

The Surprising Bias of PPML Estimates of Structural Gravity Models with Two-Way Fixed Effects^{*}

Ben Shepherd, Developing Trade Consultants.

Tom Zylkin, University of Richmond.[†]

August 8th, 2025

Abstract: The previous literature has shown that the Poisson Pseudo-Maximum Likelihood (PPML) estimator provides consistent and asymptotically unbiased estimates of the parameters of structural gravity models with two-way fixed effects, although their standard errors need correction. We show that, notwithstanding this result, PPML suffers from bias in small samples. Surprisingly, that bias is evident in the types of samples typically used for gravity models, which can involve tens of thousands of observations. We explore the nature and extent of the bias both theoretically and through simulations, and we show that bias corrections based on the bootstrap and split-sample jackknife can attenuate but not necessarily eliminate it. An empirical application suggests that, especially at the sectoral level, the bias we have identified could be significant, particularly in settings where point estimates are of primary importance, such as when gravity model estimates of the impact of a trade agreement or some other non-tariff policy are used in a second stage general equilibrium impact assessment of a counterfactual policy change.

Key words: Distance elasticity; Cross-sectional trade data; Finite-sample bias

JEL codes: C13, C50, F10

^{*}*Acknowledgements:* We would like to thank Peter Egger, two anonymous referees, Joao Santos Silva, Amrei Stammann, and Joschka Wanner for useful feedback and discussions.

[†]*Contact information:* Shepherd: ben@developing-trade.com. Zylkin (corresponding author): tzylkin@richmond.edu.

1 INTRODUCTION

Santos Silva and Tenreyro (2006, 2021) show that Poisson Pseudo-Maximum Likelihood (PPML) is a reasonable workhorse estimator for structural gravity models. It provides consistent estimates under the relatively weak assumption that the conditional mean function is correctly specified. It performs well in the presence of zero trade flows, and it is robust to heteroskedasticity in the error (unobserved trade costs) term. In addition, because gravity models typically include fixed effects, a further principal advantage of the PPML estimator is its unique robustness to the incidental parameter problem. As shown in Wooldridge (1999) and Weidner and Zylkin (2021), PPML is often consistent when other fixed effects estimators are not. Moreover, as the results of Fernandez-Val and Weidner (2016) imply, PPML estimates of the standard gravity model with exporter and importer fixed effects have the unique property that they are “asymptotically unbiased” (approach the true parameter values without any bias in their distributions) as the number of countries becomes large.

Given these favorable large sample properties, the literature since Santos Silva and Tenreyro (2006) has taken up PPML estimation and made it the standard approach for estimating structural gravity models. But relatively little attention has been paid to its *small* sample properties in the gravity context. While gravity samples are not “small” in an absolute sense—they can easily number tens of thousands of observations—the behavior of the estimator in finite samples nonetheless deserves consideration. After all, though it only occasionally receives attention in the gravity literature, PPML estimates are generally biased in finite samples, and it is interesting to wonder just how large these biases are in practice and how large a trade sample is needed for a researcher to comfortably disregard them.

To fix ideas, Head and Mayer (2014) and Shepherd (2025) have each shown that, under certain assumptions, PPML estimates of structural gravity models with two-way fixed effects can display noticeable bias for the sample sizes typically used in that context. These findings suggest that while PPML has desirable asymptotic properties, the properties of its finite sample bias may also be of interest in applied contexts in international trade. Building off of Fernandez-Val and Weidner (2016)’s results for PPML estimates of two-way fixed effects models, this finite sample bias is not due to the incidental parameter problem per se but is more closely related to the small sample bias generally exhibited by nonlinear maximum likelihood estimators. Whereas Head and Mayer (2014) report major bias in PPML estimates, Shepherd (2025) undertakes simulations over a range of parameter values, focusing on changes in sample size and the amount of unexplained variance. He documents that the degree of bias depends negatively on sample size, as expected for a small-sample issue, but he also finds it depends positively on the amount of noise in the data. Repeating the Head and Mayer (2014) simulations with a calibrated variance assumption that closely matches observation results in bias of around 5% in parameter estimates, which, though much smaller than the bias reported by Head and Mayer (2014), is still larger than one might expect strictly based on the sample size.

Given the possibility of non-negligible bias of two-way fixed effects PPML estimates, we expect that researchers will benefit from additional guidance about what can and/or should be done about this issue. Therefore, this paper builds on and extends previous work in a number of directions, focusing on the finite sample properties of PPML estimates. First, we extend the existing results of Fernandez-Val and Weidner (2016) to characterize the bias of two-way fixed effects PPML estimates. In particular, we clarify using Taylor expansions that the bias of PPML estimates with two-way fixed effects is of order $1/N^2$ asymptotically, where N is the number of countries. Moreover, we show what this $1/N^2$ bias depends on—in particular how it depends on the estimation noises of all the estimated

fixed effects. This is a stronger result than the one in Fernandez-Val and Weidner (2016), who only show that the fixed effects do not induce a bias of order $1/N$, such that its limit distribution is unbiased; they do not otherwise investigate the order of PPML’s bias or examine ways in which the fixed effects could contribute to it. Their analysis does not, for example, rule out the possibility that the bias could instead be of order $N^{-3/2}$. Therefore, by clarifying that the order of the bias is indeed $1/N^2$, we provide a necessary theoretical justification for the use of jackknife-based bias corrections to remove the bias. Due to the small order of the bias, such corrections should not be expected to improve inferences asymptotically; however, in small enough samples they still might be useful for reducing bias.

Second, we provide additional simulation evidence on the nature and extent of the finite sample bias in two-way fixed effects PPML estimates. It is more persistent than might be expected of a “small sample” bias, and it also depends on the nature of the assumed error variance. Our simulation evidence is more systematic than the earlier Monte Carlos put forward in Head and Mayer (2014) and Shepherd (2025) in that we present results for different combinations of the sample size, the amount of unexplained variance, and how the heteroskedasticity of the conditional variance varies with the conditional mean. As would be expected, different assumptions about heteroskedasticity result in different bias behaviors, and we confirm that main result of Shepherd (2025) that the bias increases with the unexplained variance irrespective of the sample size. We also use a bootstrapping procedure based on actual trade data in order to undertake additional simulations that more accurately resemble real world use cases.

Third, we show that the bootstrap and split-sample jackknife each provide a straightforward means for attenuating, though not necessarily eliminating, the bias. Gains are modest for simulation settings with a smaller number of countries and/or a larger amount of variance, but they become increasingly reliable as the number of countries increases, and they tend to be more reliable when the conditional error variance accords with the Poisson assumption. Though the bootstrap- and jackknife-based approaches perform similarly, we generally find the overall performance of the bootstrap is weakly superior to that of the jackknife. Unlike with asymptotic bias corrections used in the incidental parameter problem literature, we do not find, however, that applying these bias corrections reliably leads to an improvement in coverage. The reason is that the bias corrections increase the standard deviation of the estimates.

Finally, we take our bias correction approach to real-world data, and we show that the difference between bias-corrected and standard PPML estimates with two-way fixed effects can sometimes be substantial in the kinds of finite samples typically encountered in applied work, especially when we examine results at the sectoral level. Encouragingly, the bootstrap and jackknife produce qualitatively similar bias corrections across almost all sectors as well as for aggregate trade data. The credibility of these bias corrections is further supported by the results from our additional bootstrap simulations based on the aggregate trade data. The latter suggest that the standard deviations of the bias-corrected estimates are only modestly larger than those of the uncorrected ones in the real data setting, allaying concerns about the potential for worsened coverage.

In applied contexts, use of bootstrap- or jackknife-based bias corrections would need to pass a benefit-cost test. On the cost side is the added computational burden as well as the potential for worsened coverage, although our empirical exercise suggests that bias corrections will likely improve inferences in practice or leave them mostly unaffected more often than not. On the benefit side is the reduced

bias of the point estimates, though not always significantly so. We can see two potential use cases for our approach against this background. One is the case where a researcher wishes to verify that the bias is “not too large”, in which case bootstrap or jackknife estimates could be compared with PPML, as in our empirical exercise. A second is a context where having less biased point estimates is key. An example of such a case is where gravity model estimates provide the first stage for a general equilibrium counterfactual simulation of a policy change. The primary emphasis there is on obtaining meaningful point estimates of parameters, not inference as such.

The paper proceeds as follows. The next section characterizes the bias of PPML with two-way fixed effects and discusses the bootstrap and jackknife as ways of attenuating it. Section 3 conducts simulations, while Section 4 examines the differences between bootstrapped and standard PPML estimates using observed trade data. Section 5 provides concluding remarks, and our Appendix adds more formal discussion of the bias results to be described in the next section, including a proof.

2 CHARACTERIZING AND CORRECTING THE BIAS

Drawing on Head and Mayer (2014), the two-way gravity model may be written as:

$$(1) y_{ij} = S_i M_j \tau_{ij}^{-\theta},$$

where $i, j \in (1, \dots, N)$ respectively are indices for exporter and importer, τ_{ij} captures bilateral trade costs, θ is a trade elasticity parameter, and S_i and M_j respectively reflect supply conditions in the exporter and demand conditions in the importer. As in Head and Mayer (2014) and much of the related literature (Eaton and Kortum 2002, Anderson and van Wincoop 2003, Yotov et al 2016), S_i and M_j need not be regarded as fixed parameters but rather are allowed to depend on each country’s full set of trade costs in general equilibrium. Nonetheless, the influence they exert on trade is country-specific, and this allows them to be estimated using exporter and importer fixed effects. Thus, a typical econometric implementation is:

$$(2) y_{ij} = \exp(\alpha_i + \gamma_j + x_{ij}\beta) \omega_{ij},$$

which parameterizes trade costs as $\tau_{ij}^{-\theta} = \exp(x_{ij}\beta) \omega_{ij}$, with x_{ij} serving as the covariates of interest and ω_{ij} serving as the error term, interpreted substantively as reflecting unobserved trade costs. The two-way fixed effects α_i and γ_j take the place of S_i and M_j . As is typical in the gravity literature, we assume that the error term ω_{ij} is cross-sectionally independent, i.e., independent across pairs.¹ When

¹ Fernandez-Val and Weidner (2016) derive bias results for the estimation of panel data models with two-way fixed effects and weak time dependence. Since we work off of Fernandez-Val and Weidner (2016), but adapt their approach to cross-sectional gravity models, we conjecture that it should be possible (although not without complicating matters) to generalize our results to allow for weakly dependent errors for observations sharing the same exporter and/or the same importer, such as through spatial decay of country-specific shocks. As in Fernandez-Val and Weidner (2016), maintaining weak dependence should not affect our high-level results regarding the order of the bias. It also should not affect the applicability of the split-sample jackknife. That said, though the literature does not provide guidance for the case of strong (non-decaying) dependence, it may still be an empirically relevant one and therefore deserving of more study.

referring to individual regressors we will use $k \in (1, \dots, K)$ as an index, such that K is the number of regressors.

As is now well known, consistent estimates of the trade cost parameter β can be obtained using PPML as well as other similar estimators. However, nonlinear maximum likelihood estimators (and by extension, pseudo-maximum likelihood estimators) generally exhibit bias in finite samples. Firth (1993), for example, provides a general consideration of the issue, and discusses analytical corrections. To examine the question rigorously, equation (3.2) in Fernandez-Val and Weidner (2016) makes it possible to characterize the bias of a nonlinear estimator with two-way fixed effects as follows:

$$(3) E(\hat{\beta}^{FE} - \beta^0) = \frac{B_{\infty}^{\beta}}{N} + \frac{D_{\infty}^{\beta}}{N} + o\left(\frac{1}{N}\right),$$

where: $\hat{\beta}^{FE}$ is the fixed effects estimator; β^0 is the true parameter; B_{∞}^{β} and D_{∞}^{β} are bias terms reflecting the incidental parameters problem that do not differ with N ; and $o(1/N)$ refers to terms of order smaller than $1/N$. Our equation (3) differs from their presentation because it considers the relevant case for a single-year gravity model, namely a square dataset of N countries rather than a typical panel with cross-sectional and time dimensions.²

Notably, Fernandez-Val and Weidner (2016) show that for Poisson (and by extension, PPML), $B_{\infty}^{\beta} = D_{\infty}^{\beta} = 0$ (see their Example 2.) This result means that PPML estimates of a model with two-way fixed effects do not suffer from bias due to the incidental parameters problem, which is captured by the B_{∞}^{β} and D_{∞}^{β} terms. This is a significant result because the standard deviation of $\hat{\beta}^{FE}$ is itself of order $1/N$. If these bias terms were not zero, the bias of the estimator would be of a similar magnitude to its standard error even as N becomes arbitrarily large, causing standard hypothesis test statistics to be inconsistent even if the coefficient estimates themselves are not. However, though it is not their focus, their analysis indicates that $\hat{\beta}^{FE}$ obtained by PPML is still a biased estimate of β^0 because of the trailing term $o(1/N)$. Importantly, this remaining bias does not affect PPML's asymptotic distribution or, in turn, the validity of standard test statistics used for inference, which are based on the asymptotic distribution. Nonetheless, addressing this bias could still be useful in empirical settings.

Our interest, therefore, is in understanding better the determinants of the remaining $o(1/N)$ bias not explored in Fernandez-Val and Weidner (2016) and giving a more precise characterization of its order. In the Appendix, we derive an expression for the bias of the PPML estimator with two-way fixed effects under cross-sectional independence and show that its bias depends on the joint distribution of all the parameter estimates, including the amount of noise in their distributions. Concretely, the Appendix shows that as $N \rightarrow \infty$, the bias of the fixed effects PPML estimate for β is given by:

² Because trade data often does not include the self-trade term y_{ii} , it could be the sample size is $N(N - 1)$ rather than N^2 . Because this difference only encumbers the notation without making any practical difference to our results, which are obtained under an asymptotic where $N \rightarrow \infty$, we always express the sample size as though it is N^2 . Our results can also be extended to cases where some trade flows are randomly missing and/or the number of exporters differs from the number of importers.

$$(4) \frac{b_1}{N^2} + \frac{b_{2\alpha}}{N^2} + \frac{b_{2\gamma}}{N^2} + \frac{b_3}{N^2} + \frac{b_4}{N^2} + o(N^{-2})$$

where:

- $b_1 = \sum_{i,j} \sum_k q_{ij}^{(\beta\beta_k)} E[(\hat{\beta} - \beta_0)(\hat{\beta}_k - \beta_{0,k})] = O(1)$
- $b_{2\alpha} = \sum_{i,j} q_{ij}^{(\beta\alpha_i)} E[(\hat{\beta} - \beta_0)(\hat{\alpha}_i - \alpha_{i0})] = O(1)$
- $b_{2\gamma} = \sum_{i,j} q_{ij}^{(\beta\gamma_j)} E[(\hat{\beta} - \beta_0)(\hat{\gamma}_j - \gamma_{j0})] = O(1)$
- $b_3 = \sum_{i,j} q_{ij}^{(\alpha_i\gamma_j)} E[(\hat{\alpha}_i - \alpha_{i0})(\hat{\gamma}_j - \gamma_{j0})] = O(1)$
- $b_4 = O(1)$ collects 3 similar terms that respectively depend on $E[(\hat{\alpha}_i - \alpha_{i0})^2(\hat{\gamma}_j - \gamma_{j0})]$, $E[(\hat{\alpha}_i - \alpha_{i0})(\hat{\gamma}_j - \gamma_{j0})^2]$, and $E[(\hat{\alpha}_i - \alpha_{i0})^2(\hat{\gamma}_j - \gamma_{j0})^2]$.
- $q_{ij}^{(\beta\beta_k)}$, $q_{ij}^{(\beta\alpha_i)}$, $q_{ij}^{(\beta\gamma_j)}$, and $q_{ij}^{(\alpha_i\gamma_j)}$ are $K \times K$ matrices whose elements are constants that depend on the PPML Hessian and on the pseudolikelihood of observation ij . The terms that make up b_4 depends on similar objects.
- $o(N^{-2})$ collects all terms of order less than N^{-2} .

