CESifo WORKING PAPERS

10145
2022

December 2022

A Generalized Poisson-Pseudo Maximum Likelihood Estimator

Ohyun Kwon, Jangsu Yoon, Yoto V. Yotov

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 10145*

A Generalized Poisson-Pseudo Maximum Likelihood Estimator

Abstract

We propose a Generalized Poisson-Pseudo Maximum Likelihood (G-PPML) estimator that relaxes the PPML estimator's assumption that the dependent variable's conditional variance is proportional to its conditional mean. Instead, we employ an iterated Generalized Method of Moments (iGMM) to estimate the conditional variance of the dependent variable directly from the data, thus encompassing the standard estimators in international trade literature (i.e., PPML, Gamma-PML, and OLS) as special cases. With conditional variance estimates, GPPML generates coefficient estimates that are more efficient and robust to the underlying data generating process. After establishing the consistency and the asymptotic properties of the G-PPML estimator, we use Monte Carlo simulations to demonstrate that G-PPML shows decent finite-sample performance regardless of the underlying assumption about the conditional variance. Estimations of a canonical gravity model with trade data reinforce the properties of G-PPML and validate the practical importance of our methods.

JEL-Codes: C130, C500, F100.

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

Ohyun Kwon
School of Economics
Drexel University
Philadelphia / PA / USA
ok85@drexel.edu

Jangsu Yoon
Department of Economics
University of Wisconsin-Milwaukee
Milwaukee / WI / USA
yoon22@uwm.edu

Yoto V. Yotov
School of Economics
Drexel University
Philadelphia / PA / USA
yotov@drexel.edu

December 6, 2022
We are very grateful to Seojeong Lee, Jee-Hyeong Park, Joao Santos Silva, Jeffrey Wooldridge, and Thomas Zylkin for many helpful comments and excellent suggestions. We also thank seminar participants at Seoul National University, Tianjin University, the Asia Meeting of the Econometric Society, the Midwest Econometrics Group, and the Symposium on Econometric Theory and Applications (SETA) for discussions.

1 Introduction

Owing to the seminal work of Santos Silva and Tenreyro (2006), who demonstrate that in the presence of heteroskedasticity (which is the case, for example, with trade data) OLS is not appropriate, the Poisson Pseudo Maximum Likelihood (PPML) estimator has firmly established itself as the leading estimator for trade gravity regressions. In fact, Santos Silva and Tenreyro's *"The Log of Gravity"* paper is one of the most influential contributions to the empirical trade literature over the past 15 years.[1] In addition to effectively handling the issue of heteroskedasticity, the multiplicative form of PPML makes this estimator appropriate for utilizing the information that is contained in zero trade flows, which often take a significant fraction of the data (both at the aggregate and, especially, at the sectoral level) but are dropped in OLS regressions. 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.[2] Consistent with the convincing performance of the PPML estimator for trade regressions, Wooldridge (2021) recently notes that the *"[p]oisson regression can get one so far with so little trouble"* and he wonders *"why so many resist [using it]"*.

Since its introduction to the trade community, PPML has attracted significant attention, it has enjoyed remarkable success, 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), and Weidner and Zylkin (2021) extend this analysis to show that PPML with three-way fixed effects is consistent albeit with asymptotic bias, for which Weidner and Zylkin provide analytic correction.[3] Building on Santos Silva and Tenreyro (2011) and motivated by practical difficulties with PPML convergence, Correia et al. (2020) introduce the *ppmlhdfe* STATA command, which simultaneously addresses the issues of computational speed and convergence with non-linear PPML. Arvis and Shepherd (2013) and Fally (2015) discover an additive property of the PPML estimator, which makes it perfectly consistent with a wide class of structural gravity models.[4] Capitalizing on this property, Anderson et

[1]In support of this claim, consider the fact that "The Log of Gravity" of Santos Silva and Tenreyro (2006) is the fourth most-cited article at the *Review of Economics and Statistics* of all times and that, according to Google Scholar, it has generated more than 7,300 citations, with more than 800 citations in 2022 alone.

[2]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.

[3]Gravity models are almost always estimated with two-way fixed effects (one on the importer and one on the exporter side), which are used to control for the unobservable theoretical multilateral resistance terms (Anderson and Van Wincoop, 2003). Very often, gravity models are also estimated with pair fixed effects, which are used to mitigate endogeneity concerns with respect to bilateral trade policies (Baier and Bergstrand, 2007) and to comprehensively control for all time-invariant bilateral trade costs (Egger and Nigai, 2015; Agnosteva et al., 2019).

[4]The additive PPML property (Fally, 2015) may not apply to G-PPML. However, the estimates that are obtained with G-PPML can still be used in a constrained PPML estimation in order to obtain general equilibrium effects directly with PPML.

al. (2018) demonstrate how the PPML estimator can be used to obtain not only partial equilibrium gravity estimates but also benchmark general equilibrium effects.

Against this background, it should not be surprising that the PPML estimator has established itself as the leading technique for gravity regressions and that it has been utilized in thousands of academic and policy analysis of trade flows and trade policies. Yet, despite all the attractive properties of PPML and its remarkable success, some researchers remain skeptical about its validity and its 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 its expected value may not be consistent with the assumptions of the PPML estimator.[5] Based on Monte Carlo simulations, Head and Mayer (2014) challenge the use of PPML, and recommend that researchers should also estimate gravity with other estimators, including OLS and Gamma-PML, for robustness checks.

The main contribution of our paper is to address this challenge by proposing and implementing a *Generalized* PPML (G-PPML) estimator, which relies on the actual data used in the estimation to inform and dictate the relationship between the variance of bilateral exports and its expected value, while simultaneously delivering the gravity estimates of interest.[6] We view the extension from PPML to G-PPML as methodologically similar to that from the OLS to the FGLS estimator. While maintaining the merits of PPML, the additional properties of our estimator include its improved estimation efficiency (as reflected in lower standard errors) and the fact that it is immune to the criticism that the heteroskedastic structure may have been misspecified (Head and Mayer, 2014). As such, G-PPML generalizes the applicability of PPML to a broader context and adds to the list of attractive properties of the PPML estimator. Our estimator reinforces the argument for using PPML as the most appropriate "workhorse" estimator for gravity equations.

The G-PPML estimator is based on three building blocks. First, from a methodological perspective, we rely on and extend the ideas of Santos Silva and Tenreyro (2006, 2011) and the latest developments in related literature, e.g., Weidner and Zylkin (2021) and Mnasri and Nechi (2021). Second, we capitalize on the cutting-edge econometric contribution of Hansen and Lee (2021), who verify the misspecification-robust property of the iterated General Method of Moments (iGMM) estimator and demonstrate that the iGMM estimator provides stable estimates regardless of the initial guess on parameters. Third, in practical terms, we extend the fast *ppmlhdfe* command of Correia et al. (2020) to offer a user-friendly STATA estimation procedure that implements our

[5] Head and Mayer (2014) show that when the conditional variance of trade volume is expressed as a monomial of its conditional mean, the exponent is 1.77 and 1.79, respectively, using two different trade datasets.

[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.

methods for gravity estimations. The proposed estimator is computationally attractive and shows a better finite-sample performance than other estimators based on similar intuition (e.g., Jochmans (2017)).

Our proposed estimation method includes three stages. The first stage is a conventional PPML estimation to obtain the preliminary estimates of regression coefficients, which serve as an input to the next stage. In the second stage, we implement iGMM that estimates the conditional variance of the dependent variable in the gravity equation. The conditional variance form is general enough to accommodate all common assumptions in the literature. In the third stage, the estimated variance parameters provide the optimally weighted moment conditions that augment the PPML moment conditions. The resulting G-PPML estimator inherits all the desirable properties of the conventional PPML estimator and provides additional efficiency benefits. We verify the consistency and asymptotic distribution of the estimator considering the existence of two-way fixed effects. Our findings show that the G-PPML estimator does not suffer from the incidental parameter problem and is free of asymptotic bias for all of data generating processes (DGP) acknowledged in the literature. In short, the G-PPML estimator estimates the gravity-type equations *more* efficiently by making *less* assumption about the associated error term.

To compare the efficiency of our approach with the established estimators in the literature, we perform a series of Monte Carlo experiments. First, we show that the iGMM estimator can estimate the true conditional variance parameters consistently across a wide range of parameter spaces. In comparison, leading existing methods are only valid for a limited range of parameters. 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 presumed values of other PML estimators; (ii) G-PPML encompasses other PML estimators as special cases when the DGP conforms to the 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, contiguous borders, 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 – broadly consistent with the assumption of Santos Silva and Tenreyro (2006) – 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 for gravity estimations; and (iv) most important for our purposes, we observe significant heterogeneity in the estimates of the conditional variance parameter across sectors, ranging between 0.52 and 1.67. The deviations of the conditional variance parameter from one imply that there can be estimation efficiency gain associated with G-PPML. As data suggests that PPML's assumption about the conditional variance is generally true and our approach generalizes the idea of PPML to a greater domain, we refer to our estimator as *generalized* PPML.

