Bootstrap for Gravity Models*

Tom Zylkin†

Richmond

November 13, 2024

I study the use of bootstrap-based bias corrections in two-way and three-way fixed effects models with unknown heteroskedasticity, focusing on gravity models used with international trade data. I show that, under the cross-sectional independence assumptions that are typically made with gravity models, the distribution of bootstrap estimates can be used to estimate the asymptotic bias of the original estimator. As a by-product, the bootstrap can also provide a less biased estimate of the standard errors. Bootstrap methods therefore offer the potential to improve inferences along two different margins using a single procedure. In practice, however, only the traditional re-sampling bootstrap performs favorably at both of these tasks. The fractional weight bootstrap, in particular, performs poorly and should be avoided. I also verify that computationally efficient approximations of the bootstrap deliver satisfactory results. Because implementing the bootstrap procedure in principle only requires that an appropriate independence assumption is satisfied, it provides a potential remedy for situations in which other bias corrections are not available.

JEL Classification Codes: F10; F14; F15

Keywords: Free trade agreements; Poisson pseudo-maximum Likelihood; Incidental parameter problem

*My thanks to Michael Pfaffermayr, Joao Santos Silva, Amrei Stammann, and conference and seminar participants at the Southern Economic Association meetings, the Gravity, Trade Agreement and Policy workshop in honor of Jeff Bergstrand, and the Online Australasian Seminar in International Economics.

†*Contact information*: Robins School of Business, University of Richmond, Richmond, VA, USA. E-mail: tzylkin@richmond.edu.

1 Introduction

Econometric models with two-way fixed effects are a staple of empirical work in Economics. For standard panel data models, the two fixed effects allow for arbitrary heterogeneity across individuals and arbitrary common trends. For data on spatial flows—such as trade, migration, or commuting—it is common to specify a "gravity model" that includes fixed effects for both the origin and destination. However, for nonlinear models, the presence of these fixed effects can lead to significant biases due to the incidental parameter problem (Neyman and Scott, 1948). A significant literature has emerged proposing and analyzing remedies for these biases (Arellano and Hahn, 2007; Fernández-Val and Weidner, 2016, 2018).

Most often, contributions to the incidental parameter problem literature focus on the case of a correctly specified model estimated via maximum likelihood. However, for the estimation of gravity models—increasingly one of the most common applications of nonlinear models with two-way fixed effects—it is typical to leave the conditional variance unspecified and obtain inferences using "robust" standard error estimates. Recent work has shown these standard error estimates are themselves biased due to the incidental parameter problem (Egger and Staub, 2015; Jochmans, 2017; Pfaffermayr, 2019), and Weidner and Zylkin (2021) have recently shown that both of these issues are salient for improving inferences in this setting. Complicating matters further, it is common in applied work to extend the basic gravity model beyond the settings in which the incidental parameter problem literature provides clear guidance. Though Weidner and Zylkin (2021) study the case of a three-way fixed effects model estimated via Poisson Pseudo-maximum Likelihood (PPML), a popular panel data variant of the gravity model, such clarity is not available for other extensions in current usage, such as when the gravity model is augmented to pool across multiple sectors (French, 2019; Breinlich et al., 2022) or when an imputation-based estimation approach is used to estimate average treatment effects under heterogeneity (Borusyak et al., 2024; Nagengast and Yotov, 2023).

This paper focuses on the usefulness of the bootstrap for improving inferences for two- and three-way gravity models (as well as further variants) estimated via Pseudo-maximum Likelihood under cross-sectional independence. Relative to bias correction methods based on analytical bias formulas or the jackknife, currently the leading approaches in the literature, the bootstrap has the following advantages.[1] First, it offers the potential for refining confi-

[1]Though bias corrections are widely used, another appealing approach is Jochmans (2017)'s GMM estimator for gravity models that constructs moments that are independent of the fixed effects. Yang and Zhang

dence intervals along two different margins using a single procedure rather than having to use different methods for correcting the point estimates and standard errors. The jackknife, for example, cannot be simultaneously applied to both tasks because the order of the bias differs from that of the variance. Second, so long as the cross-sectional independence assumption is satisfied, it is conceptually easy to implement even when the application of other methods is uncertain, needing only an appropriate assumption about the sampling process for re-sampling the data. Third, the computational burden of the standard bootstrap can be greatly reduced by using a "k-step" bootstrap approach without sacrificing efficacy (Kim and Sun, 2016; Davidson and MacKinnon, 1999; Andrews, 2002, 2005).

The idea of the bootstrap as a bias correction method dates back to Efron (1982). Intuitively, the bootstrap re-samples from the observed data as though it is the population. Therefore, it is generally expected to replicate the asymptotic sampling distribution of the estimator, causing each bootstrap estimate to have two times the asymptotic bias of the original estimate. The application of the bootstrap to fixed effects models estimated via pseudo-maximum likelihood has not yet been considered, however, and I show that its effectiveness in this setting requires a surprisingly straightforward proof that sheds new light on why and when the bootstrap works to correct bias. In effect, each bootstrap estimate can be thought of as maximizing a weighted pseudo-likelihood with randomly generated weights. So long as cross-sectional independence holds, the interactions between these weights that arise from a second-order Taylor expansion of the score turn out to double the bias. I also find that bootstrap standard errors do not inherit the bias of the usual "sandwich"-based standard error estimates, confirming earlier results from Pfaffermayr (2021). Generally, my simulations show the bootstrap performs comparably to analytical and jacknife-based methods. In an empirical application, I show that bootstrap methods give near-identical bias corrections and standard errors for the three-way PPML estimator as the analytical corrections recommended in Weidner and Zylkin (2021).

In addition to validating the bootstrap as a viable remedy, I also demonstrate that *how* you bootstrap matters. In my exposition of the bootstrap, I find it useful to nest two common variants of the bootstrap within the same theoretical treatment. As shown in Gotwalt et al. (2018), the traditional re-sampling bootstrap can be thought of as weighting each unit by a randomly generated integer, whereas the "fractional weight" bootstrap, or "Bayesian" bootstrap (Rubin, 1981), allows these weights to be continuous. In the gravity literature, a recent

(2023) derive a similar estimator for three-way gravity models.

contribution by Chowdhry et al. (2023) advocates for the latter method on the basis that each bootstrap trial always preserves the same trade network as the original data, such that the multilateral resistances from structural gravity that are embedded in the fixed effects are always estimated using the same set of partner countries. However, surprisingly, simulations show that the re-sampling bootstrap, not the seemingly more flexible fractional-weight bootstrap, offers superior bias correction performance. Furthermore, standard errors based on the fractional-weight bootstrap often turn out to offer no improvement over using standard uncorrected estimates—and sometimes can even make the inference problem significantly worse.

In the panel data literature, bootstrap corrections for nonlinear models with a single fixed effect have been studied in Kim and Sun (2016) and Higgins and Jochmans (2024). Both of these contributions focus on models estimated via maximum likelihood, though they each provide notable results in this setting. Kim and Sun (2016) demonstrate the asymptotic validity of the parametric bootstrap as a bias correction method and show that a computationally efficient k-step bootstrap procedure retains the same effectiveness as the full bootstrap. Higgins and Jochmans (2024) further show that the parametric bootstrap not only duplicates the bias of the original estimate, it also can be used to simulate the entire distribution of the estimator, such that inferences can be obtained directly from the bootstrap estimates using a percentile method in spite of the bias. In other related work, Hahn et al. (2024) show that the non-parametric bootstrap bias correction has the same higher-order expansion as corrections based on analytical and jackknife methods for cross-sectional models without fixed effects. I differ from these existing contributions by considering a gravity model setting with multiple fixed effects and where the full likelihood is left unspecified. Collectively, these considerations motivate the use of new arguments clarifying both why and when the bootstrap should be expected to double the bias of the original estimator.

In the gravity literature, it is widely recognized that rigorous estimation of the gravity model requires a demanding fixed effects specification. The two-way fixed effects that often features in gravity models—either origin and destination or origin-time and destination-time fixed effects—are derived from theoretical restrictions governing the endogeneity of prices to bilateral trade frictions. Following Baier and Bergstrand (2007), the appropriate extension of the gravity model to a panel data setting involves adding a third fixed effect specific to each pair, in which case the gravity model becomes a "three-way gravity model" on account of its three-way fixed effects structure (origin-time, destination-time, and origin-destination). For the estimation of the effects of trade policies in particular, leading practitioner's guides currently recommend estimating three-way gravity models using PPML (Yotov et al., 2016), and Weidner and Zylkin (2021) have recently clarified how Fernández-Val and Weidner (2016)'s earlier results for the estimation of two-way models carries over to three-way models when the estimator is PPML. However, as noted above, Weidner and Zylkin (2021)'s analysis does not encompass the full set of possibilities that applied researchers may encounter. In principle, the bootstrap approach should be applicable to other estimation settings with two-way fixed effects where the units (either observations or pairs) are cross-sectionally independent.

Pfaffermayr (2021) has previously shown that the bootstrap produces less biased standard error estimates for the two-way PPML estimator used with cross-sectional trade data. Relative to Pfaffermayr (2021), my examination of this issue provides new results regarding the relative performance of the fractional weight bootstrap and k-step bootstrap and evaluates the bootstrap's effectiveness for two estimators not studied in Pfaffermayr (2021)—the Gamma PML two-way gravity estimator and the three-way PPML estimator commonly used to estimate panel data gravity models. More fundamentally, my analysis focuses on the bootstrap as a method for re-centering confidence intervals in addition to estimating their width. Pfaffermayr (2021) focuses only on the latter issue.

The following section documents the different ways in which the incidental parameter problem affects the estimation of gravity models. Section 3 describes the implementation of the bootstrap and explains why, as well as when, it can be used to estimate bias. Section 4 presents simulation results, Section 5 carries out a real data exercise, and Section 6 concludes. An appendix adds further technical details and proofs.

2 Gravity Models and the Incidental Parameter Problem

This section describes gravity models as they are commonly used in the study of international trade and their connection to the incidental parameter problem. Importantly, typical approaches to estimating the gravity model can suffer from two different incidental parameter biases. First, the two fixed effects that feature in gravity models induce an "asymptotic bias" problem where the bias of the estimator converges slowly as compared to its standard error, causing its asymptotic distribution to be incorrectly centered. Second, commonly used "heteroskedasticity robust" estimates of the standard errors are themselves biased, again due to the estimation of the fixed effects.