(See the Appendix for an explanation of more details regarding the derivation of b_1 , $b_{2\alpha}$, $b_{2\gamma}$, b_3 , and b_4 .)

Because the 5 b -terms shown in (4) are each of order-1 as $N \rightarrow \infty$ (i.e., they do not vary with N), the combined leading bias term $(b_1 + b_{2\alpha} + b_{2\gamma} + b_3 + b_4)/N^2$ is of order N^{-2} . Like in Fernandez-Val and Weidner (2016) and other similar work, the expressions for b_1 , $b_{2\alpha}$, $b_{2\gamma}$, b_3 , and b_4 can be derived from a Taylor expansion of the PPML first-order condition for $\hat{\beta}$ around the true parameter values $(\beta_0, \alpha_0, \gamma_0)$. The first term b_1/N^2 can be shown to be roughly equivalent to the typical bias of a PPML estimator without fixed effects (see, e.g., Firth 1993 or Fernandez-Val and Weidner 2018, section 2.1.) The next 4 terms are present because, as we show in the Appendix, the covariances between $\hat{\beta}$ and $\hat{\alpha}_i$, $\hat{\beta}$ and $\hat{\gamma}_j$, and $\hat{\alpha}_i$ and $\hat{\gamma}_j$ all turn out to be of order N^{-2} , and the same turns out to be true for the higher-order interactions between $\hat{\alpha}_i$ and $\hat{\gamma}_j$ that give rise to b_4 .

For readers less interested in technical details, we suggest emphasizing two main takeaways. First, by showing that the bias of PPML is of order N^{-2} , we also clarify what it is *not*. Namely, we rule the possibility that the bias is, e.g., order- $N^{-3/2}$, which not only would indicate a more pernicious bias but also would have implications for how to correctly implement jackknife-based bias corrections.³ In the

³ The possibility of an order- $N^{-3/2}$ bias is worth focusing on because the full expansion of $\hat{\beta}$'s first-order condition includes terms that depend on, e.g., $(\hat{\beta} - \beta_0)(\hat{\alpha}_i - \alpha_{i0})$, $(\hat{\alpha}_i - \alpha_{i0})^3$, $(\hat{\alpha}_i - \alpha_{i0})^2(\hat{\gamma}_j - \gamma_{j0})$, etc.. Since $\hat{\beta}$ is estimated based on all N^2 observations, while each $\hat{\alpha}_i$ and $\hat{\gamma}_j$ is estimated using only N observations, a simple rule of thumb suggests each of these terms could produce a bias of order $N^{-3/2}$, i.e., the product of the order of their standard deviations.

Appendix, we show that all terms from $\hat{\beta}$'s full bias expansion that could conceivably contribute biases of order $N^{-3/2}$ are either equal to 0 or turn out to contribute biases of order N^{-2} or less.

Second, since the order of the bias is the same as that of standard PPML without fixed effects, it is interesting to note how the fixed effects estimates nonetheless do still contribute to the bias. Interestingly, it turns out that they only do so *indirectly*, i.e., through their interactions with the main parameter estimates and with each other. That is, the bias of two-way PPML does not depend on the variances of either $\hat{\alpha}_i$ or $\hat{\gamma}_j$, as was effectively shown by Fernandez-Val and Weidner (2016), and it turns out it does not depend on any of their higher-order central moments either (e.g., $E[(\hat{\alpha}_i - \alpha_{i0})^3]$, $E[(\hat{\alpha}_i - \alpha_{i0})^4]$, etc.), extending Fernandez-Val and Weidner (2016)'s earlier result. This helps to clarify what the special bias properties of PPML in this context are and are not: it eliminates any direct contribution of each individual fixed effect estimate to the bias, which also eliminates the incidental parameter bias, but it does not eliminate any indirect bias contributions through how the estimation noises of all the various estimates interact with one another.

Readers should also note that the bias considered here is different from the bias analyzed in Weidner and Zylkin (2021). They show that the PPML estimator with three-way fixed effects for panel data gravity models does not suffer from an inconsistency problem, but that confidence intervals are not centered on the true values due to an asymptotic bias problem similar to the one characterized in our equation (3). In this paper, we limit consideration to a cross-sectional gravity model estimated with two-way fixed effects and focus on the finite sample bias we have just characterized rather than on an incidental parameter problem. Put another way, the order of the bias in our setting is small enough that the asymptotic distribution of estimates is not biased and hypothesis should be expected to remain valid asymptotically, unlike in Weidner and Zylkin (2021). Nonetheless, the results in Shepherd (2025) make clear that bias can still be an issue, albeit not necessarily a major one, in two-way fixed effects settings with the PPML estimator, and that it does not only arise in the three-way case.

As in Hahn, Hughes, Kuersteiner, and Newey (2024), there are three common approaches to obtaining estimates corrected for this bias. Two rely on resampling, namely the bootstrap and the jackknife. The other alternative is an analytical correction. In the spirit of Firth (1993), the literature discloses a number of contributions in the third area specifically for Generalized Linear Models, such as Cordeiro and McCullough (1991) and Kosimidis et al. (2020). However, even with an appropriate specification of the bias in hand, implementing this type of correction requires relying on sample estimates of difficult-to-estimate population objects and should not generally be expected to deliver superior performance to less conceptually demanding alternatives (see Hahn, Hughes, Kuersteiner, and Newey, 2024.)

This paper therefore focuses on the bootstrap and jackknife as potential remedies for the bias. Both rely on intuitive arguments. For the bootstrap method, for R bootstrap replications, a bias corrected parameter estimate of a parameter is:

$$(5) \hat{\beta}^{BC} = 2\hat{\beta}^{PPML} - \frac{1}{R} \sum_r \hat{\beta}^r$$

Where $\hat{\beta}^r$ is a coefficient obtained from bootstrap replication r . We use an ordinary re-sampling bootstrap (as opposed to a fractional-weighting scheme), relying on the results of Zylkin (2024). The

standard intuition behind the bootstrap bias correction is that the bootstrap replicates the bias of the estimator because it effectively treats the original sample as the population (Efron 1982).⁴ Zylkin (2024) provides another interpretation in which the random weighting effectively applied by the bootstrap amplifies the noisiness of the score in a way that turns out to double the bias.

For jackknife bias corrections, there are different approaches available. The “leave one out” jackknife works by holding out one observation at a time. Given that the original estimator has a $1/N^2$ bias, and ignoring any higher-order bias terms, each jackknife sample estimate has a bias equal to $N^2/(N^2 - 1)$ times the original bias, allowing the original bias to be estimated. However, when experimenting with the leave-one-out jackknife in our initial simulations, we found that it had a very large standard deviation and overall poor performance.⁵ Thus, we do not consider it further here. Instead, we use a “split sample” jackknife that randomly and repeatedly partitions the sample in half and produces estimates for each half-sample (see Dhaene and Jochmans, 2015; Hahn, Hughes, Kuersteiner, and Newey, 2024.) The bias-corrected estimate is then:

$$(6) \hat{\beta}^{JC} = 2\hat{\beta}^{PPML} - \frac{1}{R} \sum_r \frac{1}{2} (\hat{\beta}^{r,1} + \hat{\beta}^{r,2})$$

Where $\hat{\beta}^{r,1}$ and $\hat{\beta}^{r,2}$ are the half-sample estimates from each random partition r .⁶ The principle behind the split-sample jackknife is that, so long as the bias can be expected to scale with $1/N^2$, each of the half-sample estimates should have twice the bias of the original estimate.⁷ This is because they are based on samples of size $N^2/2$. Again, this ignores any higher-order bias terms of order less than $1/N^2$.

⁴ For some additional intuition, imagine drawing an infinite number of observations from the original sample using bootstrap resampling and call this the “population”. In this population, all the original observations will appear the same amount of times; therefore, the “population estimate” will be the original estimate, and bootstrap samples drawn from this population should approximately replicate the finite sample bias of the original sample estimate, meaning that each bootstrap estimate $\hat{\beta}^r$ in (5) should have twice the bias of the original estimate $\hat{\beta}^{PPML}$. Hahn, Hughes, Kuersteiner, and Newey (2024) prove this intuition formally for nonlinear maximum likelihood estimators (and by extension for pseudo-maximum likelihood estimators as well).

⁵ The problem appears to arise in part because of the large size of the sample. Since N^2 is the sample size, the leave-one-out jackknife is constructed as $N^2\hat{\beta}^{PPML} - (N^2 - 1)\bar{\beta}^J$, where $\bar{\beta}^J$ is the average estimate across all N^2 jackknife samples. N^2 in this case is a large number, so only a small amount of noise is needed in either $\hat{\beta}^{PPML}$ or $\bar{\beta}^J$ to produce large distortions in the implied jackknife correction.

⁶ We use random and repeated partitions following the recommendations of Fernandez-Val and Weidner (2016) and Weidner and Zylkin (2021). The reason for using random partitions is that, in contrast to the panel data applications studied in Dhaene and Jochmans (2015), the data in our case does not have a natural ordering.

⁷ The bias may not be exactly twice that of the original estimate because each of the half-samples will randomly include more trade partners for some countries than others. Because we average over many half-samples, we expect this should only make a minor difference.

3 SIMULATION EVIDENCE

We use simulations to examine three questions. First, we investigate whether PPML's bias in the two-way fixed effects case indeed scales with its asymptotic rate of N^{-2} for reasonably finite sample sizes. Second, we use different data generating processes (DGPs) to highlight the varying performance of PPML under different assumptions for the ratio of the conditional variance to the conditional mean as well as magnitude of the variance itself. Third, we examine and compare the usefulness of bootstrapping and the split-sample jackknife as bias corrections in this context. The purpose of these simulations is to provide stylized evidence. Later in the paper, we also examine these issues in the context of real-world trade data, where we do not make any assumptions about the distributions of observed or unobserved variables.

Building on the approach in Weidner and Zylkin (2021) and our equation (2), we consider the following model:

$$(7) y_{ij} = \lambda_{ij} \omega_{ij} = \exp(\alpha_i + \gamma_j + x_{ij}\beta) \omega_{ij}$$

where $\lambda_{ij} = E(y_{ij}|x_{ij}, \alpha, \gamma)$ is the conditional mean and x_{ij} is a single covariate, generated using $x_{ij} = \alpha_i + \gamma_j + v_{ij}$. We make the following assumptions in all simulations:

- $\alpha_i \sim N(0, 0.0625)$
- $\gamma_j \sim N(0, 0.0625)$
- $v_{ij} \sim N(0, 0.5)$

We also set $\beta = 1$. We then consider four cases for the error (unobserved trade costs) term ω_{ij} , assuming that errors are uncorrelated across country pairs.

1. $\log \omega_{ij}^{(1)} \sim N\left(\log \frac{1}{\sqrt{1+\lambda_{ij}^{-1}}}, \log(1 + \lambda_{ij}^{-1})\right)$
2. $\log \omega_{ij}^{(2)} \sim N\left(\log \frac{1}{\sqrt{2}}, \log 2\right)$
3. $\log \omega_{ij}^{(3)} \sim N\left(\log \frac{1}{\sqrt{1+100\lambda_{ij}^{-1}}}, \log(1 + 100\lambda_{ij}^{-1})\right)$
4. $\log \omega_{ij}^{(4)} \sim N\left(\log \frac{1}{\sqrt{101}}, \log 101\right)$

Errors 1 and 3 above result in outcome variables that have conditional variance proportional to their conditional means, while errors 2 and 4 results in outcome variables that have conditional variance proportional to the square of the conditional mean. Errors 3 and 4 are constructed in the same way as errors 1 and 2 but have larger variance. In all cases, the conditional mean of ω_{ij} is 1. These four examples make it possible to see how the performance of PPML with two-way fixed effects varies depending on both the amplitude of the error variance (as emphasized in Shepherd 2025) as well as how closely it keeps to the assumptions of PPML.

Table 1 presents results for uncorrected PPML estimates as well as the bootstrap bias correction, using 5000 simulated data sets for each of the 4 error variances. Each column shows estimated bias, standard

deviation, and coverage (percentage of estimates for which the 95% confidence interval includes the true parameter value, based on analytical “heteroskedasticity robust” standard errors) for the two estimators, organized by error assumption and sample size in terms of the approaches set out above. The rationale for using analytical standard errors to compute coverage both for the standard PPML estimator and for the bias-corrected estimators is that it emphasizes the effect of any re-centering in the bias-corrected case, without confusing it with the effect of bias correction in the standard errors themselves, as already analyzed by Weidner and Zylkin (2021) and Pfaffermayr (2021).⁸ Because we do not correct the standard errors, we should generally expect to see less than the nominal 95% coverage rate in all cases.

Table 1: Bias, standard deviation, and coverage of standard PPML and bootstrap bias corrected PPML, based on 5000 simulations.

	N	$\omega_{ij}^{(1)}$		$\omega_{ij}^{(2)}$		$\omega_{ij}^{(3)}$		$\omega_{ij}^{(4)}$	
		$\hat{\beta}^{PPML}$	$\hat{\beta}^{BC}$	$\hat{\beta}^{PPML}$	$\hat{\beta}^{BC}$	$\hat{\beta}^{PPML}$	$\hat{\beta}^{BC}$	$\hat{\beta}^{PPML}$	$\hat{\beta}^{BC}$
Bias	50	0.0002	-0.0007	-0.0016	-0.0001	0.0183	-0.0017	-0.0170	-0.0176
	100	0.0002	0.0000	-0.0005	0.0003	0.0078	-0.0007	-0.0093	-0.0058
	150	0.0000	-0.0001	-0.0006	0.0000	0.0020	-0.0030	-0.0098	-0.0060
SD	50	0.0237	0.0239	0.0499	0.0546	0.1736	0.2012	0.2360	0.2854
	100	0.0115	0.0115	0.0263	0.0280	0.0990	0.1116	0.1510	0.1803
	150	0.0075	0.0075	0.0181	0.0190	0.0663	0.0723	0.1139	0.1344
Cov	50	0.9320	0.9326	0.8812	0.8540	0.8678	0.8254	0.7880	0.7224
	100	0.9434	0.9414	0.9126	0.8974	0.9054	0.8812	0.8268	0.7782
	150	0.9452	0.9466	0.9252	0.9142	0.9230	0.9078	0.8418	0.8218

Note: Superscript PPML refers to uncorrected PPML estimates. Superscript BC refers to bias corrected bootstrap PPML estimates. Column groups identify different error assumptions, based on the notation in the main text. N is sample size. Rows list bias, standard deviation, and coverage of each estimator for different combinations of sample size and error assumptions. The true parameter is 1. The standard deviation is the standard deviation across simulations.