Comparisons between the PPML and G-PPML gravity estimates lead to four conclusions. First, consistent with our finding that most of the estimated conditional variance parameters are close to one, we also find that many of the PPML and G-PPML gravity estimates are very similar to each other. However, second, we also observe that a significant fraction of the G-PPML gravity estimates are different from the corresponding PPML estimates. Moreover, third, and consistent with our theory, we confirm that the further away the estimates of the conditional variance parameter are from one, the larger is the difference between the PPML and G-PPML coefficient 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. For example, the G-PPML standard errors are on average over 20% lower than those obtained with PPML across gravity variables and sectors.

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 proposes an intuitive modification 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 estimation efficiency improvement. Given the prominence of PPML for estimating trade gravity equations, we specify the following econometric model as our departing 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. The two-way fixed effects 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 $\mathbb{E}\left[\epsilon_{ijt}|x_{ijt}, \gamma_{it}, \eta_{jt}\right] = 0$.[7] 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 PPML method particularly assumes that the conditional variance of y_{ijt} is proportional to the conditional mean of y_{ijt}. Specifically, given a general form of the conditional variance of y_{ijt},

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

PPML assumes $\lambda = 1$ and $h > 0$. The specification also includes all the well-known PML estimators (i.e., Gamma-PML ($\lambda = 2$), Negative Binomial PML ($\lambda = 1$)[8], and Gaussian PML ($\lambda = 0$)) as special cases. 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 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 almost all of the conditional variance forms adopted in the relevant literature.[9] 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

[7] Santos Silva and Tenreyro (2006) 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.

[8] Although the parametric assumption of Negative Binomial PML is less demanding than that of PPML, it is generally not recommended for empirical application since the estimation results are unit-sensitive. See Head and Mayer (2014) for detailed discussions.

[9] To our knowledge, the only two exceptions to this statement is Santos Silva and Tenreyro (2011) and Weidner and Zylkin (2021). Santos Silva and Tenreyro (2011) obtains PPML coefficients under potential misspecification and Weidner and Zylkin (2021) obtain robust standard errors even when the conditional variance does not follow equation (2). Nevertheless, we 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.

estimate the gravity equation (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 estimate $\hat{\lambda}$. This provides a feasible routine to estimate gravity equations efficiently 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 out-performs 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 $\mathbb{E}\left[x_{ijt}x_{ijt}'\left(\epsilon_{ijt}^2 - h\cdot\mu_{ijt}^{\lambda}\right)^2\right]$ 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

$$\mathbb{E}\left[\epsilon_{ijt}^{2}-h\cdot\mu_{ijt}^{\lambda}\middle|x_{ijt},\gamma_{it},\eta_{jt}\right]=0, \tag{3}$$

and 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 $\left(\hat{\beta}^{PPML},\hat{\gamma}_{it}^{PPML},\hat{\eta}_{jt}^{PPML}\right)$. Proposition 2 of Weidner and Zylkin (2021) shows that the PPML estimator is consistent even with three-way fixed effects. 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.

Instead, we propose an iGMM estimator to obtain the conditional variance parameters (h,λ). The conditional moment of equation (3) implies k-dimensional unconditional moments,

$$\mathbb{E}\left[x_{ijt}\left(\epsilon_{ijt}^{2}-h\cdot\mu_{ijt}^{\lambda}\right)\right]=0, \tag{4}$$

and the model is identified if $k\geq 2$. Since the covariates x_{ijt} generally exceed two dimensions, we focus on the over-identified case of $k>2$. As demonstrated by Hansen and Lee (2021), the iGMM method is robust to the possible moment misspecification and the estimates do not fluctuate much depending on 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. The GMM criterion function is

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

where $\phi=(h_{\phi},\lambda_{\phi})$ is the initial guess on the parameter value. For a given $(h_{\phi},\lambda_{\phi})$, the next step estimator is to minimize 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 misspecification. 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). Even if the conditional variance deviates from the form in equation (2), $\hat{\theta}$ still converges to the *pseudo-true* parameter θ minimizing the population GMM criterion function. 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.

Define three matrices

$$\begin{aligned} W &= \mathbb{E}\left[x_{ijt}x_{ijt}'\left(\epsilon_{ijt}^2 - h \cdot \mu_{ijt}^{\lambda}\right)^2\right] \\ Q &= -\mathbb{E}\begin{bmatrix} \mu_{ijt}^{\lambda}x_{ijt}' \\ h \cdot \mu_{ijt}^{\lambda}\log\left(\mu_{ijt}\right)x_{ijt}' \end{bmatrix}' \\ V &= \mathbb{E}\left[x_{ijt}v'\left(x_{ijt}, \beta\right)' V_{PPML} v'\left(x_{ijt}, \beta\right)x_{ijt}'\right], \end{aligned}$$

where $v\left(x_{ijt}, \beta\right) = \epsilon_{ijt}^2 - h \cdot \mu_{ijt}^{\lambda}$ and V_{PPML} is the asymptotic variance of the PPML estimator $\hat{\beta}^{PPML}$. If the conditional variance of the model is correctly specified (i.e., it satisfies any of the class of PML estimators widely used to date), the asymptotic variance of $N\left(\hat{\theta} - \theta\right)$ is

$$\left(Q'W^{-1}Q\right)^{-1}Q'W^{-1}\left(W + V\right)W^{-1}Q\left(Q'W^{-1}Q\right)^{-1},$$

where V is generated from the approximation error of $\hat{\epsilon}_{ijt}$ and $\hat{\mu}_{ijt}$. We provide the derivation of V in Appendix. The asymptotic variance of $\hat{\lambda}$ is the second diagonal element of the asymptotic variance matrix.

Proposition 2.2. *Under Assumption 2.1,*

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

The asymptotic normality informs the construction of 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 fitting the conditional variance of the gravity model.

The result of Proposition 2.2 follows the classical efficient GMM estimator's asymptotic distribution except that V presents the approximation error from the first-stage estimator $\beta^{P\hat{P}ML}$. Without V, the asymptotic variance is equivalent to $(Q'W^{-1}Q)^{-1}$. The result is a special case of Theorem 4 in Hansen and Lee (2021) under the correctly specified conditional variance. If the conditional variance structure in equation (2) does not hold, i.e., $\mathbb{E}\left[x_{ijt}\left(\epsilon_{ijt}^2 - h \cdot \mu_{ijt}^{\lambda}\right)\right] \neq 0$, the asymptotic variance formula changes considering the misspecified moment condition. We provide the limiting distribution encompassing the mildly-misspecified cases in Appendix. The practitioners can still estimate λ to test the variance-to-mean ratio best describing the data generating process.

The iGMM estimator $\hat{\theta}$ has a zero asymptotic bias since the initial PPML estimator does not suffer from the asymptotic bias. Although other initial estimators, including Gaussian PML and Gamma PML, may replace the PPML and provide the consistency of $\hat{\theta}$, they may cause a problem in the inference on θ.

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

$$\begin{aligned}
\hat{\beta} &: \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T} x_{ijt}\left(y_{ijt} - \hat{\mu}_{ijt}\right)\hat{\mu}_{ijt}^{1-\hat{\lambda}} = 0, \\
\hat{\gamma}_{it} &: \sum_{j=1}^{N}\left(y_{ijt} - \hat{\mu}_{ijt}\right)\hat{\mu}_{ijt}^{1-\hat{\lambda}} = 0, \\
\hat{\eta}_{jt} &: \sum_{i=1}^{N}\left(y_{ijt} - \hat{\mu}_{ijt}\right)\hat{\mu}_{ijt}^{1-\hat{\lambda}} = 0,
\end{aligned}$$

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). According to Proposition 2 of Weidner and Zylkin (2021), G-PPML estimator is consistent even with three-way fixed effects as long as the conditional variance is correctly specified as equation (2).

2.3 Asymptotic Distribution

The proposed *generalized* PPML estimator is consistent for β, but the estimator may not be immune to the asymptotic bias due to 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 $1/N$ 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 between the two, however, is that G-PPML does not suffer from the IPP *without* having to impose a particular variance-to-mean ratio of the dependent variable.

Define the following elements

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

The defined 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 $\bar{H}_{ij} = \mathbb{E}[H_{ij}|x_{ij}]$ and $\bar{G}_{ij} = \mathbb{E}[G_{ij}|x_{ij}]$. Next, let $x_{ij} = (x_{ij,1}, \ldots, x_{ij,k})$ and 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 and 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 weak serial dependence across time. 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 will discuss the issue in Appendix A.4.

Proposition 2.4. *(Asymptotic Distribution) Suppose the model satisfies equation (3). Under Assumption 2.1,*

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

where Ω_N is a $k \times k$ matrix

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

The derived asymptotic distribution is simple and has zero asymptotic bias under the two-way fixed effects setup.[10] The asymptotic bias following Fernández-Val and Weidner (2016) in general case is $\frac{\Omega_N^{-1}(B_N + D_N)}{N}$, where $N\left(\hat{\beta} - \beta - \frac{\Omega_N^{-1}(B_N + D_N)}{N}\right) \xrightarrow{d} N\left(0, \Omega_N^{-1}\right)$. B_N and D_N are k-dimensional

[10] Under the three-way fixed effects, however, the asymptotic bias does not disappear for the G-PPML. As Weidner and Zylkin (2021) demonstrated, PPML has a non-zero asymptotic bias under the three-way fixed effects. As $\hat{\theta}$ uses the PPML as an initial estimator, $\hat{\theta}$ causes a non-zero asymptotic bias of $\hat{\beta}$.

vectors with their mth elements defined by

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

and $B_N = D_N = 0$ if we use the G-PPML estimator under equation (3). The asymptotic bias components B_N^m and D_N^m are both 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 are generally non-zeros unless $\lambda = 1$, but become zeros when the conditional variance of y_{ijt} is correctly specified.