2.1 Gravity Models

Let N be the number of countries and let i and j be indices for country that respectively index the origin and destination. Adopting the generality of Head and Mayer (2014), a typical gravity model assumes that trade flows from country i to country j are given by

$$y_{ij} = S_i M_j \phi_{ij}. \tag{1}$$

Broadly, S_i may be thought of as capturing supply conditions in country i, M_j correspondingly reflects the overall demand environment in country j, and ϕ_{ij} reflects the bilateral trade costs between i and j. In a more fully articulated trade model, it it is usually made explicit that S_i and M_j depend endogenously on trade costs in general equilibrium through how trade costs affect prices in each country. The trade data may or may not include self-trade ($i = j$) observations. To be consistent with the empirical examples I will use, I will assume that it does not, in which case the number of observations is $N(N-1)$.[2]

The researcher's empirical focus is assumed to be on the determinants of the bilateral component ϕ_{ij}, such as the geographic distance between i and j or whether they have a free trade agreement (FTA). Following Weidner and Zylkin (2021), I parameterize ϕ_{ij} as

$$\phi_{ij} = \exp(x_{ij}\beta)\omega_{ij}, \tag{2}$$

where x_{ij} are the covariates of interest and $\omega_{ij} \geq 0$ reflects the "unobserved" portion of trade costs and is treated as a random disturbance. The covariates x_{ij} are assumed to be exogenous to ω_{ij} after conditioning on the fixed effects. To tie (2) more specifically to an underlying trade model, several such trade models specify $\phi_{ij} = \tau_{ij}^{-\theta}$, where τ_{ij} is the bilateral trade cost and θ is an elasticity parameter reflecting the ease of switching between suppliers. In that case, the coefficient vector β can be interpreted as $-\rho\theta$, where ρ reflects the elasticities of trade costs with respect to each regressor.

To facilitate estimation, the conditional expectation for trade flows can be written as

$$E\left(y_{ij}|x_{ij}, \alpha_i, \gamma_j\right) = \exp\left(\alpha_i + \gamma_j + x_{ij}\beta\right), \tag{3}$$

where α_i and γ_j are origin and destination fixed effects taking the place of S_i and M_j in (1).

[2] It should be noted that trade data sets that include self-trade are increasingly available. For example, the ITPD-E database created by Borchert et al. (2021) includes consistently constructed internal trade values covering 243 countries for the years 2000-2016.

Because of the two fixed effects, and to make a needed distinction in what follows, I will refer to the estimating equation in (3) as a "two-way" gravity model.

A common criticism of the two-way gravity model in (3) is that it is purely cross-sectional in nature and thus leaves a significant amount of variation to the error term. As explained in Baier and Bergstrand (2007), this is especially a problem for evaluating the effects of FTAs and other similar variables because country pairs that are more likely to form these agreements may have systematically different initial trade cost levels than other pairs, leading to biased estimates. Consequently, researchers interested in quantifying effects of trade policy changes often use a panel data version of the gravity model with an added time dimension. Letting $t \in 1...T$ denote the time period, trade costs in this case are specified as $\phi_{ijt} = \exp(\eta_{ij} + x_{ijt}\beta)\omega_{ijt}$,with η_{ij} an added "pair" fixed effect that accounts for all time-invariant trade costs for each country-pair. The resulting "three-way" gravity model is then

$$E\left(y_{ijt}|x_{ijt}, \alpha_{it}, \gamma_{jt}\right) = \exp\left(\alpha_{it} + \gamma_{jt} + \eta_{ij} + x_{ijt}\beta\right), \quad (4)$$

where η_{ij} is the added pair fixed effect and all other terms now have an added t subscript. The α_{it} and γ_{jt} parameters therefore may be thought of as absorbing changes over time in the "gravitational pull" each country exerts on trade, such that the empirical focus remains on within-pair variation.

Finally, following typical conventions in the gravity literature, I assume that trade flows are cross-sectionally independent from one another conditional on the fixed effects and regressors. For two-way gravity models applied to a single year, this simply means that the error terms of different observations are independent of one another. For panel data, observations belonging to the same ij pair are allowed to be weakly correlated due to serial correlation.[3]

2.2 Biased Inference

Because gravity models usually assume an exponential conditional mean, it has become conventional in recent decades to estimate them using pseudo-maximum likelihood (PML), following the recommendations of Santos Silva and Tenreyro (2006). In the absence of fixed

[3]Like in Weidner and Zylkin (2021), the analysis to be presented here does still extend to mild violations of the cross-sectional independence assumption, such as when both directions of trade within the same pair are correlated with one another. However, the bias expressions to be presented in the next subsection are not valid if, for example, there is correlation across pairs sharing the same exporter and/or correlation across pairs sharing the same importer. Furthermore, as I document in the Appendix, it turns out the bootstrap fails to provide a consistent estimate of the bias in these cases.

effects, PML estimators have the advantage that they only require correct specification of the conditional mean for consistency.

However, because gravity models usually feature fixed effects, standard inference is complicated by the incidental parameter problem, whereby estimation noise in the estimated fixed effects biases the estimates of the other parameters. In the two-way gravity model, for example, it is important to recognize that the fixed effects estimates $\widehat{\alpha}_i$ and $\widehat{\gamma}_j$ are each only estimated off of $N-1$ observations—the trade partners of each country—rather than the full set of $N(N-1)$ observations. $\widehat{\alpha}_i$ and $\widehat{\gamma}_j$ are thus estimated consistently as $N \to \infty$ but at a slower-than-usual rate, and their slow rate of convergence affects the rate at which the bias of the score for β converges.

To add further intuition, suppose we knew the true parameter values β^0, α^0, and γ^0 for the two-way gravity model. Letting ℓ_{ij} denote the (log-)pseudolikelihood of a single observation, the score function evaluated at these true parameters has expectation zero:

$$\mathbb{E}\left[\sqrt{N(N-1)}\frac{\partial \ell_{ij}}{\partial \beta}\left(\beta^0, \alpha_i^0, \gamma_i^0\right)\right] = 0,$$

However, if we use only the true value for β^0, and let the PML estimator determine the estimated values for the fixed effects using their first-order conditions, then the resulting "profile score" no longer has expectation zero:

$$\mathbb{E}\left[\sqrt{N(N-1)}\frac{\partial \ell_{ij}}{\partial \beta}\left(\beta^0, \widehat{\alpha}_i(\beta^0), \widehat{\gamma}_j(\beta^0)\right)\right] \neq 0.$$

The bias occurs because the estimation noise in the fixed effects estimates $\widehat{\alpha}_i$ and $\widehat{\gamma}_j$ generally enters this expectation nonlinearly—in essence, a Jensen's inequality issue.[4] Because the score for β is biased, it follows directly that the PML estimate $\widehat{\beta}$ will be biased. Moreover, it can be shown that commonly used "heteroskedasticity robust" estimates of the standard errors, which use the square of the estimated score, are biased as well. Setting aside the latter issue for now, the slow convergence of the fixed effects estimates implies the bias in $\widehat{\beta}$ will be of order $1/N$ instead of the usual $1/N^2$, such that it has the same order as the standard error. As demonstrated in Fernández-Val and Weidner (2016), PML estimators of gravity models should therefore be expected to have an "asymptotic bias" problem, where the estimator achieves consistency but the asymptotic distribution of estimates nonetheless

[4]For example, for the Gamma PML estimator, the profile score evaluated at β^0 is $y_{ij}/e^{\widehat{\alpha}_i(\beta^0)+\widehat{\gamma}_j(\beta^0)+x_{ij}\beta^0} - 1$. Because $\mathbb{E}[1/e^{\widehat{\alpha}_i(\beta^0)+\widehat{\gamma}_j(\beta^0)}] \neq \mathbb{E}[1/e^{\alpha_i^0+\gamma_i^0}]$, the Gamma PML profile score is biased.

fails to be correctly centered even in large samples.[5]

To provide more precise illustration, I draw from the exposition of Fernández-Val and Weidner (2016, 2018). Continuing to let $\ell_{ij}(\beta, \alpha_i, \gamma_j)$ be the pseudolikelihood of a given observation, let the estimates $\widehat{\beta}$, $\widehat{\alpha}$, and $\widehat{\gamma}$ be given by

$$(\widehat{\beta}, \widehat{\alpha}, \widehat{\gamma}) := \arg\max_{\beta, \alpha\gamma} \; \mathcal{L}(\beta, \alpha, \gamma) = \sum_{i,j} \ell_{ij}(\beta, \alpha_i, \gamma_j),$$

where $\widehat{\alpha} := (\widehat{\alpha}_1, ..., \widehat{\alpha}_N)$ and $\widehat{\gamma} := (\widehat{\gamma}_1, ..., \widehat{\gamma}_N)$ are vectors collecting the fixed effects estimates. Further define $\mathcal{L}^*$ and ℓ^*_{ij} as the "information orthogonal" versions of $\mathcal{L}$ and ℓ_{ij} in the sense described by Fernández-Val and Weidner (2018). In practice, this step involves re-parameterizing the model so that x_{ij} is orthogonalized with respect to the fixed effects. Like in Weidner and Zylkin (2021), this orthogonalized version of x_{ij} will be given by $\widetilde{x}_{ij}$. When convenient, superscripts or leading ∂ symbols will be used to denote partial derivatives of the pseudolikelihood, e.g., $\ell^{\beta}_{ij} = \partial_{\beta}\ell_{ij} = \partial\ell_{ij}/\partial\beta$, $\ell^{\alpha_i\alpha_i}_{ij} = \partial_{\alpha_i\alpha_i}\ell_{ij} = \partial^2\ell_{ij}/\partial\alpha_i\partial\alpha_i$. Finally, let a "bar" serve as shorthand for an expectation, e.g., $\overline{\ell}^*_{ij} = \mathbb{E}(\ell^*_{ij})$.[6]

The expansion of $\widehat{\beta} - \beta^0$ for the two-way gravity model, including its asymptotic bias, is then

$$(\widehat{\beta} - \beta^0) \approx \frac{1}{N(N-1)} H_N^{-1} \sum_{i,j} \ell^{*\beta}_{ij} + \frac{1}{N-1} H_N^{-1} \left(B_N + D_N\right)$$

where

$$H_N = -\frac{1}{N\left(N-1\right)} \sum_{i,j=1}^{N} \overline{\ell}^{*\beta\beta}_{ij}$$

[5]See Weidner and Zylkin (2021), Section 2, for a taxonomic discussion of inconsistency versus asymptotic bias. A concise illustration is provided in figure 1 of that paper.