Looking at the estimates for standard PPML, there is a clear bias in the estimates, but it is generally declining in sample size, as per the analysis above. The bias is more pronounced for errors 2 and 4, where the Poisson assumption of conditional variance equal to the conditional mean is violated. It is also more persistent in smaller samples, particularly in the high variance cases 3 and 4. Moreover, the results for cases 3 and 4 show that the bias is generally larger in cases where the variance of the error term is larger, which is consistent with the main arguments of Shepherd (2025). As we have shown, the bias depends on the variances and covariances of the various parameter estimates. When the data

⁸ Though it is not our focus here, another advantage of the bootstrap in this context is that it can also be used to obtain more accurate standard errors, as shown in Pfaffermayr (2021) and Zylkin (2024). Thus, in practice, researchers who implement the bootstrap bias correction will likely also choose to use bootstrap standard errors. As we show in the appendix, using bootstrap standard errors improves coverage significantly across all cases in our simulations but does not otherwise change the basic takeaways regarding coverage.

is noisier, the scores for all of these estimates will be noisier, inflating their variances and covariances—and thus, the bias—irrespective of the sample size.

Turning to how the bias varies with N , the expected asymptotic rate of convergence of N^{-2} is most clearly seen for the 3rd error variance assumption. In this case, the bias is relatively large to start with and falls by about the expected rate of 9 times when N increases by 3 times from 50 to 150. However, for other cases, especially cases 2 and 4 that deviate from the constant mean-variance ratio assumption, though the bias falls substantial with N , it is not as clear that it does so at a rate of N^{-2} . For cases 2 and 4, this likely indicates that the higher-order bias terms (those with an order less than N^{-2}) can in principle continue to make non-negligible contributions to the bias even in moderately large samples. In further simulations provided in the Appendix, we find evidence supporting that the sample size needed for the bias to reach its asymptotic convergence rate in noisier data sets can be substantial.⁹

The remaining columns show bias corrected estimates based on bootstrapping (250 replications). While the results generally suggest that the bootstrap reduces the degree of bias, in line with expectations, the improvement is not necessarily drastic. Bootstrapping is generally more effective in larger samples and when the variance is smaller, but it does not show clear bias reduction when the bias itself becomes small. Nonetheless, these results provide comfort for the idea both that the analysis of the bias problem presented above is correct, i.e. that it is a small sample property of nonlinear maximum likelihood estimators, and that the bootstrap has some degree of effectiveness in attenuating that bias.

The lower panels of the table show the relevant standard deviations of estimates and coverage ratios. Performance in terms of coverage is somewhat disappointing for the bootstrap, though it depends on the error variance. Coverage improves with larger samples, and it is closer to the correct level for errors 1 and 3, where the Poisson assumption on the relationship between conditional mean and conditional variance holds. Coverage is substantially different from the ideal level for errors 2 and 4, especially in small samples. Bootstrapping does not substantially improve coverage in those cases, and indeed sometimes it worsens it. This occurs because, as can be seen in the middle 3 rows of the table, the bias-corrected estimates always have larger standard deviations than the uncorrected ones in spite of their reduced bias. Interestingly, though the bootstrap did not always seem to have a clear positive impact on the small biases found for error 1, it does seem to improve coverage even when its average reported bias is larger than PPML. This suggests overall favorable performance for error variance 1.

Table 2 shows comparable results for the split-sample jackknife, again based on 250 replications. The results are qualitatively similar to those of the bootstrap, usually about the same though slightly worse in some cases. Like the bootstrap, the effects of the jackknife vary according to the sample size, the conditional variance/mean assumption, and the amplitude of the error variance. Coverage is again a concern, particularly for errors 2 and 4. Overall, the split-sample jackknife, like the bootstrap, displays

⁹ In the simulations in the Appendix, we examine how the bias behaves in each data generating process as N continues to get larger. These simulations confirm that bias eventually converges at a rate of N^{-2} across all 4 data generating processes, though we do not begin to observe this pattern until $N \geq 300$ in the cases where the conditional variance is not proportional to the conditional mean. Of course, by this point, the bias is always quite small.

an ability to correct PPML's bias, but it does not substantially improve coverage in the best of cases and can even worsen it in others.

Table 2: Bias, standard deviation, and coverage of standard PPML and bias corrected split-sample jackknife PPML, based on 5000 simulations.

	N	$\omega_{ij}^{(1)}$		$\omega_{ij}^{(2)}$		$\omega_{ij}^{(3)}$		$\omega_{ij}^{(4)}$	
		$\hat{\beta}^{PPML}$	$\hat{\beta}^{JC}$	$\hat{\beta}^{PPML}$	$\hat{\beta}^{JC}$	$\hat{\beta}^{PPML}$	$\hat{\beta}^{JC}$	$\hat{\beta}^{PPML}$	$\hat{\beta}^{JC}$
Bias	50	0.0002	-0.0010	-0.0016	-0.0003	0.0183	-0.0116	-0.0170	-0.0213
	100	0.0002	-0.0001	-0.0005	0.0001	0.0078	-0.0050	-0.0093	-0.0077
	150	0.0000	-0.0001	-0.0006	0.0000	0.0020	-0.0030	-0.0098	-0.0060
SD	50	0.0237	0.0239	0.0499	0.0558	0.1736	0.2109	0.2360	0.3034
	100	0.0115	0.0115	0.0263	0.0283	0.0990	0.1152	0.1510	0.1901
	150	0.0075	0.0075	0.0181	0.0191	0.0663	0.0735	0.1139	0.1406
Cov	50	0.9320	0.9338	0.8812	0.847	0.8678	0.8112	0.7880	0.6970
	100	0.9434	0.9432	0.9126	0.8924	0.9054	0.8792	0.8268	0.7594
	150	0.9452	0.9460	0.9252	0.9128	0.9230	0.9058	0.8418	0.7850

Note: Superscript PPML refers to uncorrected PPML estimates. Superscript JC refers to bias corrected jackknife PPML estimates. Column groups identify different error assumptions, based on the notation in the main text. N is sample size. Rows list bias, standard deviation, and coverage for different combinations of sample size and error assumptions. The true parameter is 1. The standard deviation is the standard deviation across simulations.

Taking Tables 1 and 2 together, there are two relevant conclusions for applied work using gravity models estimated by PPML with two-way fixed effects. The first is that PPMLs bias is persistent even with numerically large samples in some cases, depending on what assumption is made for the relationship between the conditional mean and the conditional variance as well as the size of the variance itself. The second is that while the bootstrap and the jackknife generally improve parameter estimates somewhat, they firstly do not always eliminate the bias, and, secondly, they do not significantly improve coverage and can even worsen it. So, the implications for inference in applied contexts are not straightforward. If the emphasis in a particular application is on unbiased parameter estimates, then the bootstrap and jackknife could potentially be valuable additions to the toolkit, with the bootstrap slightly preferred because of its modestly superior performance in the simulations. However, if the emphasis is on inference, these methods may prove inadequate to the task, at least not without making additional adjustments to the standard errors to reflect the additional variation introduced by the bias corrections.

Of course, a limitation of the analysis in this section is that it is based on conceptual simulations. For the simulated data, PPML's bias in terms of parameter estimates is small, so the issue of correcting it is not major. However, that finding does not rule out the fact that in empirical contexts, the bias could be more substantial due to how it depends on both the heteroskedasticity and the amplitude of the error variance. The next section therefore applies our bias correction methods to real world trade data to gauge what types of results they produce.

4 EMPIRICAL APPLICATION

The simulation evidence in the previous section helps both to highlight the surprising bias of PPML estimates in the two-way fixed effects case and to make the case that the bootstrap and split-sample jackknife are simple ways of attenuating it. While their performance varies, and improvements are focused on point estimates rather than inferences, the evidence suggests that they may have a role in confirming the robustness of gravity model estimates.

In the simulations, both bias correction approaches make a modest difference to point estimates. But how does this approach play out in an empirical setting using real-world trade data? This information is important for applied researchers assessing whether or not they should systematically apply corrections to estimates from gravity models with two-way fixed effects. Whereas the previous section focused on bias, i.e. deviations of estimated parameters from known values, this section focuses on the difference between bias corrected and standard PPML estimates. If the difference between standard PPML and the bias corrected results is relatively large in terms of point estimates, it is an indication that bias is important in a particular empirical context.

To perform this analysis, we use the dataset from Weidner and Zylkin (2021). The data in that paper comes from the CEPII BACI dataset and considers international trade only (not intra-national trade). Data on geographical distance are from the CEPII Distance Dataset, and a dummy for Free Trade Agreement membership was constructed by Weidner and Zylkin (2021) based on the NSF-Kellogg dataset, supplemented with information from the WTO. We consider 28 ISIC rev. 3 industries as well as aggregate trade. The number of countries is 167, which, without intra-national trade, gives us a sample size of 27,722. For full details of dataset construction, see Weidner and Zylkin (2021).

Concretely, we estimate the following model separately for each sector by PPML, taking a single year from the source data (2015, the latest year in the data set):

$$(9) y_{ij} = \exp(\alpha_i + \gamma_j + \text{Indist}_{ij}\beta_1 + \text{FTA}_{ij}\beta_2) \omega_{ij}$$

One of the ways this approach goes beyond the simulations discussed above because it considers multiple sectors, each of which has different, generally unknown heteroskedasticity properties. Thus, we can easily gain a sense of what bias-corrected estimates are likely to look like for a wide variety of heteroskedasticity patterns found in real trade data.

[Tables 3 and 4 roughly here]

Table 3 shows results using the bootstrap bias correction. The first and fifth columns have raw PPML estimates, with their standard errors in columns two and six. The third and seventh columns have bias corrected estimates obtained by bootstrapping (250 replications). The remaining columns report percentage differences between the uncorrected and corrected estimates.

Unsurprisingly, most of the bias corrected results are not too far from the raw estimates. However, the differences are not necessarily small either. Two sectors—other mining and quarrying (no. 14) and manufactured wood products (no. 20)—have differences for their distance coefficient estimates that are greater than 10% of the original estimate. In total, six sectors have distance coefficients that differ than more than five percent relative to the bias-corrected version. Turning to the FTA coefficient, the changes are generally larger. There are nine sectors where the difference is greater than 10%, and a total of 14 where the difference is greater than five percent. There are six sectors where the difference is greater than 20%. The difference in the extent of bias for the distance and FTA coefficients may be linked to differing amounts of noise in their estimates, as captured by their PPML standard errors. This association reflects our finding that the bias depends on the variance of the coefficient estimates and on their covariance with those of the fixed effects, as described in Section 2 and discussed further in the Appendix. Moreover, for real data settings, the bias could also depend in complicated ways on the empirical distributions of the regressors, which need not be normally distributed as we assumed in our earlier simulations.

Our takeaways from Table 3 are not simply rooted in the finding that some sectoral estimates appear to have larger amounts of bias than others. Taking a step back, it is notable that, with a sample of more than 27,000 observations, these finite-sample corrections make any non-negligible difference at all. For aggregate trade, for example, a difference in the FTA coefficient estimate of 0.4723 versus 0.4489 is not one that we would expect an applied researcher to necessarily overlook. In fact, this is larger than the absolute difference in the bias-corrected FTA estimate that Weidner and Zylkin (2021) report for the correction of the three-way fixed effects PPML estimator using a panel data version of this same data set.

Table 4 shows comparable results for the split-sample jackknife. Encouragingly, these results correspond closely to those reported for the bootstrap in the previous table. Indeed, for aggregate trade, the jackknife corrections are very similar to the bootstrapped ones in both their direction and magnitude. Moreover, the sectoral results are qualitatively similar as well. When the bootstrap corrections are large, the jackknife corrections tend to be large as well (see wood, metal ores, or coke and refined petroleum, for example), and, likewise, when one method suggests a small bias, the other does also (see agriculture, manufactured food products, textiles, or plastics). The signs of the implied biases are almost always the same as well, except when the magnitudes of the corrections are close to zero. In general, the close correspondence between the corrections produced by each method gives us confidence they are each picking up similar features of the data.

Finally, as we have noted, though the corrected estimates reported in Tables 3 and 4 are likely to be less biased estimates of the two trade cost coefficients, this does not necessarily mean they should be preferred to the uncorrected estimates (or, to give a much better recommendation, to the uncorrected estimates paired with appropriately corrected standard errors). In our simulations, we found that using bias corrections can increase the standard deviation of the estimates to the point where inference can worsen in spite of the reduction in bias. How large an issue is this when using real data?

To investigate this question further, we performed one last set of simulations using bootstrap replications of the aggregate trade data as the simulated data. By construction, these simulated data sets should have close resemblance to the actual data while allowing us to obtain bootstrapped distributions for each estimate. Furthermore, using the properties of the bootstrap we have noted, we know the “true parameters” for these simulations are just the original sample estimates (-0.8042 for the distance coefficient and 0.4489 for the FTA coefficient). Because these simulations are produced using the same procedure as the bootstrap bias correction, the simulated PPML estimates will have an average bias roughly equal to the implied bias already reported in the last row of Table 3. However, that is not the point here. Instead, we want to focus on three things: (i) whether the simulated *bias-corrected* estimates are close to the original uncorrected ones; (ii) how the standard deviations of the bias-corrected estimates compare with those of the uncorrected ones; and (iii) whether the bias corrections improve coverage or worsen it. Importantly, the design of these experiments will allow us to account for how both the heteroskedasticity and the empirical distributions of the regressors affect each of these outcomes.

In Table 5, we report preliminary results for this last experiment, using 10,000 bootstrap simulations. These results are encouraging. In contrast to what one might worry about based on the earlier simulations, both the bootstrap bias correction and the split-sample jackknife only introduce a mild increase in the standard deviations of each coefficient estimate compared with using uncorrected PPML. In addition, both bias corrections work well to eliminate a majority of the uncorrected estimator’s bias. Consequently, coverage either improves modestly or decreases only slightly in all cases, with the one case where it decreases being the effect of the split-sample jackknife on the distance coefficient.¹⁰

[Table 5 roughly here]

Obviously, these last results should not be taken to indicate that these corrections can generally be expected to improve inference in all real data settings. Nonetheless, for our own application, they serve to further support the credibility of the results reported for aggregate trade in the previous two

¹⁰ As a by-product, we can also use these results to calculate the bootstrap bias-corrections we would obtain for the last row of Table 3 if we used 10,000 bootstrap replications instead of 250. Using $2 \times -0.8042 - -0.8154$ and $2 \times 0.4489 - 0.4260$, we obtain -0.7930 and 0.4718, which are close to the values reported in Table 3 (-0.7924 and 0.4723). This confirms that using 250 replications reasonably approximates what we would get with a much larger number. However, in general, the choice of the number of replications is a consequential one that should be tailored to the empirical setting being studied. See Andrews and Buchinsky (2000) for methods for choosing the number of replications. Also see Poi (2004) for Stata code implementing their methods.

tables. At the very least, this type of bootstrap simulation procedure could serve as an additional check for researchers interested in the issue, albeit a potentially expensive one. As an interesting by-product, it also produces a bootstrapped estimate of the standard error of the bias-corrected estimates, which could then could potentially be used to further improve inferences (as opposed to using the standard error of the uncorrected estimates). We leave the latter issue for further research.