We estimate the asymptotic variance of $\hat{\beta}$ by $\hat{\Omega}_N = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\tilde{x}'_{ij}\hat{\bar{H}}_{ij}\tilde{x}_{ij}$, where $\left[\hat{\bar{H}}_{ij}\right]_{ts} = \hat{\mu}_{ijt}^{2-\hat{\lambda}}1\{t=s\}$. 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 $\mathrm{Var}\left(S_{ij}|x_{ij}\right)$. 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).[11]

3 Monte Carlo Simulation Analysis

This section provides simulation-based evidence on the consistency and the estimation efficiency of G-PPML. After describing the data generating process, we compare the performance of our method with established estimators in terms of estimating conditional variance and regression coefficients. We also demonstrate the robustness of iGMM and G-PPML under misspecification.

[11]The downward bias returns when the conditional variance does not follow the equation (2).

3.1 Data Generating Process

We start by generating the vector of independent variables, $x = \{1, x_{1,ijt}, x_{2,ijt}, \dots x_{6,ijt}, \iota_{it}, \iota_{jt}\}$, 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, $x_{3,ijt} \dots, x_{6,ijt}$, 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 $x_{1,ijt}, \dots, x_{6,ijt}$ 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).[12] 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 $\theta = (\beta', \gamma_{it}, \eta_{jt})'$. All Monte Carlo results are based on 500 independent simulations.

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:[13]

$$\begin{aligned} &\mathbb{E}\left(\varepsilon|x\right) = 1 \quad \text{and} \\ &\text{Var}(\varepsilon|x) = \left[h_1 \exp\left(\lambda_1 x'\theta\right) + h_2 \exp\left(\lambda_2 x'\theta\right)\right] / \exp\left(2x'\theta\right). \end{aligned} \tag{5}$$

This implies the following first and second moments of the dependent variable $y = \exp(x'\theta) \cdot \varepsilon$, where y refers to y_{ijt}:

$$\begin{aligned} &\mathbb{E}\left(y|x\right) = \exp(x'\theta) \quad \text{and} \\ &\text{Var}(y|x) = h_1 \mathbb{E}\left(y|x\right)^{\lambda_1} + h_2 \mathbb{E}\left(y|x\right)^{\lambda_2}. \end{aligned} \tag{6}$$

The first line in equation (6) 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, focus on the first term of $\text{Var}(y|x)$ by assuming $h_1 > 0$ and $h_2 = 0$. If $\lambda_1 = 1$, the conditional variance of y is proportional to its conditional mean – the exact working assumption of the PPML estimator (Santos Silva and Tenreyro, 2006). 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,

[12]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.

[13]Mechanically, to generate the random variable ε from the lognormal distribution specified in (5), 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 (5).

2014). Under the same condition $\lambda_1 = 2$, $\mathrm{Var}(\varepsilon|x)$ reduces to h_1 according to (5); 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. In sum, simply by changing the values of our parameters, our DGP accommodates the assumptions of all common PML estimators and the OLS.[14]

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.[15] 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 would capitalize on this value of λ to construct moment conditions that are better tuned than other PML estimators, which may assume incorrect λ values. In the kinfe-edge case where iGMM suggests $\hat{\lambda} = 1$, G-PPML becomes identical to PPML.

Note that beyond determining the conditional variance, (h_1, λ_1) also dictates the degree of data noise. The second line in equation (5) shows that when $\lambda_2 = 0$, h_1 dictates the extent of data noise that is orthogonal to the conditional mean, whereas λ_1 dictates the extent of data noise so much as it correlates with the conditional mean. h_2 and λ_2 admit similar interpretations. Thus, different combinations of parameters allow us to compare G-PPML with other estimators under different levels of data noise.

We also test the performance of G-PPML under misspecification. Specifically, when we assume that both h_1 and h_2 in equation (6) are positive, the conditional variance becomes a polynomial of the conditional mean and none of the mainstream assumptions for ε are consistent. An 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 compares the performance of our iGMM method for estimating λ with two alternative leading approaches. Following the existing literature (e.g., Head and Mayer (2014)), we assume

[14] Under an alternative DGP where the dependent variable is $y = \exp(x'\theta) + \varepsilon$, 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.

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

that $h_2 = 0$ and use h and λ to refer to h_1 and λ_1, respectively.[16] 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:

$$\ln\left(\hat{\epsilon}^2\right) = \text{constant} + \lambda x'\hat{\theta}, \tag{7}$$

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).[17] Let $\hat{y}$ denote $\hat{y} \equiv \exp(x'\hat{\theta})$. Santos Silva and Tenreyro approximate the expression

$$(y - \hat{y})^2 = h\hat{y}^{\lambda} \tag{8}$$

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

$$(y - \hat{y})^2/\sqrt{\hat{y}} = h\sqrt{\hat{y}} + h(\lambda - 1)\ln(\hat{y})\sqrt{\hat{y}}. \tag{9}$$

Equation (9) can be estimated with OLS and using $\sqrt{\hat{y}}$ and $\ln(\hat{y})\sqrt{\hat{y}}$ 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 the estimate of the first coefficient $\hat{h}$ and adding one.[18] We refer to this method as "SST".[19]

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, while the alternative values of λ cover a reasonable range from the existing literature.[20] 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.

We draw three main conclusions based on the results in Figure 1. First, and most important, our

---

[16] Results under potential misspecification ($h_2 > 0$) are provided in subsection 3.4.

[17] 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).

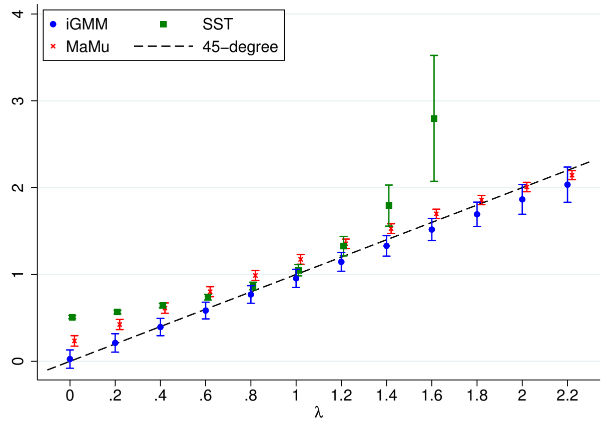
[18] Equation (9) can be estimated both with and without adding a constant term as the third regressor. In general, we find that estimating equation (9) without the constant term provides estimates of λ closer to the true value. Thus, we only present the results without the additional constant term.

[19] We remind readers that although the SST approach to estimate 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 PPML in gravity estimation. See Santos Silva and Tenreyro (2006) for details.

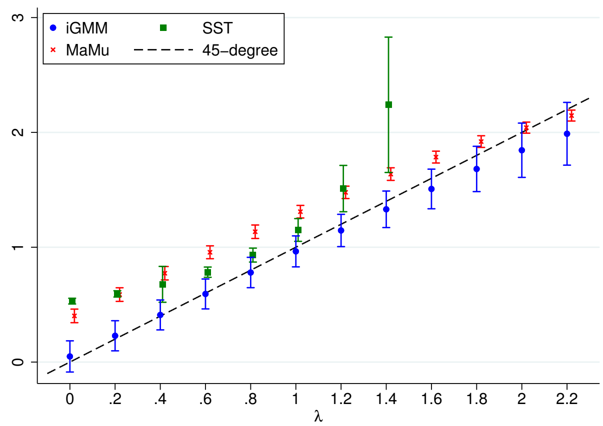
[20] For example, our own sectoral estimates of λ in Section 4 lie in the interval $(0.5, 1.7)$.

Figure 1: Comparing Different Methods to Estimate λ

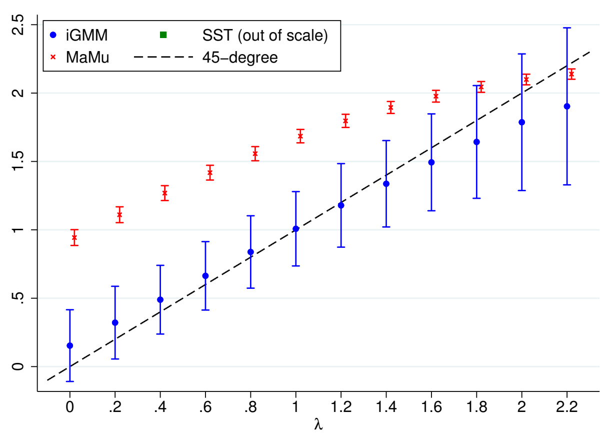
(a) $h = 0.5$



(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.

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.[21] 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.[22] 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 spaces.

3.3 Coefficient Estimates

Table 1 compares the performance of G-PPML with other leading estimators. We compare both the mean of absolute bias (column Bias) and the mean of standard errors (column S.E.) 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 6 cases. In cases 1 though 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 PPML implicitly assumes $\lambda = 1$, and Gamma-PML 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 500 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

[21] We do note, however, that the confidence intervals become wider when the noise in 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.