[6]What this accomplishes is making it so the asymptotic variance of $\widehat{\beta}$ depends on the second derivatives of $\overline{\ell}^*_{ij}$ with respect to β only rather than depending on the full Hessian. In the examples that follow, the way in which $\widetilde{x}_{ij}$ is obtained depends on the estimator. For PPML, the elements of $\widetilde{x}_{ij}$ are the residuals from regressing each $\widetilde{x}_{ij,k}$ on i- and j-specific fixed effects and weighting by the conditional mean. For Gamma PML the procedure is similar except there is no weighting.

is the expected Hessian and B_N and D_N are vectors with individual elements given by

$$B_N^m = -\frac{1}{N}\sum_{i=1}^{N}\frac{\sum_j \mathbb{E}\left(\ell_{ij}^{*\beta_m\alpha_i}\,\ell_{ij}^{\alpha_i}\right)}{\sum_{j\neq i}^{N}\overline{\ell}_{ij}^{\alpha_i\alpha_i}} + \frac{1}{2\,N}\sum_{i=1}^{N}\frac{\left(\sum_{j\neq i}\overline{\ell}_{ij}^{*\beta_m\alpha_i\alpha_i}\right)\left[\sum_{j\neq i}\mathbb{E}\left(\ell_{ij}^{\alpha_i}\,\ell_{ij}^{\alpha_i}\right)\right]}{\left(\sum_{j}^{N}\overline{\ell}_{ij}^{\alpha_i\alpha_i}\right)^2}$$

$$D_N^m = -\frac{1}{N}\sum_{j=1}^{N}\frac{\sum_i \mathbb{E}\left(\ell_{ij}^{*\beta_m\gamma_j}\,\ell_{ij}^{\gamma_j}\right)}{\sum_{i\neq j}\overline{\ell}_{ij}^{\gamma_j\gamma_j}} + \frac{1}{2\,N}\sum_{j=1}^{N}\frac{\left(\sum_{i\neq j}\overline{\ell}_{ij}^{*\beta_m\gamma_j\gamma_j}\right)\left[\sum_{i\neq j}\mathbb{E}\left(\ell_{ij}^{\gamma_j}\,\ell_{ij}^{\gamma_j}\right)\right]}{\left(\sum_{i\neq j}\overline{\ell}_{ij}^{\gamma_j\gamma_j}\right)^2}. \tag{5}$$

Notably, as $N \to \infty$, the score bias terms B_N and D_N do not themselves vary with N, confirming that the bias in $\widehat{\beta}$ is of order $1/N$. From the way these formulas are written, we can see B_N reflects the bias contribution of the estimation noise in $\widehat{\alpha}_i$, while D_N reflects the contribution of the noise in $\widehat{\gamma}_j$.

To give two examples to be used throughout the paper, I turn now to applying the above bias formulas to the Poisson PML (PPML) and Gamma PML (GPML) estimators. PPML is by far the most popular PML estimator in this context, but Head and Mayer (2014) recommend also using GPML as a diagnostic in their influential handbook chapter. Unlike with PPML, the bias properties of GPML have not received much attention, and, as the following discussion shows, it makes for an interesting contrast between the two.

Example 1 (Two-way Poisson PML) For PPML, Fernández-Val and Weidner (2016) document that $\mathbb{E}(\ell_{ij}^{*\beta_m\alpha_i}\,\ell_{ij}^{\alpha_i}) = \mathbb{E}[-\widehat{\mu}_{ij}\widetilde{x}_{ij}\,(y_{ij} - \widehat{\mu}_{ij})] = 0.$ and $\sum_{i\neq j}\overline{\ell}_{ij}^{*\beta_m\gamma_j\gamma_j} = \sum_{i\neq j}\mu_{ij}\widetilde{x}_{ij,m} = 0.$ Therefore, $B_N^m = D_N^m = 0$ and PPML is asymptotically unbiased.

Example 2 (Two-way Gamma PML) For GPML, we have that $\ell_{ij}^{\alpha_i} = y_{ij}/\mu_{ij} - 1$, $\ell_{ij}^{*\beta_m} = (y_{ij}/\mu_{ij} - 1)\widetilde{x}_{ij,m}$, $\ell_{ij}^{*\beta_m\alpha_i} = -(y_{ij}/\mu_{ij})\widetilde{x}_{ij,m}$, $\ell_{ij}^{\beta_m\alpha_i\alpha_i} = (y_{ij}/\mu_{ij})\widetilde{x}_{ij,m}$, $\overline{\ell}_{ij}^{\beta_m\alpha_i\alpha_i} = \widetilde{x}_{ij,m}$, and $\overline{\ell}_{ij}^{\alpha_i\alpha_i} = -1$. Then the α-specific bias term B_N^m can be simplified to

$$B_N^m = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j\neq i}\frac{1}{N-1}\left(\mathbb{E}(y_{ij}^2)/\mu_{ij}^2 - 1\right)\widetilde{x}_{ij,m},$$

and an analogous expression holds for D_N^m after exchanging i and j in the two summations.

Both the no-bias results of PPML and the fact that it is highly unique in this regard are generally known. The fact that GPML generally has an asymptotic bias is therefore not surprising. What is interesting about GPML is that the degree of bias comes directly from the degree of misspecification. If the variance of y_{ij} is proportional to μ_{ij}^2, as is assumed by GPML, all bias terms turn out to be zero.

Bias of robust standard errors

Though PPML with two-way fixed effects turns out to not have an asymptotic bias problem, this does not mean that inference using PPML is unaffected by the incidental parameter problem. For PML estimation, unlike with maximum likelihood estimation, confidence intervals and hypothesis tests are typically constructed using "heteroskedasticity robust" estimates of the variance matrix that correct for misspecification of the assumed variance. For gravity models, the bias of these standard errors can be significant even in large samples due to the slow convergence of the fixed effects estimates.

Intuitively, under cross-sectional independence and heteroskedasticity, the asymptotic variance of $\widehat{\beta}$ is given by

$$Var(\widehat{\beta}-\beta^0)=\frac{1}{N(N-1)}H_N^{-1}\Omega_N H_N^{-1}, \qquad \Omega_N := \sum_{i,j} S_{ij}^2 \widetilde{x}_{ij}' \widetilde{x}_{ij},$$

where $S_{ij} := \ell_{ij}^{\alpha_i} = \ell_{ij}^{\gamma_j}$. For example, for PPML, $S_{ij} = y_{ij} - \mu_{ij}$. The standard heteroskedasticity-corrected variance estimator (henceforth HC1) is then formed by using the square of the estimated scores $\widehat{S}_{ij}^2$ as a plugin estimate for S_{ij}^2, i.e.,

$$\widehat{Var}_{HC1}(\widehat{\beta}-\beta^0)=\frac{1}{N(N-1)-1}\widehat{H}_N^{-1}\widehat{\Omega}_N \widehat{H}_N^{-1}, \qquad \widehat{\Omega}_N := \sum_{i,j} \widehat{S}_{ij}^2 \widetilde{x}_{ij}' \widetilde{x}_{ij}.$$

Since the estimation noise in the fixed effects necessarily enters the estimated squared term $\widehat{S}_{ij}^2$ in a nonlinear fashion, it dominates the bias of the HC1 variance estimator and causes it to vanish at a rate of $1/N$ as $N \to \infty$ rather than the usual rate of $1/N^2$, similar to how the fixed effects affect the bias of the point estimates.[7] More formally, a first-order expansion of $\widehat{S}_{ij}$ under the assumption of correct specification of the PML model variance can be used to show that

$$\mathbb{E}\left[\widehat{S}_{ij}^2\right] \approx S_{ij}^2 - \underbrace{O(1/N^2) \text{ bias term}}_{\text{from noise in } \widehat{\beta}} - \underbrace{O(1/N) \text{ bias term}}_{\text{from noise in FEs}},$$

[7] Though the HC1 estimator is generally known to be downward biased in finite samples, it is nonetheless the most commonly used "robust" variance estimator in the literature and serves a default implementation for most regression commands in Stata. In most applications, the bias of HC1 will be negligible for a sufficiently large number of observations. However, PML estimates of gravity models have proven to be an exception. The name "HC1" comes from MacKinnon and White (1985). Some alternatives include HC0, which does not include a degrees of freedom correction, and HC2, which includes a correction for the bias in $\widehat{S}_{ij}^2$. Weidner and Zylkin (2021) supply HC2-type corrections for the two-way and three-way PPML gravity estimators.

such that $\widehat{Var}_{HC1}(\widehat{\beta}-\beta^0)$ should be expected to have a $1/N$ bias as $N \to \infty$. The details behind this derivation are shown in Weidner and Zylkin (2021)'s appendix. Importantly, the bias is generally downward, implying that estimated confidence intervals will be too narrow in addition to being off-center.

2.3 Three-way Gravity Models

For the three-way gravity model in (4), we might expect that the addition of the pair fixed effect leads to a more complicated set of results. However, Weidner and Zylkin (2021) show that, when the estimator is PPML, the three-way model can be recast as a two-way multinomial model where the only incidental parameters are α_{it} and γ_{jt}. It remains to be the case that these parameters are only estimated off of $N-1$ observations each. Thus, for fixed T at while $N \to \infty$, PPML with three-way fixed effects is consistent but has the same order-$1/N$ asymptotic bias problems as though it were a two-way estimator, including for its estimated "robust" variance matrix, which in this case would be the cluster-robust (CR) variance matrix.

For the three-way gravity model, PPML estimates are obtained using

$$(\widehat{\beta},\widehat{\alpha},\widehat{\gamma},\widehat{\eta}) := \arg\max_{\beta,\alpha\gamma,\eta}\ \mathcal{L} = \sum_{i,j,t}\ell_{ijt} = \sum_{i,j,t} y_{ijt}\log\mu_{ijt} - \mu_{ijt} \qquad \text{where} \quad \mu_{ijt} := e^{\alpha_{it}+\gamma_{jt}+\eta_{ij}+x_{ijt}\beta}.$$

To make this setting conform more closely to the previous discussion, Weidner and Zylkin (2021) show that it is possible to re-cast the three-way PPML estimator as a two-way multinomial estimator where α_{it} and γ_{jt} are the only fixed effects that need to be estimated. In particular, an equivalent estimator for (β,α,γ) is given by

$$\left(\widehat{\beta},\widehat{\alpha},\widehat{\gamma}\right) = \arg\max_{\beta,\alpha,\gamma}\mathcal{L} := \sum_{i,j}\ell_{ij} = \sum_{i,j}\sum_{t=1}^{T} y_{ijt}\ln\left(\frac{\exp\left[\alpha_{it}+\gamma_{jt}+x'_{ijt}\beta\right]}{\sum_{s=1}^{T}\exp\left[\alpha_{it}+\gamma_{jt}+x'_{ijt}\beta\right]}\right), \tag{6}$$

which is derived by profiling out the pair fixed effect η using its PPML first-order condition.

