \gamma_{it}, \eta_{jt}\right) = h \cdot \mu_{ijt}^{\lambda}, \tag{2}$$

PPML's distributional assumption assumes $\lambda = 1$ and $h > 0$. The specification also includes all the well-known PML estimators (i.e., Gamma-PML ($\lambda = 2$) and Gaussian PML ($\lambda = 0$)) as special cases.[10] From equation (2), it is evident that the PPML assumption is equivalent to assuming a constant variance-to-mean ratio. As will become clear in Section 3, the special case

[9] Santos Silva and Tenreyro (2022) and other gravity model papers specify the model by $y_{ijt} = \exp\left(x'_{ijt}\beta + \gamma_{it} + \eta_{jt}\right)\varepsilon_{ijt}$. The model assumes a multiplicative error ε_{ijt} to prevent a negative value of y_{ijt}. Since our estimation strategy relies on the conditional moment $\mathbb{E}\left[y_{ijt} - \mu_{ijt}|x_{ijt}, \gamma_{it}, \eta_{jt}\right] = 0$, the gravity model specification with an additive error ϵ_{ijt} is theoretically equivalent to the conventional model.

[10] We do not consider Negative Binomial PML, as it is generally not recommended for empirical application due to unit-sensitivity. See Head and Mayer (2014) for detailed discussions.

where $\lambda = 2$ is also consistent with the underlying assumption of OLS when the error term appears multiplicatively in equation (1). As such, our generic representation of the conditional variance form in equation (2) encompasses all of the PML-based conditional variance forms adopted in the relevant literature.[11] Notably, since we do not impose any assumption on the value of λ, our approach is immune to the criticism raised by Head and Mayer (2014).

To demonstrate how the information about λ can be crucial for G-PPML, it is informative to spell out the first order conditions (FOCs) with respect to the parameters to be estimated. Suppose that a researcher observes a random sample $\{y_{ijt}, x_{ijt}\}$, $i, j = 1, \ldots, N$ and $t = 1, \ldots, T$ to estimate the gravity equation (1). Considering the specified conditional variance in equation (2), the corresponding pseudo-likelihood function is

$$\mathcal{L}\left(\beta, \gamma_{it}, \eta_{jt}\right) = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\left(y_{ijt}\int \mu_{ijt}^{-\lambda} d\mu_{ijt} - \int \mu_{ijt}^{1-\lambda} d\mu_{ijt}\right), \tag{3}$$

which reduces to the PPML's objective function when $\lambda = 1$. The FOCs considering heteroskedasticity in equation (2) are

$$\tilde{\beta} : \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T} x_{ijt}\left(y_{ijt} - \tilde{\mu}_{ijt}\right)\tilde{\mu}_{ijt}^{1-\lambda} = 0,$$

$$\tilde{\gamma}_{it} : \sum_{j=1}^{N}\left(y_{ijt} - \tilde{\mu}_{ijt}\right)\tilde{\mu}_{ijt}^{1-\lambda} = 0,$$

$$\tilde{\eta}_{jt} : \sum_{i=1}^{N}\left(y_{ijt} - \tilde{\mu}_{ijt}\right)\tilde{\mu}_{ijt}^{1-\lambda} = 0,$$

where $\tilde{\mu}_{ijt} = \exp\left(x_{ijt}'\tilde{\beta} + \tilde{\gamma}_{it} + \tilde{\eta}_{jt}\right)$, $i, j = 1, \ldots, N$, and $t = 1, \ldots, T$. Appendix A.5 of Weidner and Zylkin (2021) shows that estimation efficiency can be improved with the knowledge of true λ. However, this is not immediately feasible since the researcher is agnostic about the true value of λ. Therefore, our first and primary goal is to propose a valid estimator for the exponent λ. The estimated λ naturally leads to a plug-in estimator of β by replacing λ in FOCs with its

[11] To our knowledge, the only two exceptions are Santos Silva and Tenreyro (2011) and Weidner and Zylkin (2021), as PPML is robust to the misspecified conditional variance. Santos Silva and Tenreyro (2011) obtain PPML coefficients under potential misspecification and Weidner and Zylkin (2021) obtain robust standard errors when the conditional variance does not necessarily follow equation (2). We still posit that it is mild to assume the conditional variance form in equation (2), as it generalizes the assumptions widely used in the literature to date. Moreover, our specification of the conditional variance can be more flexible, and the suggested theory can be applied to any other specification with a finite-dimensional parameter. Further discussions are provided in Section 2.4.

consistent estimator $\hat{\lambda}$. This provides a feasible routine to efficiently estimate gravity equations with two-way fixed effects.

2.1 Estimation of the Conditional Variance

The existing literature (e.g., Santos Silva and Tenreyro (2006) and Head and Mayer (2014)) suggests two methods to estimate the parameter $\theta = (h, \lambda) \in \Theta$ in equation (2), where Θ is the parameter space of θ. Santos Silva and Tenreyro (2006)'s approach linearly approximates the nonlinear conditional variance at $\lambda = 1$, while Head and Mayer (2014) follow Manning and Mullahy (2001) to log-linearize the conditional variance expression. Our proposed iGMM estimator of λ preserves the nonlinearity of the conditional variance function, and it outperforms the nonlinear least-squares (NLLS) estimator. To this end, we assume the following regularity conditions to estimate the conditional variance.

Assumption 2.1. *(Regularity Conditions)*

1. *The dependent variable $y_{ij} = (y_{ij1}, \ldots, y_{ijT})'$ is i.i.d. across i and j conditional on $x = (x_{ijt})$, $\gamma = (\gamma_{it})$, and $\eta = (\eta_{jt})$ for $i, j = 1, \ldots, N$ and $t = 1, \ldots, T$.*
2. *The support of $(x_{ijt}, \gamma_{it}, \eta_{jt})$ is compact, and $\mathbb{E}\left[y_{ijt}^{8+\nu}|x_{ijt}, \gamma_{it}, \eta_{jt}\right]$ is uniformly bounded over i, j, t for some $\nu > 0$.*
3. *The parameter space Θ is compact, and $W \equiv \mathbb{E}\left[x_{ijt}x_{ijt}'\left(\epsilon_{ijt}^2 - h \cdot \mu_{ijt}^{\lambda}\right)^2\right]$ is positive definite uniformly over Θ.*

Since $\text{Var}\left(\epsilon_{ijt}|x_{ijt}, \gamma_{it}, \eta_{jt}\right) = \mathbb{E}\left[\epsilon_{ijt}^2|x_{ijt}, \gamma_{it}, \eta_{jt}\right] = h \cdot \mu_{ijt}^{\lambda}$, equation (2) generates a conditional moment

$$\mathbb{E}\left[\epsilon_{ijt}^2 - h \cdot \mu_{ijt}^{\lambda}\middle|x_{ijt}, \gamma_{it}, \eta_{jt}\right] = 0, \tag{4}$$

and the conditional moment identifies θ if $\bar{\theta} \equiv (\bar{h}, \bar{\lambda}) \neq \theta$ on Θ implies $\bar{h} \cdot \mu_{ijt}^{\bar{\lambda}} \neq h \cdot \mu_{ijt}^{\lambda}$ for almost all μ_{ijt}. As suggested by Section 2.2.2 of Newey and McFadden (1994), one primitive sufficient condition for identification is that $\mu_{ijt} = \exp\left(x_{ijt}'\beta + \gamma_{it} + \eta_{jt}\right)$ has positive density on any interval of $\mathbb{R}^+$ or that μ_{ijt} has positive probability mass for more than three points. Thus, the NLLS estimator is theoretically valid if there are preliminary estimates for ϵ_{ijt}^2 and μ_{ijt}. For example, let $\hat{\mu}_{ijt}^{PPML} = \exp\left(x'_{ijt}\hat{\beta}^{PPML} + \hat{\gamma}_{it}^{PPML} + \hat{\eta}_{jt}^{PPML}\right)$ denote the fitted value of y_{ijt} using the PPML estimator $\left(\hat{\beta}^{PPML}, \hat{\gamma}_{it}^{PPML}, \hat{\eta}_{jt}^{PPML}\right)$. Then the NLLS estimator takes a regression of $\left(y_{ijt} - \hat{\mu}_{ijt}^{PPML}\right)^2$ with respect to the nonlinear conditional mean function $h \cdot \hat{\mu}_{ijt}^{\lambda}$. Unfortunately, our simulation exercise finds that the NLLS method does not perform well in practice, even without fixed effects.[12]

Instead, we propose an iGMM estimator to obtain the conditional variance parameters (h, λ).[13] The conditional moment of equation (4) implies k-dimensional unconditional moments,

$$\mathbb{E}\left[x_{ijt}\left(\epsilon_{ijt}^2 - h \cdot \mu_{ijt}^{\lambda}\right)\right] = 0, \tag{5}$$

and a necessary condition of identification is $k \geq 2$. Note that the GMM identification condition is that equation (5) has a unique solution at $\theta = (h, \lambda)$. Following the idea of Newey and McFadden (1994), we verify the local identification condition since demonstrating global identification is challenging for nonlinear models. Provided that the moment condition is (5), the conditional variance parameter θ is locally identified if $Q \equiv \mathbb{E}\left[x_{ijt}\mu_{ijt}^{\lambda},\ h \cdot x_{ijt}\mu_{ijt}^{\lambda}\log\left(\mu_{ijt}\right)\right]$ has full column rank in the neighborhood of θ. That is, the rank condition holds unless $\log\left(\mu_{ijt}\right)$ remains constant.[14] Since the covariates x_{ijt} generally exceed two dimensions, we focus on the over-identified case of $k > 2$.[15] As demonstrated by Hansen and Lee (2021), the iGMM method is robust to possible moment misspecification, and the estimates do not fluctuate depending on

[12]For all simulation exercises with different DGPs, the NLLS estimation with the STATA package `nl` frequently failed to capture the true λ, while the suggested iGMM method showed stable performance.

[13]If the conditional variance follows equation (2), taking a PPML estimation with $\hat{\epsilon}_{ijt}^2$ as a dependent variable can also provide conditional variance estimates. In the current paper, however, one of our primary contributions is to provide a valid specification test for the underlying conditional variance assumptions of PML estimators. We can further generalize the specification in equation (2) by assuming more parameters if the practitioner wants to test different conditional variance assumptions. As long as the conditional variance is correctly specified, the parametrization does not influence the asymptotic property of the G-PPML.

[14]A similar primitive identification condition is also found in Example 1.3 of Newey and McFadden (1994).

[15]The unconditional moment (5) is not the only moment condition derived by the conditional mean assumption of equation (4). For example, following Newey (1990), the use of optimal instruments may improve the performance of the proposed iGMM estimator. Still, estimating optimal instruments as a function of x_{ijt}, γ_{it}, and η_{jt} can be computationally demanding in practice, as it may require nonparametric estimation with high-dimensional covariates. Efficient estimation of θ is beyond the scope of our paper and will be left for future research.

different initial guesses of parameters.

Define the sample moment and the efficient weight matrix for $\bar{\theta} = \left(\bar{h}, \bar{\lambda}\right) \in \Theta$:

$$\bar{m}_N\left(\bar{\theta}\right) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T} x_{ijt}\left(\hat{\epsilon}_{ijt}^2 - \bar{h}\cdot\hat{\mu}_{ijt}^{\bar{\lambda}}\right)$$

$$\bar{W}_N\left(\bar{\theta}\right) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T} x_{ijt}x_{ijt}'\left(\hat{\epsilon}_{ijt}^2 - \bar{h}\cdot\hat{\mu}_{ijt}^{\bar{\lambda}}\right)^2,$$

where $\hat{\epsilon}_{ijt} = y_{ijt} - \hat{\mu}_{ijt}^{PPML}$ and $\hat{\mu}_{ijt} = \hat{\mu}_{ijt}^{PPML}$ are PPML estimates. Given that $\phi = (h_\phi, \lambda_\phi)$ is the initial guess or preliminary estimates on the parameter value, the GMM criterion function is

$$\bar{J}_N\left(\bar{\theta}, \phi\right) = \bar{m}_N\left(\bar{\theta}\right)' \bar{W}_N^{-1}\left(\phi\right)\bar{m}_N\left(\bar{\theta}\right),$$

For a fixed value of ϕ, the next-step estimator minimizes the GMM sample criterion function. Let $\bar{g}_N\left(\phi\right)$ denote the minimizer of the sample criterion function:

$$\bar{g}_N\left(\phi\right) = \arg\min_{\bar{\theta}\in\Theta} \bar{J}_N\left(\bar{\theta}, \phi\right),$$

where Θ is a closed and bounded subspace of $\mathbb{R}^+ \times \mathbb{R}$. Starting from the initial value $\hat{\theta}_0 = \left(\hat{h}_0, \hat{\lambda}_0\right)$, we define the one-step GMM estimator by $\hat{\theta}_1 = \bar{g}_N\left(\hat{\theta}_0\right)$. Similarly, the s-step GMM estimator is $\hat{\theta}_s = \bar{g}_N\left(\hat{\theta}_{s-1}\right)$. The iGMM estimator for θ is

$$\hat{\theta} = \lim_{s\to\infty}\hat{\theta}_s.$$

The convergence leads to the true parameter θ, or the values that provide the best fit to the conditional variance under mild misspecification. By mild misspecification, we follow the definition of Hansen and Lee (2021): the degree of misspecification is bounded, or more specifically, the magnitude of the population GMM criterion function is bounded. We show the existence of $\hat{\theta}$ and the consistency of the estimator by verifying Theorem 3 of Hansen and Lee (2021).

Proposition 2.1. *Under Assumption 2.1, $\hat{\theta} \xrightarrow{p} \theta$ as $N \to \infty$.*

The proposition implies that the practitioners can recover the conditional variance of y_{ijt} as far as the conditional variance form follows equation (2). Next, we establish the asymptotic normality of $\hat{\theta}$ that can be helpful for testing whether the PPML's assumption on the variance-to-mean ratio is valid. Recall $W = \mathbb{E}\left[x_{ijt}x_{ijt}'\left(\epsilon_{ijt}^2 - h\cdot\mu_{ijt}^{\lambda}\right)^2\right]$ and $Q = \mathbb{E}\left[x_{ijt}\mu_{ijt}^{\lambda},\ h\cdot x_{ijt}\mu_{ijt}^{\lambda}\log\left(\mu_{ijt}\right)\right]$, and the conditional variance of the model is correctly specified (i.e., it satisfies any class of widely used PML estimators). Define $\zeta_{ijt}(\beta) = \epsilon_{ijt}^2 - h \cdot \mu_{ijt}^{\lambda}$. Then, the asymptotic variance of $N\left(\hat{\theta} - \theta\right)$ is $\left(Q'W^{-1}Q\right)^{-1}\left(Q'W^{-1}(W+V)W^{-1}Q\right)\left(Q'W^{-1}Q\right)^{-1}$, where $W + V$ is the asymptotic variance of

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T} x_{ijt}\left(\zeta_{ijt}(\beta) + \zeta_{ijt}'(\beta)'\left(\hat{\beta}^{PPML} - \beta\right)\right).$$

In the proof, we provide a detailed derivation of V, which is generated from the approximation errors of $\hat{\epsilon}_{ijt}$ and $\hat{\mu}_{ijt}$. The asymptotic variance of $\hat{\lambda}$ is the second diagonal element of the asymptotic variance matrix.

Proposition 2.2. *Under Assumption 2.1,*

$$N\left(\hat{\theta} - \theta\right) \xrightarrow{d} \mathcal{N}\left(0, \left(Q'W^{-1}Q\right)^{-1}\left(Q'W^{-1}(W+V)W^{-1}Q\right)\left(Q'W^{-1}Q\right)^{-1}\right).$$

The result of Proposition 2.2 follows the classic efficient GMM estimator's asymptotic distribution except that V presents the approximation error from the first-stage estimator $\hat{\beta}^{PPML}$. Without V, the asymptotic variance is equivalent to $\left(Q'W^{-1}Q\right)^{-1}$, the efficient variance matrix. The result is a special case of Theorem 4 in Hansen and Lee (2021) under the correctly specified conditional variance. The asymptotic normality informs how to construct the confidence interval for the exponent component λ. The Monte Carlo simulations in Section 3 confirm that the suggested method is valid to test the true λ value that fits the conditional variance of the gravity model.

We emphasize that the asymptotic properties of $\hat{\theta}$ can be different from Proposition 2.2 depending on the preliminary estimator to compute $\hat{\epsilon}_{ijt}^2$ and $\hat{\mu}_{ijt}$ as well as potential misspecification in the conditional variance. First, if the conditional variance structure in equation (2) does not hold, i.e., $\mathbb{E}\left[x_{ijt}\left(\epsilon_{ijt}^2 - \bar{h} \cdot \mu_{ijt}^{\bar{\lambda}}\right)\right] \neq 0$ for all $\left(\bar{h}, \bar{\lambda}\right) \in \Theta$, as long as the degree of misspecification is mild, $\hat{\theta}$ still converges to the *pseudo-true* parameter θ^* minimizing the population GMM criterion function. Practitioners can still estimate λ^*, which approximates the variance-to-mean ratio that best describes the data generating process. Although the asymptotic variance changes due to the misspecified moment condition, we can still test whether the conditional variance is close to the CVMR assumption. In Appendix A, we provide the asymptotic distribution of $\hat{\theta}$

considering the mildly misspecified cases (Appendices A.1.1 and A.2.1).

Second, the iGMM estimator $\hat{\theta}$ does not exhibit asymptotic bias because the initial PPML estimator is free from asymptotic bias in the gravity model with two-way fixed effects (Weidner and Zylkin, 2021). Although other initial estimators, such as Gaussian PML and Gamma PML, can replace the PPML and ensure the consistency of $\hat{\theta}$, they may introduce an IPP problem in the inference on θ and necessitate additional bias correction methods. The selection of the initial estimator becomes particularly critical when extending the gravity model to include three-way fixed effects, as discussed in Weidner and Zylkin (2021). We discuss more details in Section 2.5.

2.2 G-PPML

The consistent estimator of λ enables us to develop a more efficient estimator than the PPML while preserving all the desirable properties of the PPML. Hence, we propose the *Generalized* PPML estimator replacing the conditional variance parameter λ with its feasible analog $\hat{\lambda}$ from the previous subsection. The FOCs with respect to the regression coefficients can be expressed as

$$\hat{\beta}: \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T} x_{ijt}\left(y_{ijt}-\hat{\mu}_{ijt}\right)\hat{\mu}_{ijt}^{1-\hat{\lambda}}=0,$$

$$\hat{\gamma}_{it}: \sum_{j=1}^{N}\left(y_{ijt}-\hat{\mu}_{ijt}\right)\hat{\mu}_{ijt}^{1-\hat{\lambda}}=0,$$

$$\hat{\eta}_{jt}: \sum_{i=1}^{N}\left(y_{ijt}-\hat{\mu}_{ijt}\right)\hat{\mu}_{ijt}^{1-\hat{\lambda}}=0,$$

where $\hat{\mu}_{ijt}=\exp\left(x_{ijt}'\hat{\beta}+\hat{\gamma}_{it}+\hat{\eta}_{jt}\right)$, $i,j=1,\ldots,N$, and $t=1,\ldots,T$. The fixed effect terms that solve FOCs satisfy

$$\exp\left(\hat{\gamma}_{it}\right)=\left(\sum_{j=1}^{N}\exp\left(\left(2-\hat{\lambda}\right)\left(x_{ijt}'\hat{\beta}+\hat{\eta}_{jt}\right)\right)\right)^{-1}\sum_{j=1}^{N}\exp\left(\left(1-\hat{\lambda}\right)\left(x_{ijt}'\hat{\beta}+\hat{\eta}_{jt}\right)\right)y_{ijt}$$

$$\exp\left(\hat{\eta}_{jt}\right)=\left(\sum_{i=1}^{N}\exp\left(\left(2-\hat{\lambda}\right)\left(x_{ijt}'\hat{\beta}+\hat{\gamma}_{it}\right)\right)\right)^{-1}\sum_{i=1}^{N}\exp\left(\left(1-\hat{\lambda}\right)\left(x_{ijt}'\hat{\beta}+\hat{\gamma}_{it}\right)\right)y_{ijt},$$

and we plug in the fixed effect estimates to the FOC for β to present the FOC as a function of β. The estimator $\hat{\beta}$ solves the system of k equations (FOCs) regarding the k-dimensional parameter $\hat{\beta}$. Under the same regularity assumptions as Proposition 2.1, we derive the consistency of the proposed estimator.

Proposition 2.3. *Under Assumption 2.1, $\hat{\beta}$ is a consistent estimator of β as $N \to \infty$.*

The consistency of the estimator $\hat{\theta}$ derived in Proposition 2.1 is the basis for the consistency of $\hat{\beta}$. The consistency of the G-PPML estimator is not surprising in the two-way fixed effects model, as PML estimators are generally consistent under the current specification (Fernández-Val and Weidner, 2016). We discuss the extension to the three-way fixed effects case in Section 2.5.

2.3 Asymptotic Distribution

The proposed *Generalized* PPML estimator is consistent for β, but we have yet to establish that it is immune to asymptotic bias in the presence of two-way fixed effects. The sample size is N^2T and the number of fixed effect terms is $2NT$. Thus, the finite sample bias of the estimator disappears with the $1/N$ rate, which generally causes the asymptotic bias for the limiting distribution of $N\left(\hat{\beta} - \beta\right)$. The closed-form expression for the asymptotic bias follows the formula derived by Fernández-Val and Weidner (2016) and Weidner and Zylkin (2021).

Given the general conditional variance form of y_{ijt}, the previous literature confirms that the PPML estimator is a unique estimator that is immune to the IPP among the class of PML estimators. In this section, we verify that the G-PPML estimator is also immune to the IPP and has no asymptotic bias if the conditional variance of y_{ijt} conforms to the form $h \cdot \mu_{ijt}^{\lambda}$ for any $h > 0$ and $\lambda \in \mathbb{R}$. A notable distinction from other PML estimators, however, is that G-PPML does not suffer from the IPP *without* having to impose a particular variance-to-mean ratio on the dependent variable.

Define the following elements

$$S_{ij,t} = (y_{ijt} - \mu_{ijt})\, \mu_{ijt}^{1-\lambda}$$

$$H_{ij,ts} = \begin{cases} \mu_{ijt}^{1-\lambda} (1-\lambda)\, y_{ijt} - \mu_{ijt}^{2-\lambda} (2-\lambda) & \text{if } t = s \\ 0 & \text{otherwise} \end{cases}$$

$$G_{ij,tsr} = \begin{cases} \mu_{ijt}^{1-\lambda} (1-\lambda)^2\, y_{ijt} - \mu_{ijt}^{2-\lambda} (2-\lambda)^2 & \text{if } t = s = r \\ 0 & \text{otherwise.} \end{cases}$$

These elements are components of the score vector $S_{ij} = (S_{ij,1}, \ldots, S_{ij,T})' \in \mathbb{R}^T$, the $T \times T$ Hessian matrix H_{ij}, and the $T \times T \times T$ cubic tensor G_{ij}. We denote $\bar{H}_{ij} = \mathbb{E}\left[H_{ij} | x_{ij}, \gamma_i, \eta_j\right]$ and $\bar{G}_{ij} = \mathbb{E}\left[G_{ij} | x_{ij}, \gamma_i, \eta_j\right]$, where $x_{ij} = (x_{ij,1}, \ldots, x_{ij,k})$ with $x_{ij,l} = (x_{ij1,l}, \ldots, x_{ijT,l})'$ including the lth element of x_{ijt} over the sample period, $\gamma_i = (\gamma_{i1}, \ldots, \gamma_{iT})'$, and $\eta_j = (\eta_{j1}, \ldots, \eta_{jT})'$. Define the normalized $T \times k$ matrix $\tilde{x}_{ij} = x_{ij} - \gamma_i^x - \eta_j^x$, where γ_i^x and η_j^x are standardized $T \times k$ matrices that minimize

$$\sum_{i=1}^{N} \sum_{j=1}^{N} Tr \left[\left(x_{ij} - \gamma_i^x - \eta_j^x \right)' \bar{H}_{ij} \left(x_{ij} - \gamma_i^x - \eta_j^x \right) \right].$$

Following Fernández-Val and Weidner (2016), our result is based on independence across country pairs with arbitrary serial dependence over the short panel.[16] The following proposition establishes the asymptotic normality of our estimator. The extension to accommodate a cluster-robust asymptotic variance is straightforward following Weidner and Zylkin (2021) and we provide corresponding discussion in Appendix A.4.