Table 3: Estimation results using observed trade data (using bootstrap bias corrections).

ISIC	Description	Distance				FTA			
		PPML	PPML SE	BC	(PPML- BC)/BC	PPML	PPML SE	BC	(PPML- BC)/BC
1	Agriculture, Hunting And Related Service Activities	-1.0323	0.0517	-1.0254	0.0067	0.5447	0.0984	0.5362	0.0159
2	Forestry, Logging And Related Service Activities	-1.1535	0.1261	-1.0844	0.0637	0.2568	0.2006	0.2905	-0.1160
5	Fishing, Operation Of Fish Hatcheries And Fish Farms; Service Activities Incidental To Fishing	-1.7138	0.0706	-1.7061	0.0045	0.4041	0.1536	0.3943	0.0249
10	Mining Of Coal And Lignite; Extraction Of Peat	-1.9625	0.1724	-1.8628	0.0535	0.3397	0.1697	0.3187	0.0659
13	Mining Of Metal Ores	-1.5563	0.0994	-1.6213	-0.0401	-0.0102	0.1755	-0.221	-0.9538
14	Other Mining And Quarrying	-0.6982	0.1175	-0.5835	0.1966	0.1206	0.2068	0.1209	-0.0025
15	Manufacture Of Food Products And Beverages	-1.0488	0.0352	-1.0506	-0.0017	0.5658	0.0797	0.5742	-0.0146
16	Manufacture Of Tobacco Products	-1.2214	0.0824	-1.1567	0.0559	1.5365	0.1853	1.6309	-0.0579
17	Manufacture Of Textiles	-1.0154	0.0400	-1.0015	0.0139	0.6238	0.0851	0.6349	-0.0175
18	Manufacture Of Wearing Apparel; Dressing And Dyeing Of Fur	-0.903	0.0701	-0.8725	0.0350	0.481	0.1197	0.5356	-0.1019
19	Tanning And Dressing Of Leather; Manufacture Of Luggage, Handbags, Saddlery, Harness And Footwear	-0.6691	0.0754	-0.6317	0.0592	0.2059	0.1204	0.2896	-0.2890
20	Manufacture Of Wood And Of Products Of Wood And Cork, Except Furniture	-1.1625	0.0933	-1.0524	0.1046	0.5742	0.1519	0.7056	-0.1862
21	Manufacture Of Paper And Paper Products	-1.1649	0.0460	-1.1371	0.0244	0.7473	0.0922	0.7577	-0.0137
22	Publishing, Printing And Reproduction Of Recorded Media	-1.0325	0.0609	-0.9935	0.0393	0.4249	0.1215	0.4667	-0.0896
23	Manufacture Of Coke, Refined Petroleum Products And Nuclear Fuel	-1.3237	0.0543	-1.2651	0.0463	0.2219	0.1173	0.2792	-0.2052
24	Manufacture Of Chemicals And Chemical Products	-0.8511	0.0351	-0.8392	0.0142	0.2748	0.0782	0.2744	0.0015
25	Manufacture Of Rubber And Plastics Products	-1.0891	0.0375	-1.0814	0.0071	0.6346	0.0693	0.6412	-0.0103
26	Manufacture Of Other Non-Metallic Mineral Products	-1.1832	0.0436	-1.142	0.0361	0.6105	0.1129	0.6507	-0.0618
27	Manufacture Of Basic Metals	-0.8616	0.0592	-0.838	0.0282	0.6829	0.1294	0.7161	-0.0464

28	Manufacture Of Fabricated Metal Products, Except Machinery And Equipment	-1.0126	0.0444	-1.0018	0.0108	0.5442	0.0769	0.5427	0.0028
29	Manufacture Of Machinery And Equipment N.E.C.	-0.7382	0.0364	-0.7295	0.0119	0.4538	0.0623	0.4534	0.0009
30	Manufacture Of Office, Accounting And Computing Machinery	-0.8757	0.0545	-0.8794	-0.0042	0.2014	0.1025	0.153	0.3163
31	Manufacture Of Electrical Machinery And Apparatus N.E.C.	-0.8631	0.0430	-0.8558	0.0085	0.5165	0.0710	0.5137	0.0055
32	Manufacture Of Radio, Television And Communication Equipment And Apparatus	-0.749	0.0468	-0.7307	0.0250	0.0435	0.1038	0.0332	0.3102
33	Manufacture Of Medical, Precision And Optical Instruments, Watches And Clocks	-0.6314	0.0568	-0.6423	-0.0170	0.0703	0.1174	0.0393	0.7888
34	Manufacture Of Motor Vehicles, Trailers And Semi-Trailers	-0.7923	0.0698	-0.7616	0.0403	1.1551	0.1181	1.2071	-0.0431
35	Manufacture Of Other Transport Equipment	-0.2851	0.0728	-0.2719	0.0485	0.641	0.1262	0.6688	-0.0416
36	Manufacture Of Furniture; Manufacturing N.E.C.	-0.8198	0.0700	-0.8008	0.0237	0.4532	0.1564	0.4819	-0.0596
100	Total	-0.8042	0.0285	-0.7924	0.0149	0.4489	0.0558	0.4723	-0.0495

Note: Each row identifies an ISIC sector in observed trade data. Columns group estimation results for the log of distance and a dummy for FTA membership. PPML is uncorrected PPML estimates, with their standard errors listed in the next column. BC gives bootstrap bias corrected PPML estimates. The third column in each group gives the percentage correction in the estimate provided by bootstrapping.

Table 4: Estimation results using observed trade data (using split-sample jackknife).

ISIC	Description	PPML	Distance			FTA			
			PPML SE	JC	(PPML- JC)/JC	PPML	PPML SE	JC	(PPML- JC)/JC
1	Agriculture, Hunting And Related Service Activities	-1.0323	0.0517	-1.0260	0.0062	0.5447	0.0984	0.5249	0.0377
2	Forestry, Logging And Related Service Activities	-1.1535	0.1261	-1.0430	0.1060	0.2568	0.2006	0.3366	-0.2369
5	Fishing, Operation Of Fish Hatcheries And Fish Farms; Service Activities Incidental To Fishing	-1.7138	0.0706	-1.6991	0.0087	0.4041	0.1536	0.3855	0.0481
10	Mining Of Coal And Lignite; Extraction Of Peat	-1.9625	0.1724	-1.7991	0.0908	0.3397	0.1697	0.3670	-0.0743
13	Mining Of Metal Ores	-1.5563	0.0994	-1.6457	-0.0543	-0.0102	0.1755	-0.2663	-0.9617
14	Other Mining And Quarrying	-0.6982	0.1175	-0.5312	0.3146	0.1206	0.2068	0.1242	-0.0291
15	Manufacture Of Food Products And Beverages	-1.0488	0.0352	-1.0464	0.0023	0.5658	0.0797	0.5860	-0.0345
16	Manufacture Of Tobacco Products	-1.2214	0.0824	-1.1313	0.0797	1.5365	0.1853	1.6441	-0.0654
17	Manufacture Of Textiles	-1.0154	0.0400	-0.9936	0.0220	0.6238	0.0851	0.6430	-0.0298
18	Manufacture Of Wearing Apparel; Dressing And Dyeing Of Fur	-0.9030	0.0701	-0.8568	0.0539	0.4810	0.1197	0.5518	-0.1282
19	Tanning And Dressing Of Leather; Manufacture Of Luggage, Handbags, Saddlery, Harness And Footwear	-0.6691	0.0754	-0.5964	0.1218	0.2059	0.1204	0.3544	-0.4190
20	Manufacture Of Wood And Of Products Of Wood And Cork, Except Furniture	-1.1625	0.0933	-0.9960	0.1672	0.5742	0.1519	0.7645	-0.2489
21	Manufacture Of Paper And Paper Products	-1.1649	0.0460	-1.1247	0.0357	0.7473	0.0922	0.7748	-0.0355
22	Publishing, Printing And Reproduction Of Recorded Media	-1.0325	0.0609	-0.9777	0.0561	0.4249	0.1215	0.4669	-0.0901
23	Manufacture Of Coke, Refined Petroleum Products And Nuclear Fuel	-1.3237	0.0543	-1.2400	0.0675	0.2219	0.1173	0.3220	-0.3110
24	Manufacture Of Chemicals And Chemical Products	-0.8511	0.0351	-0.8334	0.0212	0.2748	0.0782	0.2804	-0.0198
25	Manufacture Of Rubber And Plastics Products	-1.0891	0.0375	-1.0770	0.0113	0.6346	0.0693	0.6485	-0.0215
26	Manufacture Of Other Non-Metallic Mineral Products	-1.1832	0.0436	-1.1236	0.0530	0.6105	0.1129	0.6653	-0.0825
27	Manufacture Of Basic Metals	-0.8616	0.0592	-0.8357	0.0310	0.6829	0.1294	0.7160	-0.0463
28	Manufacture Of Fabricated Metal Products, Except Machinery And Equipment	-1.0126	0.0444	-0.9931	0.0196	0.5442	0.0769	0.5553	-0.0200
29	Manufacture Of Machinery And Equipment N.E.C.	-0.7382	0.0364	-0.7256	0.0174	0.4538	0.0623	0.4750	-0.0446

30	Manufacture Of Office, Accounting And Computing Machinery	-0.8757	0.0545	-0.8899	-0.0159	0.2014	0.1025	0.1442	0.3964
31	Manufacture Of Electrical Machinery And Apparatus N.E.C.	-0.8631	0.0430	-0.8617	0.0017	0.5165	0.0710	0.5184	-0.0038
32	Manufacture Of Radio, Television And Communication Equipment And Apparatus	-0.7490	0.0468	-0.7194	0.0411	0.0435	0.1038	0.0376	0.1578
33	Manufacture Of Medical, Precision And Optical Instruments, Watches And Clocks	-0.6314	0.0568	-0.6538	-0.0342	0.0703	0.1174	0.0222	2.1674
34	Manufacture Of Motor Vehicles, Trailers And Semi-Trailers	-0.7923	0.0698	-0.7687	0.0307	1.1551	0.1181	1.2125	-0.0474
35	Manufacture Of Other Transport Equipment	-0.2851	0.0728	-0.2736	0.0418	0.6410	0.1262	0.6558	-0.0227
36	Manufacture Of Furniture; Manufacturing N.E.C.	-0.8198	0.0700	-0.7941	0.0323	0.4532	0.1564	0.4784	-0.0526
100	Total	-0.8042	0.0285	-0.7882	0.0203	0.4489	0.0558	0.4811	-0.0671

Note: Each row identifies an ISIC sector in observed trade data. Columns group estimation results for the log of distance and a dummy for FTA membership. PPML is uncorrected PPML estimates, with their standard errors listed in the next column. JC gives split sample jackknife PPML estimates. The third column in each group gives the percentage correction in the estimate provided by bootstrapping.

Table 5: Bias, standard deviation, and coverage of standard PPML and bias corrected PPML estimates, using 10000 bootstrap simulations drawn from aggregate trade data

	<i>Indist_{ij}</i>			<i>FTA_{ij}</i>		
	PPML	Boot. BC	SSJ	PPML	Boot. BC	SSJ
Average estimate	-0.8154	-0.8029	-0.7982	0.4260	0.4444	0.4524
Bias	-0.0112	0.0013	0.0060	-0.0229	-0.0048	0.0035
SD	0.0388	0.0393	0.0397	0.0706	0.0718	0.0726
Coverage	0.8430	0.8494	0.8408	0.8654	0.8744	0.8658

Note: Rows provide statistics on each estimation approach. Columns are grouped into estimates of the coefficient on the log of distance, and a dummy for FTA membership. The first column in each group reports uncorrected PPML estimates. The second reports bias corrected bootstrap estimates. The third reports split sample jackknife estimates. The standard deviation is the standard deviation across simulations. Coverage is the share of simulations in which the original estimates are contained in the 95% confidence interval.

5 CONCLUSION

Even though the PPML estimator with two-way fixed effects does not suffer from the incidental parameters problem, it is nonetheless biased in small samples. Surprisingly, the types of samples used in structural gravity modeling appear in some cases to be “small”, in the sense that the bias is persistent even in cases with tens of thousands of observations. A theoretical examination demonstrates that the bias is generally declining with the sample size, suggesting a negligible amount of bias, but simulations show that its extent and persistence depend on the assumed form of the error term. Moreover, experiments with real-world sectoral-level and aggregate-level real trade sets indicate we cannot rule out the possibility of surprisingly large biases even with data sets with tens of thousands of observations.

We show that the bootstrap and split-sample jackknife can be used to attenuate, but not always eliminate, the bias. They are straightforward to implement and not overly burdensome from a computational standpoint. Our simulations suggest that they perform better when the number of countries become moderately large but cannot necessarily be relied in to improve inference because of how they increase the variance of the estimates. Encouragingly, when we used a further bootstrap procedure to simulate data that closely resembles actual data, we found only mild increases in variance.

In applied work, there is a cost-benefit analysis in terms of whether or not to implement a bias correction such as the bootstrap or split-sample jackknife in this setting. In terms of cost, there is computational burden and the fact that coverage does not always substantially improve, although our realistic simulation suggests that it can improve somewhat. In terms of benefit, there is the ability to obtain less-biased parameter estimates and, in the best of cases, a modest improvement in inference. The balance between these factors likely depends on the application. If inference is the primary objective, then considerations about coverage would dominate, and the researcher’s effort is likely to be better spent obtaining appropriately corrected standard error estimates (see Pfaffermayr 2021 or Weidner and Zylkin 2021, Section 3.3.) If, on the other hand, the primary objective is to obtain unbiased parameter estimates, then the case for bias correction is stronger. Given that the computational burden is not unreasonable, many researchers may choose to bootstrap PPML gravity models with two-way fixed effects as a way of diagnosing bias, as we did in the empirical exercise here: there is evidence that at the sectoral level, bias can sometimes be substantial.

One application of our work is in settings where gravity model estimates are used as the basis for a counterfactual simulation in general equilibrium. In that case, the point estimates of relevant coefficients are of particular interest, as they determine (jointly with the trade elasticity) the response of trade flows to changes in trade costs. This area is one in which the accuracy of point estimates is important for the interpretation of counterfactual results, so the cost-benefit balance for undertaking bias correction via the bootstrap or split-sample jackknife is more positive than in settings where the primary concern is establishing a confidence interval for a parameter of interest to enable hypothesis testing. While the balance between attenuating bias and obtaining correct coverage will differ in particular applications, there will likely be some cases in which researchers find it beneficial to implement bootstrap or jackknife corrections as a precautionary measure, to identify cases in which they differ substantially from standard PPML estimates—which could plausibly be linked to larger than usual bias.

REFERENCES

Anderson, J. E., & Van Wincoop, E. 2003. “Gravity with gravitas: A solution to the border puzzle.” *American Economic Review*, 93(1): 170-192.