[22] 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 (7), can explain why the MaMu estimator leads to biased estimates of λ.

Table 1: Main Mote 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.	Bias	S.E.		Bias	S.E.	Bias	S.E.
Case 1: $\mathrm{Var}(y_i\|x_i) = 1 \cdot [\mathbb{E}(y_i\|x_i)]^0$										
G-PPML	0.0469	0.0203	0.0240	0.0043	0.0054	0.0298	0.0133	0.0169	0.0030	0.0038
PPML		0.0221	0.0274	0.0047	0.0059		0.0147	0.0191	0.0033	0.0042
Gamma-PML		0.0315	0.0343	0.0119	0.0069		0.0211	0.0251	0.0066	0.0050
OLS		0.1111	0.0340	0.1077	0.0069		0.1108	0.0238	0.1078	0.0048
Case 2: $\mathrm{Var}(y_i\|x_i) = 1 \cdot [\mathbb{E}(y_i\|x_i)]$										
G-PPML	0.9641	0.0335	0.0410	0.0068	0.0084	0.9811	0.0241	0.0290	0.0049	0.0060
PPML		0.0334	0.0419	0.0068	0.0086		0.0240	0.0293	0.0049	0.0060
Gamma-PML		0.0370	0.0425	0.0109	0.0085		0.0268	0.0311	0.0065	0.0062
OLS		0.0824	0.0415	0.0823	0.0083		0.0823	0.0291	0.0818	0.0058
Case 3: $\mathrm{Var}(y_i\|x_i) = 1 \cdot [\mathbb{E}(y_i\|x_i)]^2$										
G-PPML	1.8370	0.0506	0.0595	0.0104	0.0119	1.9066	0.0357	0.0431	0.0074	0.0086
PPML		0.0580	0.0716	0.0111	0.0140		0.0402	0.0503	0.0080	0.0098
Gamma-PML		0.0505	0.0562	0.0098	0.0112		0.0357	0.0416	0.0071	0.0083
OLS		0.0437	0.0537	0.0085	0.0107		0.0305	0.0376	0.0061	0.0075
Case 4: $\mathrm{Var}(y_i\|x_i) = 4 \cdot [\mathbb{E}(y_i\|x_i)]^0$										
G-PPML	0.1575	0.0461	0.0526	0.0101	0.0120	0.1153	0.0287	0.0350	0.0063	0.0079
PPML		0.0436	0.0547	0.0094	0.0118		0.0299	0.0380	0.0066	0.0083
Gamma-PML		0.0621	0.0577	0.0351	0.0116		0.0414	0.0439	0.0198	0.0088
OLS		0.2405	0.0561	0.2381	0.0113		0.2379	0.0392	0.2378	0.0079
Case 5: $\mathrm{Var}(y_i\|x_i) = 4 \cdot [\mathbb{E}(y_i\|x_i)]$										
G-PPML	1.0098	0.0669	0.0813	0.0140	0.0168	1.0174	0.0476	0.0577	0.0099	0.0119
PPML		0.0667	0.0834	0.0140	0.0172		0.0475	0.0585	0.0099	0.0120
Gamma-PML		0.0695	0.0712	0.0273	0.0143		0.0502	0.0543	0.0168	0.0109
OLS		0.1629	0.0673	0.1617	0.0135		0.1630	0.0471	0.1619	0.0094
Case 6: $\mathrm{Var}(y_i\|x_i) = 4 \cdot [\mathbb{E}(y_i\|x_i)]^2$										
G-PPML	1.7689	0.0875	0.1016	0.0210	0.0204	1.8401	0.0629	0.0767	0.0149	0.0153
PPML		0.1118	0.1380	0.0225	0.0274		0.0782	0.0995	0.0158	0.0195
Gamma-PML		0.0856	0.0880	0.0172	0.0176		0.0622	0.0685	0.0130	0.0137
OLS		0.0654	0.0819	0.0127	0.0164		0.0456	0.0573	0.0093	0.0115

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.

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. Yet, Weidner and Zylkin (2021) show that PPML is still a consistent estimator in this setting and is robust to the IPP. 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 (0.0241 vs. 0.0274 and 0.0054 vs. 0.0059).[23] 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% and 200% greater than G-PPML, respectively. Moreover, Gamma-PML's standard errors are also much greater than those of G-PPML. A natural explanation for this result is that case 1 is inconsistent with the assumption of Gamma-PML and, as shown in Weidner and Zylkin (2021), Gamma-PML is subject to the IPP in this setting. Finally, consistent with Santos Silva and Tenreyro (2006), the OLS estimation bias is much greater than other estimators. The key message in our analysis in case 1 is similar if we experiment with a greater sample size (the right half of Table 1).

In case 2 ($\lambda = 1$), the DGP becomes consistent with the underlying assumption of PPML. Thus, not surprisingly, G-PPML and PPML deliver very similar estimates, while out-performing Gamma-PML and OLS both in terms of mean bias and standard errors. That G-PPML can perform as well as the PPML without the 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 featuring around 15% (0.0119 vs. 0.0140) to 17% (0.0594 vs. 0.0716) lower standard errors than PPML.[24] However, G-PPML does not perform as well as Gamma-PML, because our iGMM estimates of λ are not sufficiently close to 2.[25] We confirm this hypothesis by increasing the sample size to 50,000 observations to find 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

[23]We do note, however, that the advantage of G-PPML in absolute term is less obvious 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 since the coefficient of the dummy variable x_2 is inherently easier to estimate.

[24]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.

[25]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.

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. Nevertheless, given the poor performance of OLS in cases 1 and 2 and based on our empirical results in Section 4 showing $\hat{\lambda}$ to be strictly less than 2, we posit that it is risky for econometricians to employ the OLS estimator without firm prior knowledge that $\lambda = 2$.

In cases 4 thorough 6, we replicate the results in the preceding cases with a higher value of h. The purpose is to gain further confidence about 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. Specifically, the G-PPML is the most efficient estimator when $\lambda = 0$ (case 4); G-PPML performs as well as PPML and better than Gamma-PML and OLS when $\lambda = 1$ (case 5); when $\lambda = 2$, G-PPML outperforms PPML, performs similarly to Gamma-PML as the sample size increases, and it is outperformed by OLS (case 6).[26]

In sum, we conclude that (i) obtaining precise values of λ are crucial for obtaining sound coefficient estimates, and (ii) without any prior knowledge on λ, G-PPML outperforms other estimators in the majority of cases.

3.4 iGMM and G-PPML under Misspecification

We conclude 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.[27] 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 2. 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).

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

[27]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 in an unfair starting point. When the DGP deviates greatly from the PPML assumption, the efficiency gain from G-PPML will only be more pronounced.

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 $h_2\mathbb{E}\left(y|x\right)^0$.

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 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 it is robust to 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.

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 estimate 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 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 distributions of

Table 2: 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.	Bias	S.E.		Bias	S.E.	Bias	S.E.
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	0.8925	0.0349	0.0427	0.0071	0.0089	0.9092	0.0246	0.0302	0.0051	0.0063
PPML		0.0349	0.0436	0.0071	0.0090		0.0245	0.0305	0.0051	0.0063
Gamma-PML		0.0400	0.0446	0.0123	0.0089		0.0280	0.0327	0.0076	0.0065
OLS		0.0982	0.0434	0.0967	0.0087		0.0961	0.0304	0.0968	0.0061
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	0.8301	0.0355	0.0444	0.0075	0.0093	0.8480	0.0258	0.0313	0.0051	0.0065
PPML		0.0353	0.0453	0.0075	0.0094		0.0257	0.0316	0.0051	0.0066
Gamma-PML		0.0409	0.0465	0.0136	0.0093		0.0300	0.0342	0.0086	0.0069
OLS		0.1110	0.0452	0.1094	0.0091		0.1105	0.0316	0.1102	0.0063
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	0.7880	0.0383	0.0461	0.0075	0.0096	0.7949	0.0268	0.0324	0.0054	0.0068
PPML		0.0379	0.0470	0.0075	0.0098		0.0268	0.0328	0.0053	0.0068
Gamma-PML		0.0444	0.0482	0.0157	0.0097		0.0314	0.0357	0.0095	0.0071
OLS		0.1251	0.0468	0.1220	0.0094		0.1230	0.0328	0.1224	0.0066
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	1.1640	0.0371	0.0460	0.0073	0.0094	1.1916	0.0272	0.0327	0.0055	0.0067
PPML		0.0376	0.0477	0.0073	0.0097		0.0275	0.0334	0.0055	0.0068
Gamma-PML		0.0395	0.0460	0.0106	0.0092		0.0289	0.0337	0.0070	0.0067
OLS		0.0793	0.0446	0.0768	0.0089		0.0770	0.0312	0.0773	0.0062
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	1.2888	0.0406	0.0502	0.0083	0.0102	1.3226	0.0292	0.0358	0.0060	0.0073
PPML		0.0418	0.0528	0.0084	0.0107		0.0298	0.0370	0.0061	0.0075
Gamma-PML		0.0427	0.0491	0.0111	0.0098		0.0305	0.0360	0.0071	0.0072
OLS		0.0751	0.0474	0.0725	0.0095		0.0734	0.0332	0.0722	0.0066
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	1.3798	0.0433	0.0538	0.0091	0.0109	1.4128	0.0314	0.0385	0.0063	0.0078
PPML		0.0449	0.0574	0.0092	0.0115		0.0324	0.0403	0.0064	0.0081
Gamma-PML		0.0452	0.0517	0.0118	0.0104		0.0325	0.0382	0.0075	0.0076
OLS		0.0737	0.0498	0.0686	0.0100		0.0698	0.0349	0.0691	0.0070

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.