Proposition 2.4. *(Asymptotic Distribution) Suppose the model satisfies equation (4). Under Assumption 2.1,*

$$N \left(\hat{\beta} - \beta \right) \xrightarrow{d} \mathcal{N} \left(0, \Omega_\infty^{-1} \right),$$

[16] Fernández-Val and Weidner (2016) examine the traditional non-dyadic panel with $N \times T$ dimensional data, whereas most trade data follow a dyadic panel structure with dimensions $N \times N \times T$. If we conceptualize a country pair as a single 'individual' in an $N \times T$ dimensional panel, the weak correlation condition in Fernández-Val and Weidner (2016) corresponds to serial dependence over time while assuming independence across country pairs.

where $\Omega_\infty = \lim_{N\to\infty} \Omega_N$ *is a* $k \times k$ *matrix with*

$$\Omega_N = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\tilde{x}'_{ij}\bar{H}_{ij}\tilde{x}_{ij}.$$

The derived asymptotic distribution is simple and has zero asymptotic bias in a gravity setup with two-way fixed effects. The asymptotic bias following Fernández-Val and Weidner (2016) in general case is $\lim_{N\to\infty}\left(\Omega_N^{-1}\left(B_N + D_N\right)\right)$, where $N\left(\hat{\beta} - \beta - \frac{\Omega_N^{-1}(B_N+D_N)}{N}\right) \xrightarrow{d} \mathcal{N}\left(0, \Omega_\infty^{-1}\right)$. B_N and D_N are k-dimensional vectors with their mth elements defined by

$$\begin{aligned}
B_N^m = &-\frac{1}{N}\sum_{i=1}^{N} Tr\left[\left(\sum_{j=1}^{N}\bar{H}_{ij}\right)^{-1}\sum_{j=1}^{N}\mathbb{E}\left[H_{ij}\tilde{x}_{ij,m}S'_{ij}|x_{ij},\gamma_i,\eta_j\right]\right] \\
&+\frac{1}{2N}\sum_{i=1}^{N} Tr\left[\left(\sum_{j=1}^{N}\bar{H}_{ij}\right)^{-1}\left(\sum_{j=1}^{N}\bar{G}_{ij}\tilde{x}_{ij,m}\right)\right], \\
D_N^m = &-\frac{1}{N}\sum_{j=1}^{N} Tr\left[\left(\sum_{i=1}^{N}\bar{H}_{ij}\right)^{-1}\sum_{i=1}^{N}\mathbb{E}\left[H_{ij}\tilde{x}_{ij,m}S'_{ij}|x_{ij},\gamma_i,\eta_j\right]\right] \\
&+\frac{1}{2N}\sum_{j=1}^{N} Tr\left[\left(\sum_{i=1}^{N}\bar{H}_{ij}\right)^{-1}\left(\sum_{i=1}^{N}\bar{G}_{ij}\tilde{x}_{ij,m}\right)\right],
\end{aligned}$$

and $\lim_{N\to\infty} B_N = \lim_{N\to\infty} D_N = 0$ if we use the G-PPML estimator under equation (4). The asymptotic bias components B_N^m and D_N^m both converge to zero regardless of the conditional variance of y_{ijt} if $\hat{\beta}$ is the PPML estimator ($\lambda = 1$). The second terms of B_N^m and D_N^m are zeros since the G-PPML estimator's $\bar{H}_{ij}$ component $\mu_{ijt}^{2-\lambda}$ is proportional to the $\bar{G}_{ij}$ component $(3-2\lambda)\,\mu_{ijt}^{2-\lambda}$ for all λ values. The first terms of B_N^m and D_N^m generally converge to non-zero unless $\lambda = 1$, but become zeros when the conditional variance of y_{ijt} is correctly specified. Note that the $\bar{G}_{ij}$ component is not necessarily proportional to the $\bar{H}_{ij}$ component if the conditional variance of ϵ_{ijt} does not follow the functional form of $h \cdot \mu_{ijt}^{\lambda}$. This highlights why the G-PPML with $\lambda \neq 1$ needs an additional functional form assumption to obtain the zero asymptotic bias property.

Following a similar logic when comparing the OLS and GLS estimators, the optimally weighted G-PPML estimator is more efficient than the other PML-class estimators. In a finite sample, however, the G-PPML may not always outperform the PPML because of the need to estimate the conditional variance parameter λ, just as the feasible GLS may not always outperform the OLS. Appendix A.4.3 demonstrates that the conditional variance of the (infeasible) G-PPML estimator is less than the conditional variances of other PML estimators when $\text{Var}\left(\epsilon_{ijt}|x_{ijt},\gamma_{it},\eta_{jt}\right)=h\cdot\mu_{ijt}^{\lambda}$ is correctly specified.

We estimate the asymptotic variance of $\hat{\beta}$ by $\hat{\Omega}_N=\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\tilde{x}}_{ij}'\hat{\tilde{H}}_{ij}\hat{\tilde{x}}_{ij}$, where $\left[\hat{\tilde{H}}_{ij}\right]_{ts}=\hat{\mu}_{ijt}^{2-\hat{\lambda}}1\left\{t=s\right\}$. Since $\hat{\lambda}$ is a consistent estimator of λ, the plug-in estimator $\hat{\Omega}_N^{-1}$ consistently approximates Ω_N^{-1}. The asymptotic variance estimator $\hat{\Omega}_N^{-1}$ does not require an additional estimator for $\text{Var}\left(S_{ij}|x_{ij},\gamma_i,\eta_j\right)$. The property brings a notable computational benefit since the Hessian matrix of the G-PPML estimator approximates the variance of the score function without relying on the presumption about the conditional variance of y_{ijt}. The simplified asymptotic variance also implies that the G-PPML estimator does not suffer from the downward bias in robust standard errors pointed out by Weidner and Zylkin (2021).[17]

2.4 Discussions

The preceding sections provided a valid test for the CVMR assumption and suggested an alternative G-PPML estimator in the case that the CVMR assumption is violated. We acknowledge that G-PPML may not always be preferred to PPML. Specifically, if the conditional variance of the model does not follow equation (2), which forms the basis of moment conditions for all commonly used estimators in the gravity literature (e.g., OLS, PPML, Gamma-PML, etc.), then the comparison between PPML and G-PPML becomes complicated. G-PPML is still a consistent estimator, and the iGMM estimates find the best fit of the conditional variance within the specification of equation (2). However, G-PPML is no longer immune to the IPP and requires additional bias correction for inference (Weidner and Zylkin, 2021). In spite of the potential efficiency gain,

[17] In Appendix A.4, we show that the downward bias arises when the conditional variance does not follow equation (2).

the IPP induced by misspecified conditional variance is a potential drawback of the G-PPML.

Despite this caveat, we adhere to the specification in equation (2) for two appealing reasons. First, discussion on PPML has revolved around equation (2) and studies (e.g., Head and Mayer (2014)) have proposed different econometric solutions when the CVMR assumption fails within the functional form in equation (2). In this sense, we provide a convenient estimator that generalizes the common assumptions made in prior literature and obviates the need to assume particular parameter values. Second, G-PPML is consistent with our iGMM estimator, which, as we demonstrate in our Monte Carlo simulations, is the most efficient specification test for CVMR to date. The misspecification-robust property of iGMM enables practitioners to capture both conventional assumptions on the conditional variance and statistical efficiency.

Regarding the computational issues, practitioners may want to apply the iGMM to the gravity model parameters β as well as the conditional variance parameters θ, where a similar procedure is observed in Lewbel and Pendakur (2009). Practitioners can start with the PPML as a preliminary estimator, then estimate θ and β for the first iteration. The second iteration can use the estimated β as a preliminary estimator, and it can continue until all parameters converge.[18] While both algorithms are theoretically valid and empirically feasible, we observe that convergence can be an issue when the iteration becomes highly non-convex. Thus, we propose a simpler algorithm to stop at the first iteration.

There are some potential extensions allowing for more flexible conditional variance specifications. First, our specification $\text{Var}\left(\epsilon_{ijt}|x_{ijt},\gamma_{it},\eta_{jt}\right) = h\cdot\mu_{ijt}^{\lambda} = \exp\left(\log h + \lambda\left(x_{ijt}'\beta + \gamma_{it} + \eta_{jt}\right)\right)$ is an exponential function with a linear index. Thus, assuming that the variance structure follows $\mathbb{E}\left[\epsilon_{ijt}^2|x_{ijt},\gamma_{it},\eta_{jt}\right] = \exp\left(\bar{\beta}_0 + x_{ijt}'\bar{\beta} + \bar{\gamma}_{it} + \bar{\eta}_{jt}\right)$, running another PPML estimation of $\hat{\epsilon}_{ijt}^2$ with respect to x_{ijt} and two-way fixed effects accommodates a more general functional form of the conditional variance compared with equation (2).

Second, a more explicit approximation of the unknown conditional variance involves esti-

[18] The authors appreciate the anonymous referee for suggesting the idea and providing references.

mating $\mathbb{E}\left[\epsilon_{ijt}^2|x_{ijt},\gamma_{it},\eta_{jt}\right]$ without assuming a specific functional form. For linear models, the previous literature on feasible GLS (FGLS) provides several methods to construct an FGLS estimator allowing for serial correlation, clustering, and high-dimensional fixed effects (e.g., Newey and West (1987), Romano and Wolf (2017), Bai et al. (2021)). Even though the results for linear models cannot be directly applied to nonlinear models, it is important to note that `ppmlhdfe` is an implementation of the Iteratively Reweighted Least Squares (IRLS) algorithm, which linearizes the model and reduces dimensionality by focusing on non-fixed effect covariates. In analogy to how OLS residuals construct the optimal weight in the FGLS estimator with an unknown conditional variance, the PPML residuals, $\hat{\epsilon}_{ijt}$, can help improve G-PPML by allowing for more flexible functional forms for the conditional variance, following the approach of Newey and West (1987) and Newey and West (1994).[19] We leave the details to future research, as the asymptotic properties of the newly suggested method are beyond the scope of the current paper.

For practitioners, we suggest using PPML as the "go-to" estimator for gravity estimation and strongly recommend our iGMM routine (with our `cvmrtest` Stata command) to test if the CVMR assumption is rejected. If the CVMR assumption is rejected, the researcher has the option to adopt our G-PPML estimation method by making the functional form assumption of the conditional variance as in equation (2).[20] These steps are subject to pre-testing bias, as one might incorrectly reject the null hypothesis $H_0 : \lambda = 1$ when it is actually true. Yet, as we demonstrate in our Monte Carlo simulations, the consequence of a wrong assumption on the conditional variance form is mitigated owing to the robustness of iGMM to potential misspecification shown by Hansen and Lee (2021). If the researcher chooses not to make any assumption on the conditional variance, PPML remains a consistent, yet perhaps not the most efficient, estimator. We provide a full comparison of commonly used methods for gravity estimation in Table D.1.

[19]For example, if the covariance structure is simple and N and T are sufficiently large, $\hat{\sigma}^2_{h,ij} = \frac{1}{T}\sum_{t=1}^{T}\hat{\epsilon}_{ijt}\hat{\epsilon}_{ij(t-h)}$ approximates $E\left[\epsilon_{ijt}\epsilon_{ij(t-h)}\right]$.

[20]Note that the G-PPML estimator remains an attractive option even if the CVMR assumption is not rejected. When $\lambda = 1$ and the CVMR assumption holds, the G-PPML estimator is equivalent to the PPML estimator. However, in this case, G-PPML entails a more computationally intensive two-stage estimation procedure, yielding only marginal differences from the PPML estimates.

2.5 Extension to the Gravity Model with Three-way Fixed Effects

So far, we have focused on the gravity model with two-way fixed effects. This section extends the G-PPML framework to the gravity model with three-way fixed effects by incorporating an importer-exporter pair fixed effect. Following Weidner and Zylkin (2021), who establish the asymptotic properties of PPML with three-way fixed effects, we extend our methodology to this setting and provide corresponding proofs for our main propositions in the appendix. Detailed proofs can be found in Appendix A.

While $\hat{\theta}$ remains consistent for the (pseudo-true) conditional variance parameter θ and is asymptotically normal, the nonzero asymptotic bias of the initial PPML estimator with three-way fixed effects induces a nonzero asymptotic bias in the distribution of $N(\hat{\theta} - \theta)$. After applying a suitable bias correction, practitioners can use this result for testing the CVMR assumption (see Appendices A.1.2 and A.2.2).

Proposition 2 of Weidner and Zylkin (2021) establishes that the PPML estimator remains consistent even in the presence of three-way fixed effects. Building on this result, we show that the G-PPML estimator is also consistent under three-way fixed effects, provided that the conditional variance is correctly specified as in equation (2) (see Appendix A.3.2).

Moreover, the asymptotic bias does not vanish for G-PPML in a three-way fixed effects setting. As demonstrated by Weidner and Zylkin (2021), PPML exhibits a nonzero asymptotic bias under three-way fixed effects. Since the asymptotic bias components of G-PPML generalize those of PPML, they are also generally nonzero. Appendix A.4.2 provides a more rigorous discussion of this issue, along with detailed proofs.

The efficiency of the G-PPML estimator under two-way fixed effects becomes less straightforward in a three-way fixed effects setting. Under three-way fixed effects, G-PPML not only inherits the nonzero asymptotic bias property of PPML but also introduces additional variance from the plug-in estimator $\hat{\lambda}$, which does not vanish asymptotically (Murphy and Topel, 2002).

3 Monte Carlo Simulation Analysis

3.1 Data Generating Process

We start by generating the vector of independent variables, $x \equiv \{1, x_{1,ijt}, x_{2,ijt}, \ldots x_{6,ijt}, \iota_{it}, \iota_{jt}\}$, where $x_{1,ijt}$ is drawn from a normal distribution $\mathcal{N}(0, 0.1)$, $x_{2,ijt}$ is a dummy variable from a Bernoulli distribution with $p = 0.5$, and the rest of the covariates, $x_{3,ijt} \ldots, x_{6,ijt}$, are independently and identically drawn from the same distribution as $x_{1,ijt}$. ι_{it} and ι_{jt} are exporter-time and importer-time indicators, respectively. Without loss of generality, we set the constant term and the coefficients of $x_{1,ijt}, \ldots, x_{6,ijt}$ as $\beta = \{0.5, -0.5, 0.5, -0.5, 0.5, -0.5, 0.5\}'$. Consistent with the structure of a standard bilateral trade dataset, we consider N countries that import from and export to all other countries (including themselves) over a period of T years. We construct and experiment with two datasets, depending on the number of countries and years: one with 50 countries and 10 years (i.e., 25,000 observations) and another with 100 countries and 5 years (i.e., 50,000 observations).[21] The coefficients of the exporter-year (γ_{it}) and importer-year (η_{jt}) fixed-effects vary between -0.5 and 0.5. We denote the parameter vector as $\theta = (\beta', \gamma, \eta)'$. All Monte Carlo results are based on 500 independent simulations.[22]

To construct the multiplicative error term ε_{ijt}, or simply ε, we introduce four parameters $\{h_1, h_2, \lambda_1, \lambda_2\}$, where the hs are assumed to be non-negative and the λs to be real numbers, and we assume that ε follows a log-normal distribution, whose mean and variance are given, respectively, by:[23]

$$\begin{aligned} &\mathbb{E}(\varepsilon|x) = 1 \quad \text{and} \\ &\text{Var}(\varepsilon|x) = [h_1 \exp(\lambda_1 x'\theta) + h_2 \exp(\lambda_2 x'\theta)] / \exp(2x'\theta). \end{aligned} \tag{6}$$

This implies the following first and second moments of the dependent variable $y = \exp(x'\theta) \cdot \varepsilon$,

[21] The idea is to simulate two commonly used data structures: one with a smaller set of countries over a longer time span, and one with a larger set of countries over a shorter period.

[22] Further increment of the number of simulations does not change the results at the conventional precision levels.

[23] Mechanically, to generate the random variable ε from the lognormal distribution specified in (6), we first generate a random variable ξ from a standard normal distribution, and then we define $\varepsilon \equiv \exp\left(-\log(\text{Var}(\varepsilon|x)+1)/2 + \sqrt{\log(\text{Var}(\varepsilon|x)+1)}\xi\right)$, which satisfies the first and second moments in (6).

where y refers to y_{ijt}:

$$\mathbb{E}\left(y|x\right) = \exp(x'\theta) \quad \text{and} \\ \text{Var}(y|x) = h_1 \mathbb{E}\left(y|x\right)^{\lambda_1} + h_2 \mathbb{E}\left(y|x\right)^{\lambda_2}. \tag{7}$$

The first line in equation (7) is a common assumption, and the second line is a very flexible representation of the conditional variance form that encapsulates all commonly held conditional variance forms in the gravity literature (cf. Head and Mayer (2014)). To see this, when $h_1 > 0$ and $h_2 = 0$, the functional form can accommodate CVMR ($\lambda_1 = 1$),[24] Gamma-PML and OLS ($\lambda_1 = 2$).[25] Allowing $h_1 > 0$ and $h_2 > 0$ introduces misspecification into the data generating process, permitting a comparison of G-PPML with other estimators under misspecification. Due to space constraints, we briefly address misspecification at the end of Section 3, with detailed results provided in Appendix B.[26]

As we will demonstrate shortly, the key advantage of the G-PPML estimator is that it relieves researchers' "burden of proof" for the value of λ associated with a particular estimator.[27] Intuitively, our iGMM approach will automatically find the value of λ that provides the best fit of the error term structure, and then the subsequent G-PPML estimator will capitalize on this value of λ to construct moment conditions that are better tuned than other PML estimators. In the kinfe-edge case where iGMM suggests $\hat{\lambda} = 1$, G-PPML becomes identical to PPML.

We also test the performance of G-PPML under misspecification. Specifically, when we assume that both h_1 and h_2 in equation (7) are positive, the conditional variance becomes a polynomial of the conditional mean, and none of the mainstream assumptions for ε are consistent. An

[24] We emphasize that the moment conditions of PPML are obtained under the CVMR assumption. Yet, it is not a necessary condition for PPML to be a consistent estimator. See Santos Silva and Tenreyro (2022) for detailed discussions.

[25] If $\lambda_1 = 2$, the conditional variance is a quadratic function of its conditional mean, and this is consistent with the assumption of the Gamma-PML estimator (Head and Mayer, 2014). Under the same condition $\lambda_1 = 2$, $\text{Var}(\varepsilon|x)$ reduces to h_1 according to equation (6); that is, the error term ε becomes *homoskedastic*. In this case, it is innocuous to take the natural logarithm of $y = \exp(x'\theta) \cdot \varepsilon$ on both sides and estimate the resulting equation with OLS. Under an alternative DGP where the dependent variable is $y = \exp\left(x'\theta\right) + \varepsilon$, the working assumptions of Gamma-PML and OLS do not coincide. A practical issue with this alternative DGP for our purpose is that y can take negative values, and the resulting simulated dataset would not be well suited for PML or log-linearized OLS estimations.

[26] A similar form of misspecification is discussed in Santos Silva and Tenreyro (2011).

[27] Our framework can even account for cases where $\lambda < 0$, i.e., the conditional variance of the dependent variable decreases with its conditional mean.

important result from Hansen and Lee (2021) is that the iGMM is robust to potential misspecification. Applied to our setting, this property of the iGMM further relieves researchers' "burden of proof" regarding the functional-form assumptions. Thus, even when the conditional variance deviates from the specific form given by equation (2), the iGMM estimator would mitigate the misspecification problem by estimating h and λ that provide the best fit to the estimates of the conditional variance.

3.2 Estimates of the Conditional Variance

This section tests the performance of our iGMM method for estimating λ. For comparison purposes, we also report corresponding results that are obtained with two alternative leading approaches. However, we emphasize that, since the two alternatives are only valid for specific values for λ, the analysis in this section should be interpreted purely as a test of the performance of our iGMM method and not as a horse race with alternatives.[28]

Following the existing literature (e.g., Head and Mayer (2014)), we assume that $h_2 = 0$ and use h and λ to refer to h_1 and λ_1, respectively.[29] The first alternative approach that we implement is the "MaMu" method, named after Manning and Mullahy (2001). For a given set of estimated $\hat{\theta}$ via PPML, we can define the residual term $\hat{\epsilon} \equiv y - \exp(x'\hat{\theta})$. Then, to infer the conditional variance, we estimate the following equation with OLS:

$$\log\left(\hat{\epsilon}^2\right) = \text{constant} + \lambda x'\hat{\theta} + \epsilon, \tag{8}$$

which is the natural logarithm of the equation $\text{Var}(y|x) = h\exp(\lambda x'\theta)$.

The second alternative approach that we implement to recover λ follows Santos Silva and Tenreyro (2006).[30] Let $\hat{y}$ denote $\hat{y} \equiv \exp(x'\hat{\theta})$. Santos Silva and Tenreyro approximate the

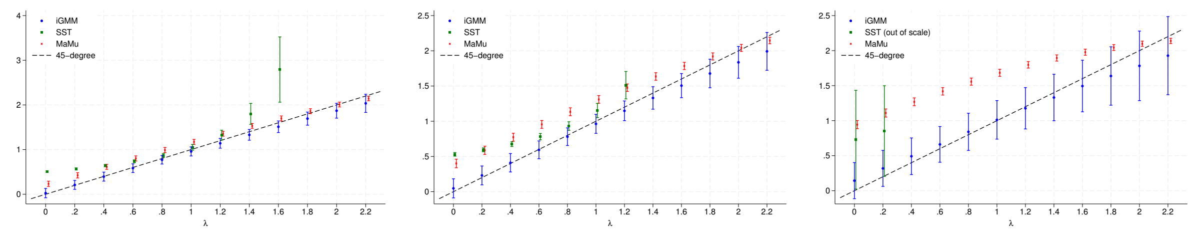
[28] We are grateful to João Santos Silva for a fruitful related exchange.

[29] Results under potential misspecification ($h_2 > 0$) are provided in Appendix B.

[30] The key purpose of the method developed by Santos Silva and Tenreyro (2006) is to test whether λ is statistically distinguishable from 1 rather than obtaining an efficient point estimate. Thus, we focus on comparing our confidence intervals with those of Santos Silva and Tenreyro (2006).

Figure 1: Comparing Different Methods to Estimate λ

(a) $h = 0.5$ (b) $h = 1$ (c) $h = 4$



Notes: These figures display the point estimates and the 95% confidence intervals of different estimators under various parametric assumptions about the data generating process. "iGMM" indicates the iterated GMM estimator proposed as the preceding step to our G-PPML estimator, "SST" is an estimator of λ proposed by Santos Silva and Tenreyro (2006), and "MaMu" is proposed by Manning and Mullahy (2001). For exposition, estimation results are omitted when the standard deviation is greater than 1 or the point estimate is negative. Full estimation results are available by request.

expression

$$(y - \hat{y})^2 = h\hat{y}^{\lambda} + \epsilon \tag{9}$$

to the first order around $\lambda = 1$. After dividing both sides by $\sqrt{\hat{y}}$, we obtain:

$$(y - \hat{y})^2/\sqrt{\hat{y}} = h\sqrt{\hat{y}} + h(\lambda - 1)\log(\hat{y})\sqrt{\hat{y}} + \epsilon^*, \tag{10}$$

where ϵ^* denotes $\epsilon/\sqrt{\hat{y}}$.[31] Equation (10) can be estimated with OLS using $\sqrt{\hat{y}}$ and $\log(\hat{y})\sqrt{\hat{y}}$ as the first and second regressors, respectively. Then, we recover $\hat{\lambda}$ by dividing the second term's coefficient estimate $\hat{h}(\hat{\lambda} - 1)$ by the estimate of the first coefficient $\hat{h}$ and adding one.[32] We refer to this method as "SST".[33]

Figure 1 reports the estimates of λ that are obtained with each of the three methods (iGMM, MaMu, and SST). To provide a more comprehensive analysis, we compare the performance of the three approaches for different values of h (varying between 0.5, 1, and 4 in panels (a), (b) and (c), respectively) and for a wide range of λs (the $[0, 2.2]$ interval on the horizontal axis in each panel). Intuitively, the alternative values of h correspond to different levels of noise in the data,

[31] Santos Silva and Tenreyro (2006) suggest that the conditional variance in equation (9) is likely to be proportional to the the mean $\hat{y}$, and thus they suggest normalizing both sides by $\sqrt{\hat{y}}$ as in equation (10) to perform weighted least squares.

[32] Equation (10) can be estimated both with and without adding a constant term as the third regressor. In general, we find that estimating equation (10) without the constant term provides estimates of λ closer to the true value. Thus, we only present the results without the additional constant term.

[33] We remind readers that although the SST approach to estimating λ was proposed in the same paper that also widely popularized the PPML for gravity models, the performance of the SST estimator is not associated with the performance of the PPML in gravity estimation. See Santos Silva and Tenreyro (2006) for details.

while the alternative values of λ cover a reasonable range from the existing literature.[34] Finally, in each panel, we plot a 45-degree reference line, which indicates that the point estimate and the assumed parameter value for λ are identical.[35]

We draw three main conclusions based on the results in Figure 1. First, and most important, our iGMM estimates (blue dots) exhibit consistent efficiency for different values of h and λ. Regardless of the level of data noise h, the λs obtained with iGMM are always close to the 45-degree line, and the associated 95 percent confidence intervals always include the true value of λ.[36] Second, the SST estimates (green squares) are very close to the 45-degree line when the true λ is close to 1, but they deviate from the 45-degree line when λ deviates from 1 and/or if there is substantial data noise. This result is expected since SST's estimating equation is obtained after applying a first-order Taylor approximation around $\lambda = 1$. Finally, the MaMu estimates (red crosses) exhibit a consistent bias toward 2. Specifically, when the true λ is less (greater) than 2, the MaMu estimates consistently overestimate (underestimate) the true λ.[37] Interestingly, the MaMu estimator delivers the narrowest confidence intervals; yet this becomes a key drawback as the confidence intervals fail to contain the true values except when $\lambda = 2$.

In sum, the Monte Carlo analysis demonstrates that the proposed iGMM method delivers reliable estimates of λ across a wide range of λ values and under different levels of noise in the data. Capitalizing on the strong performance of iGMM and the corresponding λ estimates, in the next subsection we demonstrate that our G-PPML estimator can deliver more efficient coefficient estimates than other leading estimators across a wide range of parameter values.

[34] For example, our own sectoral estimates of λ in Section 4 lie in the interval $(0.5, 1.7)$.

[35] We report and discuss the coverage of the confidence intervals in Appendix C.

[36] We do note, however, that the confidence intervals become wider when the noise in the data becomes more severe (towards higher values of λ in each figure and especially in panel (c)), reflecting the enhanced difficulty in estimating the underlying parameters.

[37] Santos Silva and Tenreyro (2006) note that due to Jensen's inequality, taking the natural logarithm on both sides of an estimating equation leads to biased coefficient estimates if the multiplicative error term features heteroskedasiticity. The same argument, applied to equation (8), can explain why the MaMu estimator leads to biased estimates of λ.

3.3 Coefficient Estimates

Table 1 compares the performance of G-PPML with other leading estimators.[38] We compare both the mean of absolute bias (column Bias), the mean of standard errors (column S.E.), and the standard deviation of the estimated coefficients (column S.D.) of various estimators under various parameter values. We also report the mean of the iGMM λs that are used in the G-PPML estimation (column $\bar{\hat{\lambda}}$). We consider six cases. In cases 1 through 3, we hold constant the level of h and experiment with different values of λ, taking values of 0, 1 and 2, respectively. In cases 4 through 6, we experiment with a higher value of h. We focus on two representative coefficient estimates. β_1 (β_2) is the coefficient of a continuous (dummy) variable $x_{1,ijt}$ ($x_{2,ijt}$). To ease the interpretation of the results, we remind readers that the Poisson distribution implicitly assumes $\lambda = 1$, and Gamma distribution implicitly assumes $\lambda = 2$. Moreover, when $\lambda = 2$, the error term becomes homoskedastic and the OLS becomes unbiased.

First, we note in column $\bar{\hat{\lambda}}$ that across different cases presented in Table 1, the average λ estimates that we obtain from 4000 Monte Carlo simulations are quite close to the assumed value of λ. As expected, the estimation is more accurate when there are more observations and the level of noise in the data, governed by h and λ, is lower.