Because the multinomial likelihood belongs to the exponential family, the resulting two-way estimator in (6) is consistent as $N \to \infty$ even as the time dimension remains fixed. It does not, however, yield the same no-asymptotic bias property as the two-way PPML estimator highlighted in the previous subsection. Because $\alpha_i := (\alpha_{i1},....,\alpha_{iT})$ and $\gamma_j := (\gamma_{j1},....,\alpha_{jT})$ are each now T-vectors instead of scalars, the appropriate bias terms in this case are given by

$$B_N^m = -\frac{1}{N}\sum_{i=1}^{N}\text{Tr}\left[\left(\sum_{j\neq i}^{N}\overline{\ell}_{ij}^{\alpha_i\alpha_i}\right)^{-1}\sum_{j\neq i}\mathbb{E}\left(\ell_{ij}^{*\beta_m\alpha_i}\,\ell_{ij}^{\alpha_i\prime}\right)\right]$$
$$+\frac{1}{2N}\sum_{i=1}^{N}\text{Tr}\left[\left(\sum_{j\neq i}^{N}\overline{\ell}_{ij}^{\alpha_i\alpha_i}\right)^{-1}\left(\sum_{j\neq i}\overline{\ell}_{ij}^{*\beta_m\alpha_i\alpha_i}\right)\left[\sum_{j\neq i}\mathbb{E}\left(\ell_{ij}^{\alpha_i}\,\ell_{ij}^{\alpha_i\prime}\right)\right]\left(\sum_{j\neq i}^{N}\overline{\ell}_{ij}^{\alpha_i\alpha_i}\right)^{-1}\right],$$
$$D_N^m = -\frac{1}{N}\sum_{j=1}^{N}\text{Tr}\left[\left(\sum_{i\neq j}^{N}\overline{\ell}_{ij}^{\gamma_j\gamma_j}\right)^{-1}\sum_{i\neq j}\mathbb{E}\left(\ell_{ij}^{*\beta_m\gamma_j}\,\ell_{ij}^{\gamma_j\prime}\right)\right]$$
$$+\frac{1}{2N}\sum_{j=1}^{N}\text{Tr}\left[\left(\sum_{i\neq j}^{N}\overline{\ell}_{ij}^{\gamma_j\gamma_j}\right)^{-1}\left(\sum_{i\neq j}\overline{\ell}_{ij}^{*\beta_m\gamma_j\gamma_j}\right)\left[\sum_{i\neq j}\mathbb{E}\left(\ell_{ij}^{\gamma_j}\,\ell_{ij}^{\gamma_j\prime}\right)\right]\left(\sum_{i\neq j}^{N}\overline{\ell}_{ij}^{\gamma_j\gamma_j}\right)^{-1}\right],$$

which has the same form as the bias of two-way PML estimators except terms such as $\ell_{ij}^{\alpha_i}$ and $\ell_{ij}^{*\beta_m\alpha_i}$ are now T-vectors instead of scalars and terms such as $\overline{\ell}_{ij}^{\alpha_i\alpha_i}$ and $\overline{\ell}_{ij}^{*\beta_m\alpha_i\alpha_i}$ are now $T \times T$ matrices. The asymptotic bias in the distribution of $\widehat{\beta}$ continues to be given by $N^{-1}H_N^{-1}(B_N + D_N)$, with H_N still given by the definition provided before. Unlike for the two-way PPML estimator, Weidner and Zylkin (2021) show that B_N and D_N are generally non-zero.

To illustrate the bias of the cluster-robust standard errors, the cross-sectional independence assumption in this case gives us

$$Var(\widehat{\beta} - \beta^0) = \frac{1}{N(N-1)}H_N^{-1}\Omega_N H_N^{-1}, \qquad\qquad \Omega_N := \sum_{i,j}\widetilde{x}_{ij}'S_{ij}S_{ij}'\widetilde{x}_{ij},$$

with the only needed modification from the previous variance formula being that S_{ij} and $\widetilde{x}_{ij}$ are now T-vectors instead of scalars. The bias problem in this case emerges because the standard cluster-robust variance estimator (henceforth CR1) uses the outer product of the estimated scores $\widehat{S}_{ij}\widehat{S}_{ij}'$ as a plugin estimate for the true score product $S_{ij}S_{ij}'$. Analogously to what happens with the HC1 variance estimator for two-way gravity models, Weidner and Zylkin (2021) show that the noise in the α- and γ-fixed effects estimates dominates the bias of the $\widehat{S}_{ij}\widehat{S}_{ij}'$ term, causing the order of the bias of the CR1 variance estimator to be $1/N$ instead of $1/N^2$.

3 Improving Inference using the Bootstrap

When referring to "the" bootstrap, it is important to acknowledge that the term "bootstrap" can refer to many different procedures: parametric versus non-parametric, re-sampling versus fractional weighting, "wild" bootstrapping versus more restrained varieties. Parametric bootstrap methods are generally not an option for PML estimators because PML estimation does not fully specify the conditional distribution of the data. Furthermore, as I will document, the Kline et al. (2018) "wild score" bootstrap is not suitable for estimating bias. Thus, in what follows, I will focus on non-parametric bootstrapping methods, using a generalized approach that encompasses both re-sampling and fractional weighting.

3.1 Implementing the Bootstrap

To describe both re-sampling and fractional weighting within the same framework, the approach I will take is the following. For each bootstrap replication $b \in 1, ..., B$, the bootstrap estimates $\widehat{\beta}^{(b)}$,$\widehat{\alpha}^{(b)}$, and $\widehat{\gamma}_j^{(b)}$ are obtained using

$$(\widehat{\beta}^{(b)}, \widehat{\alpha}^{(b)}, \widehat{\gamma}^{(b)}) := \arg\max_{\beta,\alpha\gamma} \ \mathcal{L}^{(b)}(\beta, \alpha, \gamma) = \sum_{i,j} W_{ij}^{(b)} \ell_{ij}(\beta, \alpha_i, \gamma_j), \tag{7}$$

where each $W_{ij}^{(b)}$ is randomly and independently generated weighting parameter. In the case of the traditional re-sampling bootstrap, each $W_{ij}^{(b)}$ is a randomly and independently generated non-negative integer such that $\sum_{i,j} W_{ij}^{(b)} = N(N-1)$ for all b. If $W_{ij}^{(b)} = 0$, pair ij is not included in the estimation because it receives no weight; if $W_{ij}^{(b)} \geq 2$, it is "as though" it has been sampled more than once, but one could equivalently infer that it is being given a larger weight than in the original estimation.[8] Though this formulation is written explicitly for the case of two-way fixed effects, we can adapt the above for the three-way PPML case by letting $\ell_{ij} = \sum_t \ell_{ijt}$ like in Weidner and Zylkin (2021). The bootstrap in this latter case becomes a cluster bootstrap that assumes clustering by pair.

An important point about the implementation of the bootstrap in (7) is that it does not require $W_{ij}^{(b)}$ to be an integer. In general, each $W_{ij}^{(b)}$ needs to be non-negative and satisfy $\mathbb{E}(W_{ij}^{(b)}) = 1$. Accordingly, the "fractional weight" bootstrap provides a more flexible alternative in which each $W_{ij}^{(b)}$ is instead drawn independently from a continuous distribution.

[8] What I call the re-sampling bootstrap, also commonly known as the "pairs" bootstrap, dates back to Freedman (1981).

Typical implementations of the bootstrap effectively restrict the weights to add up to the sample size—here, $\sum_{i,j} W_{ij}^{(b)} = n = N(N-1)$ for all $b \in 1, ..., B$. For the integer case, the weights are therefore drawn from a multinomial distribution; when the weights are continuous, they are drawn from a Dirichlet distribution. In either case, the stipulation that $\mathbb{E}(W_{ij}^{(b)}) = 1$ guarantees that $Var(W_{ij}^{(b)}) \to 1$ as n becomes large. For the multinomial distribution, $Var(W_{ij}^{(b)}) = (n-1)/n$. For the Dirichlet distribution, $Var(W_{ij}^{(b)}) = (n-1)/(n+1)$. Because trade data sets generally have large n, I will henceforth take $Var(W_{ij}^{(b)})$ to be 1. Allowing $Var(W_{ij}^{(b)})$ to instead depend on n does not affect any of the theoretical results I will present.[9]

From the elementary properties of random variables, if we maintain that $\mathbb{E}(W_{ij}^{(b)}) = Var(W_{ij}^{(b)}) = 1$, it then holds that

$$\mathbb{E}\left[(W_{ij}^{(b)})^2\right] = Var(W_{ij}^{(b)}) + \left[\mathbb{E}(W_{ij}^{(b)})\right]^2 = 2.$$

This simple observation turns out to crucial to why the bootstrap doubles the bias of the original estimator.