PPML and G-PPML gravity estimates for 105 manufacturing sectors.[28]

Capitalizing on the developments in the voluminous gravity literature, we specify the following benchmark estimating gravity equation. Due to separability of the theoretical gravity model (Anderson and Van Wincoop, 2004), our econometric model applies to individual sectors. However, for simplicity, we omit the sectoral notation:

$$
\begin{aligned}
y_{ijt} \;=\;& \exp[\beta_1 DIST_{ij} + \beta_2 DIST\text{-}IN_{ij} + \beta_3 CNTG_{ij} + \beta_4 CLNY_{ij} + \beta_5 LANG_{ij}] \times \\
& \exp[\beta_6 RTA_{ijt} + \beta_7 EU_{ijt} + \beta_8 WTO_{ijt} + \beta_9 BRDR_{ij} + \gamma_{it} + \eta_{jt}] \times \varepsilon_{ijt}. \qquad (10)
\end{aligned}
$$

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 (10) in levels (Santos Silva and Tenreyro, 2006, 2011). The covariates in (10) 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.[29] 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).

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 (10) for each individual sector.[30] Instead, without going into details, we

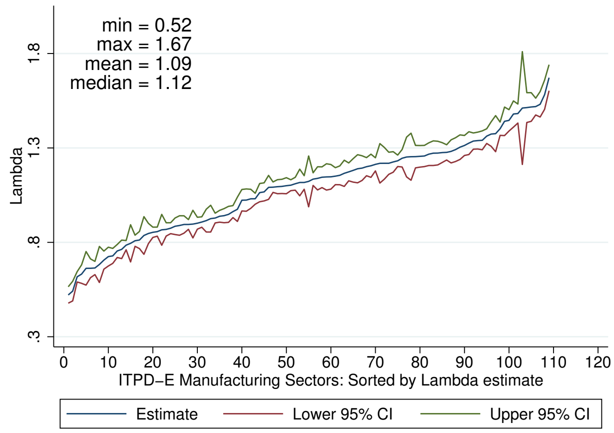
[28] ITPD-E-R02 is suitable for our purposes because it is constructed from raw/administrative data that have not been manipulated with statistical methods. In addition, ITPD-E-R02 includes a large number of sectors. Given our purposes, we only focus on 118 manufacturing sectors from ITPD-E-R02, and 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 stay 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 UN's Comtrade database.

[29] The two distance variables are consistently constructed using population-weighted distances between the major cities in each of the countries in our sample. The distance variables, as well as all other gravity covariates in our model, come from the USITC's *Dynamic Gravity Database* (Gurevich and Herman, 2018).

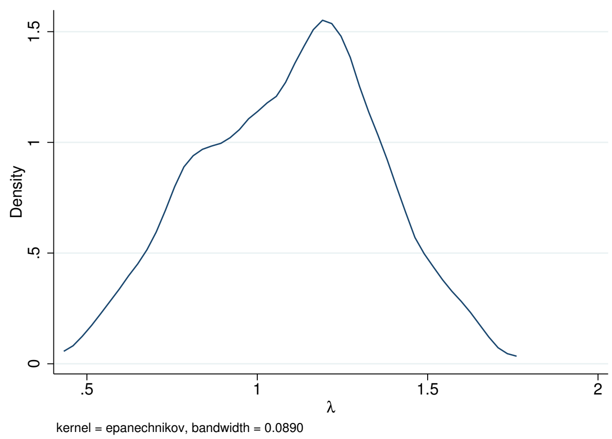
[30] All gravity estimates are available by request.

Figure 2: Estimates of Lambda: ITPD-E Manufacturing

(a) λ Estimates and Confidence Intervals



(b) Kernel Density λ



Notes: Panel (a) of this figure visualizes the estimates of λ along with their confidence intervals, which are obtained from specification (10) using all covariates from equation (10) as instruments. Panel (b) reports the kernel density of the λ estimates.

note that our PPML estimates are in line with the literature. Specifically, we obtain large, negative, and statistically significant estimates of the effects of international distance, domestic distance, and international borders and positive and statistically significant estimates of the impact of contiguous borders, colonial ties, common language, RTAs, EU membership, and WTO membership.

More important for our purposes, panel (a) of Figure 2 reports λs along with their confidence intervals, while panel (b) reports the kernel density of the distribution of λs. Four salient findings stand out from Figure 2. First, even though our framework does not impose any parameter range on λ and can accommodate negative λs if they are implied by the data, we find that all estimates of λ are strictly positive. Consistent with Santos Silva and Tenreyro's assumption behind PPML, 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.

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.[31] Panel (a) compares the PPML vs. G-PPML distance estimates directly against each other and reveals that most of the estimates are clustered around the 45-degree line, i.e., most of the PPML and G-PPML estimates of the effects of distance are very similar. This is an expected result because, as discussed earlier, most of the λ values that we obtained in the previous step were close to one. In addition, however, panel (a) 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. 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.

To test this hypothesis, we calculate the absolute value of the percentage difference between the sectoral PPML and 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. Then, for each gravity variable, we plot the sectoral percentage difference against the corresponding sectoral estimates of λ. Our results are reported in panel (b) of Figure 3, where, for exposition, we drop the top 5 percent of the observations in $\%\Delta\beta^{k,DIST}$.[32] The message from panel (b) is clear and exactly as expected. The U-shaped fitted curve reveals that the more λ deviates from one, the larger the differences between PPML and G-PPML. In extreme cases, i.e., when λ is around 0.6 or 1.6, the percentage difference between the PPML and G-PPML distance estimates can be as high as 40%. Figure 4 confirms this pattern for each of the other gravity variables in our model.

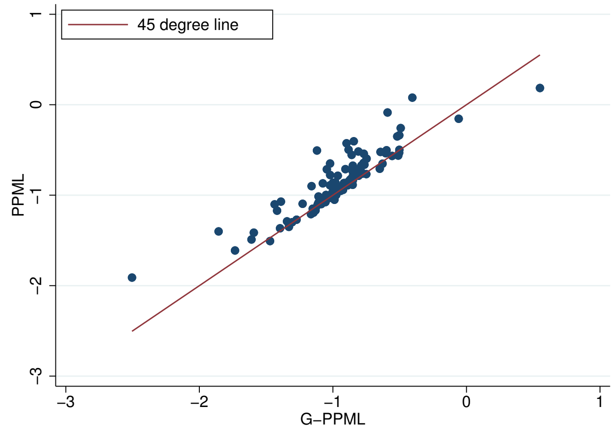
Thus far, we demonstrated that PPML and G-PPML may deliver quite different point estimates.

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

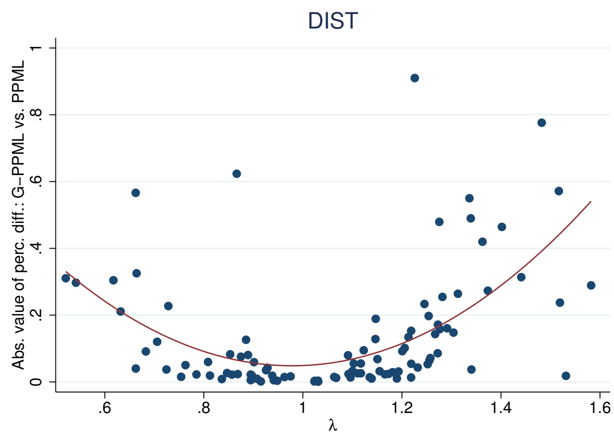
[32] The outliers in our analysis are due to the inherent difficulties of percentage differences to deal with small denominator values. Specifically, when the absolute value of $\hat{\beta}^{k,v}_{PPML}$ is very small, any difference in two estimates translates into a huge percentage difference.