In case 1 ($\lambda = 0$), G-PPML outperforms all other estimators by delivering both lower mean absolute bias and lower standard errors. The reason is that $\lambda = 0$ is not consistent with the working assumptions of any other estimators, whereas G-PPML does not preemptively assume a particular value of λ. Compared with PPML, we note that the mean bias is lower for G-PPML both in terms of β_1 and β_2, and G-PPML's standard errors are approximately 10% lower (0.0240 vs. 0.0274 and 0.0054 vs. 0.0059).[39] The Gamma-PML estimator does not perform well relative to G-PPML and PPML in this scenario – its mean biases for β_1 and β_2 are approximately 50%

[38]The coverage of β estimates are reported in Appendix Tables D.2 and D.3.

[39]We note that the advantage of G-PPML in absolute terms is less pronounced for the coefficient β_2. Intuitively, both PPML and G-PPML are *consistent* estimators, and the estimation *efficiency*, the key benefit of G-PPML, does not play a crucial role when the coefficient estimate of β_2 is already rather precise (as suggested by its substantially lower standard errors than β_1 for all cases).

and 200% greater than G-PPML, respectively. A natural explanation for this result is that case 1 is inconsistent with Gamma-PML's conditional variance assumption. Finally, consistent with Santos Silva and Tenreyro (2006), the OLS estimation bias is much greater than that of other estimators. The key message in our analysis in case 1 is similar if we experiment with a larger sample size (the right half of Table 1).

In case 2 ($\lambda = 1$), the DGP renders PPML as the most efficient estimator. Thus, not surprisingly, G-PPML and PPML deliver very similar estimates, while outperforming Gamma-PML and OLS both in terms of mean bias and standard errors. That G-PPML can perform as well as the PPML without prior knowledge about λ relies critically on the preceding iGMM estimation to be reliable.

In case 3 ($\lambda = 2$), the DGP becomes consistent with the underlying assumption of Gamma-PML. While G-PPML, PPML, and Gamma-PML are all consistent in this setting, the Gamma-PML outperforms both G-PPML and PPML. Due to improved estimation efficiency, G-PPML outperforms PPML, e.g., G-PPML features around 15% (0.0119 vs. 0.0140) to 17% (0.0594 vs. 0.0715) lower standard errors than PPML.[40] However, G-PPML does not perform as well as Gamma-PML, because our iGMM estimates of λ are not sufficiently close to 2.[41] We confirm this hypothesis by increasing the sample size to 50,000 observations and finding that G-PPML and Gamma-PML deliver more similar estimates.

Finally, turning to the OLS results, as discussed in subsection 3.1, when $\lambda = 2$, the error is homoskedastic, and taking the natural logarithm of the estimating equation does not introduce biases. Thus, OLS is the most efficient estimator, and it exhibits the lowest bias and standard errors in this case.

[40] Apparently, the extent of efficiency improvement hinges on various factors, such as the number of observations and regressors. Thus, we defer a more quantitative assessment of G-PPML's efficiency gain to Section 4 where we employ real trade data to estimate a gravity model that is widely adopted in the literature.

[41] This reflects an inherent challenge to the iGMM method – the greater the underlying λ value is, the more data noise there is, and, therefore, it becomes more challenging to infer the parameters that govern the structure of the error term. This naturally leaves open a path for future research that would lead to even more efficient estimates of λ.

Table 1: Main Monte Carlo Results

	$J = 50,\ T = 10,$ Obser. = 25 000							$J = 100,\ T = 5,$ Obser. = 50 000						
	$\bar{\hat{\lambda}}$	β_1			β_2			$\bar{\hat{\lambda}}$	β_1			β_2		
Estimator		Bias	S.E.	S.D.	Bias	S.E.	S.D.		Bias	S.E.	S.D.	Bias	S.E.	S.D.
Case 1: $\mathrm{Var}(y_i\|x_i) = 1 \cdot [\mathbb{E}(y_i\|x_i)]^0$														
G-PPML	0.0469	0.0201	0.0240	0.0250	0.0044	0.0054	0.0055	0.0312	0.0137	0.0169	0.0173	0.0031	0.0038	0.0038
PPML		0.0218	0.0274	0.0272	0.0047	0.0059	0.0059		0.0152	0.0191	0.0192	0.0033	0.0042	0.0041
Gamma-PML		0.0306	0.0343	0.0367	0.0119	0.0069	0.0073		0.0220	0.0251	0.0267	0.0067	0.0050	0.0052
OLS		0.1100	0.0340	0.0335	0.1078	0.0069	0.0067		0.1109	0.0238	0.0240	0.1077	0.0048	0.0047
Case 2: $\mathrm{Var}(y_i\|x_i) = 1 \cdot [\mathbb{E}(y_i\|x_i)]$														
G-PPML	0.9635	0.0336	0.0409	0.0422	0.0069	0.0084	0.0086	0.9834	0.0233	0.0289	0.0291	0.0049	0.0060	0.0061
PPML		0.0335	0.0419	0.0421	0.0069	0.0086	0.0086		0.0232	0.0293	0.0291	0.0049	0.0060	0.0062
Gamma-PML		0.0379	0.0426	0.0465	0.0110	0.0085	0.0092		0.0264	0.0311	0.0325	0.0067	0.0062	0.0066
OLS		0.0837	0.0415	0.0419	0.0823	0.0083	0.0084		0.0822	0.0290	0.0290	0.0820	0.0058	0.0058
Case 3: $\mathrm{Var}(y_i\|x_i) = 1 \cdot [\mathbb{E}(y_i\|x_i)]^2$														
G-PPML	1.8444	0.0503	0.0594	0.0626	0.0102	0.0119	0.0126	1.9055	0.0356	0.0431	0.0443	0.0073	0.0086	0.0091
PPML		0.0575	0.0715	0.0720	0.0110	0.0140	0.0137		0.0407	0.0504	0.0507	0.0080	0.0098	0.0100
Gamma-PML		0.0501	0.0562	0.0625	0.0098	0.0112	0.0122		0.0355	0.0416	0.0443	0.0071	0.0083	0.0089
OLS		0.0430	0.0537	0.0539	0.0085	0.0108	0.0106		0.0305	0.0376	0.0381	0.0061	0.0075	0.0076
Case 4: $\mathrm{Var}(y_i\|x_i) = 4 \cdot [\mathbb{E}(y_i\|x_i)]^0$														
G-PPML	0.1533	0.0462	0.0525	0.0622	0.0098	0.0119	0.0128	0.1117	0.0299	0.0352	0.0446	0.0065	0.0080	0.0090
PPML		0.0435	0.0546	0.0547	0.0094	0.0118	0.0117		0.0306	0.0381	0.0383	0.0066	0.0083	0.0084
Gamma-PML		0.0613	0.0577	0.0665	0.0349	0.0116	0.0133		0.0423	0.0439	0.0486	0.0199	0.0088	0.0098
OLS		0.2399	0.0561	0.0558	0.2382	0.0113	0.0111		0.2386	0.0392	0.0390	0.2380	0.0079	0.0078
Case 5: $\mathrm{Var}(y_i\|x_i) = 4 \cdot [\mathbb{E}(y_i\|x_i)]$														
G-PPML	1.0092	0.0665	0.0812	0.0838	0.0140	0.0168	0.0175	1.0207	0.0477	0.0577	0.0595	0.0095	0.0119	0.0120
PPML		0.0663	0.0833	0.0835	0.0140	0.0172	0.0175		0.0478	0.0585	0.0596	0.0095	0.0121	0.0120
Gamma-PML		0.0690	0.0712	0.0830	0.0274	0.0143	0.0170		0.0508	0.0543	0.0616	0.0167	0.0109	0.0120
OLS		0.1616	0.0673	0.0669	0.1617	0.0135	0.0137		0.1610	0.0471	0.0473	0.1620	0.0094	0.0092
Case 6: $\mathrm{Var}(y_i\|x_i) = 4 \cdot [\mathbb{E}(y_i\|x_i)]^2$														
G-PPML	1.7745	0.0888	0.1016	0.1113	0.0204	0.0204	0.0251	1.8273	0.0649	0.0768	0.0814	0.0147	0.0154	0.0182
PPML		0.1131	0.1394	0.1431	0.0217	0.0274	0.0272		0.0800	0.0995	0.1008	0.0159	0.0195	0.0200
Gamma-PML		0.0867	0.0879	0.1088	0.0173	0.0176	0.0217		0.0639	0.0685	0.0802	0.0130	0.0137	0.0162
OLS		0.0655	0.0818	0.0821	0.0131	0.0164	0.0165		0.0465	0.0573	0.0584	0.0093	0.0115	0.0117

Notes: This table shows the Monte Carlo results that compare different estimators with various sample sizes and under different assumptions about the structure of the error term. We report the average λ estimates, mean absolute bias and the standard error of the coefficient estimates. G-PPML indicates the generalized PPML estimator proposed in this paper, PPML denotes Poisson-Pseudo Maximum Likelihood estimator, Gamma-PML denotes Gamma Pseudo Maximum Likelihood, and OLS denotes ordinary least squares estimation after taking the natural logarithm of the dependent variable. β_1 and β_2 are the coefficients for a continuous variable and a dummy variable, respectively.

In cases 4 through 6, we replicate the results in the preceding cases with a higher value of h. The purpose is to gain further confidence in the G-PPML estimator when there is greater data noise. Without going into details, we note that the key conclusions that we drew based on the results from cases 1 through 3 remain the same.[42]

We conclude the Monte Carlo analysis with several experiments that investigate the potential misspecification of the error term's distribution. In Appendix B, we consider various cases in which the conditional variance takes a polynomial form in the style of Santos Silva and Tenreyro (2011). Consistent with the misspecification-robust property studied in Hansen and Lee (2021), we find that iGMM is able to estimate a conditional variance form that provides the best fit to the data noise structure and that the resulting G-PPML performs well when compared with conventional estimation methods.

4 Empirical Evidence

To demonstrate the validity and practical importance of our methods, we proceed with an empirical application in four steps. First, we set up a representative econometric gravity model, which we estimated with PPML. Then, we estimate values of λ at the sectoral level. Third, we obtain gravity estimates with G-PPML. Finally, we compare the PPML vs. G-PPML estimates and their corresponding standard errors and z-statistics. To perform the empirical analysis, we rely on sectoral trade data from the latest edition of the USITC's *International Trade and Production Database for Estimations* (ITPD-E-R02) (Borchert et al., 2022), which enables us to obtain a distribution (across sectors) of the estimated conditional variances (λs), together with corresponding

[42]Based on the analysis in Table 1, we expect that the G-PPML estimator should outperform all other estimators when $\lambda < 0$ or $\lambda > 2$.

distributions of PPML and G-PPML gravity estimates for 105 manufacturing sectors.[43]

We specify the following benchmark for estimating the gravity equation at the sectoral level:

$$
\begin{aligned}
y_{ijt} \;=\;& \exp\left(\beta_1 DIST_{ij} + \beta_2 DIST\text{-}IN_{ij} + \beta_3 CNTG_{ij} + \beta_4 CLNY_{ij} + \beta_5 LANG_{ij}\right) \times \\
& \exp\left(\beta_6 RTA_{ijt} + \beta_7 EU_{ijt} + \beta_8 WTO_{ijt} + \beta_9 BRDR_{ij} + \gamma_{it} + \eta_{jt}\right) \times \varepsilon_{ijt}. \qquad (11)
\end{aligned}
$$

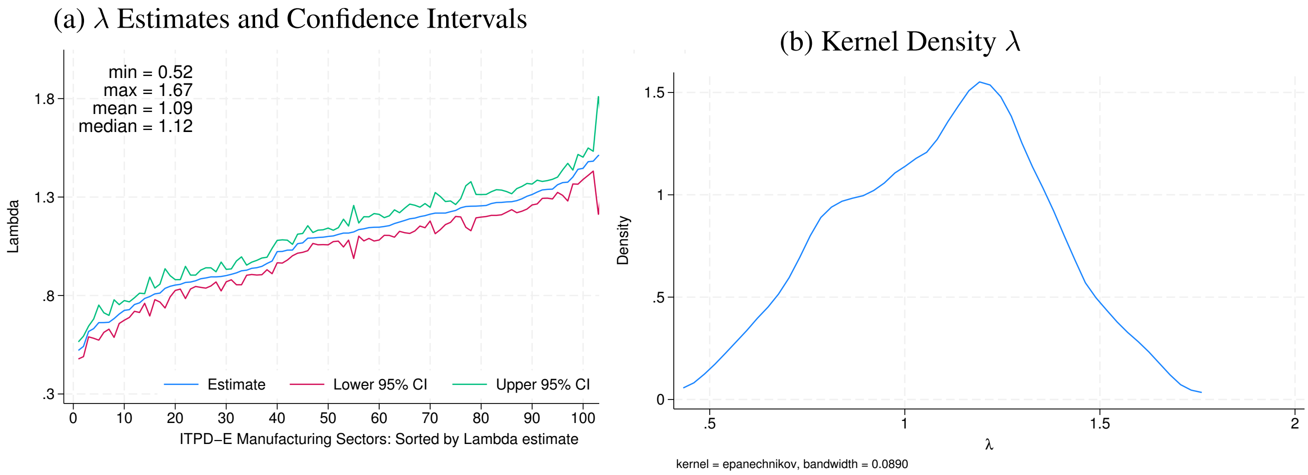
Here, y_{ijt} denotes nominal trade flows from exporter i to importer j in year t (Egger et al., 2022), including domestic trade flows (Yotov, 2022). Consistent with the multiplicative form of PPML, y_{ijt} enters (11) in levels (Santos Silva and Tenreyro, 2006, 2011). The covariates in (11) include the most widely used proxies for bilateral trade costs. $DIST$ is the logarithm of bilateral distance between i and j and $DIST\text{-}IN_{ij}$ is the corresponding variable for domestic distance. The rest of the covariates are indicator variables for common borders ($CNTG_{ij}$), colonial relationships ($CLNY_{ij}$), common official language ($LANG_{ij}$), the presence of regional trade agreements (RTA_{ijt}), EU membership (EU_{ijt}), and WTO membership (WTO_{ijt}). $BRDR_{ijt}$ is a dummy variable that takes a value of one for international trade and is equal to zero for domestic trade, which is designed to capture border/home bias effects. To control for the multilateral resistance terms of Anderson and van Wincoop (2003), as well as for any other country-specific determinants of bilateral trade flows (e.g., size), we use exporter-time (γ_{it}) and the importer-time (η_{jt}) fixed effects. Finally, we implement the finite sample bias correction of the standard errors following Weidner and Zylkin (2021).[44]

For our purposes, panel (a) of Figure 2 reports the estimates of λ along with their confidence

[43] ITPD-E-R02 is suitable for our purposes because it is constructed from raw/administrative data that has not been manipulated with statistical methods. In addition, ITPD-E-R02 includes a large number of sectors. Given our purposes, we only focused on 118 manufacturing sectors from ITPD-E-R02, and we were able to obtain estimates for 105 of them. Specifically, PPML, G-PPML, and λ estimation procedures faced convergence issues in 3, 2 and 8 different sectors, respectively. We believe that the convergence performance can be further improved by ruling out data outliers in the subset of problematic sectors. For consistency, we decided to stick with the raw data. We limit the analysis to the period 2010-2019, as robustness checks reveal that our conclusions do not depend on time coverage. Finally, we take advantage of the fact that, consistent with gravity theory, ITPD-E-R02 includes international and domestic trade flows. However, our main conclusions remain robust when we only use the international trade observations from ITPD-E-R02, which are based on the UN's Comtrade database.

[44] Since the focus of our paper is on the possible differences between the PPML and G-PPML estimates rather than on the level of the gravity coefficients *per se*, we do not report the PPML estimation results for equation (11) for each individual sector. Instead, without going into details, we note that our PPML estimates are in line with the literature. All gravity estimates are available by request.

Figure 2: Estimates of Lambda: ITPD-E Manufacturing



Notes: Panel (a) of this figure visualizes the estimates of λ along with their confidence intervals, which are obtained from specification (11) using all covariates from equation (11) as instruments. Panel (b) reports the kernel density of the λ estimates.

intervals, while panel (b) reports the kernel density of the distribution of λs. Four salient findings stand out from Figure 2. First, we find that all estimates of λ are strictly positive. Consistent with the CVMR structure of Poisson distribution, this result suggests that the conditional variance of trade volume indeed increases with its conditional mean. Second, most λ values are close to one. The practical implication of this result is that, in many cases, the PPML estimator should perform quite well. Third, all estimates of λ are smaller than two. In combination with our Monte Carlo simulations, this implies that the Gamma-PML estimator may not be very appropriate for gravity estimations. Finally, we observe significant heterogeneity in the λ estimates, which range between 0.52 and 1.67. The deviations of λ from one suggest that there is scope for gains from using G-PPML.[45]

Armed with the distribution of λ values, we use G-PPML to obtain a new set of gravity estimates for each of the sectors in our sample. Given the well-established role of bilateral distance as the most important, robust, and widely-used gravity variable, in Figure 3 we zoom in on the difference in our distance estimates.[46] Panel (a) compares the PPML vs. G-PPML distance es-

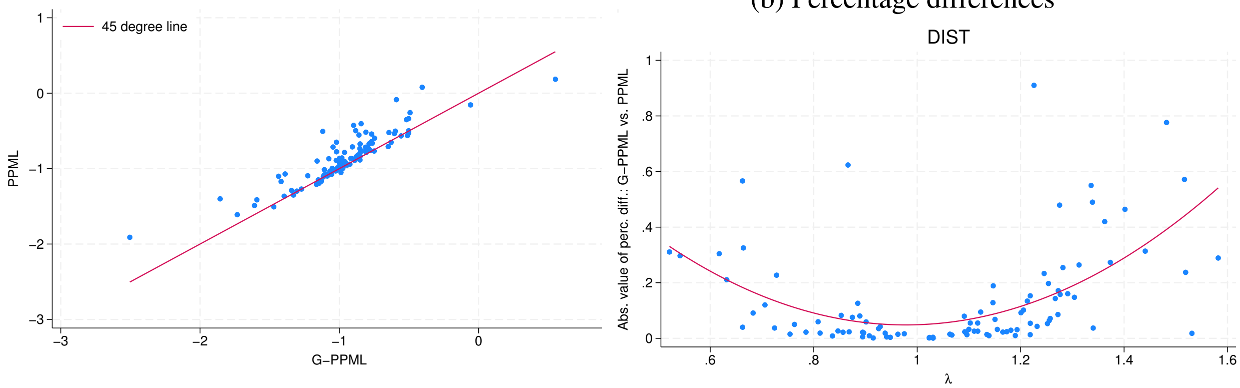
[45] The estimated value of λ using aggregate trade data is 1.31. Gravity estimation as well as the iGMM results using the aggregate trade data are reported in Appendix Tables D.4 and D.5.

[46] Comparisons between the PPML vs. G-PPML estimates for the other gravity variables in our model deliver the same message.

Figure 3: Distance estimates: PPML vs. G-PPML

(a) Coefficient estimates: PPML vs. G-PPML

(b) Percentage differences



Notes: This figure plots PPML vs. G-PPML estimates of the coefficients of distance and their percentage differences. Panel (a) plots the PPML vs. G-PPML distance estimates against each other. Panel (b) plots the sectoral percentage difference in the DIST coefficients ($\%\Delta\beta^{k,DIST}$) estimated with PPML and G-PPML against the corresponding sectoral estimates of λ. To construct panel (b) we drop the top 5 percent of $\%\Delta\beta^{k,DIST}$. See text for further details.

timates directly against each other and reveals that a significant fraction of the estimates are off the 45-degree line, i.e., a significant fraction of the PPML and G-PPML estimates of the effects of distance are different from each other.[47]

To test this hypothesis, we calculate the absolute value of the percentage difference between the sectoral PPML and the corresponding G-PPML estimates for each gravity variable as follows:

$$\%\Delta\beta^{k,v} = \left| \frac{\hat{\beta}^{k,v}_{G\text{-}PPML} - \hat{\beta}^{k,v}_{PPML}}{\hat{\beta}^{k,v}_{PPML}} \right|,$$

where $\hat{\beta}^{k,v}_{PPML}$ is the PPML estimate of the coefficient of gravity variable v for sector k, and $\hat{\beta}^{k,v}_{G\text{-}PPML}$ is the corresponding G-PPML estimate. Our results are reported in panel (b) of Figure 3, which reveals that the more λ deviates from one, the larger the differences between PPML and G-PPML.[48]

Thus far, we have demonstrated that PPML and G-PPML may deliver quite different point

[47]Appendix Figure D.1 confirms this pattern for each of the other gravity variables in our model. According to our theory, the further away the estimates of λ from one, the greater the potential difference between the PPML and G-PPML coefficient estimates.

[48]For exposition, we drop the top 5 percent of the observations in $\%\Delta\beta^{k,DIST}$. The outliers in our analysis are due to the inherent difficulties of percentage differences when dealing with small denominator values. Specifically, when the absolute value of $\hat{\beta}^{k,v}_{PPML}$ is very small, any difference in two estimates translates into a huge percentage difference.
Appendix Figure D.2 confirms this pattern for each of the other gravity variables in our model.

estimates. Another implication of our methods and Monte Carlo analysis is that G-PPML may lead to improved estimation efficiency relative to PPML. To test this hypothesis, we compare the standard errors obtained with the two methods. In addition, since a reduction in standard errors may be less crucial for hypothesis testing if the coefficient estimates experience a similar change, we also examine how relevant G-PPML is for more efficient hypothesis testing by comparing the z-statistics obtained with PPML and G-PPML. To this end, we construct two additional indices to compare the efficiency of the two methods:

$$\%\Delta SE^{k,v} = \frac{SE^{k,v}_{G\text{-}PPML} - SE^{k,v}_{PPML}}{SE^{k,v}_{PPML}} \quad \text{and} \quad \%\Delta z^{k,v} = \frac{\left|z^{k,v}_{G\text{-}PPML}\right| - \left|z^{k,v}_{PPML}\right|}{\left|z^{k,v}_{PPML}\right|}.$$

We do not take the absolute value for $\%\Delta SE^{k,v}$, since the term is negative (positive) when the standard errors of G-PPML are less (greater) than those of PPML. However, we do take the absolute value for $\%\Delta z^{k,v}$, since the sign of z-statistics is not meaningful for two-sided hypothesis testing.[49]

The results regarding the standard errors and z-statistics are presented in Table 2, and we visualize the full distribution of $\%\Delta SE^{k,v}$ and $\%\Delta z^{k,v}$ in Appendix Figure D.3. In the column labeled "lower SE" in panel A of Table 2, we show the percentage of sectors for which the G-PPML standard errors of the corresponding variable are lower as compared to PPML. Overall, the G-PPML standard errors are lower than the PPML standard errors for about 80% to 90% of the sectors in our sample. The subsequent columns in Table 2 show the key distributional statistics for $\%\Delta SE^{k,v}$. On average, G-PPML standard errors are more than 20% smaller than PPML standard errors, whereas in some extreme cases they are 50% smaller (p10). Panel (a) of Appendix Figure D.3 reinforces these results and offers further evidence for G-PPML's improved estimation efficiency vis-à-vis PPML.

Our findings regarding the z-statistics are presented in panel B of Table 2 and in panel (b)

[49]The formula of $\%\Delta z^{k,v}$ is sensible if $z^{k,v}_{GPPML}$ and $z^{k,v}_{PPML}$ have the same sign. This is true for 96% ($= 901/943$) of the coefficients. Due to the nature of fraction, some z-statistics can attain extremely large values. Thus, we drop top 10% values of $\%\Delta z^{k,v}$ for reporting. All of the results, which are available by request, remain robust even if we use the full distribution of $\%\Delta z^{k,v}$.

Table 2: Comparison of Estimation Efficiency, PPML vs. G-PPML

Panel A. sectoral comparison of standard errors, PPML vs. G-PPML						
variable	lower SE	mean	median	std. dev.	p10	p90
BRDR	93.27%	-0.261	-0.288	0.178	-0.435	-0.102
CLNY	91.43%	-0.202	-0.215	0.154	-0.375	-0.021
CNTG	91.43%	-0.224	-0.241	0.162	-0.430	-0.028
DIST	78.10%	-0.210	-0.244	0.246	-0.519	0.077
DIST-IN	93.27%	-0.265	-0.302	0.179	-0.435	-0.100
EU	86.67%	-0.257	-0.314	0.258	-0.509	0.021
LANG	88.57%	-0.230	-0.240	0.204	-0.481	0.013
RTA	84.76%	-0.234	-0.226	0.216	-0.510	0.036
WTO	81.90%	-0.140	-0.187	0.202	-0.333	0.112
Panel B. sectoral comparison of z-statistics, PPML vs. G-PPML						
variable	greater $\|z\|$	mean	median	std. dev.	p10	p90
BRDR	84.62%	0.298	0.340	0.358	-0.113	0.712
CLNY	73.33%	0.463	0.274	0.922	-0.499	1.531
CNTG	80.00%	0.125	0.176	0.230	-0.142	0.373
DIST	80.95%	0.450	0.339	0.524	-0.079	1.107
DIST-IN	92.31%	0.436	0.505	0.350	0.042	0.853
EU	74.29%	0.095	0.181	0.364	-0.476	0.493
LANG	74.29%	0.385	0.264	0.637	-0.330	1.372
RTA	80.95%	0.524	0.350	0.781	-0.259	1.661
WTO	84.76%	0.131	0.157	0.199	-0.085	0.364

Notes: This table compares the standard errors (panel A) and z-statistics (panel B) obtained with PPML and G-PPML. We use the ITPD-E-R02 data to estimate the coefficients of gravity variables for each of 105 sectors. In panel A, the first column (lower SE) shows the percentage of sectors for which the G-PPML standard errors for the corresponding variables are lower than the PPML standard errors. Subsequent columns in panel A show the distributional statistics for the percentage difference in standard errors. In panel B, the first column (greater $|z|$) shows the percentage of sector for which the G-PPML z-statistics for the corresponding variables are greater than the PPML z-statistics. Subsequent columns in panel B are similar to those in panel A. "std. dev.", "p10" and "p90" denote the standard deviation, 10th percentile value and 90th percentile value of the corresponding distribution, respectively.

of Appendix Figure D.3. Column "greater $|z|$" of Table 2 reveals that, for the majority of cases (ranging from 73% for CLNY to 92% for DIST-IN), the G-PPML z-statistics are greater than the corresponding PPML values. This is consistent with our expectations and with the results regarding the differences between the standard errors of the two estimators. Subsequent columns in panel B show that the z-statistics obtained with G-PPML are, in general, much greater than the corresponding PPML z-statistics – the range varying from 10% for EU to 52% for RTA. The G-PPML z-statistics also feature greater median values. Panel (b) of Appendix Figure D.3 confirms these findings and suggests that G-PPML may help with more efficient hypothesis testing.