- Andrews, D.W. and Buchinsky, M., 2000. "A three-step method for choosing the number of bootstrap repetitions." *Econometrica*, 68(1): 23-51.
- Cordeiro, G., and P. McCullagh. 1991. "Bias Correction in Generalized Linear Models." *Journal of the Royal Statistical Society. Series B (Methodological)*, 53(3): 629-643.
- Dhaene, G., and K. Jochmans. 2015. "Split-panel jackknife estimation of fixed-effect models." *The Review of Economic Studies* 82(3): 991-1030.
- Eaton, J., & Kortum, S. 2002. "Technology, geography, and trade." *Econometrica*, 70(5): 1741-1779.
- Efron, B. 1982. "The jackknife, the bootstrap and other resampling plans." *Society for Industrial and Applied Mathematics*.
- Fernandez-Val, I., and M. Weidner. 2016. "Individual and Time Effects in Nonlinear Panel Models with Large N, T." *Journal of Econometrics*, 192(1): 291-312.
- Fernández-Val, I., and M. Weidner. 2018. "Fixed effects estimation of large-T panel data models." *Annual Review of Economics* 10(1): 109-138.
- Firth, D. 1993. "Bias Reduction of Maximum Likelihood Estimates." *Biometrika*, 80(1): 27-38.
- Hahn, J., Hughes, D. W., Kuersteiner, G., & Newey, W. K. 2024. "Efficient Bias Correction for Cross-section and Panel Data." arXiv preprint arXiv:2207.09943.
- Head, K., and T. Mayer. 2014. "Gravity Equations: Workhorse, Toolkit, and Cookbook." In G. Gopinath, E. Helpman, and K. Rogoff (eds.) *Handbook of International Economics*, 4: 131-195.
- Kosimidis, I., E. Pagui, and N. Sartori. 2020. "Mean and Median Bias Reduction in Generalized Linear Models." *Statistics and Computing*, 30: 43-59.
- Pfaffermayr, M., 2019. "Gravity models, PPML estimation and the bias of the robust standard errors." *Applied Economics Letters*, 26(18): 1467-1471.
- Pfaffermayr, M. 2021. "Confidence intervals for the trade cost parameters of cross-section gravity models." *Economics Letters*, 201: 109787.
- Poi, B.P., 2004. "From the help desk: Some bootstrapping techniques." *Stata Journal*, 4(3): 312-328.
- Santos Silva, J., and S. Tenreyro. 2006. "The Log of Gravity." *Review of Economics and Statistics*, 88(4): 641-658.
- Santos Silva, J., and S. Tenreyro. 2011. "Further Simulation Evidence on the Performance of the Poisson Pseudo-Maximum Likelihood Estimator." *Economics Letters*, 112(2): 220-222.
- Santos Silva, J., and S. Tenreyro. 2021. "The Log of Gravity at 15." *Portuguese Economic Journal*, 21: 423-437.
- Shepherd, B., 2025. "Bias in Poisson Pseudo-Maximum Likelihood Estimation of Structural Gravity Models: How Much of a Problem for Applied Research?" *Applied Economics Letters*, 32(2): 222-227.
- Weidner, M., and T. Zylkin. 2021. "Bias and Consistency in Three-Way Gravity Models." *Journal of International Economics*, 132: 103513.
- Wooldridge, J. M. (1999). Distribution-free estimation of some nonlinear panel data models. *Journal of Econometrics*, 90(1), 77-97.

Yotov, Y. V., Piermartini, R., & Larch, M. 2016. “An advanced guide to trade policy analysis: The structural gravity model.” WTO iLibrary.

T. Zylkin. 2024. “Bootstrap for Gravity Models.” Mimeo.

APPENDIX

This first part of this Appendix contains additional technical details on the bias results stated in Section 2 of the main text. These details include a more formal statement of the bias of two-way fixed effects PPML for cross-sectional gravity models (phrased as a theorem), some remarks to guide understanding, and a proof. Because these details require a more in-depth treatment of the PPML estimator, we first provide some additional notation to be used later in the theorem and proof. Some auxiliary results needed for the proof are provided in a [Supplementary Appendix](#). The second part of the Appendix adds some further simulation results supplementing those from the main text.

ADDITIONAL NOTATION

The fixed effects PPML estimator maximizes

$$\sum_{i,j} \ell_{ij} = \sum_{i,j} y_{ij} \log \lambda_{ij} - \lambda_{ij}, \quad \lambda_{ij} = \exp(\alpha_i + \gamma_j + x_{ij}\beta),$$

Where $\ell_{ij} = y_{ij} \log \lambda_{ij} - \lambda_{ij}$ is the PPML pseudolikelihood. To save space where needed, we will use superscripts to denote partial derivatives and higher-order partial derivatives where needed. For example, $\ell_{ij}^\beta = \frac{\partial \ell_{ij}}{\partial \beta}$, $\ell_{ij}^{\beta\beta} = \frac{\partial^2 \ell_{ij}}{\partial \beta \partial \beta}$, $\ell_{ij}^{\beta\alpha_i} = \frac{\partial^2 \ell_{ij}}{\partial \beta \partial \alpha_i}$, etc. An “E” preceding one of these derivatives denotes a conditional expectation, e.g., $E\ell_{ij}^{\beta\beta} = E(\ell_{ij}^{\beta\beta} | x_{ij}, \alpha_i, \gamma_j)$. Also, let $\mathbf{b} = (\beta, \alpha, \gamma)$ refer to the full vector of parameters to be estimated, including all of the fixed effects.

Next, let ℓ_{ij}^* be the “orthogonalized” pseudolikelihood that re-parameterizes the model so that $\sum_j E\ell_{ij}^{*\beta\alpha_i} = \sum_i E\ell_{ij}^{*\beta\gamma_j} = 0$. In practice, this means re-parameterizing the model by replacing x_{ij} with $\tilde{x}_{ij}$, where $\tilde{x}_{ij}$ is a residualized version of x_{ij} obtained after projecting it on the fixed effects and weighting by λ_{ij} . That is, for each element of $\tilde{x}_{ij}$, we use the residuals from the following minimization:

$$\min_{\alpha_{i,k}^x, \gamma_{j,k}^x} \frac{1}{2} \sum_{i,j} \lambda_{ij} (x_{ij,k} - \alpha_{i,k}^x - \gamma_{j,k}^x)^2.$$

After replacing x_{ij} with $\tilde{x}_{ij}$ in the re-parameterized pseudolikelihood ℓ_{ij}^* , the first-order conditions from this minimization problem give us $\sum_j \lambda_{ij} \tilde{x}_{ij,k} = 0 \Rightarrow \sum_j E\ell_{ij}^{*\beta_k\alpha_i} = 0$, $\sum_i \lambda_{ij} \tilde{x}_{ij,k} = 0 \Rightarrow \sum_i E\ell_{ij}^{*\beta_k\gamma_j} = 0$ for each k .

After implementing this orthogonalization, what we will refer to as the expected Hessian matrix is given by

$$H := \sum_{i,j} -E\ell_{ij}^{*bb} = \begin{pmatrix} -\sum_{i,j} E\ell_{ij}^{*\beta\beta} & 0 & 0 \\ 0 & H_{(\alpha\alpha)} & H_{(\alpha\gamma)} \\ 0 & [H_{(\alpha\gamma)}]' & H_{(\gamma\gamma)} \end{pmatrix},$$

where $H_{(\alpha\alpha)} := \text{diag}(\sum_j -E\ell_{ij}^{*\alpha_i\alpha_i})$, $H_{(\gamma\gamma)} := \text{diag}(\sum_i -E\ell_{ij}^{*\gamma_j\gamma_j})$, and $H_{(\alpha\gamma)}$ is a matrix of non-zero terms with each individual element given by $H_{(\alpha\gamma)ij} = -E\ell_{ij}^{*\alpha_i\gamma_j}$. An important point that arises from this presentation is that the non-zero elements of $H_{(\alpha\alpha)}$ and $H_{(\gamma\gamma)}$ that lie along their diagonals are each of order N , whereas all of the elements of $H_{(\alpha\gamma)}$ are of order 1.

Finally, the inverse expected Hessian is H^{-1} , which we similarly decompose as

$$(A.1) \quad H^{-1} = \begin{pmatrix} [H^{-1}]_{(\beta\beta)} & 0 & 0 \\ 0 & [H^{-1}]_{(\alpha\alpha)} & [H^{-1}]_{(\alpha\gamma)} \\ 0 & [H^{-1}]_{(\alpha\gamma)}' & [H^{-1}]_{(\gamma\gamma)} \end{pmatrix}.$$

Note that $[H^{-1}]_{(\beta\beta)} = \left(-\sum_{i,j} E\ell_{ij}^{*\beta\beta}\right)^{-1}$. However, otherwise, the individual blocks of H^{-1} shown in this decomposition are not simply the inverses of the corresponding blocks of H . The individual (row i , column j) elements of each of these blocks will be given by $[H^{-1}]_{(\alpha\alpha)ij}$, $[H^{-1}]_{(\alpha\gamma)ij}$, and $[H^{-1}]_{(\gamma\gamma)ij}$.

THEOREM

Consider the two-way gravity model $y_{ij} = \exp(\alpha_i + \gamma_j + x_{ij}\beta) \omega_{ij}$, with $i, j \in (1, \dots, N)$, $E(\omega_{ij}|x, \alpha, \gamma) = 1$. In addition, assume that we have cross-sectional independence, in the sense that $E(\omega_{ij}\omega_{i'j'}|x, \alpha, \gamma) = 0$ if $i' \neq i$ or $j' \neq j$. We also require that assumption (ii) from Theorem 1 of Weidner and Zylkin (2021) on the boundedness of ℓ_{ij} 's derivatives is satisfied, except our proof requires that all of ℓ_{ij} 's partial derivatives are bounded up to the 6th order instead of up to the 4th order as in that paper.

As $N \rightarrow \infty$, the bias of the fixed effects PPML estimate for β is given by:

$$\frac{b_1}{N^2} + \frac{b_{2\alpha}}{N^2} + \frac{b_{2\gamma}}{N^2} + \frac{b_3}{N^2} + \frac{b_{4\alpha}}{N^2} + \frac{b_{4\gamma}}{N^2} + \frac{b_5}{N^2} + o(N^{-2}),$$

where:

- $b_1 = \frac{1}{2} \left(-\frac{1}{N^2} \sum_{i,j} E\ell_{ij}^{*\beta\beta}\right)^{-1} \sum_{i,j} \sum_k E\ell_{ij}^{*\beta\beta\beta_k} E[(\hat{\beta} - \beta_0)(\hat{\beta}_k - \beta_{0,k})] = O(1),$
- $b_{2\alpha} = \left(-\frac{1}{N^2} \sum_{i,j} E\ell_{ij}^{*\beta\beta}\right)^{-1} \sum_{i,j} E\ell_{ij}^{*\beta\beta\alpha_i} E[(\hat{\beta} - \beta_0)(\hat{\alpha}_i - \alpha_{i0})] = O(1),$
- $b_{2\gamma} = \left(-\frac{1}{N^2} \sum_{i,j} E\ell_{ij}^{*\beta\beta}\right)^{-1} \sum_{i,j} E\ell_{ij}^{*\beta\beta\gamma_j} E[(\hat{\beta} - \beta_0)(\hat{\gamma}_j - \gamma_{j0})] = O(1),$
- $b_3 = \left(-\frac{1}{N^2} \sum_{i,j} E\ell_{ij}^{*\beta\beta}\right)^{-1} \sum_{i,j} E\ell_{ij}^{*\beta\beta\alpha_i\gamma_j} E[(\hat{\alpha}_i - \alpha_{i0})(\hat{\gamma}_j - \gamma_{j0})] = O(1).$

- $b_{4\alpha} = \frac{1}{2} \left(-\frac{1}{N^2} \sum_{i,j} E \ell_{ij}^{*\beta\beta} \right)^{-1} \sum_{i,j} E \ell_{ij}^{*\beta\alpha_i\alpha_i\gamma_j} E [(\hat{\alpha}_i - \alpha_{i0})^2 (\hat{\gamma}_j - \gamma_{j0})] = O(1),$
- $b_{4\gamma} = \frac{1}{2} \left(-\frac{1}{N^2} \sum_{i,j} E \ell_{ij}^{*\beta\beta} \right)^{-1} \sum_{i,j} E \ell_{ij}^{*\beta\alpha_i\gamma_j\gamma_j} E [(\hat{\alpha}_i - \alpha_{i0}) (\hat{\gamma}_j - \gamma_{j0})^2] = O(1),$
- $b_5 = \frac{1}{2} \left(-\frac{1}{N^2} \sum_{i,j} E \ell_{ij}^{*\beta\beta} \right)^{-1} \sum_{i,j} E \ell_{ij}^{*\beta\alpha_i\alpha_i\gamma_j\gamma_j} E [(\hat{\alpha}_i - \alpha_{i0})^2 (\hat{\gamma}_j - \gamma_{j0})^2] = O(1),$
- $o(N^{-2})$ collects all terms of order less than N^{-2} .

REMARKS

1. Because each of the b -terms is of order-1 as $N \rightarrow \infty$ (i.e., they do not vary with N), the combined leading bias term $(b_1 + b_{2\alpha} + b_{2\gamma} + b_3 + b_{4\alpha} + b_{4\gamma} + b_5)/N^2$ is of order N^{-2} .
2. Like in Fernandez-Val and Weidner (2016), Weidner and Zylkin (2021), and other similar work, the expressions for b_1 , $b_{2\alpha}$, $b_{2\gamma}$, b_3 , $b_{4\alpha}$, $b_{4\gamma}$, and b_5 can be derived from a Taylor expansion of the PPML first-order condition for $\hat{\beta}$ around the true parameter values $(\beta_0, \alpha_0, \gamma_0)$. However, unlike in Fernandez-Val and Weidner (2016) and Weidner and Zylkin (2021), we will need to do this expansion up to the fourth-order terms instead of only to the second order. This explains why we need a stricter assumption on the boundedness of the derivatives of the likelihood than the one in Weidner and Zylkin (2021). In addition to these assumptions, our proof will also assume one of the fixed effect values is taken as given in order to enable identification of all of the parameters. For example, we can assume γ_1 , the importer fixed effect for country 1, is known to be 0, such that it can be treated as redundant. However, for simplicity of exposition, we will refrain from making any such redundancies explicit in articulating the proof.
3. To relate these results back to the notation used in Section 2, note that, for example, $q_{ij}^{(\beta\beta_k)} = \frac{1}{2} \left(-\frac{1}{N^2} \sum_{i,j} E \ell_{ij}^{*\beta\beta} \right)^{-1} E \ell_{ij}^{*\beta\beta\beta_k}$, with similar constructions used to constitute $q_{ij}^{(\beta\alpha_i)}$, $q_{ij}^{(\beta\gamma_j)}$, and $q_{ij}^{(\alpha_i\gamma_j)}$. The “ b_4 ” term used in Section 2 consists of $b_{4\alpha} + b_{4\gamma} + b_5$.
4. As discussed in the main text, one of the notable implications of the theorem is that there are no bias terms that have an order greater than N^{-2} . In particular, there are 6 terms from the expansion of $\hat{\beta}$ ’s first-order condition not discussed in Fernandez-Val and Weidner (2016) that, from behind the veil of ignorance, could ostensibly contribute a bias of order $N^{-3/2}$:
 - one term each depending on $\sum_{i,j} E \ell_{ij}^{*\beta\beta\alpha_i} (\hat{\beta} - \beta_0)(\hat{\alpha}_i - \alpha_{i0})$ and $\sum_{i,j} E \ell_{ij}^{*\beta\beta\gamma_j} (\hat{\beta} - \beta_0)(\hat{\gamma}_j - \gamma_{j0})$,
 - a second set of two terms depending on $\sum_{i,j} E \ell_{ij}^{*\beta\alpha_i\alpha_i\alpha_i} (\hat{\alpha}_i - \alpha_{i0})^3$ and $\sum_{i,j} E \ell_{ij}^{*\beta\gamma_j\gamma_j\gamma_j} (\hat{\gamma}_j - \gamma_{j0})^3$,

- a third set depending on $\sum_{i,j} E \ell_{ij}^{*\beta\alpha_i\alpha_i\gamma_j} (\hat{\alpha}_i - \alpha_{i0})^2 (\hat{\gamma}_j - \gamma_{j0})$ and $\sum_{i,j} E \ell_{ij}^{*\beta\alpha_i\gamma_j\gamma_j} (\hat{\alpha}_i - \alpha_{i0})(\hat{\gamma}_j - \gamma_{j0})^2$.