3.2 Why (and When) Bootstrapping Doubles the Bias

Recall from the previous section that the bias of PML fixed effects estimators comes derives from a complex expression that involves the partial derivatives and higher-order derivatives of the likelihood. Consider in particular the bias of the score of the two-way estimator presented in (5). If the asymptotic bias of the original estimator is given by (5), then, for a given set of randomly generated bootstrap weights, the bias of the weighted estimator described in (7) as $N \to \infty$ is $(N-1)^{-1} H_N^{-1} \left(B_N^{(b)} + D_N^{(b)}\right)$, where each element of $B_N^{(b)}$ may be written as

$$B_N^{(b),m} = -\frac{1}{N}\sum_{i=1}^{N} \frac{\sum_{j\neq i}\left(W_{ij}^{(b)}\right)^2 \mathbb{E}\left(\ell_{ij}^{*\beta_m\alpha_i}\ell_{ij}^{\alpha_i}\right)}{\sum_{j\neq i}^{N} W_{ij}^{(b)}\overline{\ell}_{ij}^{\alpha_i\alpha_i}} + \frac{1}{2N}\sum_{i=1}^{N} \frac{\left(\sum_{j\neq i} W_{ij}^{(b)}\overline{\ell}_{ij}^{*\beta_m\alpha_i\alpha_i}\right)\left[\sum_{j\neq i}\left(W_{ij}^{(b)}\right)^2 \mathbb{E}\left(\ell_{ij}^{\alpha_i}\ell_{ij}^{\alpha_i}\right)\right]}{\left(\sum_{j\neq i}^{N} W_{ij}^{(b)}\overline{\ell}_{ij}^{\alpha_i\alpha_i}\right)^2},$$

with a corresponding expression following for each element of $D_N^{(b)}$, only interchanging j with i and α_i with γ_j where appropriate. The derivation of this expressions follows from letting

[9] This is because the bias expansions from Fernández-Val and Weidner (2016) and Weidner and Zylkin (2021) are valid under an asymptotic where $N \to \infty$, which is sufficient to ensure that $Var(W_{ij}^{(b)}) \to 1$. Alternatively, one could draw n random weighting variables that always satisfy $E(W_{ij}^{(b)}) = Var(W_{ij}^{(b)}) = 1$ by removing the restriction that $\sum_{i,j} W_{ij}^{(b)} = n$.

$\ell_{ij}^{(b)} = W_{ij}^{(b)} \ell_{ij}$, where ℓ_{ij} is the original pseudolikelihood, and following the same steps as in Fernández-Val and Weidner (2016) with $\ell_{ij}^{(b)}$ in place of ℓ_{ij}. Whereever a derivative of ℓ_{ij} appears in the original score bias, a $W_{ij}^{(b)}$ now also appears in the score bias of the weighted estimator. Necessarily, this means that all terms of the form $\mathbb{E}\left(\ell_{ij}^{*\beta_k \alpha_i}\, \ell_{ij}^{\alpha_i}\right)$, $\mathbb{E}\left(\ell_{ij}^{\alpha_i}\, \ell_{ij}^{\alpha_i}\right)$, etc. are now multiplied by $(W_{ij}^{(b)})^2$.

The next step is to recognize that, as $N \to \infty$, $(N-1)^{-1} \sum_{j \neq i} W_{ij}^{(b)}$ converges to $\mathbb{E}\left(W_{ij}^{(b)}\right)$; likewise, $\lim_{N\to\infty}(N-1)^{-1} \sum_{j\neq i}(W_{ij}^{(b)})^2 = \mathbb{E}[(W_{ij}^{(b)})^2]$. Recalling that we have stipulated that $\mathbb{E}(W_{ij}^{(b)}) = 1$ while $\mathbb{E}[(W_{ij}^{(b)})^2] = 2$, and using the fact that $W_{ij}^{(b)}$ is independent of ℓ_{ij} and all of its derivatives, we then have the following result:

Proposition 1 *Suppose that, for an estimator of a two-way fixed effects model with its pseudolikelihood described by ℓ_{ij}, the bias of the score of the estimator is given by (5). Furthermore, suppose that we have a set of weights for each observation satisfying the following properties: (i) $\mathbb{E}(W_{ij}^{(b)}) = Var(W_{ij}^{(b)}) = 1$; (ii) $\mathbb{E}(W_{ij}^{(b)} W_{i'j'}^{(b)}) = 1$ if $i \neq i'$ or $j \neq j'$;(iii) $W_{ij}^{(b)}$ is independent of ℓ_{ij} and all of its derivatives. Then the weighted estimator that maximizes $\sum_{i,j} W_{ij}^{(b)} \ell_{ij}$ has an asymptotic bias equal to twice that of the original unweighted estimator as $N \to \infty$.*

The appendix provides more details on the derivation of the bias of the bootstrap-weighted estimator. Simply put, terms from the original bias that now depend on $(W_{ij}^{(b)})^2$ effectively become multiplied by 2. Because each bootstrap replication turns out to double the bias of the original estimator as $N \to \infty$, it follows that the average bootstrap estimate $\overline{\beta}^B := B^{-1} \sum_b \widehat{\beta}^{(b)}$ also has twice the bias of the original estimate $\widehat{\beta}$. The bootstrap bias-corrected estimate is then given by

$$\widehat{\beta}^{BBC} := 2\widehat{\beta} - \overline{\beta}^B.$$

Moreover, the asymptotic bias of $\widehat{\beta}$ can be estimated using $\overline{\beta}^B - \widehat{\beta}$. More formally, the bias of $\overline{\beta}^B$ is given by

$$\mathbb{E}(\overline{\beta}^B - \beta^0) = \frac{2}{N-1} H_N^{-1} (B_N + D_N) + o(N^{-1}) = 2 \times \mathbb{E}(\widehat{\beta} - \beta^0) + o(N^{-1}),$$

where $o(N^{-1})$ indicates terms of order smaller than N^{-1}. Therefore, the bias-corrected estimate $\widehat{\beta}^{BBC}$ is asymptotically unbiased as $N \to \infty$, in the sense that its bias decreases faster than its standard deviation. Furthermore, if we also allow the number of bootstrap trials B to become large, we obtain the following result:

Proposition 2 *Suppose that the asymptotic bias of the score of a two-way fixed effects estimator is given by (5). Further, suppose that, for each bootstrap trial $b = 1, ..., B$, the set of bootstrap weights $\{W_{ij}^{(b)}\}$ satisfy the conditions described in Proposition 1. Then the asymptotic distribution of the bootstrap bias-corrected estimate $\widehat{\beta}^{BBC} = 2\widehat{\beta} - \overline{\beta}^B$ as both $B, N \to \infty$ is described by*

$$\sqrt{N(N-1)}\widehat{\beta}^{BBC} \to \mathcal{N}(0, H_N^{-1}\Omega_N H_N^{-1}).$$

That is, $\widehat{\beta}^{BBC}$ is asymptotically unbiased and has the same asymptotic variance as the original estimate $\widehat{\beta}$.

This proposition makes it clear that re-centering estimates using the bootstrap should not otherwise affect their asymptotic distributions, similar to what has been found for jackknife and analytical bias corrections. However, in practice, bias-corrected estimates should be expected to have at least some additional variance relative to uncorrected estimators in finite samples.

At this point, it is worth clarifying how this approach to demonstrating the ability of the bootstrap to estimate bias differs from those used elsewhere in the literature. The approach used by Hahn et al. (2024), in particular, could conceivably be extended to this setting to demonstrate the results given thus far, at least for the re-sampling bootstrap. In their analysis of the bootstrap, which applies only to settings without fixed effects, the bias of the bootstrap estimates is shown by expanding $\overline{\beta}^B$ around $\widehat{\beta}$ instead of β^0 and then by proving that $\overline{\beta}^B - \widehat{\beta}$ converges to the same asymptotic distribution as $\widehat{\beta} - \beta^0$. This differs from the approach I have used, which instead expands each individual bootstrap estimate $\widehat{\beta}^{(b)}$ around β^0 in order to show how the weighting used by the bootstrap inflates its bias.

Though Hahn et al. (2024)'s treatment of the bootstrap is intuitively appealing, being consistent with Efron (1982)'s original explanation for how the bootstrap amplifies bias, it is known that this logic does not extrapolate to the full spectrum of settings where a researcher may wish to apply a bias correction.[10] Indeed, an advantage of my own approach is that it can transparently be used to determine how bootstrapping impacts the bias in fairly general cases,

[10]For example, it is known that the re-sampling bootstrap does not correct the bias in dynamic panel data models (Gonçalves and Kaffo 2015) or in panel data models with serial correlation in the time dimension (Kaffo 2015, ch. 2).

requiring only a generic characterization of the form of the bias such as the one in (5). For two-way fixed effects models with uncorrelated errors, my results demonstrate the bootstrap does have the expected effect of doubling the bias. However, for other settings, matters can differ. In the appendix, I give two simple examples: a panel data model with weak time dependence and a gravity model in which the cross-sectional independence assumption is violated. In both examples, the dependence of the error structure means that the bias now depends on how the scores of different observations can interact with one another, whereas under independence the bias only depends on terms such as $\mathbb{E}(\ell_{ij}^{*\beta_k\gamma_j}\,\ell_{ij}^{\gamma_j})$ and $\mathbb{E}(\ell_{ij}^{\gamma_j}\,\ell_{ij}^{\gamma_j})$. Because different observations will have different bootstrap weights, the bootstrap fails to double the bias when the expected score product of any two observations sharing the same fixed effect is non-zero. Furthermore, using a cluster bootstrap that treats dependent observations as belonging to the same cluster does not provide a valid bias correction for these cases either.

3.3 The k-step bootstrap

One of the known disadvantages of the bootstrap relative to using analytical methods is that it requires re-computing the estimate for B additional bootstrap samples in order to obtain complete results. In principle, a k-step bootstrap approach can alleviate this problem by allowing the researcher to obtain asymptotically equivalent results in a fraction of the computing time.