5 Conclusion

Owing to the seminal work of Santos Silva and Tenreyro (2006), the PPML estimator has firmly established itself as the leading estimator for trade gravity regressions. Despite the success and popularity of PPML, some researchers have remained skeptical about its efficiency when the CVMR assumption fails (Head and Mayer, 2014). We contribute to this debate by using an iGMM method to estimate the conditional variance à la Hansen and Lee (2021) and propose a new *Generalized* PPML estimator that capitalizes on the estimated conditional variance. Using Monte Carlo analysis and an application with real sectoral trade data, we demonstrate that G-PPML can be an alternative to PPML when the CVMR assumption fails.

References

Anderson, James E. and Eric van Wincoop, "Gravity with Gravitas: A Solution to the Border Puzzle," *American Economic Review*, March 2003, *93* (1), 170–192.

_ , **Mario Larch, and Yoto V. Yotov**, "GEPPML: General Equilibrium Analysis with PPML," *The World Economy*, 2018, *41* (10), 2750–2782.

Bai, Jushan, Sung Hoon Choi, and Yuan Liao, "Feasible Generalized Least Squares for Panel Data with Cross-Sectional and Serial Correlations," *Empirical Economics*, January 2021, *60* (1), 309–326.

Borchert, Ingo, Mario Larch, Serge Shikher, and Yoto V. Yotov, "The International Trade and Production Database for Estimation - Release 2 (ITPD-E-R02)," *Working Paper*, 2022.

Correia, Sergio, Paulo Guimarães, and Tom Zylkin, "Fast Poisson Estimation with High-Dimensional Fixed Effects," *Stata Journal*, 2020, *20* (1), 95–115.

Egger, Peter H., Mario Larch, and Yoto V. Yotov, "Gravity Estimations with Interval Data: Revisiting the Impact of Free Trade Agreements," *Economica*, 2022, *89* (353), 44–61.

Fally, Thibault, "Structural Gravity and Fixed Effects," *Journal of International Economics*, September 2015, *97* (1), 76–85.

Fernández-Val, Iván and Martin Weidner, "Individual and Time Effects in Nonlinear Panel Models with Large N, T," *Journal of Econometrics*, 2016, *1* (192), 291–312.

Gurevich, Tamara and Peter Herman, "The Dynamic Gravity Dataset: 1948-2016," *US International Trade Commission, Office of Economics Working Paper*, 2018.

Hansen, Bruce E. and Seojeong Lee, "Inference for Iterated GMM Under Misspecification," *Econometrica*, 2021, *89* (3), 1419–1447.

Head, Keith and Thierry Mayer, "Gravity Equations: Workhorse,Toolkit, and Cookbook," in Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff, eds., *Handbook of International Economics*, Vol. 4 of *Handbook of International Economics*, Elsevier, January 2014, pp. 131–195.

Jochmans, Koen, "Two-way models for gravity," *Review of Economics and Statistics*, 2017, *99* (3), 478–485.

Lewbel, Arthur and Krishna Pendakur, "Tricks with Hicks: The EASI demand system," *American Economic Review*, 2009, *99* (3), 827–863.

Manning, Willard G and John Mullahy, "Estimating Log Models: To Transform or Not to Transform?," *Journal of Health Economics*, July 2001, *20* (4), 461–494.

Murphy, Kevin M and Robert H Topel, "Estimation and inference in two-step econometric models," *Journal of Business & Economic Statistics*, 2002, *20* (1), 88–97.

Newey, Whitney K., "Efficient Instrumental Variables Estimation of Nonlinear Models," *Econometrica*, 1990, *58* (4), 809–837.

— **and Daniel McFadden**, "Large Sample Estimation and Hypothesis Testing," in "Handbook of Econometrics," Vol. 4, Elsevier, January 1994, pp. 2111–2245.

— **and Kenneth D. West**, "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 1987, *55* (3), 703–708.

— **and** — , "Automatic Lag Selection in Covariance Matrix Estimation," *The Review of Economic Studies*, October 1994, *61* (4), 631–653.

Romano, Joseph P. and Michael Wolf, "Resurrecting Weighted Least Squares," *Journal of Econometrics*, March 2017, *197* (1), 1–19.

Silva, J. M. C. Santos and Silvana Tenreyro, "The Log of Gravity," *The Review of Economics and Statistics*, November 2006, *88* (4), 641–658.

Silva, J. M. C. Santos and Silvana Tenreyro, "The Log of Gravity," *The Review of Economics and Statistics*, November 2006, *88* (4), 641–658.

Silva, J. M. C. Santos and Silvana Tenreyro, "Further Simulation Evidence on the Performance of the Poisson Pseudo-Maximum Likelihood Estimator," *Economics Letters*, August 2011, *112* (2), 220–222.

— **and** — , "The Log of Gravity at 15," *Portuguese Economic Journal*, September 2022, *21* (3), 423–437.

Weidner, Martin and Thomas Zylkin, "Bias and Consistency in Three-Way Gravity Models," *Journal of International Economics*, September 2021, *132*, 103513.

Yang, Yimin and Huili Zhang, "Three-way gravity models with multiplicative unobserved effects," *The Econometrics Journal*, 2023, *26* (3), 422–443.

Yotov, Yoto V., "On the Role of Domestic Trade Flows for Estimating the Gravity Model of Trade," *Contemporary Economic Policy*, 2022, *40* (3), 526–540.

A Proofs

A.1 Proof of Proposition 2.1

The moment equation is $\mathbb{E}\left[x_{ijt}\left(\epsilon_{ijt}^{2}-h\cdot\mu_{ijt}^{\lambda}\right)\right]=0$ following equation (5). The generalized PML estimator's conditional variance depends on the two-dimensional parameter $(h,\lambda)\in\mathbb{R}^{2}$ and $x_{ijt}\in\mathbb{R}^{k}$, thus the model is generally overidentified assuming that the parameters of the conditional mean function $\mu_{ijt}=\exp\left(x_{ijt}'\beta+\gamma_{it}+\eta_{jt}\right)$ are known. Define $m\left(\theta\right)=\mathbb{E}\left[x_{ijt}\left(\epsilon_{ijt}^{2}-h\cdot\mu_{ijt}^{\lambda}\right)\right]$, $W\left(\theta\right)=\mathbb{E}\left[x_{ijt}x_{ijt}'\left(\epsilon_{ijt}^{2}-h\cdot\mu_{ijt}^{\lambda}\right)^{2}\right]$, and $\nu_{ijt}\equiv\log\left(\mu_{ijt}\right)=x_{ijt}'\beta+\gamma_{it}+\eta_{jt}$. Θ is the support of (h,λ) and is compact by Assumption 2.1. The population GMM criterion function is

$$J\left(\bar{\theta},\phi\right)=m\left(\bar{\theta}\right)'W\left(\phi\right)^{-1}m\left(\bar{\theta}\right),$$

and let $g\left(\phi\right)=\arg\min_{\bar{\theta}\in\Theta}J\left(\bar{\theta},\phi\right)$. Since the FOC is

$$\frac{\partial J\left(\bar{\theta},\phi\right)}{\partial\bar{\theta}}=-2\mathbb{E}\begin{bmatrix}\mu_{ijt}^{\bar{\lambda}}x_{ijt}'\\ \bar{h}\cdot\mu_{ijt}^{\bar{\lambda}}\nu_{ijt}x_{ijt}'\end{bmatrix}W\left(\phi\right)^{-1}m\left(\bar{\theta}\right)=0, \tag{A.1}$$

the solution $g\left(\phi\right)=\theta$ uniquely satisfies the FOC under the correctly specified conditional variance. The infeasible sample GMM criterion function is

$$\bar{J}_{N,0}\left(\bar{\theta},\phi\right)=\bar{m}_{N,0}\left(\bar{\theta}\right)'\bar{W}_{N,0}^{-1}\left(\phi\right)\bar{m}_{N,0}\left(\bar{\theta}\right),$$

where

$$\bar{m}_{N,0}\left(\bar{\theta}\right)=\frac{1}{N^{2}}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}x_{ijt}\left(\epsilon_{ijt}^{2}-\bar{h}\cdot\mu_{ijt}^{\bar{\lambda}}\right)$$

$$\bar{W}_{N,0}\left(\phi\right)=\frac{1}{N^{2}}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}x_{ijt}x_{ijt}'\left(\epsilon_{ijt}^{2}-h_{\phi}\cdot\mu_{ijt}^{\lambda_{\phi}}\right)^{2}.$$

We verify that our model setup with Assumption 2.1 satisfies Assumptions 1 and 2 of Hansen and Lee (2021). Assumption 1 of Hansen and Lee (2021) is verified in the following steps. First, the parameter space Θ and the support of x_{ijt} are compact by assumption. Second, $g\left(\phi\right)$ is well-defined to satisfy the FOC (A.1) since $\bar{h}>0$ and $\mu_{ijt}^{\bar{\lambda}}>0$ on Θ. Third, x_{ijt}^{k} denotes the kth element of x_{ijt} and $m^k\left(\bar{\theta}\right)=\mathbb{E}\left[x_{ijt}^k\left(\epsilon_{ijt}^2-\bar{h}\cdot\mu_{ijt}^{\bar{\lambda}}\right)\right]$. Then,

$$\frac{\partial m^k\left(\bar{\theta}\right)}{\partial\bar{\theta}}=\mathbb{E}\begin{bmatrix}\mu_{ijt}^{\bar{\lambda}}x_{ijt}^k\\ \bar{h}\cdot\mu_{ijt}^{\bar{\lambda}}\nu_{ijt}x_{ijt}^k\end{bmatrix}$$

$$\frac{\partial^2 m^k\left(\bar{\theta}\right)}{\partial\bar{\theta}\partial\bar{\theta}'}=\mathbb{E}\begin{bmatrix}0 & \mu_{ijt}^{\bar{\lambda}}\nu_{ijt}x_{ijt}^k\\ \mu_{ijt}^{\bar{\lambda}}\nu_{ijt}x_{ijt}^k & \bar{h}\cdot\mu_{ijt}^{\bar{\lambda}}\nu_{ijt}^2x_{ijt}^k\end{bmatrix},$$

and all elements are uniformly bounded by compactness of x_{ijt}, γ_{it}, and η_{jt}. Fourth, $W\left(\bar{\theta}\right)=\mathbb{E}\left[x_{ijt}x_{ijt}'\left(\epsilon_{ijt}^2-\bar{h}\cdot\mu_{ijt}^{\bar{\lambda}}\right)^2\right]$ is continuously differentiable with respect to $\bar{h}$ and $\bar{\lambda}$. Since

$$\frac{\partial W\left(\bar{\theta}\right)}{\partial\bar{h}}=-\mathbb{E}\left[x_{ijt}x_{ijt}'\left(\epsilon_{ijt}^2-\bar{h}\cdot\mu_{ijt}^{\bar{\lambda}}\right)2\mu_{ijt}^{\bar{\lambda}}\right]$$
$$\frac{\partial W\left(\bar{\theta}\right)}{\partial\bar{\lambda}}=-\mathbb{E}\left[x_{ijt}x_{ijt}'\left(\epsilon_{ijt}^2-\bar{h}\cdot\mu_{ijt}^{\bar{\lambda}}\right)2\bar{h}\mu_{ijt}^{\bar{\lambda}}\nu_{ijt}\right],$$

the derivatives are uniformly bounded by compactness of x_{ijt} and μ_{ijt}. Assumption 2.1-2 states $\mathbb{E}\left[y_{ijt}^{8+\nu}|x_{ijt},\gamma_{it},\eta_{jt}\right]<\infty$, which implies $\mathbb{E}\left[\epsilon_{ijt}^{8+\nu}|x_{ijt},\gamma_{it},\eta_{jt}\right]<\infty$ for some $\nu>0$. Fifth, $W\left(\bar{\theta}\right)$ is positive definite unless there is multicollinearity in x_{ijt}. No multicollinearity is a sufficient condition of Assumption 2.1-3. Then,

$$\begin{aligned}W\left(\bar{\theta}\right)&=\mathbb{E}\left[x_{ijt}x_{ijt}'\left(\epsilon_{ijt}^2-h\cdot\mu_{ijt}^{\lambda}+h\cdot\mu_{ijt}^{\lambda}-\bar{h}\cdot\mu_{ijt}^{\bar{\lambda}}\right)^2\right]\\&=\mathbb{E}\left[x_{ijt}x_{ijt}'\left(Var\left(\epsilon_{ijt}^2|x_{ijt},\gamma_{it},\eta_{jt}\right)+\mathbb{E}\left[\left(h\cdot\mu_{ijt}^{\lambda}-\bar{h}\cdot\mu_{ijt}^{\bar{\lambda}}\right)^2|x_{ijt},\gamma_{it},\eta_{jt}\right]\right)\right]\\&\geq\mathbb{E}\left[x_{ijt}x_{ijt}'Var\left(\epsilon_{ijt}^2|x_{ijt}\right)\right],\end{aligned}$$

and the lower bound's smallest eigenvalue λ_{min} is always strictly positive since the conditional variance $Var\left(\epsilon_{ijt}^2|x_{ijt},\gamma_{it},\eta_{jt}\right)$ is uniformly bounded above zero. Since λ_{min} does not depend on $\bar{\theta}$, $W\left(\bar{\theta}\right)$ is always positive definite. Last, $J\left(g\left(\phi\right),\phi\right)=0$ under the correctly specified conditional variance $h\cdot\mu_{ijt}^{\lambda}$.

Assumption 2.1 directly implies Assumption 2 of Hansen and Lee (2021). First, x_{ijt} are independent across different county pairs (i,j). Second, The compact support of $(x_{ijt},\gamma_{it},\eta_{jt})$ implies that μ_{ijt} is uniformly bounded by a finite number. $m\left(\bar{\theta}\right)$ and $\frac{\partial m\left(\bar{\theta}\right)}{\partial\bar{\theta}}$ are all uniformly bounded on compact support, hence uniformly integrable. Since $\frac{\partial^2 m^k\left(\bar{\theta}\right)}{\partial\bar{\theta}\partial\bar{\theta}'}$ is also uniformly bounded on Θ for all k, $\mathbb{E}\left[\sup_{\bar{\theta}\in\Theta}\left\|\frac{\partial^2}{\partial\bar{\theta}\partial\bar{\theta}'}m\left(\bar{\theta}\right)\right\|^2\right]<\infty$. Assumption 2-3 of Hansen and Lee (2021) holds since

$$\mathbb{E}\left[\sup_{\bar{\theta}\in\Theta}\left\|\frac{\partial^3 m^k\left(\bar{\theta}\right)}{\partial\bar{h}\partial\bar{\lambda}\partial\bar{\theta}'}\right\|\right]=\mathbb{E}\left[\sup_{\bar{\theta}\in\Theta}\left\|\mu_{ijt}^{\bar{\lambda}}\nu_{ijt}^2x_{ijt}^k\right\|\right]<\infty,$$

where both μ_{ijt} and ν_{ijt} are uniformly bounded over $\bar{\theta} \in \Theta$. As our model specification satisfies both Assumptions 1 and 2 of Hansen and Lee (2021), the iterated GMM estimator based on $\mathbb{E}\left[x_{ijt}\left(\epsilon_{ijt}^2 - h \cdot \mu_{ijt}^{\lambda}\right)\right] = 0$ is consistent to θ conditional on true gravity equation parameters β, γ_{it}, and η_{jt}.

Next, we show that

$$\bar{m}_N\left(\bar{\theta}\right) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T} x_{ijt}\left(\hat{\epsilon}_{ijt}^2 - \bar{h} \cdot \hat{\mu}_{ijt}^{\bar{\lambda}}\right)$$

$$\bar{W}_N\left(\phi\right) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T} x_{ijt}x_{ijt}'\left(\hat{\epsilon}_{ijt}^2 - h_\phi \cdot \hat{\mu}_{ijt}^{\lambda_\phi}\right)^2$$

approximate the infeasible sample moment $\bar{m}_{N,0}\left(\bar{\theta}\right)$ and weight matrix $\bar{W}_{N,0}\left(\phi\right)$. $\bar{m}_N\left(\bar{\theta}\right)$ and $\bar{W}_N\left(\phi\right)$ replace the unobservable error term ϵ_{ijt} and conditional mean μ_{ijt} with $\hat{\epsilon}_{ijt}$ and $\hat{\mu}_{ijt} = \exp\left(x_{ijt}'\hat{\beta}^{PPML} + \hat{\gamma}_{it}^{PPML} + \hat{\eta}_{jt}^{PPML}\right)$. The goal is to show that the minimizer of $\bar{J}_N\left(\bar{\theta}, \phi\right) = \bar{m}_N\left(\bar{\theta}\right)'\bar{W}_N^{-1}\left(\phi\right)\bar{m}_N\left(\bar{\theta}\right)$ is not different from that of $\bar{J}_{N,0}\left(\bar{\theta}, \phi\right) = \bar{m}_{N,0}\left(\bar{\theta}\right)'\bar{W}_{N,0}^{-1}\left(\phi\right)\bar{m}_{N,0}\left(\bar{\theta}\right)$. The feasible estimator $\hat{\epsilon}_{ijt} = y_{ijt} - \exp\left(x_{ijt}'\hat{\beta}^{PPML} + \hat{\gamma}_{it}^{PPML} + \hat{\eta}_{jt}^{PPML}\right)$ is close to ϵ_{ijt} as $\hat{\mu}_{ijt}$ approximates the conditional mean $\exp\left(x_{ijt}'\beta + \gamma_{it} + \eta_{jt}\right)$. Using the FOCs of the PPML estimation, we present $\hat{\gamma}_{it}^{PPML} = \hat{r}_1\left(z_{it}; \hat{\beta}^{PPML}\right)$ and $\hat{\eta}_{jt}^{PPML} = \hat{r}_2\left(z_{it}; \hat{\beta}^{PPML}\right)$ for some functions $\hat{r}_1$ and $\hat{r}_2$ of $z_{it} = \left(z_{i1t}', \ldots, z_{iNt}'\right)'$ where $z_{ijt} = \left(y_{ijt}, x_{ijt}'\right)'$. Then,

$$\begin{aligned}
\hat{\epsilon}_{ijt} =& \epsilon_{ijt} + \exp\left(x_{ijt}'\hat{\beta}^{PPML} + \hat{\gamma}_{it}^{PPML} + \hat{\eta}_{jt}^{PPML}\right) - \exp\left(x_{ijt}'\beta + \gamma_{it} + \eta_{jt}\right) \\
=& \epsilon_{ijt} - \mu_{ijt}\left(1 - \exp\left(x_{ijt}'\left(\hat{\beta}^{PPML} - \beta\right) + \left(\hat{\gamma}_{it}^{PPML} - \gamma_{it}\right) + \left(\hat{\eta}_{jt}^{PPML} - \eta_{jt}\right)\right)\right) \\
=& \epsilon_{ijt} - \mu_{ijt}\left(1 - \exp\left(O_P\left(\frac{1}{N}\right)\right)\right),
\end{aligned}$$

and $\hat{\mu}_{ijt} = \mu_{ijt}\left(1 - \exp\left(O_P\left(\frac{1}{N}\right)\right)\right)$. Therefore,

$$\begin{aligned}
\bar{m}_N\left(\bar{\theta}\right) &= \bar{m}_{N,0}\left(\bar{\theta}\right) + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T} x_{ijt}\left(\left(\hat{\epsilon}_{ijt}^2 - \epsilon_{ijt}^2\right) + \bar{h}\left(\hat{\mu}_{ijt}^{\bar{\lambda}} - \mu_{ijt}^{\bar{\lambda}}\right)\right) \\
&= \bar{m}_{N,0}\left(\bar{\theta}\right) + O_P\left(\frac{1}{N}\right),
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial \bar{m}_N\left(\bar{\theta}\right)}{\partial \bar{\theta}'} &= \frac{\partial \bar{m}_{N,0}\left(\bar{\theta}\right)}{\partial \bar{\theta}'} + \begin{bmatrix} \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T} x_{ijt}\left(\hat{\mu}_{ijt}^{\bar{\lambda}} - \mu_{ijt}^{\bar{\lambda}}\right) \\ \frac{\bar{h}}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T} x_{ijt}\left(\hat{\nu}_{ijt}\hat{\mu}_{ijt}^{\bar{\lambda}} - \nu_{ijt}\mu_{ijt}^{\bar{\lambda}}\right) \end{bmatrix} \\
&= \frac{\partial \bar{m}_{N,0}\left(\bar{\theta}\right)}{\partial \bar{\theta}'} + O_P\left(\frac{1}{N}\right).
\end{aligned}$$

Since $\bar{W}_N\left(\phi\right)$ converges to the same limit of $\bar{W}_{N,0}\left(\phi\right)$, $\bar{W}_N\left(\phi\right) \xrightarrow{p} W\left(\phi\right)$. The sample criterion function satisfies $\bar{J}_N\left(\bar{\theta},\phi\right)=J_{N,0}\left(\bar{\theta},\phi\right)+O_P\left(\frac{1}{N}\right)$ since

$$
\begin{aligned}
\bar{J}_N\left(\bar{\theta},\phi\right)&=\bar{m}_N\left(\bar{\theta}\right)'\bar{W}_N^{-1}\left(\phi\right)\bar{m}_N\left(\bar{\theta}\right)\\
&=\left(\bar{m}_{N,0}\left(\bar{\theta}\right)+O_P\left(\frac{1}{N}\right)\right)'\left(\bar{W}_{N,0}\left(\phi\right)+O_P\left(\frac{1}{N}\right)\right)^{-1}\left(\bar{m}_{N,0}\left(\bar{\theta}\right)+O_P\left(\frac{1}{N}\right)\right)\\
&=\left(\bar{m}_{N,0}\left(\bar{\theta}\right)+O_P\left(\frac{1}{N}\right)\right)'\left(\bar{W}_{N,0}^{-1}\left(\phi\right)+O_P\left(\frac{1}{N}\right)\right)\left(\bar{m}_{N,0}\left(\bar{\theta}\right)+O_P\left(\frac{1}{N}\right)\right)\\
&=\bar{J}_{N,0}\left(\bar{\theta},\phi\right)+O_P\left(\frac{1}{N}\right),
\end{aligned}
$$

where the third equality follows the Woodbury matrix identity. The result implies that

$$
\begin{aligned}
\sup_{\bar{\theta}\in\Theta}\left|\bar{J}_N\left(\bar{\theta},\phi\right)-J\left(\bar{\theta},\phi\right)\right|&=\sup_{\bar{\theta}\in\Theta}\left|\bar{J}_N\left(\bar{\theta},\phi\right)-\bar{J}_{N,0}\left(\bar{\theta},\phi\right)+\bar{J}_{N,0}\left(\bar{\theta},\phi\right)-J\left(\bar{\theta},\phi\right)\right|\\
&\leq\sup_{\bar{\theta}\in\Theta}\left|\bar{J}_N\left(\bar{\theta},\phi\right)-\bar{J}_{N,0}\left(\bar{\theta},\phi\right)\right|+\sup_{\bar{\theta}\in\Theta}\left|\bar{J}_{N,0}\left(\bar{\theta},\phi\right)-J\left(\bar{\theta},\phi\right)\right|\\
&\xrightarrow{p}0,
\end{aligned}
$$

where the uniform convergence properties follow Theorem 2.6 of Newey and McFadden (1994). The estimator $\hat{\theta}=\left(\hat{h},\hat{\lambda}\right)$ approximates the minimizer of $\bar{J}_{N,0}\left(\bar{\theta},\phi\right)$ for a given ϕ, so $\hat{\theta}$ is a consistent estimator by Theorem 3 of Hansen and Lee (2021).