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

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



(b) Percentage differences

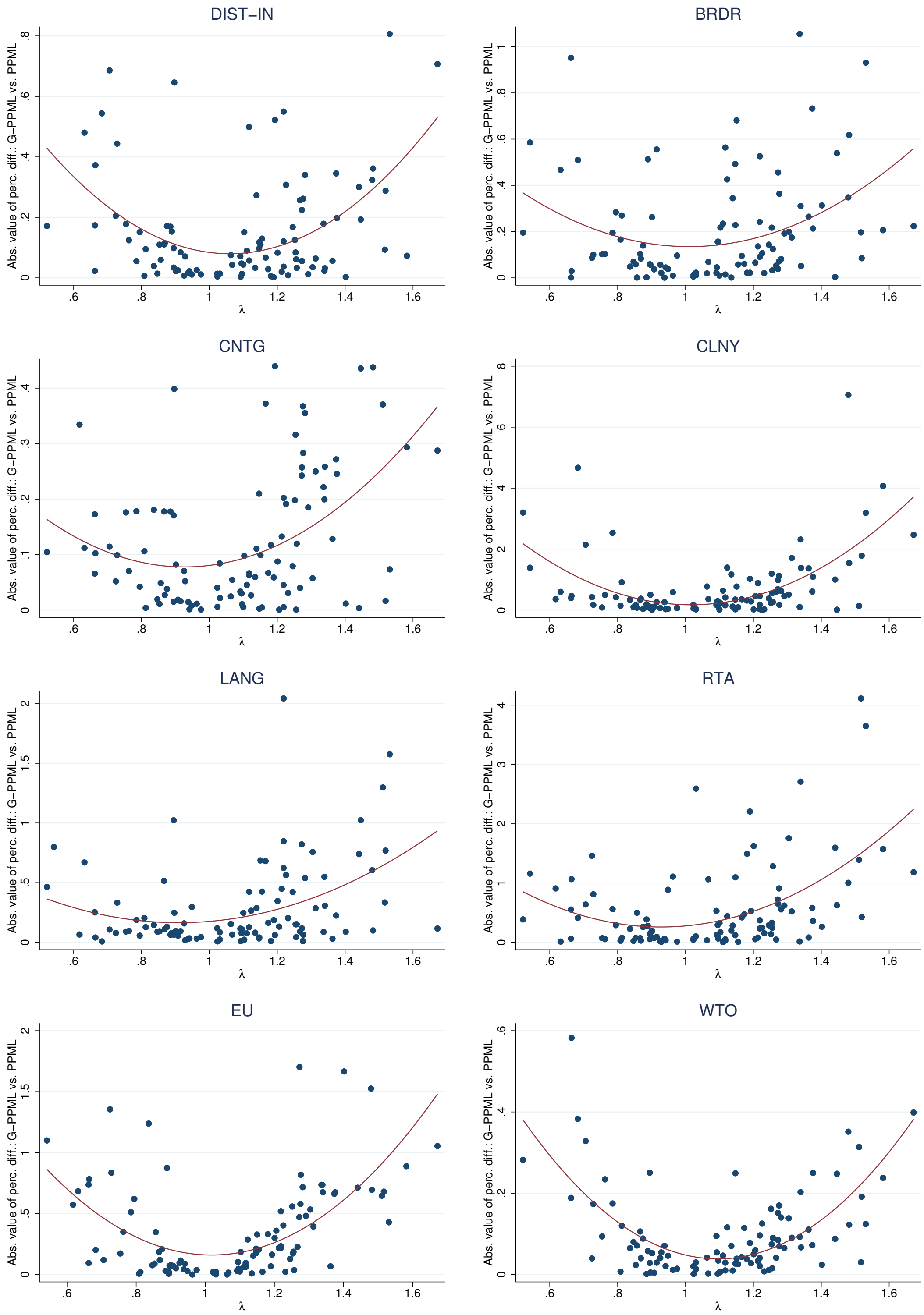


Notes: This figure plots PPML vs. G-PPML estimates of the coefficients of distance and their percentage differences. Panel (a) plots the PPML vs. G-PPML distance estimates against each other. Panel (b) plots the sectoral percentage difference in the DIST coefficients ($\%\Delta\beta^{k,DIST}$) estimated 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.

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 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:

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

Figure 4: 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.

Table 3: 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	87.50%	0.372	0.366	0.358	-0.065	0.788
CLNY	71.43%	0.415	0.260	1.120	-0.568	1.543
CNTG	84.76%	0.181	0.187	0.188	-0.099	0.404
DIST	84.76%	0.565	0.385	0.641	-0.019	1.487
DIST-IN	97.12%	0.531	0.532	0.322	0.126	0.875
EU	74.29%	0.098	0.184	0.410	-0.476	0.504
LANG	75.24%	0.444	0.287	0.719	-0.298	1.529
RTA	78.10%	0.459	0.342	1.087	-0.522	1.859
WTO	89.52%	0.184	0.177	0.162	-0.015	0.392

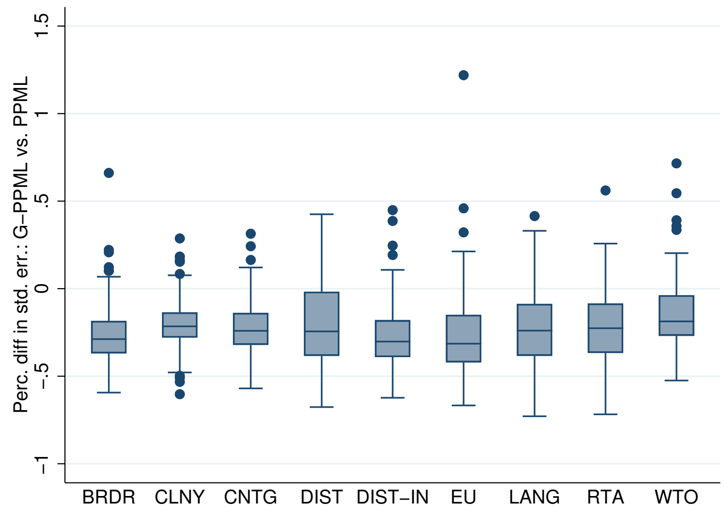
Notes: This table compares the standard errors (panel A) and z-statistics (panel B) obtained with PPML and G-PPML. We use the ITPD-E-R02 data to estimate the coefficients of gravity variables for each of 105 sectors. In panel A, the first column (lower SE) shows the percentage of sectors for which the 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.

We do not take the absolute value for $\%\Delta SE^{k,v}$, since the term is negative (positive) when standard errors of G-PPML is 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.

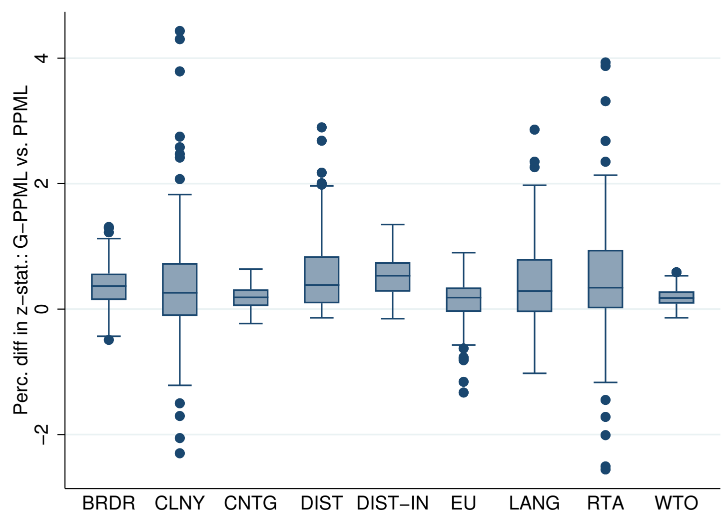
The results regarding the standard errors and z-statistics are presented in Table 3, and we visualize the full distribution of $\%\Delta SE^{k,v}$ and $\%\Delta z^{k,v}$ in Figure 5. In the column labeled 'lower SE' in panel A of Table 3, we show the percentage of sectors for which the G-PPML standard 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 3 show the key distributional statistics for $\%\Delta SE^{k,v}$. On average,

Figure 5: 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.

G-PPML standard errors are more than 20% smaller than the PPML standard errors, whereas in some extreme cases they are 50% smaller (p10). Panel (a) of Figure 5 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 3 and in panel (b) of Figure 5. Column 'greater $|z|$' of Table 3 reveals that, for the majority of cases (ranging from 71% for CLNY to 97% 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 57% for DIST. The G-PPML z-statistics also feature greater median values. Panel (b) of Figure 5 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 and a number of papers have added to the list of its attractive properties. Despite the success and popularity of PPML, some researchers have remained skeptical about its validity due to the assumption that the conditional mean of trade flows should to be proportional to its conditional variance (Head and Mayer, 2014). We contribute to this debate by capitalizing on the iGMM estimator of Hansen and Lee (2021) to propose a new *Generalized* PPML estimator that relies on actual data to estimate the conditional model variance. Using Monte Carlo analysis and an application with real sectoral trade data, we demonstrate the benefits of G-PPML in expanding the applicable domain of PPML. Our paper naturally leads to the following two questions: (i) What drives the underlying determinants of the conditional variance term as well as its variation across different dimensions (e.g., over time)? and (ii) What is the statistical property of G-PPML under three-way fixed effects à la Weidner and Zylkin (2021)? We leave the answers for future research.

References

Agnosteva, Delina E., James E. Anderson, and Yoto V. Yotov, "Intra-national Trade Costs: Assaying Regional Frictions," *European Economic Review*, 2019, *112* (C), 32–50.

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

_ and _ , "Trade costs," *Journal of Economic literature*, 2004, *42* (3), 691–751.

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

Arvis, Jean-Francois and Ben Shepherd, "The Poisson Quasi-Maximum Likelihood Estimator: A Solution to the "Adding Up" Problem in Gravity Models," *Applied Economics Letters*, April 2013, *20* (6), 515–519.

Baier, Scott L. and Jeffrey H. Bergstrand, "Do Free Trade Agreements Actually Increase Members' International Trade?," *Journal of International Economics*, 2007, *71* (1), 72–95.