The first two of these terms give rise to $b_{2\alpha}$ and $b_{2\gamma}$ in the theorem and will be shown to contribute bias terms of order N^{-2} , not $N^{-3/2}$. We will also show that the second set of terms are each equal to 0 for PPML and that the third set contributes bias terms of order N^{-2} .

5. b_1 can be thought of as analogous to the standard bias term one would normally obtain for the PPML estimator without fixed effects. Note when β is a scalar, the combined term b_1/N^2 can be reduced to:

$$\frac{1}{2} \left(-\frac{1}{N^2} \sum_{i,j} E \ell_{ij}^{*\beta\beta} \right)^{-1} \left(\frac{1}{N^2} \sum_{i,j} E \ell_{ij}^{*\beta\beta\beta} \right) \text{Var}(\hat{\beta} - \beta_0),$$

where we ignore the (negligible) difference between $\text{Var}(\hat{\beta} - \beta_0)$ and $E[(\hat{\beta} - \beta_0)^2]$.¹¹ This presentation makes it clear that the order of b_1/N^2 comes from $\text{Var}(\hat{\beta} - \beta_0)$, which has order $1/N^2$. It also follows that $b_1 = O(1)$.

6. Compared with b_1 , it is less immediate to see why some of the other b -terms are $O(1)$. Though the full proof requires more details, some insights can be gained by studying the off-diagonal terms of the inverse-Hessian matrix, which is closely related to the asymptotic covariances of all of the parameter estimates. For example, each element of the off-diagonal inverse-Hessian block term $[H^{-1}]_{(\alpha\gamma)}$ turns out to be of order $1/N^2$ as $N \rightarrow \infty$, whereas the diagonal elements of $[H^{-1}]_{(\alpha\alpha)}$ and $[H^{-1}]_{(\gamma\gamma)}$ are of order $1/N$, a result that is closely related to Lemma D.1 of Fernandez-Val and Weidner (2016). For our purposes, this observation has three important implications:

- Because the order of the asymptotic covariance between $\hat{\alpha}_i$ and $\hat{\gamma}_j$ is determined by the order of either $[H^{-1}]_{(\alpha\gamma)ij}$ (under correct specification of the PPML variance) or $[H^{-1}]_{(\alpha\alpha)ii}[H^{-1}]_{(\gamma\gamma)jj}$ (in more general cases), the order of $\text{Cov}(\hat{\alpha}_i, \hat{\gamma}_j)$ can be shown to be $1/N^2$, indicating that the order of b_3 is likely 1.
- Somewhat similarly, the order of $\text{Cov}(\hat{\beta}, \hat{\alpha}_i)$ is generally determined by that of $[H^{-1}]_{(\alpha\alpha)ii}[H^{-1}]_{\beta\beta} \times \sum_j E(\ell_{ij}^{*\alpha_i} \ell_{ij}^{*\beta})$, indicating it is $O(1/N^2)$, with a similar argument following for $\text{Cov}(\hat{\beta}, \hat{\gamma}_j)$. This suggests that both $b_{2\alpha}$ and $b_{2\gamma}$ are likely to be order-1.

¹¹ The reason for this difference is the bias of $\hat{\beta}$. One can write $E[(\hat{\beta} - \beta_0)^2] = \text{Var}(\hat{\beta} - \beta_0) + \text{bias}(\hat{\beta})^2$. Since we will show the bias of $\hat{\beta}$ is $O(N^{-2})$, the latter term is $O(N^{-4})$.

- Though Fernandez-Val and Weidner (2016) found that the estimation noise in the fixed effects estimates $\hat{\alpha}_i$ and $\hat{\gamma}_j$ do not induce a $1/N$ asymptotic bias term for the PPML estimator—a result that is unique to PPML—it turns out that the noisiness of $\hat{\alpha}_i$ and $\hat{\gamma}_j$ still contributes in general to the bias of two-way PPML. Moreover, it also turns out that the noise in these fixed effects contributes to the bias of PPML in a different way than the noise of the estimated slope parameters, i.e., strictly through their interactions with each other and with the slope estimates rather than through the variances or other central moments of either term.
7. Because we rely on Lemma D.1 of Fernandez-Val and Weidner (2016), it is useful to clarify how their result translates to our setting given the differences in notation. In particular, because we express the Hessian and its inverse differently, their Lemma D.1 in our setting implies that

$$\left\| \begin{pmatrix} [H^{-1}]_{(\alpha\alpha)} & [H^{-1}]_{(\alpha\gamma)} \\ [H^{-1}]_{(\alpha\gamma)}' & [H^{-1}]_{(\gamma\gamma)} \end{pmatrix} - \begin{pmatrix} H_{(\alpha\alpha)}^{-1} & 0 \\ 0 & H_{(\gamma\gamma)}^{-1} \end{pmatrix} \right\|_{\max} = O_p(N^{-2}),$$

where, recall, $H_{(\alpha\alpha)} := \text{diag}(\sum_j -E\ell_{ij}^{*\alpha_i\alpha_i}), H_{(\gamma\gamma)} := \text{diag}(\sum_i -E\ell_{ij}^{*\gamma_j\gamma_j})$. The reason for the difference with the statement in their Lemma is, first, $\sqrt{NT}$ in their setting equates to N in ours; second, their Hessian is scaled so that their $H_{(\alpha\alpha)}^{-1}$ and $H_{(\gamma\gamma)}^{-1}$ are each of order-1, whereas in our presentation they are order- $1/N$.

8. Some auxiliary results that needed for the complete proof are provided in our [Supplementary Appendix](#). These auxiliary results include deriving appropriate approximations for $(\hat{\alpha}_i - \alpha_{i0})$ and $(\hat{\gamma}_j - \gamma_{j0})$ and deriving the order of some of the trickier terms from the full expansion of $(\hat{\beta} - \beta_0)$. In particular, we prove that $(\hat{\alpha}_i - \alpha_{i0})$ and $(\hat{\gamma}_j - \gamma_{j0})$ are generally $O_p(N^{-1/2})$ and are $O(N^{-1})$ in expectation. That is, each fixed effect estimate can be decomposed into an $O(N^{-1})$ bias and an $O_p(N^{-1/2})$ stochastic part. We further show how the stochastic part can itself be decomposed into larger and smaller parts, with the larger parts depending only on the scores of each fixed effect. These decompositions can then be used to rigorously bound terms that depend on combinations of $(\hat{\alpha}_i - \alpha_{i0})$ and $(\hat{\gamma}_j - \gamma_{j0})$.

PROOF

Consider the following PPML first-order condition for $\hat{\beta}$:

$$\sum_{i,j} \ell_{ij}^{*\beta}(\hat{\beta}, \hat{\alpha}_i, \hat{\gamma}_j) = 0.$$

An initial second-order Taylor expansion around $(\beta_0, \alpha_0, \gamma_0)$ gives us

$$(A.2) \quad 0 = \sum_{i,j} \ell_{ij}^{*\beta}(\hat{\beta}, \hat{\alpha}_i, \hat{\gamma}_j) = \sum_{i,j} \ell_{ij}^{*\beta} + \sum_{i,j} \ell_{ij}^{*\beta\beta}(\hat{\beta} - \beta_0) + \sum_{i,j} \ell_{ij}^{*\beta\alpha_i}(\hat{\alpha}_i - \alpha_{0i})$$

$$\begin{aligned}
& + \sum_{i,j} \ell_{ij}^{*\beta\gamma_j} (\hat{\gamma}_j - \gamma_{0j}) + \frac{1}{2} \sum_{i,j} \sum_k \ell_{ij}^{*\beta\beta\beta_k} (\hat{\beta} - \beta_0) (\hat{\beta}_k - \beta_{0,k}) + \frac{1}{2} \sum_{i,j} \ell_{ij}^{*\beta\alpha_i\alpha_i} (\hat{\alpha}_i - \alpha_{0i})^2 \\
& + \sum_{i,j} \ell_{ij}^{*\beta\beta\alpha_i} (\hat{\beta} - \beta_0) (\hat{\alpha}_i - \alpha_{0i}) + \sum_{i,j} \ell_{ij}^{*\beta\beta\gamma_j} (\hat{\beta} - \beta_0) (\hat{\gamma}_j - \gamma_{0j}) \\
& + \sum_{i,j} \ell_{ij}^{*\beta\alpha_i\gamma_j} (\hat{\alpha}_i - \alpha_{0i}) (\hat{\gamma}_j - \gamma_{0j}) + \frac{1}{2} \sum_{i,j} \ell_{ij}^{*\beta\gamma_j\gamma_j} (\hat{\gamma}_j - \gamma_{0j})^2 + R,
\end{aligned}$$

where R is the remainder, which we will elaborate on later. The first term $\sum_{i,j} \ell_{ij}^{*\beta}$ is the score evaluated at the true parameter values and thus contributes only to the variance of $(\hat{\beta} - \beta_0)$ and does not contribute to its bias. The remaining terms therefore dictate the bias of the PPML score when evaluated at the estimated parameter values. Before breaking down these remaining terms further, it is useful to specify the following derivatives of the PPML pseudo-likelihood:

- $\ell_{ij}^{*\beta\beta} = -\lambda_{ij} \tilde{x}_{ij} \tilde{x}'_{ij}$
- $\ell_{ij}^{*\beta\alpha_i} = \ell_{ij}^{*\beta\gamma_j} = \ell_{ij}^{*\beta\alpha_i\alpha_i} = \ell_{ij}^{*\beta\gamma_j\gamma_j} = \ell_{ij}^{*\beta\alpha_i\gamma_j} = \dots = -\lambda_{ij} \tilde{x}_{ij}$
- $\ell_{ij}^{*\beta\beta\beta_k} = -\lambda_{ij} \tilde{x}_{ij} \tilde{x}'_{ij} \tilde{x}_{ij,k}$

Note that all of the second- and higher-order derivatives of the PPML pseudolikelihood do not depend on the random parts of the model (i.e., y_{ij} , ω_{ij}). Therefore, we have $\ell_{ij}^{*\beta\beta} = E\ell_{ij}^{*\beta\beta}$, $\ell_{ij}^{*\beta\alpha_i} = E\ell_{ij}^{*\beta\alpha_i}$, $\ell_{ij}^{*\beta\beta\beta_k} = E\ell_{ij}^{*\beta\beta\beta_k}$, $\ell_{ij}^{*\beta\alpha_i\alpha_i} = E\ell_{ij}^{*\beta\alpha_i\alpha_i}$, etc.. Also, note that the definitions of ℓ_{ij}^* and $\tilde{x}_{ij}$ guarantee that

$$(A.3) \quad \sum_j E\ell_{ij}^{*\beta\alpha_i} = \sum_j E\ell_{ij}^{*\beta\alpha_i\alpha_i} = \dots = \sum_j -\lambda_{ij} \tilde{x}_{ij} = 0 \quad \forall i,$$

$$(A.4) \quad \sum_i E\ell_{ij}^{*\beta\gamma_j} = \sum_i E\ell_{ij}^{*\beta\gamma_j\gamma_j} = \dots = \sum_i -\lambda_{ij} \tilde{x}_{ij} = 0 \quad \forall j.$$

Returning to the Taylor expansion from (A.2), we start by incorporating the existing results from Fernandez-Val and Weidner (2016) that imply $\sum_{i,j} \ell_{ij}^{*\beta\alpha_i} (\hat{\alpha}_i - \alpha_{0i})$, $\sum_{i,j} \ell_{ij}^{*\beta\gamma_j} (\hat{\gamma}_j - \gamma_{0j})$, $\sum_{i,j} \ell_{ij}^{*\beta\alpha_i\alpha_i} (\hat{\alpha}_i - \alpha_{0i})^2$, and $\sum_{i,j} \ell_{ij}^{*\beta\gamma_j\gamma_j} (\hat{\gamma}_j - \gamma_{0j})^2$ all drop out of the bias when the estimator is PPML; see their Example 2 and Remark 3. The most straightforward way to show this for our purposes is using (A.3) and (A.4). Namely, these 4 terms are all equal to zero due to the properties noted in (A.3) and (A.4) after making the innocuous replacement of $\ell_{ij}^{*\beta\alpha_i}$ with $E\ell_{ij}^{*\beta\alpha_i}$, $\ell_{ij}^{*\beta\gamma_j}$ with $E\ell_{ij}^{*\beta\gamma_j}$, $\ell_{ij}^{*\beta\alpha_i\alpha_i}$ with $E\ell_{ij}^{*\beta\alpha_i\alpha_i}$ and $\ell_{ij}^{*\beta\gamma_j\gamma_j}$ with $E\ell_{ij}^{*\beta\gamma_j\gamma_j}$. Note this is a slightly different argument than the one used in Fernandez-Val and Weidner (2016) since they first derive a general formula for the order- N^{-1} bias of fixed effects estimators and then specialize it to PPML. Their demonstration effectively only requires that the leading parts of the first two terms $\sum_{i,j} \ell_{ij}^{*\beta\alpha_i} (\hat{\alpha}_i - \alpha_{0i})$ and

$\sum_{i,j} \ell_{ij}^{*\beta\gamma_j} (\hat{\gamma}_j - \gamma_{j0})$ have expectation 0 and that their remaining parts are small enough in order to be disregarded, whereas our approach produces the stronger result that both of these terms are equal to zero. In any case, the absence of these four terms ensures that PPML does not have any order- $1/N$ bias terms that would be present for other similar estimators, in accordance with Fernandez-Val and Weidner (2016). Moreover, these 4 terms do not contribute to any $1/N^2$ bias either.