A.1.1 The case with mild misspecification

This subsection considers the case that the conditional variance of ϵ_{ijt} on x_{ijt}, γ_{it}, and η_{jt} is mildly misspecified, i.e., $\mathbb{E}\left[x_{ijt}\left(\epsilon_{ijt}^2-h_\phi\cdot\mu_{ijt}^{\lambda_\phi}\right)\right]\neq 0$ for any $\phi\in\Theta$. The misspecification includes the case of our Monte Carlo simulation in Appendix B. Suppose that $\left|\mathbb{E}\left[x_{ijt}\left(\epsilon_{ijt}^2-h_\phi\cdot\mu_{ijt}^{\lambda_\phi}\right)\right]\right|\geq\mathfrak{M}$ for all $\phi\in\Theta$, so $\mathfrak{M}$ is the degree of misspecification. Note that Assumption 1.6 of Hansen and Lee (2021) provides a sufficient condition for the existence of the iGMM estimator:

$$
\sup_{\phi\in\Theta}J\left(g\left(\phi\right),\phi\right)<\frac{C_3^2}{4C_1C_2},
$$

where

$$
\begin{aligned}
C_1&=\sup_{\phi\in\Theta}\left\|Q\left(g\left(\phi\right)\right)'W\left(\phi\right)^{-1}Q\left(g\left(\phi\right)\right)\right\|\\
C_2&=\sup_{\phi\in\Theta}\left\|S\left(g\left(\phi\right)\right)'\left(W\left(\phi\right)^{-1}\otimes W\left(\phi\right)^{-1}\right)S\left(g\left(\phi\right)\right)\right\|\\
C_3&=\inf_{\phi\in\Theta}\left\|\frac{\partial}{\partial\bar{\theta}\partial\bar{\theta}'}J\left(\bar{\theta},\phi\right)|_{\bar{\theta}=g(\phi)}\right\|,
\end{aligned}
$$

and the corresponding components in the current paper are

$$Q\left(g\left(\phi\right)\right)=\mathbb{E}\left[\begin{array}{c}\mu_{ijt}^{\lambda_{\phi}}x_{ijt}'\\ h_{\phi}\cdot\mu_{ijt}^{\lambda_{\phi}}\nu_{ijt}x_{ijt}'\end{array}\right]'$$

$$S\left(g\left(\phi\right)\right)=\mathbb{E}\left[\begin{array}{c}2\left(\epsilon_{ijt}^{2}-h_{\phi}\cdot\mu_{ijt}^{\lambda_{\phi}}\right)\mu_{ijt}^{\lambda_{\phi}}vec\left(x_{ijt}x_{ijt}'\right)'\\ 2\left(\epsilon_{ijt}^{2}-h_{\phi}\cdot\mu_{ijt}^{\lambda_{\phi}}\right)h_{\phi}\mu_{ijt}^{\lambda_{\phi}}\nu_{ijt}vec\left(x_{ijt}x_{ijt}'\right)'\end{array}\right]',$$

and

$$\begin{aligned}&vec\left(\frac{\partial}{\partial\bar{\theta}\partial\bar{\theta}'}J\left(\bar{\theta},\phi\right)|_{\bar{\theta}=g(\phi)}\right)\\ &=2\left[\begin{array}{c}\mathbb{E}\left[x_{ijt}\mu_{ijt}^{\lambda_{\phi}}\right]'\mathbb{E}\left[x_{ijt}\mu_{ijt}^{\lambda_{\phi}}\right]\\ \mathbb{E}\left[x_{ijt}\mu_{ijt}^{\lambda_{\phi}}\nu_{ijt}\right]'\mathbb{E}\left[x_{ijt}\left(\mu_{ijt}^{\lambda_{\phi}}+\left(\epsilon_{ijt}^{2}-h_{\phi}\cdot\mu_{ijt}^{\lambda_{\phi}}\right)\right)\right]\\ \mathbb{E}\left[x_{ijt}\mu_{ijt}^{\lambda_{\phi}}\nu_{ijt}\right]'\mathbb{E}\left[x_{ijt}\left(\mu_{ijt}^{\lambda_{\phi}}+\left(\epsilon_{ijt}^{2}-h_{\phi}\cdot\mu_{ijt}^{\lambda_{\phi}}\right)\right)\right]\\ \mathbb{E}\left[x_{ijt}\mu_{ijt}^{\lambda_{\phi}}\nu_{ijt}\right]'\mathbb{E}\left[x_{ijt}\mu_{ijt}^{\lambda_{\phi}}\nu_{ijt}\right]+\mathbb{E}\left[x_{ijt}h_{\phi}\mu_{ijt}^{\lambda_{\phi}}\nu_{ijt}^{2}\right]\mathbb{E}\left[x_{ijt}\left(\epsilon_{ijt}^{2}-h_{\phi}\cdot\mu_{ijt}^{\lambda_{\phi}}\right)\right]\end{array}\right].\end{aligned}$$

Under the uniform boundedness assumption from Assumption 2.1, we know that $C_1>0$ is well-defined. The sensitivity of the weight matrix C_2 and the Hessian matrix C_3 depends on the moments $\mathbb{E}\left[x_{ijt}\mu_{ijt}^{\lambda_{\phi}}\right]$, $\mathbb{E}\left[x_{ijt}\mu_{ijt}^{\lambda_{\phi}}\nu_{ijt}\right]$, and $\mathbb{E}\left[x_{ijt}\mu_{ijt}^{\lambda_{\phi}}\nu_{ijt}^{2}\right]$, which are all uniformly bounded by Assumption 2.1. Thus, the iGMM estimator exists even with a certain level of misspecification.

For an illustration, assume that $\mathbb{E}\left[x_{ijt}\left(\epsilon_{ijt}^{2}-h_{\phi}\cdot\mu_{ijt}^{\lambda_{\phi}}\right)\right]=\mathfrak{M}$ for a fixed weight matrix. Let $\mathbb{E}\left[x_{ijt}\mu_{ijt}^{\lambda_{\phi}}\right]=L_1\left(\mu_{ijt}\right)=L_1$, $\mathbb{E}\left[x_{ijt}\mu_{ijt}^{\lambda_{\phi}}\nu_{ijt}\right]=L_2\left(\mu_{ijt}\right)\leq c_1L_1$ and $\mathbb{E}\left[x_{ijt}\mu_{ijt}^{\lambda_{\phi}}\nu_{ijt}^{2}\right]=L_3\left(\mu_{ijt}\right)\leq c_2L_1$ for some positive constants c_1 and c_2, $h_\phi=1$, and $W\left(\phi\right)=I$ satisfy the the iGMM existence conditions C_1, C_2 and C_3. Then, a lower bound of $C_3^2/\left(4C_1C_2\right)$ is presented by a function of L_1. Since $C_3^2\geq4\left(L_1'L_1\right)^2$ and $C_1C_2\leq\left(\left(1+c_1^2\right)L_1'L_1\right)\left(4\mathfrak{M}^2\left(1+c_1^2\right)L_1'L_1\right)$, the iGMM estimator converges to a limit if

$$\mathfrak{M}^2<\frac{1}{2+2c_1^2}$$

and the upper bound for the degree of misspecification $\mathfrak{M}$ depends on the upper bound of ν_{ijt}. The example confirms that if the moment condition is too far from zero, it may not guarantee the convergence of the proposed iGMM estimator. In general, as L_1, L_2, and L_3 are all functions of μ_{ijt}, the maximum allowable degree of misspecification also depends on μ_{ijt}.

A.1.2 The case with three-way fixed effects

The recent strand of literature extends the gravity model with two-way fixed effects to the one with three-way fixed effects, considering the importer-exporter pair fixed effects. The subsection explores the applicability of the results from the two-way fixed effects specification to the three-way fixed effects case. In summary, the extension to the three-way fixed effects requires additional attention for both testing the CVMR assumption and deriving the asymptotic properties of β.

Following Weidner and Zylkin (2021), consider the model with $\mathbb{E}\left[y_{ijt}|x_{ijt},\gamma_{it},\eta_{jt},\delta_{ij}\right] = \exp\left(x'_{ijt}\beta+\gamma_{it}+\eta_{jt}+\delta_{ij}\right)$, where δ_{ij} is the importer-exporter pair fixed effects. Proposition 1 of Weidner and Zylkin (2021) demonstrates the PPML estimator's consistency with three-way fixed effects under the general assumptions addressed in the current paper. The PPML estimator $\hat{\delta}_{ij}^{PPML}$ for δ_{ij} is given by a function of $(\beta,\gamma_{it},\eta_{jt})$, so the same set of parameters used for two-way fixed effects also works for three-way fixed effects. Simply put, the parameters used for two-way fixed effects are the same set of parameters used for three-way fixed effects.

Thus, redefine $\hat{\mu}_{ijt} = \exp\left(x'_{ijt}\hat{\beta}^{PPML}+\hat{\gamma}_{it}^{PPML}+\hat{\eta}_{jt}^{PPML}+\hat{\delta}_{ij}^{PPML}\right)$ and $\hat{\epsilon}_{ijt} = y_{ijt}-\hat{\mu}_{ijt}$, and $\hat{\theta} = \arg\min_{\bar{\theta}\in\Theta} \bar{J}_{N,0}\left(\bar{\theta},\phi\right)$. Then, $\hat{\theta}$ is still a consistent estimator of θ, which includes the information on the exponent λ describing the conditional variance $Var\left(\epsilon_{ijt}|x_{ijt},\gamma_{it},\eta_{jt},\delta_{ij}\right) = h\cdot\mu_{ijt}^{\lambda}$. The case of mildly misspecified conditional variance also applies to the three-way fixed effects case. $\hat{\theta}$ converges to the pseudo-true value $\theta^{*} = (h^{*},\lambda^{*})$ that minimizes the weighted distance between $\hat{\epsilon}_{ijt}^{2}$ and $h^{*}\cdot\mu_{ijt}^{\lambda^{*}}$.

A.2 Proof of Proposition 2.2

The asymptotic normality of $\hat{\theta}$ follows from the standard theory on the large-sample properties of an overidentified GMM estimator. Since the initial PPML estimator does not suffer from the IPP issue and remains consistent regardless of the conditional variance, our moment condition focuses on the conditional variance parameters, $\theta = (h,\lambda)$, rather than the gravity model parameters. Considering potential misspecification in conditional variance, we adopt the asymptotic theory from Hansen and Lee (2021). If the conditional variance is correctly specified, $m(\theta^*) = 0$ uniquely at $\theta^* = \theta \in \Theta$. The iterated GMM estimator $\hat{\theta}$ satisfies

$$0 = \frac{1}{2} \frac{\partial J\left(\bar{\theta}, \hat{\theta}\right)}{\partial \bar{\theta}} \Bigg|_{\bar{\theta}=\hat{\theta}} = \bar{F}_N\left(\hat{\theta}\right) = \bar{Q}_N\left(\hat{\theta}\right)' \bar{W}_N^{-1}\left(\hat{\theta}\right) \bar{m}_N\left(\hat{\theta}\right),$$

where

$$\bar{Q}_N\left(\bar{\theta}\right) = \begin{bmatrix} -\frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^T \hat{\mu}_{ijt}^{\bar{\lambda}} x_{ijt}' \\ -\frac{\bar{h}}{N^2} \sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^T \hat{\mu}_{ijt}^{\bar{\lambda}} \hat{\nu}_{ijt} x_{ijt}' \end{bmatrix}.$$

We borrow some notation of Hansen and Lee (2021), considering potential misspecification of the conditional variance with $m(\theta^*) \neq 0$. Define $\bar{R}_N\left(\bar{\theta}\right) = \frac{\partial}{\partial \bar{\theta}} vec\left(\bar{Q}_N\left(\bar{\theta}\right)'\right)$, $\mathfrak{M} = m(\theta^*)$, $Q = Q(\theta^*)$, $W = W(\theta^*)$, $R\left(\bar{\theta}\right) = \frac{\partial}{\partial \theta'} vec\left(Q\left(\bar{\theta}\right)'\right)$, $R = R(\theta^*)$, $S = S(\theta^*) = \frac{\partial}{\partial \theta'} vec\left(W\left(\theta^*\right)\right)$, $\bar{m}_N = \bar{m}_N(\theta^*)$, $\bar{Q}_N = \bar{Q}_N(\theta^*)$, and $\bar{W}_N = \bar{W}_N(\theta^*)$. Then, $N\left(\hat{\theta} - \theta^*\right) \approx -\bar{H}_N(\theta^*)^{-1} N\bar{F}_N(\theta^*)$, where

$$\begin{aligned} \bar{H}_N\left(\bar{\theta}\right) &= \bar{Q}_N\left(\bar{\theta}\right)' \bar{W}_N^{-1}\left(\bar{\theta}\right) \bar{Q}_N\left(\bar{\theta}\right) + \left(\bar{m}_N\left(\bar{\theta}\right)' \bar{W}_N^{-1}\left(\bar{\theta}\right) \otimes I_k\right) \bar{R}_N\left(\bar{\theta}\right) \\ &\quad - \left(\bar{m}_N\left(\bar{\theta}\right)' \bar{W}_N^{-1}\left(\bar{\theta}\right) \otimes \bar{Q}_N\left(\bar{\theta}\right)' \bar{W}_N^{-1}\left(\bar{\theta}\right)\right) \bar{S}_n\left(\bar{\theta}\right) \\ N\bar{F}_N\left(\bar{\theta}\right) &= N\left(Q'W^{-1}\bar{m}_N\left(\bar{\theta}\right) + \bar{Q}_N\left(\bar{\theta}\right)' W^{-1}\mathfrak{M} - Q'W^{-1}\bar{W}_N\left(\bar{\theta}\right) W^{-1}\mathfrak{M}\right) + o_p(1). \end{aligned}$$

The correctly specified conditional variance simplifies a lot of notation. As verified in Section 2.1, $\bar{H}_N(\theta) \xrightarrow{p} Q'W^{-1}Q$ as $\bar{m}_N(\theta) \xrightarrow{p} 0$ and $\mathfrak{M} = 0$. In the same way, $\mathfrak{M} = 0$ implies $N\bar{F}_N(\theta) = N\left(Q'W^{-1}\bar{m}_N(\theta)\right) + o_P(1)$. Recall that the PPML estimator $\hat{\beta}^{PPML}$ is a consistent estimator, and $\left(\hat{\gamma}_{it}^{PPML}, \hat{\eta}_{jt}^{PPML}\right)$ are functions of $\hat{\beta}^{PPML}$. Then,

$$\begin{aligned} N\bar{m}_N(\theta) &= N\bar{m}_{N,0}(\theta) + \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^T x_{ijt}\left(\left(\hat{\epsilon}_{ijt}^2 - \epsilon_{ijt}^2\right) - h\left(\hat{\mu}_{ijt}^{\lambda} - \mu_{ijt}^{\lambda}\right)\right) \\ &= N\bar{m}_{N,0}(\theta) + \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^T x_{ijt}\left(\zeta_{ijt}\left(\hat{\beta}^{PPML}\right) - \zeta_{ijt}(\beta)\right) \\ &= N\bar{m}_{N,0}(\theta) + \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^T x_{ijt}\left(\zeta_{ijt}'(\beta)'\left(\hat{\beta}^{PPML} - \beta\right)\right) + o_P(1), \end{aligned}$$

where $\zeta_{ijt}(\beta) = \epsilon_{ijt}^2 - h \cdot \mu_{ijt}^{\lambda}$ is a smooth function of β and $\mathbb{E}\left[\zeta_{ijt}(\beta) | x_{ijt}, \gamma_{it}, \eta_{jt}\right] = 0$. The existence of high-order moments following Assumption 2.1-2 and the smoothness of $\zeta_{ijt}(\beta)$ implies that we can linearize $\zeta_{ijt}\left(\hat{\beta}^{PPML}\right) - \zeta_{ijt}(\beta)$ by $\zeta_{ijt}'(\beta)'\left(\hat{\beta}^{PPML} - \beta\right) + o_P\left(\frac{1}{N}\right)$. We denote that $N\left(\hat{\beta}^{PPML}-\beta\right)\xrightarrow{d}\mathcal{N}\left(0,V_{PPML}\right)$ and

$$W_1\equiv\mathbb{E}\left[\sum_{t=1}^{T}x_{ijt}\zeta_{ijt}'\left(\beta\right)'\right]$$

$$R_1\equiv\mathbb{E}\left[\left(\sum_{t=1}^{T}x_{ijt}\zeta_{ijt}'\left(\beta\right)'\left(\hat{\beta}^{PPML}-\beta\right)\right)\left(\sum_{t=1}^{T}x_{ijt}\left(\epsilon_{ijt}^2-h\cdot\mu_{ijt}^{\lambda}\right)\right)'\right],$$

where V_{PPML} is the asymptotic variance matrix of the PPML estimator. Following Theorem 2 of Murphy and Topel (2002), $N\bar{m}_N\left(\theta\right)\xrightarrow{d}\mathcal{N}\left(0,W+W_1V_{PPML}W_1'+R_1+R_1'\right)\sim\mathcal{N}\left(0,W+V\right)$, where $V=W_1V_{PPML}W_1'+R_1+R_1'$ is a function of the PPML estimator. Note that R_1 is simply zero if the PPML estimator is estimated by using an independent sample. The resulting asymptotic distribution is

$$N\left(\hat{\theta}-\theta\right)\xrightarrow{d}N\left(0,\left(Q'W^{-1}Q\right)^{-1}\left(Q'W^{-1}\left(W+V\right)W^{-1}Q\right)\left(Q'W^{-1}Q\right)^{-1}\right),$$

if the conditional variance is correctly specified, i.e., $\mathfrak{M}=0$. The asymptotic variance formula is simplified to the efficient variance estimator $\left(Q'W^{-1}Q\right)^{-1}$ without the effect of the preliminary estimator $\hat{\beta}^{PPML}$.

The asymptotic distribution enables researchers to conduct a test of the CVMR assumption. If the t-statistic for $H_0:\lambda=1$ is far from zero, we can reject the CVMR assumption, and the G-PPML is likely to perform better. The correctly specified case implies that the asymptotic variance of the G-PPML estimator is similar to the asymptotic variance of the efficient GMM estimator $\left(Q'W^{-1}Q\right)^{-1}$, but not exactly the same due to approximation errors of $\hat{\epsilon}_{ijt}$ and $\hat{\mu}_{ijt}$ from the first stage estimator. The zero asymptotic bias property of $\hat{\beta}^{PPML}$ directly contributes to the zero asymptotic bias of $\hat{\theta}$.

The proof highlights that the existence of the preliminary estimator $\hat{\beta}^{PPML}$ is essential to presenting the asymptotic properties of $\hat{\theta}$. The main model in this paper looks at two-way fixed effects, so there are a few different options for the preliminary estimator of β, including the Gaussian PML and Gamma PML estimators. Note that flexibility in choosing the preliminary estimator does not work in the case of three-way fixed effects. As we did in Section A.1, we discuss more generalized cases, considering the mild misspecification and three-way fixed effects.

A.2.1 The case with mild misspecification

We also allow for some degree of conditional variance misspecification. Under the mild misspecification discussed in Section A.1, the asymptotic variance includes the misspecification-related terms. We additionally assume that the degree of misspecification $\mathfrak{M}$ satisfies the condition in Section A.1.1. That is, $\mathfrak{M}$ is small enough to ensure that the population criterion function converges to a limit. Then,

$$\bar{H}_N(\theta) \xrightarrow{p} Q'W^{-1}Q + \left(\mathfrak{M}'W^{-1} \otimes I_2\right) R - \left(\mathfrak{M}'W^{-1} \otimes Q'W^{-1}\right) S = H_{\mathfrak{M}},$$

where $\mathfrak{M}$ is the degree of misspecification. Next,

$$N\bar{F}_N(\theta) = N\left(Q'W^{-1}\bar{m}_N(\theta) + \bar{Q}_N(\theta)' W^{-1}\mathfrak{M} - Q'W^{-1}\bar{W}_N(\theta) W^{-1}\mathfrak{M}\right),$$

when the mild misspecification problem exists. Define a new matrix $\Omega_{\mathfrak{M}} = \frac{1}{N^2}\sum_{i,j,t}\mathbb{E}\left[\psi_{ijt}\psi'_{ijt}\right]$, where $Q(x_{ijt}, \theta^*) = \left[x_{ijt}\hat{\mu}_{ijt}^{\lambda^*}, x_{ijt}h^* \cdot \hat{\mu}_{ijt}^{\lambda^*}\log(\hat{\mu}_{ijt})\right]$,

$$\psi_{ijt} = Q'W^{-1}x_{ijt}\left(\hat{\epsilon}_{ijt}^2 - h^* \cdot \hat{\mu}_{ijt}^{\lambda^*}\right) + Q(x_{ijt}, \theta^*)' W^{-1}\mathfrak{M}$$
$$- Q'W^{-1}x_{ijt}x'_{ijt}\left(\hat{\epsilon}_{ijt}^2 - h^* \cdot \hat{\mu}_{ijt}^{\lambda^*}\right)^2 W^{-1}\mathfrak{M},$$

and $V_{\mathfrak{M}} = H_{\mathfrak{M}}^{-1}\Omega_{\mathfrak{M}}H_{\mathfrak{M}}^{-1}$. Note that assumptions for asymptotic normality are already verified. Then, following Theorem 4 of Hansen and Lee (2021),

$$NV_{\mathfrak{M}}^{-1/2}\left(\hat{\theta} - \theta^*\right) \xrightarrow{d} N(0, I_2).$$

The provided asymptotic distribution enables practitioners to test whether the CVMR assumption has empirical evidence, even if there is mild misspecification in the conditional variance of ϵ_{ijt}. For example, suppose the true conditional variance follows $0.5\mu_{ijt}^{0.9} + 0.5\mu_{ijt}^{1.1}$, which violates the assumption of PML class estimators. Still, in spite of the misspecification, the researcher can fit the conditional variance function to the conventional formula from PML estimators and check if the conditional variance is close to the one following the CVMR assumption. If λ^* is close to one, the PPML estimator is not only consistent but also close to an efficient estimator. However, if λ^* is far from one, the G-PPML estimator can be a more efficient estimator, even though the potential misspecification in the conditional variance may require additional bias correction steps.

A.2.2 The case with three-way fixed effects

The proof of Proposition 2.2 applies to the same model with three-way fixed effects. Suppose the preliminary estimator $\hat{\beta}^{PPML}$ is the Weidner and Zylkin (2021)'s PPML estimator with three-way fixed effects. The asymptotic properties of the PPML estimator with two-way fixed effects, $\left(\hat{\beta}^{PPML}, \hat{\gamma}_{it}^{PPML}, \hat{\eta}_{jt}^{PPML}\right)$, hold for the case with three-way fixed effects, since the additional pair fixed effect estimator $\hat{\delta}_{ij}^{PPML}$ is presented by a function of $\left(\hat{\beta}^{PPML}, \hat{\gamma}_{it}^{PPML}, \hat{\eta}_{jt}^{PPML}\right)$. As $\hat{\beta}^{PPML} = \beta + O_P\left(\frac{1}{N}\right)$, we can still find that $N\bar{m}_N(\theta)$ and $N\bar{m}_{N,0}(\theta)$ share the same asymptotic distribution under three-way fixed effects.

Note that the asymptotic normality of $\hat{\theta}$ is more sensitively affected by the preliminary estimator to generate $\bar{m}_N\left(\bar{\theta}\right)$ and $\bar{W}_N(\phi)$. For the gravity model with two-way fixed effects, where other PML class estimators for β are all consistent estimators, any existing estimator can work as the preliminary estimator. However, Proposition 2 of Weidner and Zylkin (2021) shows that the gravity model with three-way fixed effects only uses the PPML estimator as the consistent estimator under the general assumption of $\mathbb{E}\left[y_{ijt}|x_{ijt}, \gamma_{it}, \eta_{jt}, \delta_{it}\right] = \exp\left(x_{ijt}'\beta + \gamma_{it} + \eta_{jt} + \delta_{it}\right)$. Recall that

$$N\bar{m}_N(\theta) = N\bar{m}_{N,0}(\theta) + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T} x_{ijt}\zeta_{ijt}'(\beta)' N\left(\hat{\beta}^{PPML} - \beta\right) + o_P(1),$$

where $N\bar{m}_{N,0}(\theta) \xrightarrow{d} \mathcal{N}(0, W)$, $\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T} x_{ijt}\zeta_{ijt}'(\beta)' \xrightarrow{p} W_1$ as defined in the two-way fixed effects case, and $N\left(\hat{\beta}^{PPML} - \beta\right) \xrightarrow{d} \mathcal{N}(B_{PPML}, V_{PPML})$ when the preliminary estimator is the PPML estimator. B_{PPML} represents the non-zero asymptotic bias. Under three-way fixed effects, other PML estimators except for the PPML are not even consistent. Theorem 2 of Murphy and Topel (2002) implies that $N\bar{m}_N(\theta) \xrightarrow{d} \mathcal{N}(B, W + V)$ for some non-zero asymptotic bias B and the additional variance V from the preliminary estimator $\hat{\beta}^{PPML}$.

Therefore, the asymptotic distribution of $N\left(\hat{\theta} - \theta\right)$ is a normal distribution that is not centered at zero. After correcting the bias using the numerical approximation of the asymptotic bias (e.g., jackknife bias correction method of Weidner and Zylkin (2021)), practitioners can still test the CVMR assumption.

A.3 Proof of Proposition 2.3

The consistency of $\hat{\beta}$ follows directly from Fernández-Val and Weidner (2016). Proposition 2 of Weidner and Zylkin (2021) establishes the consistency of the infeasible estimator $\tilde{\beta}$ in the three-way fixed effects case, while Section A.3 of the supplementary appendix of Weidner and Zylkin (2021) examines the consistency of $\tilde{\beta}$ in the two-way fixed effects case. $\tilde{\beta}$ is defined as:

$$\tilde{\beta} : \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} x_{ijt} \left(y_{ijt} - \tilde{\mu}_{ijt} \right) \tilde{\mu}_{ijt}^{1-\lambda} = 0,$$

$$\tilde{\gamma}_{it} : \sum_{j=1}^{N} \left(y_{ijt} - \tilde{\mu}_{ijt} \right) \tilde{\mu}_{ijt}^{1-\lambda} = 0,$$

$$\tilde{\eta}_{jt} : \sum_{i=1}^{N} \left(y_{ijt} - \tilde{\mu}_{ijt} \right) \tilde{\mu}_{ijt}^{1-\lambda} = 0,$$

$\tilde{\beta}$ is consistent under the assumption that the conditional variance component λ is known. Based on the FOCs for λ, the proposed G-PPML estimator $\hat{\beta}$ satisfies the FOCs:

$$\hat{\beta} : \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} x_{ijt} \left(y_{ijt} - \hat{\mu}_{ijt} \right) \hat{\mu}_{ijt}^{1-\hat{\lambda}} = 0,$$

$$\hat{\gamma}_{it} : \sum_{j=1}^{N} \left(y_{ijt} - \hat{\mu}_{ijt} \right) \hat{\mu}_{ijt}^{1-\hat{\lambda}} = 0,$$

$$\hat{\eta}_{jt} : \sum_{i=1}^{N} \left(y_{ijt} - \hat{\mu}_{ijt} \right) \hat{\mu}_{ijt}^{1-\hat{\lambda}} = 0,$$

and since $\hat{\mu}_{ijt}^{1-\lambda}$ is uniformly bounded,

$$\begin{aligned} \hat{\mu}_{ijt}^{1-\hat{\lambda}} &= \exp\left(\left(1 - \hat{\lambda}\right) \left(x'_{ijt}\hat{\beta} + \hat{\gamma}_{it} + \hat{\eta}_{jt} \right) \right) \\ &= \exp\left(\left(1 - \lambda - O_P\left(\frac{1}{N}\right) \right) \left(x'_{ijt}\hat{\beta} + \hat{\gamma}_{it} + \hat{\eta}_{jt} \right) \right) \\ &= \exp\left(-O_P\left(\frac{1}{N}\right) \right) \hat{\mu}_{ijt}^{1-\lambda} = \hat{\mu}_{ijt}^{1-\lambda} + O_P\left(\frac{1}{N}\right), \end{aligned}$$

which approximates the infeasible PML estimator. The consistency of $\hat{\lambda}$ implies the consistency of the plug-in estimator $\hat{\beta}$, following Theorem 2.5 of Newey and McFadden (1994).