Borchert, Ingo, Mario Larch, Serge Shikher, and Yoto V. Yotov, "The International Trade and Production Database for Estimation R02 (ITPD-E-R02)," *USITC,* forthcoming., 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. and Sergey Nigai, "Energy Demand and Trade in General Equilibrium," *Environmental and Resource Economics*, 2015, *60* (2), 191–213.

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

Fally, Thibault, "Structural Gravity and Fixed Effects," *Journal of International Economics*, 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, *192* (1), 291–312.

Gurevich, Tamara and Peter Herman, "The Dynamic Gravity Dataset: 1948-2016," 2018. USITC Working Paper 2018-02-A.

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 "Handbook of international economics," Vol. 4, Elsevier, 2014, pp. 131–195.

Jochmans, Koen, "Two-Way Models for Gravity," *The Review of Economics and Statistics*, 07 2017, *99* (3), 478–485.

Manning, Willard G and John Mullahy, "Estimating log models: to transform or not to transform?," *Journal of health economics*, 2001, *20* (4), 461–494.

Mnasri, Ayman and Salem Nechi, "New nonlinear estimators of the gravity equation," *Economic Modelling*, 2021, *95*, 192–202.

Newey, Whitney K. and Daniel McFadden, "Chapter 36 Large sample estimation and hypothesis testing," in "Handbook of Econometrics," Vol. 4, Elsevier, 1994, pp. 2111–2245.

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

_ and _ , "Further Simulation Evidence on the Performance of the Poisson Pseudo-Maximum Likelihood Estimator," *Economics Letters*, 2011, *112* (2), 220–222.

_ and _ , "The Log of Gravity At 15," *Portuguese Economic Journal*, 2022.

Weidner, Martin and Thomas Zylkin, "Bias and consistency in three-way gravity models," *Journal of International Economics*, 2021, *132*, 103513.

Yotov, Yoto V., "On the role of domestic trade flows for estimating the gravity model of trade," *Contemporary Economic Policy*, 2022.

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 (4). 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 if the parameters in $\mu_{ijt}=\exp\left(x_{ijt}'\beta+\gamma_{it}+\eta_{jt}\right)$ 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 correct specification. 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,

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

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

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

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,

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

and the lower bound's smallest eigenvalue λ_{min} is always strictly positive. Since 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

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

approximate the infeasible sample moment $\bar{m}_{0,N}\left(\bar{\theta}\right)$ 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 $\hat{\gamma}_{it}^{PPML} = \hat{r}_1\left(z_{ijt};\hat{\beta}^{PPML}\right)$ and $\hat{\eta}_{jt}^{PPML} = \hat{r}_2\left(z_{ijt};\hat{\beta}^{PPML}\right)$ for some functions $\hat{r}_1$ and $\hat{r}_2$ of $z_{ijt} = \left(y_{ijt}, x'_{ijt}\right)'$. Then,

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

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

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

and

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

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 $\bar{J}_N\left(\bar{\theta},\phi\right) = J_{N,0}\left(\bar{\theta},\phi\right) + O_P\left(\frac{1}{N}\right)$, since

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

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

Consider the case that the conditional variance form is misspecified, i.e., $\inf_{\phi\in\Theta}\mathbb{E}\left[x_{ijt}\left(\epsilon_{ijt}^2-h_\phi\cdot\mu_{ijt}^{\lambda_\phi}\right)\right]=\mathfrak{M}\neq 0$. 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 components are

$$\begin{aligned}
Q\left(g\left(\phi\right)\right) &= \mathbb{E}\begin{bmatrix}\mu_{ijt}^{\lambda_\phi}x_{ijt}' \\ h_\phi\cdot\mu_{ijt}^{\lambda_\phi}\log\left(\mu_{ijt}\right)x_{ijt}'\end{bmatrix}' \\
S\left(g\left(\phi\right)\right) &= \mathbb{E}\begin{bmatrix}2\left(\epsilon_{ijt}^2-h_\phi\cdot\mu_{ijt}^{\lambda_\phi}\right)\mu_{ijt}^{\lambda_\phi}vec\left(x_{ijt}x_{ijt}'\right)' \\ 2\left(\epsilon_{ijt}^2-h_\phi\cdot\mu_{ijt}^{\lambda_\phi}\right)h_\phi\mu_{ijt}^{\lambda_\phi}\log\left(\mu_{ijt}\right)vec\left(x_{ijt}x_{ijt}'\right)'\end{bmatrix}',
\end{aligned}$$

and

$$vec\left(\frac{\partial}{\partial\bar{\theta}\partial\bar{\theta}'}J\left(\bar{\theta},\phi\right)|_{\bar{\theta}=g(\phi)}\right)$$

$$=2\begin{bmatrix}\mathbb{E}\left[x_{ijt}\mu_{ijt}^{\lambda_\phi}\right]'\mathbb{E}\left[x_{ijt}\mu_{ijt}^{\lambda_\phi}\right]\\ \mathbb{E}\left[x_{ijt}\mu_{ijt}^{\lambda_\phi}\log\left(\mu_{ijt}\right)\right]'\mathbb{E}\left[x_{ijt}\left(\mu_{ijt}^{\lambda_\phi}+\left(\epsilon_{ijt}^2-h_\phi\cdot\mu_{ijt}^{\lambda_\phi}\right)\right)\right]\\ \mathbb{E}\left[x_{ijt}\mu_{ijt}^{\lambda_\phi}\log\left(\mu_{ijt}\right)\right]'\mathbb{E}\left[x_{ijt}\left(\mu_{ijt}^{\lambda_\phi}+\left(\epsilon_{ijt}^2-h_\phi\cdot\mu_{ijt}^{\lambda_\phi}\right)\right)\right]\\ \mathbb{E}\left[x_{ijt}\mu_{ijt}^{\lambda_\phi}\log\left(\mu_{ijt}\right)\right]'\mathbb{E}\left[x_{ijt}\mu_{ijt}^{\lambda_\phi}\log\left(\mu_{ijt}\right)\right]+\mathbb{E}\left[x_{ijt}h_\phi\mu_{ijt}^{\lambda_\phi}\log^2\left(\mu_{ijt}\right)\right]\mathbb{E}\left[x_{ijt}\left(\epsilon_{ijt}^2-h_\phi\cdot\mu_{ijt}^{\lambda_\phi}\right)\right]\end{bmatrix}.$$

Under the uniform boundedness 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 rely on $\mathbb{E}\left[x_{ijt}\mu_{ijt}^{\lambda_\phi}\right]$, $\mathbb{E}\left[x_{ijt}\mu_{ijt}^{\lambda_\phi}\log\left(\mu_{ijt}\right)\right]$, and $\mathbb{E}\left[x_{ijt}\mu_{ijt}^{\lambda_\phi}\log^2\left(\mu_{ijt}\right)\right]$, 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 $\mathbb{E}\left[x_{ijt}\left(\epsilon_{ijt}^2-h_\phi\cdot\mu_{ijt}^{\lambda_\phi}\right)\right]=\mathfrak{M}$, $\mathbb{E}\left[x_{ijt}\mu_{ijt}^{\bar{\lambda}}\right]=L_1\left(\mu_{ijt}\right)=L_1$, $\mathbb{E}\left[x_{ijt}\mu_{ijt}^{\bar{\lambda}}\log\left(\mu_{ijt}\right)\right]=L_2\left(\mu_{ijt}\right)=L_2$, $\mathbb{E}\left[x_{ijt}\mu_{ijt}^{\bar{\lambda}}\log^2\left(\mu_{ijt}\right)\right]=L_3\left(\mu_{ijt}\right)=L_3$, $\bar{h}=1$ and $W\left(\phi\right)=I$ satisfy the the iGMM existence conditions C_1, C_2 and C_3. Then, the iGMM estimator converges to a limit if

$$\|\mathfrak{M}\|^2<\frac{\left\|\left(L_1'L_1,L_2'\left(L_1+\mathfrak{M}\right),L_2'\left(L_1+\mathfrak{M}\right),L_2'L_2+L_3'\mathfrak{M}\right)'\right\|^2}{\left\|\left(L_1,L_2\right)'\right\|^2\|\mathfrak{M}\|^2},$$

and the upper bound for the degree of misspecification $\mathfrak{M}$ depends on L_1, L_2, and L_3. 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.2 Proof of Proposition 2.2

The asymptotic normality of $\hat{\theta}$ follows the standard theory for an overidentified GMM estimator's large sample properties. Considering potential misspecification in conditional variance, we adopt the asymptotic theory from Hansen and Lee (2021). We start from the correct specification case. 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

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

Define some more notation following Hansen and Lee (2021). Define $\bar{R}_N\left(\bar{\theta}\right)=\frac{\partial}{\partial\theta}vec\left(\bar{Q}_N\left(\bar{\theta}\right)'\right)$,

$\mathfrak{M} = m(\theta)$, $Q = Q(\theta)$, $W = W(\theta)$, $R(\bar{\theta}) = \frac{\partial}{\partial \theta'} vec\left(Q(\bar{\theta})'\right)$, $R = R(\theta)$, $S = S(\theta) = \frac{\partial}{\partial \theta'} vec(W(\theta))$, $\bar{m}_N = \bar{m}_N(\theta)$, $\bar{Q}_N = \bar{Q}_N(\theta)$, 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