Here, I will adapt Kim and Sun (2016)'s k-step bootstrap for panel data models to the case of a two-way fixed effects model. Recall that $\mathcal{L}^{(b)} = \sum_{i,j} W_{ij}^{(b)}\ell_{ij}$ is the weighted pseudolikelihood for a given bootstrap draw. Extending some other earlier notation, let $\mathcal{L}^{*(b)}$ be the information-orthogonalized version of $\mathcal{L}^{(b)}$ that orthogonalizes the main regressors with respect to the fixed effects. Analogous to the presentation in Kim and Sun (2016), for each bootstrap draw b, let $(\widehat{\beta}_{[k]}^{(b)}, \widehat{\alpha}_{[k]}^{(b)}, \widehat{\gamma}_{[k]}^{(b)})$ be the kth set of estimates obtained from the following recursive procedure:

$$\widehat{\beta}_{[k]}^{(b)} = \widehat{\beta}_{[k-1]}^{(b)} - \left(\partial_{\beta\beta}\mathcal{L}_{[k-1]}^{*(b)}\right)^{-1}\partial_{\beta}\mathcal{L}_{[k-1]}^{*(b)}, \tag{8}$$

$$\begin{pmatrix} \widehat{\alpha}_{[k]}^{(b)} \\ \widehat{\gamma}_{[k]}^{(b)} \end{pmatrix} = \begin{pmatrix} \widehat{\alpha}_{[k-1]}^{(b)} \\ \widehat{\gamma}_{[k-1]}^{(b)} \end{pmatrix} - \begin{pmatrix} \partial_{\alpha\alpha}\mathcal{L}_{[k-1]}^{(b)} & \partial_{\alpha\gamma}\mathcal{L}_{[k-1]}^{(b)} \\ \partial_{\gamma\alpha}\mathcal{L}_{[k-1]}^{(b)} & \partial_{\gamma\gamma}\mathcal{L}_{[k-1]}^{(b)} \end{pmatrix}^{-1} \begin{pmatrix} \partial_{\alpha}\mathcal{L}_{[k-1]}^{(b)} \\ \partial_{\gamma}\mathcal{L}_{[k-1]}^{(b)} \end{pmatrix}, \tag{9}$$

where $\partial_{\beta\beta}\mathcal{L}_{[k-1]}^{*(b)}$, $\partial_{\beta}\mathcal{L}_{[k-1]}^{*(b)}$, $\partial_{\alpha\alpha}\mathcal{L}_{[k-1]}^{*(b)}$, etc. serve as shorthand for derivatives of $\mathcal{L}^{*(b)}$ evaluated at the prior estimate values $(\widehat{\beta}_{[k-1]}^{(b)}, \widehat{\alpha}_{[k-1]}^{(b)}, \widehat{\gamma}_{[k-1]}^{(b)})$. The procedure is initialized using $\widehat{\beta}_{[0]}^{(b)} = \widehat{\beta}$,

$\widehat{\alpha}_{[0]}^{(b)} = \widehat{\alpha}$, $\widehat{\alpha}_{[0]}^{(b)} = \widehat{\alpha}$. The k-step bootstrap for bootstrap draw b is given by $\widehat{\beta}_{[k]}^{(b)}$. In words, the k-step bootstrap uses a set number of iterations (k) to approximate the difference between the original estimates and each bootstrap estimate. For panel data models, Kim and Sun (2016) show the k-step bootstrap bias correction is asymptotically equivalent to the full bootstrap bias correction if $k \geq 2$.

Since Kim and Sun (2016) focus only on models with a single fixed effect, it is worth noting that my implementation of the k-step bootstrap introduces two additional refinements to aid computation with multi-way fixed effects. First, the orthogonalization of the likelihood means that the updating of the main parameter estimates can be partitioned from the updating of the fixed effects estimates. Second, as in Correia et al. (2020), a version of the above procedure can be implemented using an iteratively re-weighted least squares (IRLS) algorithm, in which case the fixed effects estimates can be updated in a computationally efficient way using within transformations rather than needing to invert any large Hessian matrices. The implementation of the IRLS version of the k-step bootstrap is discussed in the Appendix.

3.4 Bootstrap standard errors

The discussion in this section thus far has focused on the potential for the bootstrap to correct the bias of PML gravity estimates. However, it is well known that the distribution of the bootstrap estimates can be used as an alternative to analytical methods for obtaining inferences. Justifying longstanding common practice, Hahn and Liao (2021) show that the variance of the different bootstrap estimates itself produces a conservative estimate of the variance of the original estimator, and evidence from Pfaffermayr (2021) accordingly suggests that bootstrap standard error estimates are less biased than typical HC1 standard errors for PML two-way gravity estimates. As discussed in Section 2.2, the bias in two-way PML standard errors is itself due to an incidental parameter problem; thus, it is worth investigating if the bootstrap has the potential to address two different incidental parameter problems using a single procedure.

Generally, there are two prevailing approaches to constructing bootstrap inferences: using the standard deviation of the bootstrap estimates as an estimate of the standard error and using percentile methods to obtain p values directly. Given the results of Hahn and Liao (2021), I will focus moreso on the use of bootstrap standard errors (also arguably the more common of the two approaches). Given a set of B bootstrap estimates $\{\widehat{\beta}^{(b)}\}_{b=1}^{B}$,the bootstrap estimate of the variance can be computed using either $B^{-1}\sum_{b=1}^{B}\left(\widehat{\beta}^{(b)}-\widehat{\beta}\right)^2$ or $B^{-1}\sum_{b=1}^{B}\left(\widehat{\beta}^{(b)}-\overline{\beta}^{B}\right)^2$. For my implementations, I use the latter of the two, which is simply the second moment of the bootstrap distribution. However, using either leads to similar results. In either case, the bootstrap standard error is obtained as the square root of the variance estimate.

Similarly, k-step bootstrap standard errors (as well as k-step percentiles) can be computed using the same bootstrap estimates used for the k-step bias correction. Here, it interesting to note the distinction between k-step standard errors computed using $k=1$ vs. $k\geq 2$. For $k=1$, no updating of the fixed effects estimates is needed, and the variance is simply estimated using (8) to produce each bootstrap estimate, i.e., $\widehat{\beta}_{[1]}^{(b)}=\widehat{\beta}-\left(\partial_{\beta\beta}\mathcal{L}_{[0]}^{*(b)}\right)^{-1}\partial_{\beta}\mathcal{L}_{[0]}^{*(b)}$. A closely related variance estimator is the "wild score" bootstrap of Kline et al. (2018), which uses the original (unweighted) sample expected Hessian $\partial_{\beta\beta}\overline{\mathcal{L}}_{[0]}^{*}$ in place of $-\partial_{\beta\beta}\mathcal{L}_{[0]}^{*(b)}$. That is,

$$\widehat{\beta}_{wild}^{(b)}=\widehat{\beta}-\left(\partial_{\beta\beta}\overline{\mathcal{L}}_{[0]}^{*}\right)^{-1}\sum_{i,j}W_{ij}^{(b)}\ell_{ij[0]}^{*\beta},\qquad \widehat{Var}_{wild}(\widehat{\beta})=B^{-1}\sum_{b=1}^{B}\left(\widehat{\beta}_{wild}^{(b)}-\widehat{\beta}\right)^2.$$

This variance estimate has the advantage of being very easy to program and compute, requiring only the original sample score and Hessian. However, because of the near equivalence between the wild bootstrap and the $k=1$ bootstrap, the wild bootstrap cannot be used to provide a bias correction for the coefficient estimates.[11]

3.5 Bootstrap for three-way gravity models

For three-way gravity model, an analogous set of results follows when using a cluster bootstrap that clusters by pair. In the Appendix, I show that the cluster bootstrap doubles the asymptotic bias of the original thee-way PPML gravity estimator, similar to how the ordinary bootstrap doubles the bias of two-way PML gravity estimates. Drawing on the discussion of three-way PPML in Section 2.3, this is because the three-way gravity model can be re-written as a multinomial model with two-way fixed effects when the estimator is PPML. Implementations of the fractional weight bootstrap, the k-step bootstrap, and boot-

[11]Indeed, because $\sum_{i,j}\ell_{ij[0]}^{*\beta}=0$ (by the FOC for $\widehat{\beta}$), it is straightforward to see that the expected average wild bootstrap estimate as $B\to\infty$ is just $\widehat{\beta}$, indicating the wild bootstrap does not introduce any additional bias. Typically for the wild bootstrap the bootstrap weights are such that $\mathbb{E}(W_{ij}^{(b)})=0$, but this is covered here by noting that $\sum_{i,j}W_{ij}^{(b)}\ell_{ij[0]}^{*\beta}=\sum_{i,j}(W_{ij}^{(b)}-1)\ell_{ij[0]}^{*\beta}$.

strap standard errors all follow by way of analogy by using cluster-specific weights in place of observation-specific weights when carrying out these procedures.

4 Simulations

This section presents simulation evidence regarding the performance of the bootstrap for improving inferences as well as a comparison with other available methods. In keeping with the scope of the preceding analysis, I consider simulations for both the two-way and three-way gravity models, using a variety of estimators (PPML and Gamma PML) for the two-way case and using different assumptions for the conditional variance.

4.1 Setup

The two-way gravity model is simulated using $\lambda_{ij}\omega_{ij}$, with $\lambda_{ij} = e^{\alpha_i+\gamma_j+x_{ij}\beta}$. Similarly, the three-way gravity model is simulated using $y_{ijt} = \lambda_{ijt}\omega_{ijt}$, with $\lambda_{ijt} = e^{\alpha_{it}+\gamma_{jt}+\eta_{ij}+x_{ijt}\beta}$. In all cases, I use $\beta = 1$ and $\alpha, \gamma \sim N(0, 1/16)$ and generate the error term ω as a heteroskedastic log-normal with mean 1. For the three-way gravity model, $\eta \sim N(0, 1/16)$ as well and x_{ijt} is generated using $x_{ijt-1}/2 + \alpha_{it} + \gamma_{jt} + \nu_{ijt}$, with $x_{ij1} = \alpha_{i1} + \gamma_{j1} + \eta_{ij} + \nu_{ijt}$. For the two-way gravity model, x_{ij} is generated using $x_{ij} = \alpha_i + \gamma_j + \nu_{ij}$. ν, the random part of x, is always drawn from $N(0, 1/2)$. Generally, these assumptions mirror those of Fernández-Val and Weidner (2016) and Weidner and Zylkin (2021).

To allow for variation in the conditional variance, I consider two cases, one in which PPML is correctly specified and one in which Gamma PML is correctly specified. Suppressing subscripts for generality, "case 1" corresponds to specifying that $Var(\omega) = \lambda^{-1}$, in which case the conditional variance of y is given by λ, as in a Poisson model. "Case 2" corresponds to correct specification of Gamma PML, in which case $Var(\omega) = 1$, implying the conditional variance of each observation is λ. Moreover, like in Weidner and Zylkin (2021), the three-way gravity model incorporates serial correlation within pairs by assuming that

$$Cov[\omega_{ijs}, \omega_{ijt}] = \exp\left[\rho^{|s-t|} \times \sqrt{\ln\left(1+\sigma^2_{ijs}\right)}\sqrt{\ln\left(1+\sigma^2_{ijt}\right)}\right] - 1,$$

where $\sigma^2_{ijs} = Var(\omega_{ijs})$, $\sigma^2_{ijt} = Var(\omega_{ijt})$ and where ρ serves as a quasi-correlation parameter. As in Weidner and Zylkin (2021), I use $\rho = 0.3$.

Panel A Table 1 describes the different bias corrections to be used in the simulations.