A.3.1 The case with mild misspecification

Under the two-way fixed effects, other "misspecified" PML estimators are consistent as well. That is, a mild misspecification we discussed in the conditional variance estimation does not affect the consistency of $\hat{\beta}$. Recall $\mathbb{E}\left[y_{ijt} | x_{ijt}, \gamma_{it}, \eta_{jt}\right] = \exp\left(x'_{ijt}\beta + \gamma_{it} + \eta_{jt}\right)$, thus $\mathbb{E}\left[g\left(\mu_{ijt}\right)\epsilon_{ijt}\right] =$

$\mathbb{E}\left[\mathbb{E}\left[g\left(\mu_{ijt}\right)\epsilon_{ijt}|x_{ijt},\gamma_{it},\eta_{jt}\right]\right] = \mathbb{E}\left[g\left(\mu_{ijt}\right)\mathbb{E}\left[\epsilon_{ijt}|x_{ijt},\gamma_{it},\eta_{jt}\right]\right] = 0$ for any function g. The unconditional moment is implied by the variation of equation (3) with a generalized weight function $g\left(\mu_{ijt}\right)$, where the pseudo-likelihood function is

$$\mathcal{L}\left(\beta,\gamma_{it},\eta_{jt}\right) = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\left(y_{ijt}\int\frac{g\left(\mu_{ijt}\right)}{\mu_{ijt}}d\mu_{ijt} - \int g\left(\mu_{ijt}\right)d\mu_{ijt}\right),$$

and the FOCs are

$$\hat{\beta}^{g}: \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}x_{ijt}\left(y_{ijt}-\hat{\mu}_{ijt}^{g}\right)g\left(\hat{\mu}_{ijt}^{g}\right) = 0,$$

$$\hat{\gamma}_{it}^{g}: \sum_{j=1}^{N}\left(y_{ijt}-\hat{\mu}_{ijt}^{g}\right)g\left(\hat{\mu}_{ijt}^{g}\right) = 0,$$

$$\hat{\eta}_{jt}^{g}: \sum_{i=1}^{N}\left(y_{ijt}-\hat{\mu}_{ijt}^{g}\right)g\left(\hat{\mu}_{ijt}^{g}\right) = 0,$$

where $\hat{\mu}_{ijt}^{g} = \exp\left(x_{ijt}'\hat{\beta}^{g}+\hat{\gamma}_{it}^{g}+\hat{\eta}_{jt}^{g}\right)$. Since the number of fixed effects terms grows with N, while the number of observations grows with N^2, $bias\left(\hat{\beta}^{g}\right) = O\left(\frac{1}{N}\right)$ even if $g\left(\mu_{ijt}\right) \neq \mu_{ijt}^{1-\lambda}$. As far as the first-stage iterated GMM estimator $\hat{\theta}$ exists and $\hat{\lambda} \xrightarrow{p} \lambda^{*}$, $\hat{\beta}$ with $g\left(\mu_{ijt}\right) = \mu_{ijt}^{1-\hat{\lambda}}$ is a consistent estimator of β.

A.3.2 The case with three-way fixed effects

Even if we consider the importer-exporter pair fixed effects δ_{ij} and extend the model with the three-way fixed effects, the G-PPML estimator is still consistent. Section A.5 in the supplementary appendix of Weidner and Zylkin (2021) provides consistency for the infeasible version of the G-PPML estimator with three-way fixed effects. That is, the estimator $\tilde{\beta}$ satisfying the FOCs

$$\tilde{\beta}: \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}x_{ijt}\left(y_{ijt}-\tilde{\mu}_{ijt}\right)\tilde{\mu}_{ijt}^{1-\lambda} = 0$$

$$\tilde{\gamma}_{it}: \sum_{j=1}^{N}\left(y_{ijt}-\tilde{\mu}_{ijt}\right)\tilde{\mu}_{ijt}^{1-\lambda} = 0$$

$$\tilde{\eta}_{jt}: \sum_{i=1}^{N}\left(y_{ijt}-\tilde{\mu}_{ijt}\right)\tilde{\mu}_{ijt}^{1-\lambda} = 0$$

$$\tilde{\delta}_{ij}: \sum_{t=1}^{T}\left(y_{ijt}-\tilde{\mu}_{ijt}\right)\tilde{\mu}_{ijt}^{1-\lambda} = 0$$

is a consistent estimator of β if $Var\left(\epsilon_{ijt}^{2}|x_{ijt},\gamma_{it},\eta_{jt},\delta_{ij}\right) = h\cdot\mu_{ijt}^{\lambda}$. The G-PPML estimator with three-way fixed effects simply extends $\tilde{\beta}$ by plugging-in $\hat{\lambda}$ instead of the unknown λ. Following the same argument as the main proof in Section A.3, $\hat{\beta}$ is a consistent estimator as far as $\hat{\lambda}$ is a consistent estimator.

A.4 Proof of Proposition 2.4

Recall that the pseudo-likelihood function is

$$\mathcal{L}\left(\beta, \gamma_{it}, \eta_{jt}; \hat{\lambda}\right) = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\left(y_{ijt}\int \mu_{ijt}^{-\hat{\lambda}} d\mu_{ijt} - \int \mu_{ijt}^{1-\hat{\lambda}} d\mu_{ijt}\right),$$

where $\hat{\lambda}$ is given by the first stage's iGMM estimation. The asymptotic distribution for $\hat{\beta}$ comes from the Taylor series approximation of the first order conditions such that,

$$\begin{aligned} 0 &= \partial_{\beta}\mathcal{L}\left(\hat{\beta}, \hat{\gamma}_{it}\left(\hat{\beta}\right), \hat{\eta}_{jt}\left(\hat{\beta}\right); \hat{\lambda}\right) \\ &\approx \partial_{\beta}\mathcal{L}\left(\beta, \hat{\gamma}_{it}\left(\beta\right), \hat{\eta}_{jt}\left(\beta\right); \hat{\lambda}\right) - \Omega_{\infty} N\left(\hat{\beta}-\beta\right) \\ &\approx \partial_{\lambda\beta}\mathcal{L}\left(\beta, \hat{\gamma}_{it}\left(\beta\right), \hat{\eta}_{jt}\left(\beta\right); \lambda\right) N\left(\hat{\lambda}-\lambda\right) + \partial_{\beta}\mathcal{L}\left(\beta, \hat{\gamma}_{it}\left(\beta\right), \hat{\eta}_{jt}\left(\beta\right); \lambda\right) - \Omega_{\infty} N\left(\hat{\beta}-\beta\right), \end{aligned}$$

where $\partial_{\beta}\mathcal{L}\left(\hat{\beta}, \hat{\gamma}_{it}\left(\hat{\beta}\right), \hat{\eta}_{jt}\left(\hat{\beta}\right); \lambda\right)$ is the first order condition of the infeasible G-PPML estimator, when λ is known to the econometrician. Without considering the estimator $\hat{\lambda}$, $\Omega_{\infty} N\left(\hat{\beta}-\beta\right) \approx \partial_{\beta}\mathcal{L}\left(\beta, \hat{\gamma}_{it}\left(\beta\right), \hat{\eta}_{jt}\left(\beta\right); \lambda\right) \xrightarrow{d} \mathcal{N}\left(0, \Omega_{\infty}\right)$ implies that $N\left(\hat{\beta}-\beta\right) \xrightarrow{d} \mathcal{N}\left(0, \Omega_{\infty}^{-1}\right)$. The derived asymptotic variance is a direct application of Theorem 4.1 of Fernández-Val and Weidner (2016).

Since we verify the asymptotic distribution of $N\left(\hat{\lambda}-\lambda\right)$ in Proposition 2.2, let

$$\partial_{\lambda\beta}\mathcal{L}\left(\beta, \hat{\gamma}_{it}\left(\beta\right), \hat{\eta}_{jt}\left(\beta\right); \lambda\right) N\left(\hat{\lambda}-\lambda\right) + \partial_{\beta}\mathcal{L}\left(\beta, \hat{\gamma}_{it}\left(\beta\right), \hat{\eta}_{jt}\left(\beta\right); \lambda\right) \xrightarrow{d} \mathcal{N}\left(\Omega_{\infty}B_{\beta}, \Omega_{\infty}+V_{\lambda}\right),$$

where $\partial_{\beta}\mathcal{L}\left(\beta, \hat{\gamma}_{it}\left(\beta\right), \hat{\eta}_{jt}\left(\beta\right); \lambda\right) \xrightarrow{d} \mathcal{N}\left(\Omega_{\infty}B_{\beta}, \Omega_{\infty}\right)$ and V_{λ} is derived by the first term of above equation. Since the asymptotic distribution of $N\left(\hat{\lambda}-\lambda\right)$ is centered at zero, the first term does not affect the mean but possibly affects the asymptotic variance. V_{λ} is a function of the asymptotic variance of $\hat{\lambda}$. Note that under two-way fixed effects, $\partial_{\lambda\beta}\mathcal{L}\left(\beta, \hat{\gamma}_{it}\left(\beta\right), \hat{\eta}_{jt}\left(\beta\right); \lambda\right) = o_P\left(1\right)$ since the derivative $\frac{1}{N^2}\sum_{i,j,t} x_{ijt}\left(y_{ijt}-\mu_{ijt}\right)\nu_{ijt}\mu_{ijt}^{1-\lambda}$ converges to zero in probability as long as $\mathbb{E}\left[y_{ijt}-\mu_{ijt}|x_{ijt}, \gamma_{it}, \eta_{jt}\right] = 0$. Since any PML estimator is basically consistent under two-way fixed effects, the change in λ does not affect the first order condition of the likelihood function $\mathcal{L}$ asymptotically. While $N\left(\hat{\lambda}-\lambda\right) = O_P\left(1\right)$, $V_{\lambda} = 0$. Thus,

$$N\left(\hat{\beta}-\beta\right) \xrightarrow{d} \mathcal{N}\left(B_{\beta}, \Omega_{\infty}^{-1}\right),$$

where B_{β} denotes the asymptotic bias.

For the next step, we show that the asymptotic bias B_β is zero in our setup. The asymptotic bias following Weidner and Zylkin (2021) is the probability limit of $\Omega_N^{-1}(B_N + D_N)/N$. Note that the mth elements of B_N and D_N are denoted by B_N^m and D_N^m, satisfying

$$B_N^m = -\frac{1}{N}\sum_{i=1}^{N} Tr\left[\left(\sum_{j=1}^{N}\bar{H}_{ij}\right)^{-}\sum_{j=1}^{N}\mathbb{E}\left[H_{ij}\tilde{x}_{ij,m}S'_{ij}|x_{ij},\gamma_i,\eta_j\right]\right]$$
$$+\frac{1}{2N}\sum_{i=1}^{N} Tr\left[\left(\sum_{j=1}^{N}\bar{G}_{ij}\tilde{x}_{ij,m}\right)\left(\sum_{j=1}^{N}\bar{H}_{ij}\right)^{-}\sum_{j=1}^{N}\mathbb{E}\left[S_{ij}S'_{ij}|x_{ij},\gamma_i,\eta_j\right]\left(\sum_{j=1}^{N}\bar{H}_{ij}\right)^{-}\right],$$

and

$$D_N^m = -\frac{1}{N}\sum_{j=1}^{N} Tr\left[\left(\sum_{i=1}^{N}\bar{H}_{ij}\right)^{-}\sum_{i=1}^{N}\mathbb{E}\left[H_{ij}\tilde{x}_{ij,m}S'_{ij}|x_{ij},\gamma_i,\eta_j\right]\right]$$
$$+\frac{1}{2N}\sum_{j=1}^{N} Tr\left[\left(\sum_{i=1}^{N}\bar{G}_{ij}\tilde{x}_{ij,m}\right)\left(\sum_{i=1}^{N}\bar{H}_{ij}\right)^{-}\sum_{i=1}^{N}\mathbb{E}\left[S_{ij}S'_{ij}|x_{ij},\gamma_i,\eta_j\right]\left(\sum_{i=1}^{N}\bar{H}_{ij}\right)^{-}\right].$$

Since Ω_N is a positive definite matrix, the size of the bias relies on B_N^m and D_N^m. In general, both B_N^m and D_N^m are non-zeros and asymptotic bias does not vanish. In the correctly specified case that $Var(\epsilon_{ijt}|x_{ijt},\gamma_{it},\eta_{jt}) = h\cdot\mu_{ijt}^{\lambda}$, $S_{ij,t} = (y_{ijt}-\mu_{ijt})\,\mu_{ijt}^{1-\lambda}$ and $\mathbb{E}\left[S_{ij,t}S_{ij,s}|x_{ij},\gamma_i,\eta_j\right] = \mu_{ijt}^{2-\lambda}1\{t=s\}$. Similarly, $\left[\bar{H}_{ij}\right]_{ts} = \mu_{ijt}^{2-\lambda}1\{t=s\}$ and $\left[\bar{G}_{ij}\right]_{tsr} = (3-2\lambda)\,\mu_{ijt}^{2-\lambda}1\{t=s=r\}$. The second terms of B_N^m and D_N^m can be simplified by

$$\frac{1}{2N}\sum_{i=1}^{N} Tr\left[\left(\sum_{j=1}^{N}\bar{H}_{ij}\right)^{-}\left(\sum_{j=1}^{N}\bar{G}_{ij}\tilde{x}_{ij,m}\right)\right],$$
$$\frac{1}{2N}\sum_{j=1}^{N} Tr\left[\left(\sum_{i=1}^{N}\bar{H}_{ij}\right)^{-}\left(\sum_{i=1}^{N}\bar{G}_{ij}\tilde{x}_{ij,m}\right)\right],$$

since $\mathbb{E}\left[S_{ij,t}S_{ij,s}|x_{ij},\gamma_i,\eta_j\right] = \left[\bar{H}_{ij}\right]_{ts}$.

Next, by definition of $\tilde{x}_{ij,m}$, which minimizes

$$\sum_{i=1}^{N}\sum_{j=1}^{N} Tr\left[\left(x_{ij}-\gamma_i^x-\eta_j^x\right)'\bar{H}_{ij}\left(x_{ij}-\gamma_i^x-\eta_j^x\right)\right]$$

with respect to γ_i^x and η_j^x, we find that $\sum_{j=1}^{N}\bar{G}_{ij}\tilde{x}_{ij,m} = \sum_{j=1}^{N}\bar{H}_{ij}\tilde{x}_{ij,m} = 0$ and $\sum_{i=1}^{N}\bar{G}_{ij}\tilde{x}_{ij,m} = \sum_{i=1}^{N}\bar{H}_{ij}\tilde{x}_{ij,m} = 0$ if the CVMR assumption holds, i.e., if $\lambda = 1$. But even if the λ value is not equal to 1, $\sum_{j=1}^{N}\bar{G}_{ij}\tilde{x}_{ij,m} = (3-2\lambda)\sum_{j=1}^{N}\bar{H}_{ij}\tilde{x}_{ij,m} = 0$ for the infeasible G-PPML estimator because $\sum_{j=1}^{N}\bar{H}_{ij}\tilde{x}_{ij,m} = 0$. Thus, as far as the DGP follows the assumption of any PML estimator, the second terms of B_N^m and D_N^m are zeros.

The first terms of B_N^m and D_N^m are also zeros if the conditional variance is correctly specified. That is, $\sum_{i=1}^N \mathbb{E}\left[H_{ij}\tilde{x}_{ij,m}S'_{ij}|x_{ij},\gamma_i,\eta_j\right] = \sum_{j=1}^N \mathbb{E}\left[H_{ij}\tilde{x}_{ij,m}S'_{ij}|x_{ij},\gamma_i,\eta_j\right] = 0$ and hence there is no asymptotic bias. Note that the mth column of the matrix $H_{ij}\tilde{x}_{ij,m}S'_{ij}$ is zero when $\lambda = 1$ (PPML). If $\lambda \neq 1$, the formula becomes

$$\mathbb{E}\left[\tilde{x}_{ij,m}\mu_{ijt}^{2-2\lambda}(1-\lambda)\left(y_{ijt} - \frac{2-\lambda}{1-\lambda}\mu_{ijt}\right)(y_{ijt}-\mu_{ijt})|x_{ij},\gamma_i,\eta_j\right], \tag{A.2}$$

so we compute whether these column elements are zeros or not. Hereafter, suppose $\lambda \neq 1$. Since $\mathbb{E}\left[(y_{ijt}-\mu_{ijt})^2|x_{ijt},\gamma_{it},\eta_{jt}\right] = h \cdot \mu_{ijt}^{\lambda}$ under the correctly specified conditional variance and $\mathbb{E}\left[y_{ijt}-\mu_{ijt}|x_{ijt},\gamma_{it},\eta_{jt}\right] = 0$, equation (A.2) becomes

$$\begin{aligned}&\tilde{x}_{ij,m}\mu_{ijt}^{2-2\lambda}(1-\lambda)\,\mathbb{E}\left[\left(y_{ijt} - \frac{2-\lambda}{1-\lambda}\mu_{ijt}\right)(y_{ijt}-\mu_{ijt})|x_{ij},\gamma_i,\eta_j\right]\\ &= h(1-\lambda)\cdot\tilde{x}_{ij,m}\mu_{ijt}^{2-\lambda}.\end{aligned}$$

Therefore,

$$\sum_{i=1}^N \mathbb{E}\left[H_{ij}\tilde{x}_{ij,m}S'_{ij}|x_{ij},\gamma_i,\eta_j\right] = h(1-\lambda)\sum_{i=1}^N \tilde{x}_{ij,m}\mu_{ijt}^{2-\lambda},$$

and similarly, $h(1-\lambda)\sum_{j=1}^N \tilde{x}_{ij,m}\mu_{ijt}^{2-\lambda} = 0$. Note that by the FOCs for the definition of $\tilde{x}_{ij}$,

$$\sum_{j=1}^N\sum_{i=1}^N \mathbb{E}\left[H_{ij}\tilde{x}_{ij,m}S'_{ij}|x_{ij},\gamma_i,\eta_j\right] = h(1-\lambda)\sum_{j=1}^N\sum_{i=1}^N \tilde{x}_{ij,m}\mu_{ijt}^{2-\lambda} = 0.$$

Thus, we confirm that the first terms of B_N^m and D_N^m are zeros.

The last equation shows that B_N^m and D_N^m are always zero, when $\lambda = 1$, so the asymptotic bias is zero regardless of the true formula of the conditional variance. But even if $\lambda \neq 1$ and $h(1-\lambda) \neq 0$, $\sum_{j=1}^N\sum_{i=1}^N \tilde{x}_{ij,m}\mu_{ijt}^{2-\lambda} = 0$ since we are assumed to know the true value of λ and use the infeasible G-PPML. Thus, under the class of PML estimators, when we have a reliable estimate of λ, we do not have asymptotic bias if λ is known, thus no bias correction is needed. For general conditional variances that do not follow the form of $h\cdot\mu_{ijt}^{\lambda}$, $\sum_{i=1}^N \mathbb{E}\left[H_{ij}\tilde{x}_{ij,m}S'_{ij}|x_{ij},\gamma_i,\eta_j\right]$ nor $\sum_{j=1}^N \mathbb{E}\left[H_{ij}\tilde{x}_{ij,m}S'_{ij}|x_{ij},\gamma_i,\eta_j\right]$ will be guaranteed to be zero, causing non-zero asymptotic bias of the estimator. The asymptotic bias also exists when the true λ and the supposed λ value selected by the researcher are different. This confirms why the existing PML estimators except for PPML estimator have asymptotic bias issues in the general form of conditional variance.

Considering potential clustering and serial correlation, the robust asymptotic variance estimator is a traditional sandwich form $W_N^{-1}\Omega_N W_N^{-1}$. In this case, the asymptotic variance estimation requires an estimator for $\mathbb{E}\left[S_{ij,t}S_{ij,s}|x_{ij},\gamma_i,\eta_j\right]$ for $t \neq s$. As $\hat{\mathbb{E}}\left[\hat{S}_{ij,t}\hat{S}_{ij,s}|x_{ij},\gamma_i,\eta_j\right]$ is the potential source of finite-sample downward bias, the (finite-sample) bias correction method proposed by Weidner and Zylkin (2021) is still beneficial.

A.4.1 The case with mild misspecification

If the conditional variance of y_{ijt} does not follow equation (2), the asymptotic distribution of $N\left(\hat{\beta}-\beta\right)$ suffers from a non-zero asymptotic bias. While a mild misspecification of the conditional variance does not affect the consistency of the G-PPML estimator, the researcher should correct the asymptotic bias for a valid inference in β. Even if the researcher correctly specifies the conditional variance, the asymptotic bias correction becomes necessary when considering the gravity model with three-way fixed effects, as we will discuss in the next subsection.

Since the first-stage plug-in estimator $\hat{\lambda}$ influences the asymptotic variance only, we focus on the case with asymptotic normality of infeasible G-PPML estimator under misspecified conditional variance.

The asymptotic bias is non-zero since $\mathbb{E}\left[H_{ij}\tilde{x}_{ij,m}S'_{ij}|x_{ij},\gamma_i,\eta_j\right]$ is non-zero under misspecification. Suppose a $1\times T$ matrix presenting the conditional variance of $\epsilon_{ij}=(\epsilon_{ij1},\ldots,\epsilon_{ijT})'$ is $\nu\left(\mu_{ij}\right)=\left(Var\left(\epsilon_{ij1}|x_{ij1},\gamma_{i1},\eta_{j1}\right),\ldots,Var\left(\epsilon_{ijT}|x_{ijT},\gamma_{iT},\eta_{jT}\right)\right)$, without loss of generality. Also, define $\mu_{ij}=Diag\left(\mu_{ij1},\ldots,\mu_{ijT}\right)$. Then,

$$\mathbb{E}\left[H_{ij}\tilde{x}_{ij,m}S'_{ij}|x_{ij},\gamma_i,\eta_j\right]=\mu_{ij}^{2-2\lambda}\left(1-\lambda\right)\tilde{x}_{ij,m}\nu\left(\mu_{ij}\right).$$

In general functional form $\nu\left(\mu_{ij}\right)$ where the conditional variance is unknown, therefore, both components $\sum_{j=1}^{N}\mu_{ij}^{2-2\lambda}\left(1-\lambda\right)\tilde{x}_{ij,m}\nu\left(\mu_{ij}\right)\neq 0$ and $\sum_{i=1}^{N}\mu_{ij}^{2-2\lambda}\left(1-\lambda\right)\tilde{x}_{ij,m}\nu\left(\mu_{ij}\right)\neq 0$, so the asymptotic bias terms do not disappear.

Based on the asymptotic bias derived by Fernández-Val and Weidner (2016) and Weidner and Zylkin (2021), we propose an inferential method with the asymptotic bias correction. The provided asymptotic bias formula offers a feasible method to correct the asymptotic bias of $\hat{\beta}$. First, we estimate $\bar{G}_{ij}$ by $\left[\hat{\bar{G}}_{ij}\right]_{tsr}=\left(3-2\hat{\lambda}\right)\hat{\mu}_{ijt}^{2-\hat{\lambda}}1\left\{t=s=r\right\}$, and similarly, $\bar{H}_{ij}$ is estimated by $\left[\hat{\bar{H}}_{ij}\right]_{ts}=\hat{\mu}_{ijt}^{1-\hat{\lambda}}\left(1-\hat{\lambda}\right)\left(y_{ijt}-\frac{2-\hat{\lambda}}{1-\hat{\lambda}}\hat{\mu}_{ijt}\right)1\left\{t=s\right\}$. Also, $\left[\hat{S}_{ij}\right]_t=\left(y_{ijt}-\hat{\mu}_{ijt}\right)\hat{\mu}_{ijt}^{1-\hat{\lambda}}$. Then, the estimated B_N and D_N's mth columns are given by replacing the population moments with their sample analogs.

$$\hat{B}_N^m = -\frac{1}{N}\sum_{i=1}^{N} Tr\left[\left(\sum_{j=1}^{N}\hat{\hat{H}}_{ij}\right)^{-}\sum_{j=1}^{N}\hat{H}_{ij}\hat{\tilde{x}}_{ij,m}\hat{S}'_{ij}\right]$$
$$+\frac{1}{2N}\sum_{i=1}^{N} Tr\left[\left(\sum_{j=1}^{N}\hat{\hat{G}}_{ij}\hat{\tilde{x}}_{ij,m}\right)\left(\sum_{j=1}^{N}\hat{\hat{H}}_{ij}\right)^{-}\sum_{j=1}^{N}\hat{S}_{ij}\hat{S}'_{ij}\left(\sum_{j=1}^{N}\hat{\hat{H}}_{ij}\right)^{-}\right],$$

and

$$\hat{D}_N^m = -\frac{1}{N}\sum_{j=1}^{N} Tr\left[\left(\sum_{i=1}^{N}\hat{\hat{H}}_{ij}\right)^{-}\sum_{i=1}^{N}\hat{H}_{ij}\hat{\tilde{x}}_{ij,m}\hat{S}'_{ij}\right]$$
$$+\frac{1}{2N}\sum_{j=1}^{N} Tr\left[\left(\sum_{i=1}^{N}\hat{\hat{G}}_{ij}\hat{\tilde{x}}_{ij,m}\right)\left(\sum_{i=1}^{N}\hat{\hat{H}}_{ij}\right)^{-}\sum_{i=1}^{N}\hat{S}_{ij}\hat{S}'_{ij}\left(\sum_{i=1}^{N}\hat{\hat{H}}_{ij}\right)^{-}\right],$$

where $\hat{\tilde{x}}_{ij} = x_{ij} - \hat{\gamma}_i^x - \hat{\eta}_j^x$ with $\hat{\gamma}_i^x$ and $\hat{\eta}_j^x$ minimizing

$$\sum_{i=1}^{N}\sum_{j=1}^{N} Tr\left[\left(x_{ij}-\gamma_i^x-\eta_j^x\right)'\hat{\hat{H}}_{ij}\left(x_{ij}-\gamma_i^x-\eta_j^x\right)\right],$$

and $\hat{H}_{ij}\hat{\tilde{x}}_{ij,m}\hat{S}'_{ij} = \hat{\mu}_{ij}^{2-2\lambda}\left(1-\lambda\right)\hat{\tilde{x}}_{ij,m}\nu\left(\hat{\mu}_{ij}\right)$.