The next step is to solve (A.2) in terms of $(\hat{\beta} - \beta_0)$ in order to obtain an expression for the bias of $\hat{\beta}$. After also removing from (A.2) the terms we have just discussed, replacing $\ell_{ij}^{*\beta\beta}$ with $E\ell_{ij}^{*\beta\beta}$, making other similar innocuous replacements, and applying expectations, the resulting expression is:

$$\begin{aligned} E(\hat{\beta} - \beta_0) = & \left(- \sum_{i,j} E\ell_{ij}^{*\beta\beta} \right)^{-1} \left(\frac{1}{2} \sum_{i,j} \sum_k E\ell_{ij}^{*\beta\beta\beta_k} E[(\hat{\beta} - \beta_0)(\hat{\beta}_k - \beta_{0,k})] \right. \\ & + \sum_{i,j} E\ell_{ij}^{*\beta\beta\alpha_i} E[(\hat{\beta} - \beta_0)(\hat{\alpha}_i - \alpha_{i0})] + \sum_{i,j} E\ell_{ij}^{*\beta\beta\gamma_j} E[(\hat{\beta} - \beta_0)(\hat{\gamma}_j - \gamma_{j0})] \\ & \left. + \sum_{i,j} E\ell_{ij}^{*\beta\alpha_i\gamma_j} E[(\hat{\alpha}_i - \alpha_{i0})(\hat{\gamma}_j - \gamma_{j0})] \right) + R_\beta, \end{aligned}$$

where $R_\beta = \left(- \sum_{i,j} E\ell_{ij}^{*\beta\beta} \right)^{-1} E(R)$ inherits the remainder from (A.1).¹²

This last expression now more closely resembles our articulation of the bias in the theorem. We turn now to establishing that $E[(\hat{\beta} - \beta_0)(\hat{\alpha}_i - \alpha_{i0})]$, $E[(\hat{\beta} - \beta_0)(\hat{\gamma}_j - \gamma_{j0})]$, and $E[(\hat{\alpha}_i - \alpha_{i0})(\hat{\gamma}_j - \gamma_{j0})]$ are each $O(1/N^2)$, which is sufficient ensure that $b_{2\alpha}$, $b_{2\gamma}$, and b_3 are all $O(1)$ as stated in the theorem. $b_{4\alpha}$, $b_{4\gamma}$, and b_5 will be extracted from the remainder, which we will turn to later.

Before unpacking each of these terms, we first find it useful to state some useful expressions to work with for $(\hat{\alpha}_i - \alpha_{i0})$ and $(\hat{\gamma}_j - \gamma_{j0})$. In our [Supplementary Appendix](#), we prove (in the form of Lemma 1), that $(\hat{\alpha}_i - \alpha_{i0})$ and $(\hat{\gamma}_j - \gamma_{j0})$ can respectively be written as

$$(A.5) \quad (\hat{\alpha}_i - \alpha_{i0}) = \psi^{(\alpha_i)} + \left(- \sum_j E\ell_{ij}^{*\alpha_i\alpha_i} \right)^{-1} \sum_j E\ell_{ij}^{*\alpha_i\gamma_j} \psi^{(\gamma_j)} + B^{(\alpha_i)} + O_p\left(N^{-\frac{3}{2}}\right),$$

$$(A.6) \quad (\hat{\gamma}_j - \gamma_{j0}) = \psi^{(\gamma_j)} + \left(- \sum_i E\ell_{ij}^{*\gamma_j\gamma_j} \right)^{-1} \sum_i E\ell_{ij}^{*\gamma_j\alpha_i} \psi^{(\alpha_i)} + B^{(\gamma_j)} + O_p\left(N^{-\frac{3}{2}}\right),$$

¹² Because $\ell_{ij}^{*\beta\beta} = E\ell_{ij}^{*\beta\beta}$, we ignore the possibility that $\left(- \sum_{i,j} E\ell_{ij}^{*\beta\beta} \right)^{-1} \sum_{i,j} E\left[\left(\ell_{ij}^{*\beta\beta} - E\ell_{ij}^{*\beta\beta} \right) (\hat{\beta} - \beta_0) \right]$ contributes any bias. A more generalized bias expansion would include this term.

where $\psi^{(\alpha_i)} = (-\sum_j E \ell_{ij}^{*\alpha_i \alpha_i})^{-1} \sum_j \ell_{ij}^{*\alpha_i}$ and $\psi^{(\gamma_j)} = (-\sum_i E \ell_{ij}^{*\gamma_j \gamma_j})^{-1} \sum_i \ell_{ij}^{*\gamma_j}$ are both $O_p(N^{-1/2})$ with mean 0 and $B^{(\alpha_i)}$ and $B^{(\gamma_j)}$ are $O(1/N)$ bias terms capturing the bias of each fixed effect estimate.

Similarly, though more simply, a corresponding expansion for $(\hat{\beta} - \beta_0)$ can be written as

$$(A.7) \quad (\hat{\beta} - \beta_0) = \psi^{(\beta)} + B^{(\beta)} + O_p(N^{-2}),$$

where $\psi^{(\beta)} = (-\sum_{i,j} E \ell_{ij}^{*\beta \beta})^{-1} \sum_{i,j} \ell_{ij}^{*\beta}$ and $B^{(\beta)} = o(N^{-1})$ is the bias of $\hat{\beta}$. Note we do not know yet that the order of $B^{(\beta)}$ is actually $O(N^{-2})$, since that is part of what we are trying to prove. We do know that it is $o(N^{-1})$ thanks to Fernandez-Val and Weidner (2016); see their Example 2 and our equation (3). The derivation of this expression for $(\hat{\beta} - \beta_0)$ is formalized in our Supplementary Appendix (as Lemma 3).

Importantly, the assumption of cross-sectional independence means that the score terms $\psi^{(\alpha_i)}$, $\psi^{(\gamma_j)}$, and $\psi^{(\beta)}$ have the following properties:

- For any non-stochastic term B , $E(\psi^{(\alpha_i)} B) = E(\psi^{(\gamma_j)} B) = E(\psi^{(\beta)} B) = 0$.
- If $i \neq i'$, $E(\psi^{(\alpha_i)} \psi^{(\alpha_{i'})}) = 0$; likewise, If $j \neq j'$, $E(\psi^{(\gamma_j)} \psi^{(\gamma_{j'})}) = 0$.
- $E(\psi^{(\alpha_i)} \psi^{(\gamma_j)}) = (-\sum_j E \ell_{ij}^{*\alpha_i \alpha_i})^{-1} (-\sum_i E \ell_{ij}^{*\gamma_j \gamma_j})^{-1} E(\ell_{ij}^{*\alpha_i} \ell_{ij}^{*\gamma_j}) = O(N^{-2})$.
- $E(\psi^{(\alpha_i)} \psi^{(\beta)}) = (-\sum_j E \ell_{ij}^{*\alpha_i \alpha_i})^{-1} (-\sum_{i,j} E \ell_{ij}^{*\beta \beta})^{-1} E(\sum_j \ell_{ij}^{*\alpha_i} \ell_{ij}^{*\beta}) = O(N^{-2})$, with a similar result holding for $E(\psi^{(\gamma_j)} \psi^{(\beta)})$.

In the third of these properties, we use the fact that $E(\sum_{j'} \ell_{ij'}^{*\alpha_i} \sum_{i'} \ell_{i'j}^{*\gamma_j})$ reduces to $E(\ell_{ij}^{*\alpha_i} \ell_{ij}^{*\gamma_j}) = O(1)$. The fourth property uses $E(\sum_j \ell_{ij}^{*\alpha_i} \sum_{i',j} \ell_{i'j}^{*\beta}) = E(\sum_j \ell_{ij}^{*\alpha_i} \ell_{ij}^{*\beta}) = O(N)$, with a similar reduction occurring for $E(\sum_i \ell_{ij}^{*\gamma_j} \sum_{i,j'} \ell_{i'j'}^{*\beta})$.

Using these properties along with (A.5), (A.6), and (A.7), we can immediately derive the following for $E[(\hat{\beta} - \beta_0)(\hat{\alpha}_i - \alpha_{i0})]$:

$$\begin{aligned} E[(\hat{\beta} - \beta_0)(\hat{\alpha}_i - \alpha_{i0})] &= E(\psi^{(\alpha_i)} \psi^{(\beta)}) + \left(-\sum_j E \ell_{ij}^{*\alpha_i \alpha_i} \right)^{-1} \sum_j E \ell_{ij}^{*\alpha_i \gamma_j} E(\psi^{(\gamma_j)} \psi^{(\beta)}) + o(N^{-2}) \\ &= O(N^{-2}). \end{aligned}$$

The middle term on the right-hand side of the first line is equal to $O(N^{-2})$ because it equates to $O(N^{-1}) \cdot \sum_j O(N^{-2}) = O(N^{-1}) \cdot O(N^{-1})$. We also have that $E[(\hat{\beta} - \beta_0)(\hat{\gamma}_j - \gamma_{j0})] = O(N^{-2})$ by a similar logic. This is sufficient to prove that $b_{2\alpha}$ and $b_{2\gamma}$ from the theorem are $O(1)$.

Deriving the order of $E[(\hat{\alpha}_i - \alpha_{i0})(\hat{\gamma}_j - \gamma_{j0})]$ requires slightly more explanation but follows from the same types of arguments. We have:

$$\begin{aligned}
& E[(\hat{\alpha}_i - \alpha_{i0})(\hat{\gamma}_j - \gamma_{j0})] \\
&= E(\psi^{(\alpha_i)}\psi^{(\gamma_j)}) + \left(-\sum_{j'} E\ell_{ij'}^{*\alpha_i\alpha_i}\right)^{-1} E\ell_{ij}^{*\alpha_i\gamma_j} E(\psi^{(\gamma_j)}\psi^{(\gamma_j)}) + \left(-\sum_{i'} E\ell_{i'j}^{*\gamma_j\gamma_j}\right)^{-1} E\ell_{ij}^{*\gamma_j\alpha_i} E(\psi^{(\alpha_i)}\psi^{(\alpha_i)}) \\
&+ \left(-\sum_{j'} E\ell_{ij'}^{*\alpha_i\alpha_i}\right)^{-1} \left(-\sum_{i'} E\ell_{i'j}^{*\gamma_j\gamma_j}\right)^{-1} \sum_{i',j'} E\ell_{ij'}^{*\alpha_i\gamma_{j'}} E\ell_{i'j}^{*\alpha_{i'}\gamma_j} E(\psi^{(\alpha_{i'})}\psi^{(\gamma_{j'})}) + O(N^{-2}) \\
&= O(N^{-2})
\end{aligned}$$

This result holds because we can write right-hand side of the first line as $O(N^{-2}) + O(N^{-1}) \cdot O(N^{-1}) + O(N^{-1}) \cdot O(N^{-1}) + O(N^{-1}) \cdot O(N^{-1}) \cdot \sum_{i,j} O(N^{-2}) + O(N^{-2})$, which equals $O(N^{-2})$. This is sufficient to show that b_3 is $O(1)$ as stated in the theorem.

Having characterized all of the second-order terms from the expansion of $E(\hat{\beta} - \beta_0)$, we now need to return to R_β , which contains the bias contributions from all of the higher-order terms. Highlighting the terms of interest in R_β that could plausibly be of order N^{-2} or greater, we have:

$$\begin{aligned}
(A.8) \ R_\beta &= \left(-\sum_{i,j} E\ell_{ij}^{*\beta\beta}\right)^{-1} \left(r^{\beta\beta\alpha_i\alpha_i} + r^{\beta\beta\alpha_i\gamma_j} + r^{\beta\beta\gamma_j\gamma_j} + r^{\beta\alpha_i\alpha_i\alpha_i} + r^{\beta\alpha_i\alpha_i\gamma_j} + r^{\beta\alpha_i\gamma_j\gamma_j} \right. \\
&\quad + r^{\beta\gamma_j\gamma_j\gamma_j} + r^{\beta\alpha_i\alpha_i\alpha_i\alpha_i} + r^{\beta\alpha_i\alpha_i\alpha_i\gamma_j} + r^{\beta\alpha_i\alpha_i\gamma_j\gamma_j} + r^{\beta\alpha_i\gamma_j\gamma_j\gamma_j} + r^{\beta\gamma_j\gamma_j\gamma_j\gamma_j} \Big) \\
&\quad + o(N^{-2}),
\end{aligned}$$

where we henceforth let $r^{\beta\beta\alpha_i\alpha_i} := \frac{1}{2} \sum_{i,j} E\ell_{ij}^{*\beta\beta\alpha_i\alpha_i} E[(\hat{\beta} - \beta_0)(\hat{\alpha}_i - \alpha_{0i})^2]$, $r^{\beta\beta\alpha_i\gamma_j} := \sum_{i,j} E\ell_{ij}^{*\beta\beta\alpha_i\gamma_j} E[(\hat{\beta} - \beta_0)(\hat{\alpha}_i - \alpha_{0i})(\hat{\gamma}_j - \gamma_{0j})]$, etc., using obvious notation.¹³ Because $\left(-\sum_{i,j} E\ell_{ij}^{*\beta\beta}\right)^{-1}$ is $O(N^{-2})$, we need to distinguish which of the r -terms in parentheses are $O(1)$ versus being $o(1)$. We proceed using the following steps, some of which are elaborated on in the Supplementary Appendix:

1. By (A.3) and (A.4), all terms of the form $r^{\beta\alpha_i\alpha_i\alpha_i}$, $r^{\beta\gamma_j\gamma_j\gamma_j}$, $r^{\beta\alpha_i\alpha_i\alpha_i\alpha_i}$, etc. are equal to 0.
2. For $r^{\beta\beta\alpha_i\alpha_i}$, $r^{\beta\beta\alpha_i\gamma_j}$, $r^{\beta\beta\gamma_j\gamma_j}$, $r^{\beta\alpha_i\alpha_i\alpha_i\gamma_j}$, and $r^{\beta\alpha_i\gamma_j\gamma_j\gamma_j}$, when we use the decompositions for $(\hat{\beta} - \beta_0)$, $(\hat{\alpha}_i - \alpha_{0i})$, and $(\hat{\gamma}_j - \gamma_{0j})$ from (A.5)-(A.7) to construct the relevant products such as $(\hat{\beta} - \beta_0)(\hat{\alpha}_i - \alpha_{0i})^2$, $(\hat{\beta} - \beta_0)(\hat{\alpha}_i - \alpha_{0i})(\hat{\gamma}_j - \gamma_{0j})$, etc., we find there is only one

¹³ The logic for focusing on these terms follows from how $(\hat{\beta} - \beta_0)$ is $O_p(N^{-1})$ while both $(\hat{\alpha}_i - \alpha_{0i})$ and $(\hat{\gamma}_j - \gamma_{0j})$ are $O_p(N^{-1/2})$.

term from each of these products that is at most $O_p(N^{-2})$ in order. In all of these cases, the expectation of this single term turns out to be $o(N^{-2})$.¹⁴ Further details are provided in the Supplementary Appendix.