Aside from the bootstrap corrections described in the previous section, I also consider analytical bias corrections based on Fernández-Val and Weidner (2016) and Weidner and Zylkin (2021), a split-panel jackknife (SPJ) inspired by Dhaene and Jochmans (2015), and a "node jackknife" that leaves out one country at a time, drawing on a similar method for two-way panel data models suggested by Fernández-Val and Weidner (2018). For the SPJ, I use Weidner and Zylkin (2021)'s adaptation of the SPJ for gravity models that randomly and repeatedly partitions the set of countries into two equal-sized country groupings and then constructs 4 different subsample estimates for each partition based on the different pair-wise combinations of the two groups.[12] For each partition, the average subsample estimate has twice the asymptotic bias of the original estimate, such that the bias can be corrected in a similar manner to the bootstrap bias correction. The node jackknife uses a similar principle: by leaving out one country at a time, the bias of the average jackknife estimate should be $(N-1)/(N-2)$ times that of the original estimate, again enabling the bias to be corrected using simple arithmetic.[13]

The different methods for computing standard errors are listed in Table 1, panel B. In addition to HC1 "uncorrected" standard errrors, implemented as CR1 cluster-robust standard errors for the three-way gravity model—I use locally de-biased HC2 and CR2 standard errors based on Weidner and Zylkin (2021), jackknife standard errors, and a variety of bootstrap-based standard errors. For simulations of the two-way gravity model, jackknife standard errors are computed by leaving out one observation at a time; for the three-way gravity model, I use a cluster jackknife that leaves out one country-pair at a time. Similarly, bootstrap standard errors for the three-way gravity model use a cluster bootstrap. For the re-sampling and fractional weight bootstraps, 2- and 3-step bootstrap standard errors are produced as by-products of the same computations used to compute the full bootstrap estimates (as are the 2- and 3-step bias corrections).

All together, including all the different bootstrap variants, the simulations incorporate 9 different methods for re-centering coefficient estimates, 10 different ways of computing standard errors for the two-way gravity model estimators, and 9 different ways of obtaining

[12]For illustration, if the two groups are group 1 and group 2, the 4 subsamples consist of group 1's exports to group 1, group 1's exports to group 2, group 2's exports to group 1, and group 2's exports to group 2. I use random and repeated sampling because the set of countries does not have a natural ordering, in contrast to the panel data settings studied in Dhaene and Jochmans (2015).

[13]Recall that the data is assumed to not include observations for which a country trades with itself. If instead it does include these observations, the bias of each jackknife estimate should be expected to be $N/(N-1)$ times that of the original one.

standard errors for the three-way PPML estimator. Thus, to limit the number of possibilities, I assume that each bootstrap-based bias correction is always paired with the standard errror (or percentile calculation) produced by that same bootstrap. For example, bias corrections using the fractional weight bootstrap are only paired with fractional weight bootstrap standard errors, 2-step re-sample bootstrap bias-corrected estimates only use 2-step bootstrap standard errors, and so on. In addition, I only consider wild bootstrap standard errors for the two-way gravity model estimators.

4.2 Results

The first set of results presents a comparison of bootstrap standard errors for the two-way gravity model. The two panes of Figure 1 use box-and-whisker plots to summarize the results of different methods for computing standard errors for two-way PPML, while Figure 2 shows analogous results for two-way Gamma PML.[14] In both figures, the horizontal lines depict the standard deviation of the estimates, which serve as a target. The left panes correspond to the case where PPML is correctly specified (case 1); the right panes show the case where Gamma PML is correctly specified (case 2). In all cases, N is equal to 50. For all bootstrap methods, the number of bootstrap replications is $B = 250$.

Since PPML is the more commonly used of the two estimators, the left pane of Figure 1 where PPML is correctly specified is worth focusing on as a benchmark. As expected, the default HC1 standard errors are downward biased, while HC2 standard errors, which are based on the assumption of correct specification, remove most of the bias. However, echoing prior results from Pfaffermayr 2021, bootstrap standard errors based on the re-sampling bootstrap outperform HC2 standard errors in matching the correct confidence interval width. Encouragingly, the 2-step and 3-step variants of the bootstrap produce near-identical results to those of the full bootstrap. In contrast, standard errors using the fractional weight bootstrap are disappointing, offering roughly the same bias as HC1 standard errors at the median but a wider interquartile range. Jackknife standard errors, meanwhile, are overall over-conservative but have the widest dispersion, while the wild bootstrap performs similarly to HC1 and the fractional weight bootstrap.

In the right pane of Figure 1, in which PPML is no longer correctly specified, results

[14]In each box-and-whisker plot, the line inside the box is the median, the bottom of the box is the first quartile, the top of the box is the third quartile, and the span of the whiskers equals 1.5 times the interquartile range.

for the different standard errors methods remain qualitatively similar. The main difference are that all standard error estimates are significantly downward biased, with the jackknife now being the least biased. It remains the case that the re-sampling bootstrap performs similarly to using HC2 standard errors while the fractional weight bootstrap continues to be disappointing, actually performing worse than using HC1 standard errors in this case. Turning to the results for Gamma PML, shown in Figure 2, we still generally see that bootstrap standard errors are generally more conservative than the uncorrected HC1 standard errors, but the difference becomes small when Gamma PML is correctly specified (right pane). Once again, the fractional weight bootstrap performs poorly relative to the re-sampling bootstrap, especially for case 2 where it again under-performs HC1. Interestingly, 2-step and 3-step bootstrap standard errors do not behave as similarly to their full bootstrap equivalents as they do when the estimator is PPML.

More detailed results, now also incorporating bias corrections for the Gamma PML coefficient estimates, are provided in Tables 2 and 3. Table 1 shows results for case 2, while Table 2 shows results for case 3. Results for PPML, which only consider corrections to the standard errors and confidence intervals, are in line with those reported in Figure 2. For Gamma PML, the uncorrected estimates in Table 2 show clear evidence of an asymptotic bias: for $N = 50$, its bias is larger than its standard deviation, and the ratio of the bias to the standard deviation stays roughly the same as the number of countries increases from $N = 50$ to $N = 100$. The node jackknife offers the best bias correction on average, but the re-sampling bootstrap still removes more than half of the bias. The same cannot be said for the fractional weight bootstrap, which seems to underperform at bias correction in addition to providing significantly under-conservative standard errors. When simultaneous corrections are applied to both the point estimates and the standard errors, using the re-sampling bootstrap for both correction steps performs respectably at improving coverage compared with the best performing methods, all of which use bootstrap standard errors.[15] Results for the percentile bootstrap, reported in square brackets, show that it generally performs no better than using bootstrap standard errors at improving coverage and actually performs worse when bias corrections are used.

The results in Table 3, where Gamma PML is correctly specified, represent a more chal-

[15]Note that one could also bootstrap the bias-corrected estimates so as to get a better estimate of their variability that includes the added variance due to the bias correction. For the analytical bias correction, this would be relatively straightforward. For the bootstrap bias correction, one would need to perform a "double bootstrap" as in Kim and Sun (2016). Given the computational burden involved, an appropriate k-step bootstrap approach would likely be necessary.

lenging scenario for bias correction. As discussed in Section 2.2, Gamma PML actually has no asymptotic bias under correct specification, though the researcher is unlikely to know this beforehand. Therefore, asymptotic bias corrections all have the effect of making coverage worse relative to the uncorrected estimator, as is shown in the middle panel of Table 3. Keeping the focus on the bootstrap, it worth noting that the re-sampling bootstrap and the fractional weight bootstrap seem to do the least amount of harm when $N = 50$ and perform about the same as the other bias correction methods when $N = 100$. When corrections to standard errors are also considered, using jackknife standard errors seems like the most important consideration. Benchmarking the re-sample bootstrap, it performs about the same or worse as using analytical corrections in terms of coverage but significantly better than the fractional weight bootstrap.

Finally, Table 4 presents corresponding results for PPML estimates of the three-way gravity model, using $N = 50$ and $T = 10$. For the case of correct specification, shown in the lefthand side of Table 4, all bias corrections under consideration, including the bootstrap, remove almost all of the bias on average and lead to similar improvements in coverage. When bootstrap standard errors are added, inferences actually become slightly overconservative, reaching a coverage ratio of 0.960 when the bootstrap bias correction is paired with standard errors produced using the same bootstrap. For case 2, shown in the righthand portion of Table 4, bias correction is generally less effective. Only the split-panel jackknife removes more than half the bias on average, and only the combination of the split-panel jackknife with jackknife SEs leads to a significant improvement in coverage, while using a bootstrap for both bias correction and standard errors is outperformed provides no improvement.[16] Still, given the different possible bootstrap methods, it is notable that the fractional weight bootstrap again performs appreciably worse than the re-sampling bootstrap, in this case causing coverage of the correct confidence interval to become significantly worse.

5 Empirical Application

My empirical illustration focuses on the three-way PPML estimator for panel data gravity models. Because three-way PPML is the typical estimator of choice in the literature for

[16] It is worth noting that the ratio of the bias to the standard deviation of the uncorrected estimator for case 2 is almost 3 times less in magnitude than that of case 1. Thus, case 2 represents a case with a much smaller asymptotic bias problem than the one in case 1, similar to the role that it plays for the Gamma PML results.

estimating the effects of trade policies, it is likely to be the most common application in which a researcher needs to consider corrections to both the point estimates and standard errors due to the fixed effects. The data I use is the same as in Weidner and Zylkin (2021), consisting of total trade flows between 167 countries observed over 5 time periods (1995, 2000, 2005, 2010, 2015). The model being estimated is

$$y_{ijt} = \exp\left(\alpha_{it} + \gamma_{jt} + \eta_{ij} + \beta FTA_{ijt}\right)\omega_{ijt},$$

where FTA_{ijt} is a 0/1 indicator for the presence of a free trade agrreement (FTA). The researcher's presumed goal is to estimate the average effect of an FTA on trade, a typical application.

In Table 5, I show results for different bias corrections and standard error methods. The results shown for the uncorrected PPML estimates and CR1 cluster-robust standard errors are the same as in Weidner and Zylkin (2021), as are the analytical- and split-panel jackknife-corrected estimates as well as the CR2-corrected standard errors proposed by Weidner and Zylkin (2021). Most of the other results shown demonstrate the effects of implementing different versions of the bootstrap. Strikingly, the re-sampling bootstrap gives an almost identical bias corrected estimate to the analytical bias correction (0.0856 vs 0.0857), suggesting that both approaches are picking up on similar features of the data. Echoing a theme from some of the simulations, the fractional weight bootstrap bias correction goes in the same direction as the other bias correction methods but provides a smaller degree of correction.