Next, in order to compute the asymptotic bias correction term, define a $k\times k$ matrix $\Xi_N = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\tilde{x}'_{ij}Var\left(S_{ij}|x_{ij},\gamma_i,\eta_j\right)\tilde{x}_{ij}$, where $Var\left(S_{ijt}|x_{ijt},\gamma_{it},\eta_{jt}\right) = \mu_{ij}^{2-2\lambda}Diag\left(\nu\left(\mu_{ij}\right)\right)$. The empirical measure of Ξ_N is

$$\hat{\Xi}_N = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\tilde{x}}'_{ij}\hat{\mu}_{ij}^{2-2\hat{\lambda}}Diag\left(\nu\left(\hat{\mu}_{ij}\right)\right)\hat{\tilde{x}}_{ij},$$

and $\Xi_\infty = \lim_{N\to\infty}\Xi_N$. Note that $\hat{\Xi}_N^{-1}\left(\hat{B}_N+\hat{D}_N\right)$ is the estimated asymptotic bias. As a result,

$$N\left(\hat{\beta}-\beta-\frac{\hat{\Xi}_N^{-1}\left(\hat{B}_N+\hat{D}_N\right)}{N}\right)\xrightarrow{d}\mathcal{N}\left(0,\Omega_\infty^{-1}\Xi_\infty\Omega_\infty^{-1}\right).$$

A.4.2 The case with three-way fixed effects

We extend the asymptotic normality of $\hat{\beta}$ for the gravity model with three-way fixed effects. As Weidner and Zylkin (2021) discussed, even the PPML estimator suffers from the asymptotic bias under three-way fixed effects. So, it is not surprising that the G-PPML inherits a non-zero asymptotic bias, considering the special case $\lambda = 1$. Another notable difference from the case of two-way fixed effects is that the consistency of $\hat{\beta}$ sensitively responds to the value of λ. This causes an additional inefficiency in the asymptotic properties of the three-way fixed effects.

The first order conditions in the three-way fixed effects satisfy

$$\begin{aligned}
0 &= \partial_{\beta}\mathcal{L}\left(\hat{\beta}, \hat{\gamma}_{it}\left(\hat{\beta}\right), \hat{\eta}_{jt}\left(\hat{\beta}\right), \hat{\delta}_{ij}\left(\hat{\beta}\right); \hat{\lambda}\right) \\
&\approx \partial_{\lambda\beta}\mathcal{L}\left(\beta, \hat{\gamma}_{it}\left(\beta\right), \hat{\eta}_{jt}\left(\beta\right), \hat{\delta}_{ij}\left(\beta\right); \lambda\right) N\left(\hat{\lambda}-\lambda\right) \\
&\quad + \partial_{\beta}\mathcal{L}\left(\beta, \hat{\gamma}_{it}\left(\beta\right), \hat{\eta}_{jt}\left(\beta\right), \hat{\delta}_{ij}\left(\beta\right); \hat{\lambda}\right) - \Omega_{\infty} N\left(\hat{\beta}-\beta\right),
\end{aligned}$$

where $\partial_{\beta}\mathcal{L}\left(\beta, \hat{\gamma}_{it}\left(\beta\right), \hat{\eta}_{jt}\left(\beta\right), \hat{\delta}_{ij}\left(\beta\right); \hat{\lambda}\right) \xrightarrow{d} \mathcal{N}\left(B_{\beta}, \Omega_{\infty}\right)$ follows Proposition 3 of Weidner and Zylkin (2021). Under three-way fixed effects, however, $\hat{\beta}$ is not a consistent estimator unless λ is the true exponent value of the conditional variance, so $\partial_{\lambda\beta}\mathcal{L}\left(\beta, \hat{\gamma}_{it}\left(\beta\right), \hat{\eta}_{jt}\left(\beta\right), \hat{\delta}_{ij}\left(\beta\right); \lambda\right) \xrightarrow{p} D_{\lambda}$ does not converge to zero. Next, section A.2.2 shows that $N\left(\hat{\theta}-\theta\right) \xrightarrow{d} \mathcal{N}\left(B_{\theta}, V_{\theta}\right)$, where $B_{\theta} = \left(Q'W^{-1}Q\right)^{-1} Q'W^{-1}B$ and $V_{\theta} = \left(Q'W^{-1}Q\right)^{-1}\left(Q'W^{-1}\left(W+V\right)W^{-1}Q\right)\left(Q'W^{-1}Q\right)^{-1}$. B is non-zero since the PPML estimator $\hat{\beta}^{PPML}$ has non-zero asymptotic bias under three-way fixed effects, and V is a function of V_{PPML}. Therefore,

$$\partial_{\lambda\beta}\mathcal{L}\left(\beta, \hat{\gamma}_{it}\left(\beta\right), \hat{\eta}_{jt}\left(\beta\right), \hat{\delta}_{ij}\left(\beta\right); \lambda\right) N\left(\hat{\lambda}-\lambda\right) \xrightarrow{d} \mathcal{N}\left(D_{\lambda}\left[B_{\theta}\right]_{2}, D_{\lambda}\left[V_{\theta}\right]_{2,2} D_{\lambda}'\right),$$

where $\left[B_{\theta}\right]_{2}$ is the second element of B_{θ} and $\left[V_{\theta}\right]_{2,2}$ is the $(2,2)$ component of the matrix V_{θ}.

The asymptotic distribution of $\hat{\beta}$ is eventually non-trivial. If the distributions of $\hat{\lambda}$ and $\hat{\beta}$ are independent,

$$N\left(\hat{\beta}-\beta\right) \xrightarrow{d} \mathcal{N}\left(\Omega_{\infty}^{-1}\left(B_{\beta}+B_{\lambda}\right), \Omega_{\infty}^{-1}\left(\Omega_{\infty}+V_{\lambda}\right)\Omega_{\infty}^{-1}\right),$$

where $B_{\lambda} = D_{\lambda}\left[B_{\theta}\right]_{2}$ and $V_{\lambda} = D_{\lambda}\left[V_{\theta}\right]_{2,2} D_{\lambda}'$. Thus, the valid inference on the gravity parameter β requires a bias correction, either a Jackknife bias correction or a numerical approximation of the bias. Consider an empirical analog of B_{λ} by $\hat{B}_{\lambda} = \hat{D}_{\lambda}\left[\hat{B}_{\theta}\right]_{2}$, and we focus on figuring out B_{β}.

Following Proposition 3 of Weidner and Zylkin (2021), we obtain a similar result as the case with mild misspecification. We simply recalculate the asymptotic bias and asymptotic variance components using the same notation but different formulae.

Recall that our pseudo-likelihood function with three-way fixed effects is

$$\mathcal{L}\left(\beta, \gamma_{it}, \eta_{jt}, \delta_{ij}\right) = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\left(y_{ijt}\int \mu_{ijt}^{-\hat{\lambda}} d\mu_{ijt} - \int \mu_{ijt}^{1-\hat{\lambda}} d\mu_{ijt}\right),$$

where $\hat{\lambda}$ is the estimated exponent of the conditional variance $Var\left(\epsilon_{ijt}|x_{ijt}, \gamma_{it}, \eta_{jt}, \delta_{ij}\right) = h \cdot \mu_{ijt}^{\lambda}$ and $\mu_{ijt} = \exp\left(x_{ijt}'\beta + \gamma_{it} + \eta_{jt} + \delta_{ij}\right)$. Using the same strategy as Weidner and Zylkin (2021), we first solve for the optimal δ_{ij} in order to present the likelihood function as a function of β, γ_{it}, and η_{jt}. Thus, we can treat the gravity model with three-way fixed effects in the same context as the one with two-way fixed effects. Computing the FOCs,

$$\exp\left(\hat{\delta}_{ij}\left(\beta,\gamma_{it},\eta_{jt}\right)\right)=\frac{\sum_{t=1}^{T}y_{ijt}\xi_{ijt}^{1-\hat{\lambda}}}{\sum_{t=1}^{T}\xi_{ijt}^{2-\hat{\lambda}}},$$

where $\xi_{ijt}=\exp\left(x_{ijt}'\beta+\gamma_{it}+\eta_{jt}\right)$. Thus, the pseudo-likelihood function becomes

$$\mathcal{L}\left(\beta,\gamma_{it},\eta_{jt}\right)=\max_{d_{ij}}\mathcal{L}\left(\beta,\gamma_{it},\eta_{jt},d_{ij}\right)=\sum_{i=1}^{N}\sum_{j=1}^{N}\ell_{ij}\left(\beta,\gamma_{it},\eta_{jt}\right),$$

with $\ell_{ij}\left(\beta,\gamma_{it},\eta_{jt}\right)=\left(\sum_{t=1}^{T}y_{ijt}\xi_{ijt}^{1-\hat{\lambda}}\right)^{2-\hat{\lambda}}/\left(\left(1-\hat{\lambda}\right)\left(2-\hat{\lambda}\right)\sum_{t=1}^{T}\xi_{ijt}^{2-\hat{\lambda}}\right)$ for any $\hat{\lambda}\neq1$ and $\hat{\lambda}\neq2$. The likelihood function follows the PPML's likelihood if $\hat{\lambda}=1$ and the Gamma PML's likelihood if $\hat{\lambda}=2$. Using the derived likelihood function component ℓ_{ij}, we define the corresponding score vector, the Hessian matrix, and the cubic tensor such that $\left[S_{ij}\right]_t=\partial\ell_{ij}/\partial\alpha_{it}$, $\left[H_{ij}\right]_{ts}=-\partial^2\ell_{ij}/\partial\alpha_{it}\partial\alpha_{is}$, and $\left[G_{ij}\right]_{tsr}=\partial^3\ell_{ij}/\partial\alpha_{it}\partial\alpha_{is}\partial\alpha_{ir}$. For example, the score vector is a T-dimensional vector with

$$\left[S_{ij}\right]_t=y_{ijt}\left(\frac{\xi_{ijt}\sum_{s=1}^{T}y_{ijs}\xi_{ijs}^{1-\hat{\lambda}}}{\sum_{s=1}^{T}\xi_{ijs}^{2-\hat{\lambda}}}\right)^{1-\hat{\lambda}}-\left(\frac{\xi_{ijt}\sum_{s=1}^{T}y_{ijs}\xi_{ijs}^{1-\hat{\lambda}}}{\sum_{s=1}^{T}\xi_{ijs}^{2-\hat{\lambda}}}\right)^{2-\hat{\lambda}},$$

and the Hessian matrix and cubic tensor can be generated accordingly. Following the same notation as the main proof, the asymptotic normality of G-PPML under the three-way fixed effects is verified with some non-zero asymptotic bias. When the conditional variance is correctly specified by $Var\left(\epsilon_{ijt}|x_{ijt},\gamma_{it},\eta_{jt},\delta_{ij}\right)=h\cdot\mu_{ijt}^{\lambda}$, we obtain

$$N\left(\hat{\beta}-\beta-\frac{\hat{\Omega}_N^{-1}\left(\hat{B}_\beta+\hat{B}_\lambda\right)}{N}\right)\xrightarrow{d}\mathcal{N}\left(0,\Omega_\infty^{-1}\left(\Omega_\infty+V_\lambda\right)\Omega_\infty^{-1}\right),$$

where $\hat{B}_\beta=\hat{B}_N+\hat{D}_N$.

Note that $\hat{\beta}$'s asymptotic variance is not exactly the same as the efficient asymptotic variance matrix, even after correcting for the asymptotic bias. The inefficiency V_λ comes from the first-stage estimator $\hat{\lambda}$'s variance, while the $\hat{\lambda}$'s variance also depends on the initial PPML estimator's asymptotic variance. The result implies that under the three-way fixed effects, the suggested G-PPML estimator might not be necessarily more efficient than the one-step PPML estimator.

A.4.3 Comparing the conditional variances of the G-PPML and other PML-class estimators

Without loss of generality, suppose $Var\left(\epsilon_{ijt}|x_{ijt},\gamma_{it},\eta_{jt}\right)=\mu_{ijt}^{\lambda}$ with $h=1$. Following the result of Proposition 2.4, the conditional variance of the (infeasible) G-PPML estimator is $\left(\tilde{x}'\mu^{2-\lambda}\tilde{x}\right)^{-1}$ and that of other PML estimators assuming some specific values of λ^* follow the functional form

$$\left(\tilde{x}'\mu^{2-\lambda^*}\tilde{x}\right)^{-1'}\left(\tilde{x}'\mu^{2+\lambda-2\lambda^*}\tilde{x}\right)\left(\tilde{x}'\mu^{2-\lambda^*}\tilde{x}\right)^{-1},$$

where $\tilde{x}'\mu^{\alpha}\tilde{x}=\sum_{i,j}\tilde{x}_{ij}'\mu_{ij}^{\alpha}\tilde{x}_{ij}$ and $\mu_{ij}^{\alpha}=diag\left(\mu_{ij1}^{\alpha},\ldots,\mu_{ijT}^{\alpha}\right)$ for some $\alpha\in\mathbb{R}_{+}$. If $\lambda^*=1$, the estimator is the PPML estimator, and if $\lambda^*=2$, the estimator is the Gamma PML estimator.

Our goal is to show that the PML estimator's conditional variance is minimized when $\lambda^*=\lambda$, i.e., when the estimator's underlying assumption on the conditional variance matches the true DGP.

Note that

$$\begin{aligned}&\left(\tilde{x}'\mu^{2-\lambda^*}\tilde{x}\right)^{-1'}\left(\tilde{x}'\mu^{2+\lambda-2\lambda^*}\tilde{x}\right)\left(\tilde{x}'\mu^{2-\lambda^*}\tilde{x}\right)^{-1}-\left(\tilde{x}'\mu^{2-\lambda}\tilde{x}\right)^{-1}\\ =\;&\left(\tilde{x}'\mu^{2-\lambda^*}\tilde{x}\right)^{-1'}\tilde{x}'\left[\mu^{2+\lambda-2\lambda^*}-\mu^{2-\lambda^*}\tilde{x}\left(\tilde{x}'\mu^{2-\lambda}\tilde{x}\right)^{-1}\tilde{x}'\mu^{2-\lambda^*}\right]\tilde{x}\left(\tilde{x}'\mu^{2-\lambda^*}\tilde{x}\right)^{-1}\\ =\;&\tilde{G}'\left(I-\mu^{1-0.5\lambda}\tilde{x}\left(\tilde{x}'\mu^{2-\lambda}\tilde{x}\right)^{-1}\tilde{x}'\mu^{1-0.5\lambda}\right)\tilde{G},\end{aligned}$$

where $\tilde{G}'=\left(\tilde{x}'\mu^{2-\lambda^*}\tilde{x}\right)^{-1'}\tilde{x}'\mu^{1+0.5\lambda-\lambda^*}$. That is, the conditional variance of a PML-class estimator is greater than that of the G-PPML estimator if $I-\mu^{1-0.5\lambda}\tilde{x}\left(\tilde{x}'\mu^{2-\lambda}\tilde{x}\right)^{-1}\tilde{x}'\mu^{1-0.5\lambda}$ in the last equality is a positive semi-definite matrix. If we define $\tilde{X}=\mu^{1-0.5\lambda}\tilde{x}$, the last equality of the above equation is $\tilde{G}'\left(I-\tilde{X}\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'\right)\tilde{G}$, where $I-\tilde{X}\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'$ is a symmetric and idempotent matrix, thus positive semi-definite. This verifies that the G-PPML's conditional variance is the lower bound of other PML estimators' conditional variances when our assumption $Var\left(\epsilon_{ijt}|x_{ijt},\gamma_{it},\eta_{jt}\right)=\mu_{ijt}^{\lambda}$ is satisfied as specified.

B iGMM and G-PPML under Misspecification

We extend the Monte Carlo analysis with several experiments that investigate the possibility that the error term could be misspecified. Specifically, we consider additional cases where the conditional variance can be described as a *perturbation* to the PPML assumption.[50] For the first three cases (M1 through M3), we consider a conditional variance structure given by $\mathrm{Var}(y|x) = \mathbb{E}(y|x) + h_2\mathbb{E}(y|x)^0$, where $h_2 \in \{0.2, 0.4, 0.6\}$. In the next three cases (M4 through M6), we specify the conditional variance as $\mathrm{Var}(y|x) = \mathbb{E}(y|x) + h_2\mathbb{E}(y|x)^2$, where $h_2 \in \{0.1, 0.2, 0.3\}$. The objective is to compare the performance of different estimators when there is model misspecification in the error term.

Our findings appear in Table B.1. For cases M1 through M3, the average λ estimates are between 0 and 1, as expected. Moreover, the estimates of λ become smaller as h_2 increases from 0.2 to 0.6. This suggests that the conditional variance increasingly resembles the $h_2\mathbb{E}(y|x)^0$ structure, and our iGMM is successful at detecting the change. Similarly, for cases M4 through M6, the λ estimates are between 1 and 2 and, as expected, the λs deviate further away from 1 as h_2 increases. The intuitive shifts in the λ estimates offer reassuring evidence for the robustness of the iGMM estimator to potential model misspecification (Hansen and Lee, 2021).

Turning to the mean bias and standard errors, cases M1 through M3 reveal that G-PPML and PPML produce very similar low mean bias, with slightly lower standard errors in favor of G-PPML. The advantages of G-PPML in terms of mean bias and standard errors become more pronounced in cases M4 and M6, especially as h_2 allows for the quadratic term to increase. A natural explanation for this result is that the performance gap between G-PPML and PPML throughout these cases can be attributed to the quadratic perturbation being much more impactful than adding a constant term $h_2\mathbb{E}(y|x)^0$.

G-PPML and PPML outperform Gamma-PML and OLS, both in terms of mean bias and standard errors, in cases M1 through M3. This is expected, as the increment of h_2 in these cases means that the assumptions of Gamma-PML and OLS are further violated. In cases M4 through M6, however, both the mean bias and the standard errors of Gamma-PML decline vis-à-vis those of G-PPML and PPML as h_2 increases. This suggests that the conditional variance increasingly conforms to the assumption of Gamma-PML. The OLS standard errors are relatively low, but the

[50]We only consider a small perturbation to the PPML assumption since we attempt to gauge the advantage of G-PPML relative to PPML without placing PPML at an unfair starting point. When the DGP deviates greatly from the PPML assumption, the efficiency gain from G-PPML will only be more pronounced.

mean bias remains high in all cases.

The Monte Carlo analysis reveals that the G-PPML estimator remains efficient for a wide parameter range and that the performance is stable with potential misspecification in the error term. These results reinforce the key benefit of G-PPML, which is to relieve researchers' "burden of proof" for a particular value of λ, which, when specified incorrectly, can lead to estimation bias and/or estimation efficiency loss.

Table B.1: Mote Carlo Results (Misspecification)

	$J = 50,\ T = 10,$ Obser. = 25 000							$J = 100,\ T = 5,$ Obser. = 50 000						
	$\bar{\hat{\lambda}}$	β_1			β_2			$\bar{\hat{\lambda}}$	β_1			β_2		
Estimator		Bias	S.E.	S.D.	Bias	S.E.	S.D.		Bias	S.E.	S.D.	Bias	S.E.	S.D.
Case M1: $\mathrm{Var}(y_i\|x_i) = 1 \cdot [\mathbb{E}(y_i\|x_i)] + 0.2 \cdot [\mathbb{E}(y_i\|x_i)]^0$														
G-PPML	0.8931	0.0348	0.0427	0.0435	0.0072	0.0089	0.0091	0.9078	0.0240	0.0302	0.0303	0.0051	0.0063	0.0064
PPML		0.0347	0.0436	0.0433	0.0072	0.0090	0.0091		0.0239	0.0305	0.0303	0.0051	0.0063	0.0064
Gamma-PML		0.0404	0.0446	0.0490	0.0127	0.0089	0.0099		0.0272	0.0328	0.0338	0.0075	0.0066	0.0069
OLS		0.0985	0.0434	0.0442	0.0969	0.0087	0.0088		0.0977	0.0304	0.0297	0.0967	0.0061	0.0062
Case M2: $\mathrm{Var}(y_i\|x_i) = 1 \cdot [\mathbb{E}(y_i\|x_i)] + 0.4 \cdot [\mathbb{E}(y_i\|x_i)]^0$														
G-PPML	0.8340	0.0373	0.0444	0.0462	0.0075	0.0093	0.0094	0.8475	0.0256	0.0313	0.0321	0.0052	0.0065	0.0066
PPML		0.0370	0.0453	0.0459	0.0075	0.0094	0.0094		0.0256	0.0317	0.0320	0.0053	0.0066	0.0066
Gamma-PML		0.0425	0.0465	0.0514	0.0140	0.0093	0.0101		0.0299	0.0343	0.0365	0.0084	0.0069	0.0073
OLS		0.1118	0.0452	0.0464	0.1099	0.0091	0.0091		0.1114	0.0316	0.0321	0.1099	0.0063	0.0063
Case M3: $\mathrm{Var}(y_i\|x_i) = 1 \cdot [\mathbb{E}(y_i\|x_i)] + 0.6 \cdot [\mathbb{E}(y_i\|x_i)]^0$														
G-PPML	0.7854	0.0377	0.0460	0.0473	0.0080	0.0096	0.0100	0.7951	0.0264	0.0324	0.0330	0.0054	0.0068	0.0068
PPML		0.0374	0.0469	0.0469	0.0080	0.0098	0.0100		0.0263	0.0328	0.0329	0.0054	0.0068	0.0068
Gamma-PML		0.0440	0.0482	0.0528	0.0156	0.0097	0.0108		0.0310	0.0356	0.0380	0.0091	0.0071	0.0075
OLS		0.1235	0.0468	0.0467	0.1216	0.0094	0.0094		0.1229	0.0328	0.0329	0.1217	0.0066	0.0064
Case M4: $\mathrm{Var}(y_i\|x_i) = 1 \cdot [\mathbb{E}(y_i\|x_i)] + 0.1 \cdot [\mathbb{E}(y_i\|x_i)]^2$														
G-PPML	1.1618	0.0372	0.0460	0.0469	0.0078	0.0094	0.0098	1.1907	0.0263	0.0327	0.0328	0.0054	0.0067	0.0069
PPML		0.0377	0.0476	0.0474	0.0078	0.0097	0.0098		0.0265	0.0334	0.0330	0.0055	0.0068	0.0069
Gamma-PML		0.0402	0.0460	0.0497	0.0110	0.0092	0.0101		0.0286	0.0337	0.0353	0.0069	0.0067	0.0071
OLS		0.0792	0.0446	0.0445	0.0770	0.0089	0.0090		0.0777	0.0312	0.0309	0.0771	0.0062	0.0063
Case M5: $\mathrm{Var}(y_i\|x_i) = 1 \cdot [\mathbb{E}(y_i\|x_i)] + 0.2 \cdot [\mathbb{E}(y_i\|x_i)]^2$														
G-PPML	1.2882	0.0403	0.0501	0.0504	0.0084	0.0102	0.0105	1.3227	0.0287	0.0358	0.0360	0.0058	0.0073	0.0073
PPML		0.0412	0.0527	0.0516	0.0084	0.0107	0.0106		0.0295	0.0370	0.0368	0.0059	0.0075	0.0074
Gamma-PML		0.0427	0.0490	0.0526	0.0113	0.0098	0.0107		0.0302	0.0360	0.0375	0.0071	0.0072	0.0075
OLS		0.0748	0.0474	0.0468	0.0727	0.0095	0.0095		0.0740	0.0332	0.0324	0.0725	0.0066	0.0067
Case M6: $\mathrm{Var}(y_i\|x_i) = 1 \cdot [\mathbb{E}(y_i\|x_i)] + 0.3 \cdot [\mathbb{E}(y_i\|x_i)]^2$														
G-PPML	1.3764	0.0436	0.0538	0.0544	0.0090	0.0109	0.0113	1.4121	0.0311	0.0384	0.0391	0.0063	0.0078	0.0079
PPML		0.0454	0.0574	0.0567	0.0092	0.0115	0.0115		0.0324	0.0403	0.0406	0.0064	0.0081	0.0080
Gamma-PML		0.0452	0.0518	0.0563	0.0116	0.0104	0.0113		0.0319	0.0382	0.0400	0.0074	0.0076	0.0080
OLS		0.0716	0.0498	0.0496	0.0687	0.0100	0.0100		0.0704	0.0349	0.0344	0.0688	0.0070	0.0069

Notes: This table shows the Monte Carlo results that compare different estimators when there is misspecification in the error term structure. We report the average λ estimates, mean absolute bias and the standard error of the coefficient estimates. G-PPML indicates the generalized PPML estimator proposed in this paper, PPML denotes Poisson-Pseudo Maximum Likelihood estimator, Gamma-PML denotes Gamma Pseudo Maximum Likelihood, and OLS denotes ordinary least squares estimation after taking the natural logarithm of the dependent variable. β_1 and β_2 are the coefficients for a continuous variable and a dummy variable, respectively.