3. For the remaining terms $r^{\beta\alpha_i\alpha_i\gamma_j}$, $r^{\beta\alpha_i\gamma_j\gamma_j}$, and $r^{\beta\alpha_i\alpha_i\gamma_j\gamma_j}$, we work with the decompositions from (A.5)-(A.7) in the Supplementary Appendix to establish the following:
 - The order of $E[(\hat{\alpha}_i - \alpha_{0i})^2(\hat{\gamma}_j - \gamma_{0j})]$ is determined by $E\left((\psi^{(\alpha_i)})^2 B^{(\gamma_j)}\right) = O(N^{-2})$.
 - Similarly, the order of $E\left[(\hat{\alpha}_i - \alpha_{0i})(\hat{\gamma}_j - \gamma_{0j})^2\right]$ is determined by $E\left((\psi^{\gamma_j})^2 B^{(\alpha_i)}\right) = O(N^{-2})$.
 - Finally, the order of $E\left[(\hat{\alpha}_i - \alpha_{0i})^2(\hat{\gamma}_j - \gamma_{0j})^2\right]$ is determined by $\left(-\sum_{j'} E\ell_{ij'}^{*\alpha_i\alpha_i}\right)^{-2} \left(-\sum_{i'} E\ell_{i'j}^{*\gamma_j\gamma_j}\right)^{-2} E\left[\sum_{j' \neq j} (\ell_{ij'}^{*\alpha_i})^2\right] E\left[\sum_{i' \neq i} (\ell_{i'j}^{*\gamma_j})^2\right] = O(N^{-2}) \cdot O(N^{-2}) \cdot O(N) \cdot O(N) = O(N^{-2})$.

This means that $r^{\beta\alpha_i\alpha_i\gamma_j} = \frac{1}{2} \sum_{i,j} E\ell_{ij}^{*\beta\beta\alpha_i\alpha_i\gamma_j} E[(\hat{\alpha}_i - \alpha_{0i})^2(\hat{\gamma}_j - \gamma_{0j})]$, $r^{\beta\alpha_i\gamma_j\gamma_j} = \frac{1}{2} \sum_{i,j} E\ell_{ij}^{*\beta\beta\alpha_i\gamma_j\gamma_j} E[(\hat{\alpha}_i - \alpha_{0i})(\hat{\gamma}_j - \gamma_{0j})^2]$, and $r^{\beta\alpha_i\alpha_i\gamma_j\gamma_j} = \frac{1}{2} \sum_{i,j} E\ell_{ij}^{*\beta\beta\alpha_i\gamma_j\gamma_j} E[(\hat{\alpha}_i - \alpha_{0i})^2(\hat{\gamma}_j - \gamma_{0j})^2]$ are each $O(1)$, ensuring in turn that $b_{4\alpha}$, $b_{4\gamma}$, and b_5 from the theorem are $O(1)$ as well.

Therefore, the bias of $\hat{\beta}$ is given by

$$\begin{aligned}
 E(\hat{\beta} - \beta_0) = & \left(-\sum_{i,j} E\ell_{ij}^{*\beta\beta} \right)^{-1} \left(\frac{1}{2} \sum_{i,j} \sum_k E\ell_{ij}^{*\beta\beta\beta_k} E[(\hat{\beta} - \beta_0)(\hat{\beta}_k - \beta_{0,k})] \right. \\
 & + \sum_{i,j} E\ell_{ij}^{*\beta\beta\alpha_i} E[(\hat{\beta} - \beta_0)(\hat{\alpha}_i - \alpha_{i0})] + \sum_{i,j} E\ell_{ij}^{*\beta\beta\gamma_j} E[(\hat{\beta} - \beta_0)(\hat{\gamma}_j - \gamma_{j0})] \\
 & + \sum_{i,j} E\ell_{ij}^{*\beta\alpha_i\gamma_j} E[(\hat{\alpha}_i - \alpha_{i0})(\hat{\gamma}_j - \gamma_{j0})] + \sum_{i,j} E\ell_{ij}^{*\beta\alpha_i\alpha_i\gamma_j} E[(\hat{\alpha}_i - \alpha_{i0})^2(\hat{\gamma}_j - \gamma_{j0})] \\
 & + \sum_{i,j} E\ell_{ij}^{*\beta\alpha_i\gamma_j\gamma_j} E[(\hat{\alpha}_i - \alpha_{i0})(\hat{\gamma}_j - \gamma_{j0})^2] \\
 & \left. + \sum_{i,j} E\ell_{ij}^{*\beta\alpha_i\alpha_i\gamma_j\gamma_j} E[(\hat{\alpha}_i - \alpha_{i0})^2(\hat{\gamma}_j - \gamma_{j0})^2] \right) + o(N^{-2}).
 \end{aligned}$$

¹⁴ This is sufficient to prove, e.g., that $E(r^{\beta\beta\alpha_i\gamma_j}) = o(1)$ because it dictates that $E(r^{\beta\beta\alpha_i\alpha_i}) = \frac{1}{2} \sum_{i,j} E\ell_{ij}^{*\beta\beta\alpha_i\alpha_i} \cdot o(N^{-2}) = o(1)$.

Furthermore, we have shown that the 4 terms in parentheses of the last expression have order 1 as $N \rightarrow \infty$, ensuring that the terms $b_1, b_{2\alpha}, b_{2\gamma}, b_3, b_{4\alpha}, b_{4\gamma}$, and b_5 defined in the theorem are of order 1 as well. ■

SIMULATIONS WITH LARGER N

These simulations explore the implications of further increasing the number of countries (N) beyond the sample sizes used in the main simulations. For added precision, and on account of the biases can become quite small, we also increase the number of repetitions to 50,000 (versus 5,000 for the simulations used in the main text). Because these simulations are already computationally intensive, and result in very small biases anyway, we do not consider bias corrections.

We present the results in Table A.1 below. To demonstrate the effects of increasing N by multiples of 2, starting with $N = 40$ up to $N = 640$. This gives us results for $N = 40, N = 80, N = 160, N = 320$, and $N = 640$. Given the presumption of order- $1/N^2$ bias, we thus expect that using a larger value for N should eventually result in the bias being divided by 4. The DGPs we use in this case use the following error assumptions:

1. $\log \omega_{ij}^{(1)} \sim N \left(\log \frac{1}{\sqrt{1+\lambda_{ij}^{-1}}}, \log(1 + \lambda_{ij}^{-1}) \right)$
2. $\log \omega_{ij}^{(2)} \sim N \left(\log \frac{1}{\sqrt{2}}, \log 2 \right)$
3. $\log \omega_{ij}^{(3)} \sim N \left(\log \frac{1}{\sqrt{1+100\lambda_{ij}^{-1}}}, \log(1 + 100\lambda_{ij}^{-1}) \right)$
4. $\log \omega_{ij}^{(4)} \sim N \left(\log \frac{1}{\sqrt{101}}, \log 101 \right)$
5. $\log \omega_{ij}^{(5)} \sim N \left(\log \frac{1}{\sqrt{1+10\lambda_{ij}^{-1}}}, \log(1 + 10\lambda_{ij}^{-1}) \right)$
6. $\log \omega_{ij}^{(6)} \sim N \left(\log \frac{1}{\sqrt{11}}, \log 11 \right)$

That is, they include intermediate cases between DGPs 1 and 3 and 2 and 4. This allows us to interpolate between these various cases to the extent that results differ for large N .

Table A.1: Simulations results for larger values of N.

		DGP					
	N	$\omega_{ij}^{(1)}$	$\omega_{ij}^{(2)}$	$\omega_{ij}^{(3)}$	$\omega_{ij}^{(4)}$	$\omega_{ij}^{(5)}$	$\omega_{ij}^{(6)}$
Bias	40	0.000702	-0.002395	0.026031	-0.018220	0.005982	-0.010411
	80	0.000278	-0.000699	0.009805	-0.012453	0.001816	-0.004873

160	0.000105	-0.000237	0.003046	-0.007574	0.000500	-0.002151
320	0.000014	-0.000128	0.000821	-0.003406	0.000088	-0.000783
640	0.000001	-0.000029	0.000206	-0.001234	0.000016	-0.000187

Note: Column groups identify different error assumptions, based on the notation used in this appendix. N is the number of countries, which equates to the square root of the sample size. Results show the bias of PPML with two-way fixed effects based on different values for N and for different error assumptions.

Broadly, the results in Table A.1 support the proposition that the bias decreases asymptotically with $1/N^2$. However, they also underline the possibility that this asymptotic rate may only emerge as the sample size gets very large. Notably, for error variance $\omega_{ij}^{(4)}$, which introduces a high degree of noise, even doubling the sample size from 320 to 640 only reduces the simulated bias by a factor closer to $1/3$ rather than $1/4$, despite giving us a sample size of over 400,000. As shown in the results for $\omega_{ij}^{(6)}$, reducing the degree of noise slightly again delivers the expected result of a bias that varies asymptotically with $1/N^2$, at least for larger sample sizes.

SIMULATION RESULTS WITH BOOTSTRAP STANDARD ERRORS

In the main text, to make comparisons as transparent as possible, we only used typical “heteroskedasticity robust” analytical standard errors. However, as our own results for coverage show, and as we have tried to carefully indicate, these standard errors are far from adequate.

Indeed, it is by now well known that these standard errors are themselves biased (see Pfaffermayr 2019, 2021; Weidner and Zylkin 2021.) Moreover, the bias is generally downwards and, unfortunately, can be persistent in large samples due to the dependence of the true variances on the fixed effects, which are each estimated from only a small portion of the sample.

Though we have not dwelt on this issue in the main text other than to note that it should be kept in mind, the recent literature has suggested that the bootstrap can provide a potential remedy (Pfaffermayr 2021, Zylkin 2024). That is, the same bootstrap we use to construct bias corrections also can be used to obtain more accurate standard errors by taking the standard deviation of the bootstrap estimates. Given that our simulations provide these bootstrap standard errors as a by-product, we report some additional results using these alternative standard errors below in Table A.2. Specifically, we report 95% coverage ratios for our prior simulation results for the bootstrap bias correction now using the standard error estimate produced by the bootstrap instead of the usual analytical standard errors. As we would expect, using these alternative standard errors results in significant improvements in coverage across the board, much moreso than applying bias corrections to the point estimates. Though we do not have much to add to the literature on this subject, it is notable that the bootstrap provides near-perfect bias correction of the standard errors in DGP 1 when PPML is correctly specified with a small amount of noise, though its performance degrades when we either move away from correct specification, as in DGP 2 and/or add more noise, as in the other DGPs.

We can also do the same exercise for our bootstrap simulations from Section 4. In Table A.3, we add to our original bootstrap bias correction results for our real data-based simulations an evaluation of the heteroskedasticity-robust standard errors plus results for coverage using bootstrap standard errors.

Interestingly, the width of robust standard errors is relatively poor, always below 0.8 of the standard deviation in spite of the relatively large number of countries (167). Bootstrap standard errors again improve width and coverage. More importantly for our purposes, when bootstrap standard errors are used, we again find that the bootstrap bias correction modestly improves coverage, as we found in the main text.

Table A.2: Bias, standard deviation, and coverage of standard PPML and bootstrap bias corrected PPML, based on 5000 simulations.

		$\omega_{ij}^{(1)}$		$\omega_{ij}^{(2)}$		$\omega_{ij}^{(3)}$		$\omega_{ij}^{(4)}$	
	N	$\hat{\beta}^{PPML}$	$\hat{\beta}^{BC}$	$\hat{\beta}^{PPML}$	$\hat{\beta}^{BC}$	$\hat{\beta}^{PPML}$	$\hat{\beta}^{BC}$	$\hat{\beta}^{PPML}$	$\hat{\beta}^{BC}$
Bias	50	0.0002	-0.0007	-0.0016	-0.0001	0.0183	-0.0017	-0.0170	-0.0176
	100	0.0002	0.0000	-0.0005	0.0003	0.0078	-0.0007	-0.0093	-0.0058
	150	0.0000	-0.0001	-0.0006	0.0000	0.0020	-0.0030	-0.0098	-0.0060
SD	50	0.0237	0.0239	0.0499	0.0546	0.1736	0.2012	0.2360	0.2854
	100	0.0115	0.0115	0.0263	0.0280	0.0990	0.1116	0.1510	0.1803
	150	0.0075	0.0075	0.0181	0.0190	0.0663	0.0723	0.1139	0.1344
Orig. SE/SD	50	0.9356	0.9297	0.7937	0.7252	0.7262	0.6268	0.6186	0.5115
	100	0.9610	0.9568	0.8684	0.8153	0.7939	0.7048	0.6652	0.5571
	150	0.9772	0.9754	0.8997	0.8554	0.8660	0.7945	0.6953	0.5889
Boot. SE/SD	50	1.0028	0.9965	0.8539	0.7802	0.8254	0.7125	0.7554	0.6246
	100	0.9951	0.9909	0.8809	0.8270	0.8185	0.7266	0.7320	0.6131
	150	1.0000	0.9982	0.8994	0.8551	0.8655	0.7940	0.7304	0.6186
Orig. Cov.	50	0.9320	0.9326	0.8812	0.8540	0.8678	0.8254	0.7880	0.7224
	100	0.9434	0.9414	0.9126	0.8974	0.9054	0.8812	0.8268	0.7782
	150	0.9452	0.9466	0.9252	0.9142	0.9230	0.9078	0.8418	0.8218
Boot. Cov.	50	0.9494	0.9464	0.9086	0.8856	0.9204	0.8864	0.8804	0.8152
	100	0.9498	0.9470	0.9192	0.9042	0.9198	0.8950	0.8778	0.7902
	150	0.9506	0.9518	0.9248	0.9146	0.9276	0.9078	0.8690	0.7948

Note: Superscript PPML refers to uncorrected PPML estimates. Superscript BC refers to bias corrected bootstrap PPML estimates. Column groups identify different error assumptions, based on the notation in the main text. N is sample size. Rows list bias, standard deviation, standard error/standard deviation ratio, and coverage for different combinations of sample size and error assumptions. “Boot. Cov.” And “Boot. SE/SD” use bootstrap standard errors. The true parameter is 1. The standard deviation is the standard deviation across simulations.

Table A.3: Bias, standard deviation, and coverage of standard PPML and bias corrected PPML estimates, using 10000 bootstrap simulations drawn from aggregate trade data (with bootstrap standard errors)

	$ln dist_{ij}$		FTA_{ij}	
	PPML	Boot. BC	PPML	Boot. BC
Average estimate	-0.8154	-0.8029	0.4260	0.4441
Bias	-0.0112	0.0013	-0.0229	-0.0048
SD	0.0388	0.0393	0.0706	0.0718
Orig. SE/SD	0.7379	0.7285	0.7900	0.7764
Boot. SE/SD	0.9747	0.9644	0.9736	0.9571
Orig. Coverage	0.8430	0.8494	0.8654	0.8744
Boot. Coverage	0.9276	0.9344	0.9250	0.9316

Note: Rows provide statistics on each estimation approach. Columns are grouped into estimates of the coefficient on the log of distance, and a dummy for FTA membership. The first column in each group reports uncorrected PPML estimates. The second reports bias corrected bootstrap estimates. The standard deviation is the standard deviation across simulations. Coverage is the share of simulations in which the original estimates are contained in the 95% confidence interval. SE/SD is the ratio of the standard error to the standard deviation. “Boot. SE/SD” and “Boot. Coverage” use bootstrap standard errors.