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

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 the estimator with no asymptotic bias, hence

$$N\left(\hat{\beta}^{PPML} - \beta\right) \xrightarrow{d} N\left(0, V^{PPML}\right)$$

under the two-way fixed effects and a smooth function of $\hat{\beta}^{PPML}$ also preserves the zero asymptotic bias, as

$$N\left(r\left(\hat{\beta}^{PPML}\right) - r(\beta)\right) \approx r'(\beta) N\left(\hat{\beta}^{PPML} - \beta\right) \xrightarrow{d} N\left(0, r'(\beta)' V^{PPML} r'(\beta)\right).$$

Thus,

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

where $\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T} x_{ijt}\left(\left(\hat{\epsilon}_{ijt}^2 - \epsilon_{ijt}^2\right) + h\left(\hat{\mu}_{ijt}^{\lambda} - \mu_{ijt}^{\lambda}\right)\right) \xrightarrow{d} N(0, V)$ for some V and $W = E\left[x_{ijt}x_{ijt}'\left(\epsilon_{ijt}^2 - h \cdot \mu_{ijt}^{\lambda}\right)^2\right]$. Let $v(\beta) = \epsilon_{ijt}^2 - h \cdot \mu_{ijt}^{\lambda}$ and $v\left(\hat{\beta}^{PPML}\right) = \hat{\epsilon}_{ijt}^2 - h \cdot \hat{\mu}_{ijt}^{\lambda}$. The resulting asymptotic distribution is

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

and

$$V = E\left[x_{ijt} v'(\beta)' V^{PPML} v'(\beta) x_{ijt}'\right].$$

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 $(Q'WQ)^{-1}$, 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 $\hat{\theta}$.

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. First,

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

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

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

when a mild misspecification problem exists, we define a new matrix $\Omega_{\mathfrak{M}} = \frac{1}{N^2}\sum_{i,j,t}\mathbb{E}\left[\psi_{ijt}\psi_{ijt}'\right]$, where

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

and $\hat{Q}\left(x_{ijt}, \theta\right) = \left[x_{ijt}\hat{\mu}_{ijt}^{\lambda}, x_{ijt}h \cdot \hat{\mu}_{ijt}^{\lambda}\log\left(\hat{\mu}_{ijt}\right)\right]$. Note that assumptions for asymptotic normality are already verified. We additionally assume that the degree of misspecification $\mathfrak{M}$ satisfies the condition in Section A.1.1. Then, following Theorem 4 of Hansen and Lee (2021),

$$N\left(\hat{\theta} - \theta\right) \xrightarrow{d} N\left(0, H_{\mathfrak{M}}^{-1}\Omega_{\mathfrak{M}}H_{\mathfrak{M}}^{-1}\right).$$

A.3 Proof of Proposition 2.3

Note that Proposition 2 of Weidner and Zylkin (2021) provides consistency of the infeasible estimator $\tilde{\beta}$

$$\begin{aligned}
\tilde{\beta} &: \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T} x_{ijt}\left(y_{ijt} - \tilde{\mu}_{ijt}\right)\tilde{\mu}_{ijt}^{1-\lambda} = 0, \\
\tilde{\gamma}_{it} &: \sum_{j=1}^{N}\left(y_{ijt} - \tilde{\mu}_{ijt}\right)\tilde{\mu}_{ijt}^{1-\lambda} = 0, \\
\tilde{\eta}_{jt} &: \sum_{i=1}^{N}\left(y_{ijt} - \tilde{\mu}_{ijt}\right)\tilde{\mu}_{ijt}^{1-\lambda} = 0,
\end{aligned}$$

even for the three-way fixed effects case. The estimator $\hat{\beta}$ solves the FOCs

$$\begin{aligned}
\hat{\beta} &: \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T} x_{ijt}\left(y_{ijt} - \hat{\mu}_{ijt}\right)\hat{\mu}_{ijt}^{1-\hat{\lambda}} = 0, \\
\hat{\gamma}_{it} &: \sum_{j=1}^{N}\left(y_{ijt} - \hat{\mu}_{ijt}\right)\hat{\mu}_{ijt}^{1-\hat{\lambda}} = 0, \\
\hat{\eta}_{jt} &: \sum_{i=1}^{N}\left(y_{ijt} - \hat{\mu}_{ijt}\right)\hat{\mu}_{ijt}^{1-\hat{\lambda}} = 0,
\end{aligned}$$

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 feasible 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).

Note that 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 $\hat{\beta}$.

A.4 Proof of Proposition 2.4

The derived asymptotic variance is a direct application of Theorem 4.1 of Fernández-Val and Weidner (2016). The asymptotic bias following Weidner and Zylkin (2021) follows the form of $W_N^{-1}\left(B_N^m+D_N^m\right)/N$. Note that

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

and

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

Since W_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, $S_{ij,t} = \left(y_{ijt}-\mu_{ijt}\right)\mu_{ijt}^{1-\lambda}$ and $\mathbb{E}\left[S_{ij,t}S_{ij,s}\right] = \mu_{ijt}^{2-\lambda}1\left\{t=s\right\}$. Similarly, $\left[\bar{H}_{ij}\right]_{ts} = \mu_{ijt}^{2-\lambda}1\left\{t=s\right\}$ and $\left[\bar{G}_{ij}\right]_{tsr}$ =$(3-2\lambda)\,\mu_{ijt}^{2-\lambda}1\left\{t=s=r\right\}$. The second terms of B_N^m and D_N^m can be reduced by $\frac{1}{2N}\sum_{j=1}^{N} Tr\left[\left(\sum_{i=1}^{N}\bar{H}_{ij}\right)^{-1}\left(\sum_{i=1}^{N}\bar{G}_{ij}\tilde{x}_{ij,m}\right)\right]$ since $\mathbb{E}\left[S_{ij,t}S_{ij,s}\right] = \left[\bar{H}_{ij}\right]_{ts}$.

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, $\sum_{j=1}^N \bar{G}_{ij}\tilde{x}_{ij,m} = \sum_{j=1}^N \bar{H}_{ij}\tilde{x}_{ij,m} = 0$ 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 $\lambda = 1$. But for any other λ values, $\sum_{j=1}^N \bar{G}_{ij}\tilde{x}_{ij,m} = (3-2\lambda)\sum_{j=1}^N \bar{H}_{ij}\tilde{x}_{ij,m} = 0$ for all PML estimators. 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 PML estimator's assumption holds. That is, $\sum_{i=1}^N \mathbb{E}\left[H_{ij}\tilde{x}_{ij,m}S'_{ij}|x_{ij,m}\right] = \sum_{j=1}^N \mathbb{E}\left[H_{ij}\tilde{x}_{ij,m}S'_{ij}|x_{ij,m}\right] = 0$ and there is no asymptotic bias. Since

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

is the mth column of the matrix $H_{ij}\tilde{x}_{ij,m}S'_{ij}$, we compute if the column elements are zeros. Since $\mathbb{E}\left[\left(y_{ijt}-\mu_{ijt}\right)^2|x_{ij}\right] = h\cdot\mu_{ijt}^{\lambda}$ under the correctly specified conditional variance and $\mathbb{E}\left[y_{ijt}-\mu_{ijt}|x_{ij}\right] = 0$,

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

Therefore,

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

and by the FOC of $\tilde{x}_{ij}$,

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

We find that B_N^m and D_N^m are always zero if $\lambda = 1$. But even if $\lambda \neq 1$ and $h\left(1-\lambda\right) \neq 0$, $\sum_{j=1}^N\sum_{i=1}^N \tilde{x}_{ij,m}\mu_{ijt}^{2-\lambda} = 0$ since the model is correctly specified. Under the class of PML estimators, we do not have asympotic bias, thus no bias correction is needed. For general conditional variances, $\sum_{i=1}^N \mathbb{E}\left[H_{ij}\tilde{x}_{ij,m}S'_{ij}|x_{ij,m}\right]$ nor $\sum_{j=1}^N \mathbb{E}\left[H_{ij}\tilde{x}_{ij,m}S'_{ij}|x_{ij,m}\right]$ will be zero, causing non-zero asymptotic bias of the estimator.

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

If the conditional variance of y_{ijt} does not follow equation (2), the provided asymptotic bias formula offers a feasible method to correct the asymptotic bias of $\hat{\beta}$. 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 $\left[\hat{H}_{ij}\right]_{ts}=\hat{\mu}_{ijt}^{1-\hat{\lambda}}\left(1-\hat{\lambda}\right)\left(y_{ijt}-\frac{2-\hat{\lambda}}{1-\hat{\lambda}}\hat{\mu}_{ijt}\right)1\left\{t=s\right\}$. Then,

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

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

where $\hat{S}_{ij,t}=\left(y_{ijt}-\hat{\mu}_{ijt}\right)\hat{\mu}_{ijt}^{1-\hat{\lambda}}$, and $\hat{\tilde{x}}_{ij}=x_{ij}-\hat{\gamma}_i^x-\hat{\eta}_j^x$. $\hat{\Omega}_N^{-1}\left(\hat{B}_N+\hat{D}_N\right)/N$ is the estimated correction term.