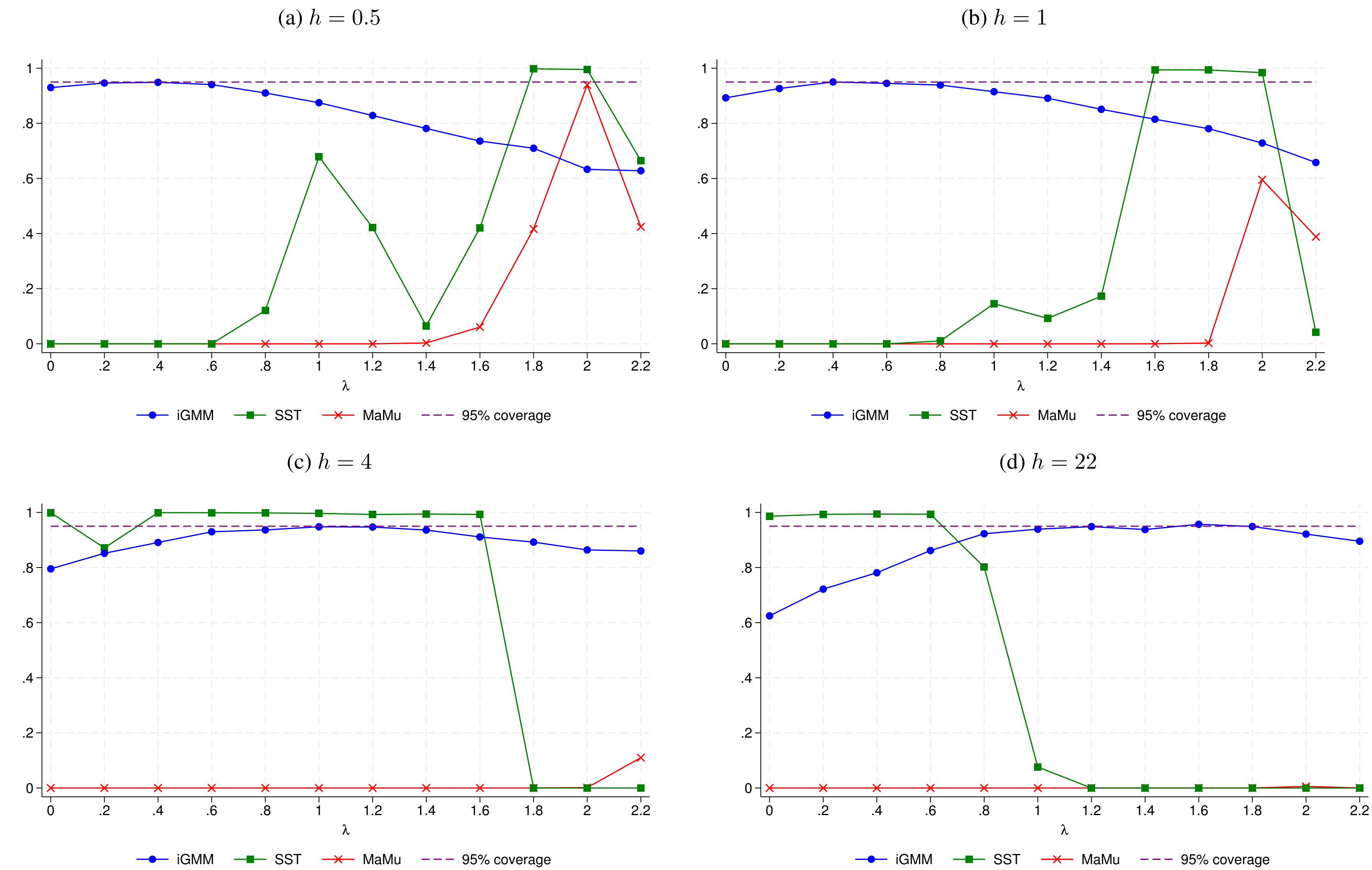
C A Comparison of the Coverage of the Confidence Intervals in Section 3.2

Figure C.1 compares the coverage of 95% confidence intervals using three different approaches: iterated GMM (iGMM), the method proposed by Silva and Tenreyro (2006), and the approach by Manning and Mullahy (2001). We use subplots to illustrate the performance of these methods under varying levels of absolute noise. Specifically, we selected $h = 0.5, 1$, and 4 to align with our main Monte Carlo analysis and included $h = 22$ to reflect the median value of h in our sectoral empirical analysis with actual trade data.

For iGMM, the confidence interval contains the true value of λ in approximately 95% of simulations when both λ and h are low ($h = 0.5, 1$). In contrast, the coverage approaches 95% for relatively higher values of λ when h becomes higher ($h = 4, 22$). This discrepancy occurs because, when both h and λ are very low, small variations in h closely mimic those in λ, making it challenging for iGMM to accurately distinguish between the two. Conversely, as the absolute data noise increases, h and λ become more distinguishable for intermediate values of λ, though they remain difficult to separate for low or high values of λ. While Figure C.1 does not explore this further, we hypothesize that for even higher values of h, it may become challenging to identify h from lower or intermediate values of λ.

Figure C.1 clearly demonstrates that the performance of iGMM is significantly more reliable than the existing methods. The SST method's coverage is either 1 (due to disproportionately large standard deviation) or 0 (due to inaccurate point estimates). The MaMu approach performs well only when the level of data noise is low ($h = 0.5$) and $\lambda = 2$. Therefore, we conclude that the iGMM method provides confidence intervals with superior coverage compared to other existing methods.

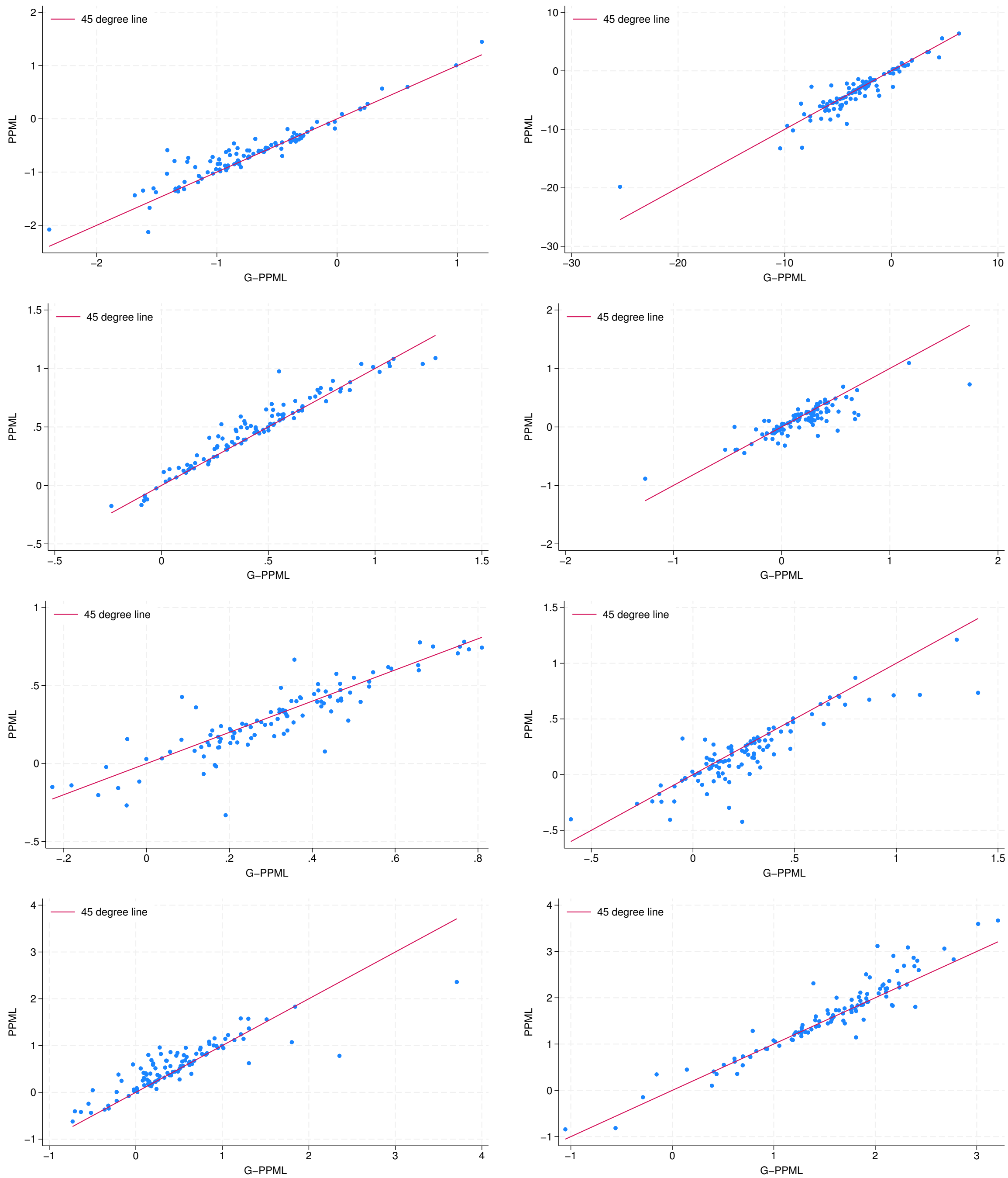
Figure C.1: Coverage of Confidence Intervals for λ



Notes: These figures compare the coverage of 95% confidence intervals constructed with different estimation methods. See text in Appendix C for details.

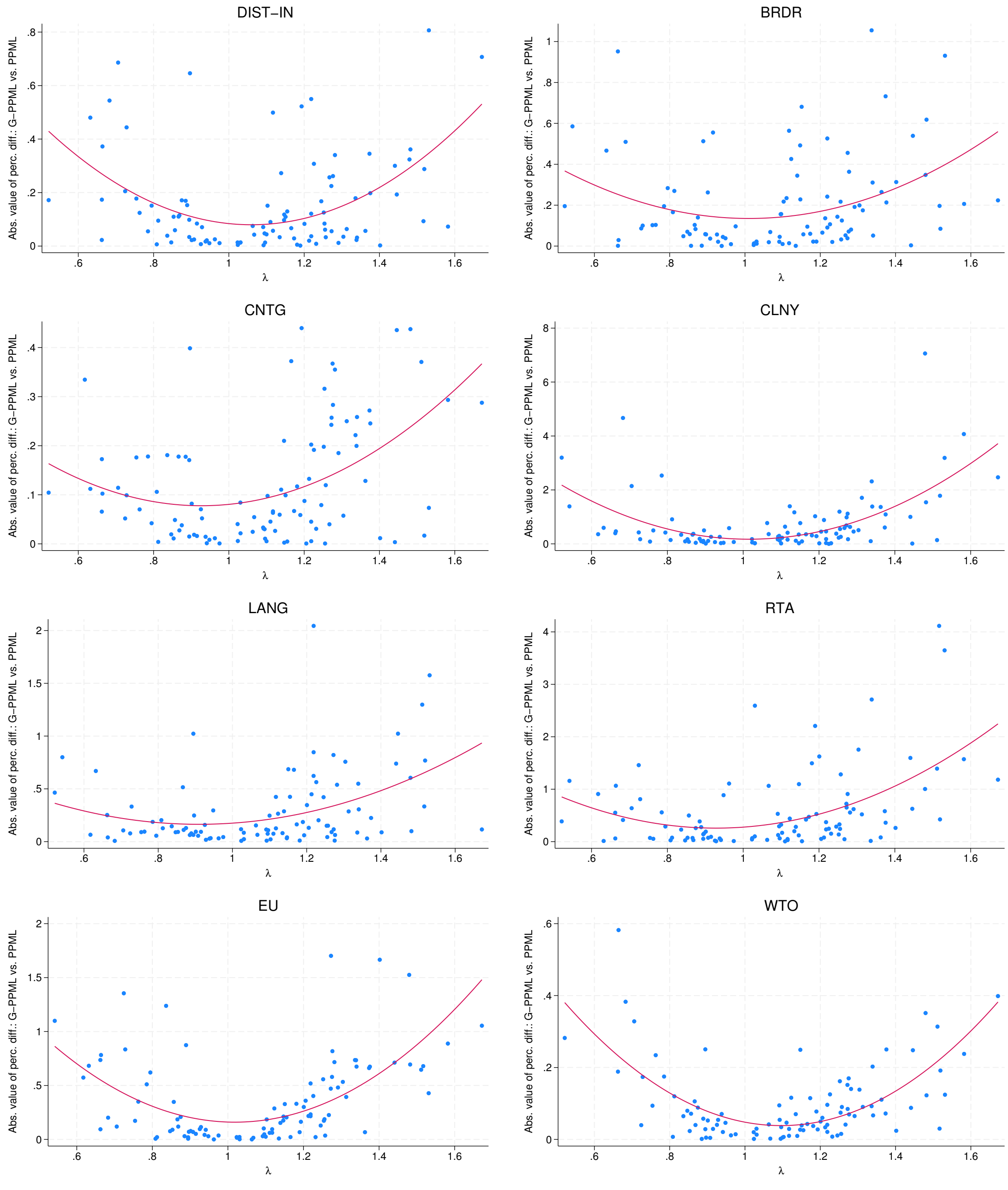
D Additional Figures and Tables

Figure D.1: Difference in Gravity Estimates: PPML vs. G-PPML.



Notes: This figure plots the sectoral difference in the gravity coefficients.

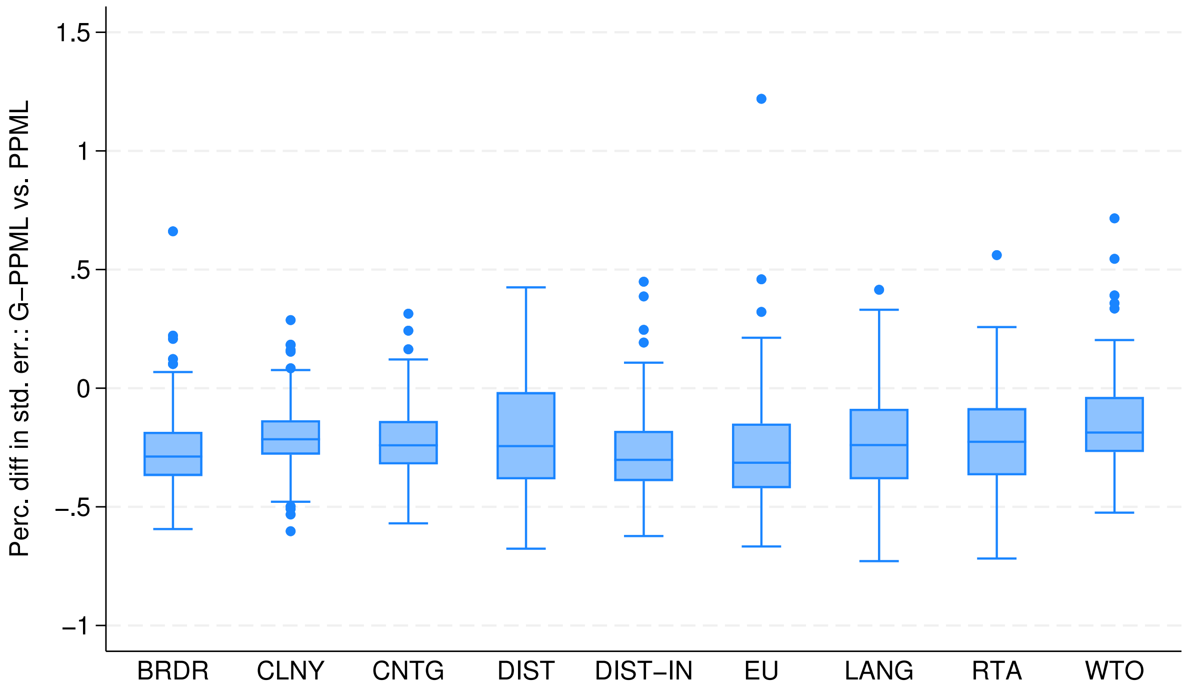
Figure D.2: Difference in Gravity Estimates: PPML vs. G-PPML.



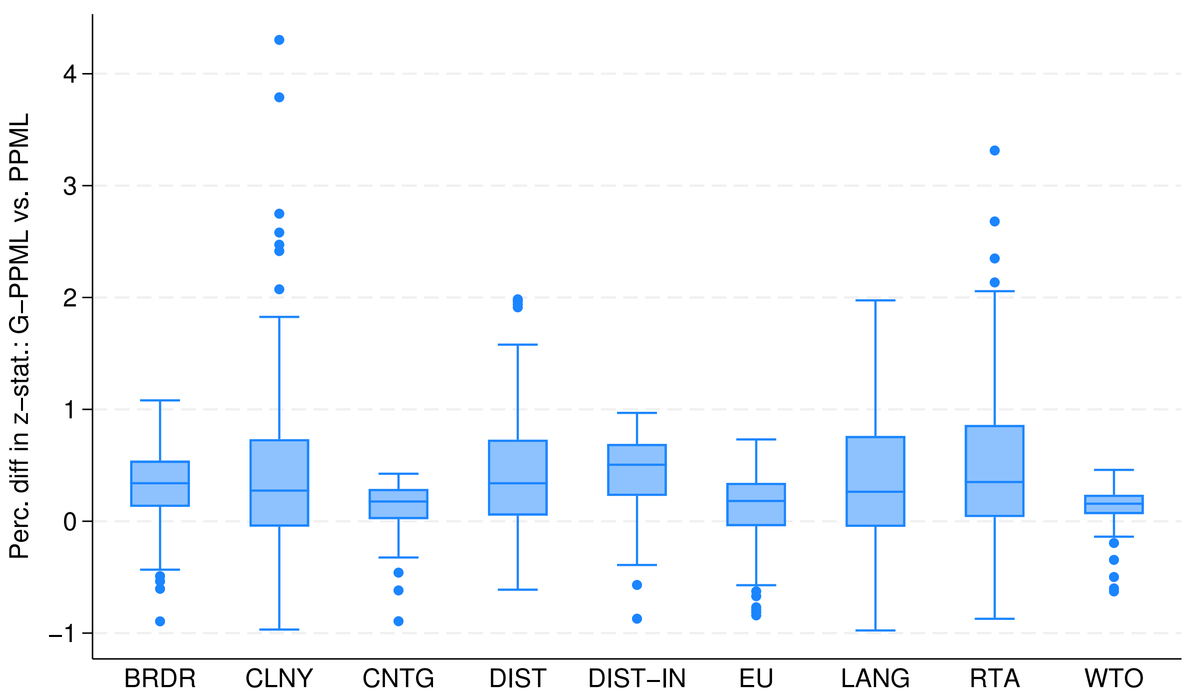
Notes: This figure plots the sectoral percentage difference in the gravity coefficients ($\%\Delta\beta^{k,v}$) estimated with PPML and G-PPML against the corresponding sectoral estimates of λ. For clarity, we drop the top 5 percent of $\%\Delta\beta^{k,v}$. See text for further details.

Figure D.3: Comparison of Estimation Efficiency, PPML vs. G-PPML

(a) Distribution of Percentage Difference in Standard Errors



(b) Distribution of Percentage Difference in z-statistics



Notes: These figures show the distribution of PPML vs. G-PPML percentage difference in standard errors and z-statistics. See text for further details.

Table D.1: Comparison of Estimators for Gravity Equations

	PPML	GPPML	OLS (in logs)	Gamma PML
	$\text{Var}(y\|x) = h \cdot \mathbb{E}(y\|x)^{\lambda}$			
Panel A. two-way fixed effects				
Consistency	Y	Y	N	Y
Asymptotic bias (IPP)	N	N		Y
Efficiency		most eff.		
Standard error bias (finite sample)	Y	N		Y
Panel B. three-way fixed effects				
Consistency	Y	Y	N	N
Asymptotic bias (IPP)	Y	Y		
Efficiency		†		
Standard error bias (finite sample)	Y	Y		Y
	conditional variance misspecification*			
Panel C. two-way fixed effects				
Consistency	Y	Y	N	Y
Asymptotic bias (IPP)	N	Y		Y
Efficiency		†		
Standard error bias (finite sample)	Y	Y		Y
Panel D. three-way fixed effects				
Consistency	Y	N	N	N
Asymptotic bias (IPP)	Y	Y		
Efficiency		†		
Standard error bias (finite sample)	Y	Y		Y

Notes: This table compares the performance of commonly used methods for gravity model estimation. While OLS is included for reference, it is generally not recommended due to its inconsistency in this context.

$*$: Misspecification refers to cases where the conditional variance of the dependent variable does not follow the form $\text{Var}(y|x) = h \cdot \mathbb{E}(y|x)^{\lambda}$, as examined in Head and Mayer (2014). Under mild misspecification, iGMM converges to a pseudo-true parameter (Hansen and Lee, 2021), mitigating the effects of misspecification.

†: Asymptotic bias complicates direct efficiency comparisons with PPML.

Table D.2: Mote Carlo Results for the Coverage of β

	$J=50,\ T=10,$ Obser. = 25 000		$J=100,\ T=5,$ Obser. = 50 000	
	β_1	β_2	β_1	β_2
Case 1: $\mathrm{Var}(y_i\|x_i) = 1 \cdot [\mathbb{E}(y_i\|x_i)]^0$				
G-PPML	0.9538	0.9503	0.9488	0.9510
PPML	0.9515	0.9490	0.9475	0.9463
Gamma-PML	0.9375	0.6375	0.9452	0.7958
OLS	0.0948	0.0000	0.0033	0.0000
Case 2: $\mathrm{Var}(y_i\|x_i) = 1 \cdot [\mathbb{E}(y_i\|x_i)]$				
G-PPML	0.9492	0.9525	0.9520	0.9530
PPML	0.9492	0.9532	0.9523	0.9538
Gamma-PML	0.9438	0.8183	0.9530	0.8810
OLS	0.4922	0.0000	0.1950	0.0000
Case 3: $\mathrm{Var}(y_i\|x_i) = 1 \cdot [\mathbb{E}(y_i\|x_i)]^2$				
G-PPML	0.9520	0.9480	0.9503	0.9470
PPML	0.9477	0.9530	0.9492	0.9477
Gamma-PML	0.9520	0.9520	0.9520	0.9473
OLS	0.9495	0.9500	0.9507	0.9528
Case 4: $\mathrm{Var}(y_i\|x_i) = 4 \cdot [\mathbb{E}(y_i\|x_i)]^0$				
G-PPML	0.9610	0.9545	0.9787	0.9693
PPML	0.9495	0.9495	0.9480	0.9470
Gamma-PML	0.9178	0.2582	0.9255	0.4798
OLS	0.0105	0.0000	0.0000	0.0000
Case 5: $\mathrm{Var}(y_i\|x_i) = 4 \cdot [\mathbb{E}(y_i\|x_i)]$				
G-PPML	0.9488	0.9548	0.9528	0.9498
PPML	0.9485	0.9542	0.9515	0.9490
Gamma-PML	0.9367	0.6535	0.9433	0.7423
OLS	0.3318	0.0000	0.0768	0.0000
Case 6: $\mathrm{Var}(y_i\|x_i) = 4 \cdot [\mathbb{E}(y_i\|x_i)]^2$				
G-PPML	0.9480	0.9452	0.9517	0.9475
PPML	0.9495	0.9477	0.9507	0.9488
Gamma-PML	0.9500	0.9485	0.9510	0.9505
OLS	0.9500	0.9530	0.9545	0.9490

Notes: This table compares the Monte Carlo results for the coverage of β estimates using different estimation methods.

Table D.3: Mote Carlo Results for the Coverage of β (Misspecification)

	$J=50,\ T=10,$ Obser. = 25 000		$J=100,\ T=5,$ Obser. = 50 000	
	β_1	β_2	β_1	β_2
Case M1: $\text{Var}(y_i\|x_i) = 1 \cdot [\mathbb{E}(y_i\|x_i)] + 0.2 \cdot [\mathbb{E}(y_i\|x_i)]^0$				
G-PPML	0.9507	0.9488	0.9448	0.9490
PPML	0.9515	0.9488	0.9457	0.9475
Gamma-PML	0.9400	0.7872	0.9427	0.8605
OLS	0.4067	0.0000	0.0970	0.0000
Case M2: $\text{Var}(y_i\|x_i) = 1 \cdot [\mathbb{E}(y_i\|x_i)] + 0.4 \cdot [\mathbb{E}(y_i\|x_i)]^0$				
G-PPML	0.9535	0.9500	0.9488	0.9457
PPML	0.9532	0.9498	0.9498	0.9452
Gamma-PML	0.9373	0.7490	0.9450	0.8310
OLS	0.3290	0.0000	0.0653	0.0000
Case M3: $\text{Var}(y_i\|x_i) = 1 \cdot [\mathbb{E}(y_i\|x_i)] + 0.6 \cdot [\mathbb{E}(y_i\|x_i)]^0$				
G-PPML	0.9517	0.9555	0.9457	0.9492
PPML	0.9523	0.9553	0.9460	0.9503
Gamma-PML	0.9425	0.7175	0.9420	0.8080
OLS	0.2490	0.0000	0.0358	0.0000
Case M4: $\text{Var}(y_i\|x_i) = 1 \cdot [\mathbb{E}(y_i\|x_i)] + 0.1 \cdot [\mathbb{E}(y_i\|x_i)]^2$				
G-PPML	0.9500	0.9510	0.9490	0.9460
PPML	0.9510	0.9492	0.9482	0.9473
Gamma-PML	0.9457	0.8540	0.9470	0.8938
OLS	0.5755	0.0000	0.2878	0.0000
Case M5: $\text{Var}(y_i\|x_i) = 1 \cdot [\mathbb{E}(y_i\|x_i)] + 0.2 \cdot [\mathbb{E}(y_i\|x_i)]^2$				
G-PPML	0.9538	0.9485	0.9505	0.9485
PPML	0.9517	0.9513	0.9510	0.9470
Gamma-PML	0.9490	0.8692	0.9465	0.9047
OLS	0.6622	0.0000	0.3775	0.0000
Case M6: $\text{Var}(y_i\|x_i) = 1 \cdot [\mathbb{E}(y_i\|x_i)] + 0.3 \cdot [\mathbb{E}(y_i\|x_i)]^2$				
G-PPML	0.9510	0.9500	0.9505	0.9490
PPML	0.9495	0.9475	0.9525	0.9480
Gamma-PML	0.9500	0.8827	0.9457	0.9080
OLS	0.7265	0.0000	0.4660	0.0000

Notes: This table compares the Monte Carlo results for the coverage of β estimates using different estimation methods.

Table D.4: Comparison of Gravity Estimates using Aggregate Trade Data

VARIABLES	(1) PPML	(2) PPML_FE_BIAS	(3) GPPML
DIST	-0.878***	-0.878***	-0.840***
	(0.036)	(0.053)	(0.025)
CNTG	0.661***	0.661***	0.867***
	(0.104)	(0.177)	(0.089)
LANG	0.033	0.033	0.171***
	(0.096)	(0.143)	(0.052)
CLNY	0.302**	0.302	0.620***
	(0.136)	(0.220)	(0.111)
EU	0.905***	0.905***	1.078***
	(0.111)	(0.181)	(0.066)
WTO	-0.097	-0.097	0.137*
	(0.157)	(0.249)	(0.075)
RTA	0.235***	0.235**	0.289***
	(0.066)	(0.097)	(0.043)
Observations	621,607	621,607	621,607
Exporter-Year FE	✓	✓	✓
Importer-Year FE	✓	✓	✓

Notes: This table compares the gravity estimation results using PPML (Santos Silva and Tenreyro, 2006), PPML_FE_BIAS (Weidner and Zylkin, 2021), and G-PPML. Clustered standard errors robust against heteroskedasticity and serial correlation at the country-pair level are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table D.5: Iterated GMM Estimates using Aggregate Trade Data

VARIABLES	(1) iGMM
λ	1.307***
	(0.0302)
h	214.6***
	(60.46)
Observations	621,607

Notes: This tables shows the iGMM estimates of the conditional variance parameters using aggregate trade data. Heteroskedasticity-robust standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.