Turning to the standard error estimates, the result that stands out is again is the close correspondence between the bootstrap standard error and the most commonly used analytical de-biasing method, in this case the CR2 standard error. Continuing with the comparison of the re-sampling and fractional weight versions of the bootstrap, most researchers interested in obtaining less biased standard errors would likely regard the standard errors produced by the fractional weighting method to be especially poor in this instance, yielding a standard error that is actually less than the original uncorrected CR1 standard errors (0.0261 vs. 0.0275). Given the general expectation that CR1 standard errors should be downward biased in this setting (see Section 2.3), plus the close resemblance between the results produced by the other methods, it would appear that the fractional weight bootstrap not only fails to reduce the bias of the uncorrected standard errors in this case but actually makes the problem worse.

6 Conclusion

This paper has shown that the bootstrap can be a credible method for correcting the bias of PML estimators for two-way and three-way gravity models. The reason why the bootstrap is suitable as a bias correction method for these settings has to do with the assumption of cross-sectional independence for trade flows not involving the same pair of countries. Because of this type of independence, the observations that are used to estimate each fixed effect parameter are uncorrelated with one another, so that the bootstrap turns out to double the contribution of each fixed effect estimate to the bias of the main parameter estimates. In other common fixed effects settings, such as typical panel data, it is usually not plausible to assume that observations for the same individual are not correlated over time, complicating the application of the bootstrap.

In showing the bootstrap's effectiveness in this context, I use new arguments for why the bootstrap works to correct bias. Instead of treating the bootstrap strictly as a random re-sampling of the data, I treat it more generally as a random re-weighting of the data. Under the latter perspective, the same proofs can be used for implementations of the bootstrap that use continuous weights instead of integer weights. Using either type of weighting, it turns out that the bootstrap-weighted estimates produced by each bootstrap replication have an asymptotic bias that depends on the expected square of these weights, causing them to have twice the bias of the original estimator. However, despite the seeming equivalence of these different bootstrap approaches, I generally find using simulations and real data evidence that the re-sampling bootstrap offers superior performance to that of the fractional weight bootstrap. The comparison between the two bootstrap methods becomes more stark when considering the standard errors they produce. As in Pfaffermayr (2021), I find that the re-sampling bootstrap generally provides less biased standard error estimates than using the heteroskedasticity-robust and cluster-robust standard errors that are ubiquitous in the literature. The fractional weight bootstrap, however, offers standard errors that are usually less conservative, and when applied to real trade data I find it seems to yield standard errors that are implausibly small.

In describing these results, it is important to acknowledge that the bootstrap does not offer a clear advantage over using analytical methods when analytical methods are available. The best use case for the bootstrap is likely to be in settings where analytical corrections cannot be easily applied, in which case it becomes a viable alternative to using jackknife-based approaches. Straightforward examples include pooled estimators for gravity models with multiple sectors and imputation-based estimators for gravity models with heterogeneous treatment affects. In comparison with the jackknife, bootstrap methods have the advantage that they provide re-centered point estimates and conservative estimates of the standard errors using a single procedure, whereas jackknife standard errors use a different jackknife method than the one used for bias correction in this setting. Though the bootstrap necessarily entails a computational burden to obtain all the bootstrap estimates, a k-step bootstrap approach can be used to greatly reduce computation time.

References

Andrews, D. W. (2002): "Higher-order improvements of a computationally attractive k-step bootstrap for extremum estimators," *Econometrica*, 70, 119–162.

——— (2005): "Higher-order improvements of the parametric bootstrap for Markov processes," *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, 171–215.

Arellano, M. and J. Hahn (2007): "Understanding Bias in Nonlinear Panel Models: Some Recent Developments," *Econometric Society Monographs*, 43, 381.

Baier, S. L. and J. H. Bergstrand (2007): "Do Free Trade Agreements actually Increase Members' International Trade?" *Journal of International Economics*, 71, 72–95.

Borchert, I., M. Larch, S. Shikher, and Y. V. Yotov (2021): "The international trade and production database for estimation (ITPD-E)," *International Economics*, 166, 140–166.

Borusyak, K., X. Jaravel, and J. Spiess (2024): "Revisiting event-study designs: robust and efficient estimation," *Review of Economic Studies*, rdae007.

Breinlich, H., D. Novy, and J. Santos Silva (2022): "Trade, gravity and aggregation," *Review of Economics and Statistics*, 1–29.

Chowdhry, S., J. Hinz, K. Kamin, and J. Wanner (2023): "Brothers in arms: The value of coalitions in sanctions regimes," .

Correia, S., P. Guimarães, and T. Zylkin (2020): "Fast Poisson estimation with high-dimensional fixed effects," *The Stata Journal*, 20, 95–115.

DAVIDSON, R. AND J. G. MACKINNON (1999): "Bootstrap testing in nonlinear models," *International Economic Review*, 40, 487–508.

DHAENE, G. AND K. JOCHMANS (2015): "Split-panel Jackknife Estimation of Fixed-effect Models," *The Review of Economic Studies*, 82, 991–1030.

EFRON, B. (1982): *The jackknife, the bootstrap and other resampling plans*, SIAM.

EGGER, P. H. AND K. E. STAUB (2015): "GLM estimation of trade gravity models with fixed effects," *Empirical Economics*, 50, 137–175.

FERNÁNDEZ-VAL, I. AND M. WEIDNER (2016): "Individual and time effects in nonlinear panel models with large N, T," *Journal of Econometrics*, 192, 291–312.

——— (2018): "Fixed effect estimation of large T panel data models," *Annual Review of Economics*, 10, 109–138.

FREEDMAN, D. A. (1981): "Bootstrapping regression models," *The annals of statistics*, 9, 1218–1228.

FRENCH, S. (2019): "Comparative advantage and biased gravity," *UNSW Business School Research Paper*.

GONÇALVES, S. AND M. KAFFO (2015): "Bootstrap inference for linear dynamic panel data models with individual fixed effects," *Journal of Econometrics*, 186, 407–426.

GOTWALT, C., L. XU, Y. HONG, AND W. Q. MEEKER (2018): "Applications of the Fractional-Random-Weight Bootstrap," *arXiv preprint arXiv:1808.08199*.

HAHN, J., D. W. HUGHES, G. KUERSTEINER, AND W. K. NEWEY (2024): "Efficient Bias Correction for Cross-section and Panel Data," *arXiv preprint arXiv:2207.09943*.

HAHN, J. AND Z. LIAO (2021): "Bootstrap standard error estimates and inference," *Econometrica*, 89, 1963–1977.

HARDIN, J. W. AND J. M. HILBE (2018): *Generalized linear models and extensions*, Stata press.

HEAD, K. AND T. MAYER (2014): "Gravity equations: workhorse, toolkit, and cookbook," *Handbook of International Economics*, 4, 131–196.

HIGGINS, A. AND K. JOCHMANS (2024): "Bootstrap Inference for Fixed-Effect Models," *Econometrica*, 92, 411–427.

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

Kaffo, M. (2015): "Essays on bootstrap in econometrics," .

Kim, M. S. and Y. Sun (2016): "Bootstrap and k-step bootstrap bias corrections for the fixed effects estimator in nonlinear panel data models," *Econometric Theory*, 32, 1523–1568.

Kline, P., R. Saggio, and M. Sølvsten (2018): "Leave-out estimation of variance components," *arXiv preprint arXiv:1806.01494*.

MacKinnon, J. G. and H. White (1985): "Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties," *Journal of econometrics*, 29, 305–325.

Nagengast, A. and Y. V. Yotov (2023): "Staggered difference-in-differences in gravity settings: Revisiting the effects of trade agreements," .

Neyman, J. and E. L. Scott (1948): "Consistent estimates based on partially consistent observations," *Econometrica*, 16, 1–32.

Pfaffermayr, M. (2019): "Gravity models, PPML estimation and the bias of the robust standard errors," *Applied Economics Letters*, 1–5.

——— (2021): "Confidence intervals for the trade cost parameters of cross-section gravity models," *Economics Letters*, 201, 109787.

Rubin, D. B. (1981): "The bayesian bootstrap," *The annals of statistics*, 130–134.

Santos Silva, J. M. C. and S. Tenreyro (2006): "The log of gravity," *Review of Economics and Statistics*, 88, 641–658.

Weidner, M. and T. Zylkin (2021): "Bias and consistency in three-way gravity models," *Journal of International Economics*, 132, 103513.

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

Yotov, Y. V., R. Piermartini, J.-A. Monteiro, and M. Larch (2016): "An Advanced Guide to Trade Policy Analysis: The Structural Gravity Model," *World Trade Organization, Geneva*.

Tables & Figures

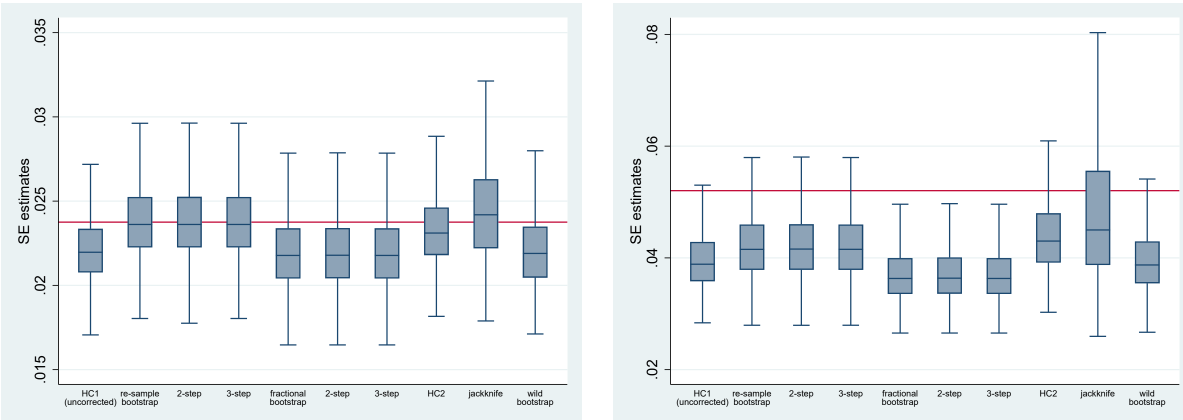

Figure 1: Summarizing different methods for computing PPML standard errors for the two-way gravity model, $N = 50$. Left pane: case where PPML is correctly specified, i.e., $Var(\omega_{ij}) = \lambda_{ij}^{-1}$. Right pane: case where Gamma PML is correctly specified, i.e., $Var(\omega_{ij}) = 1$.

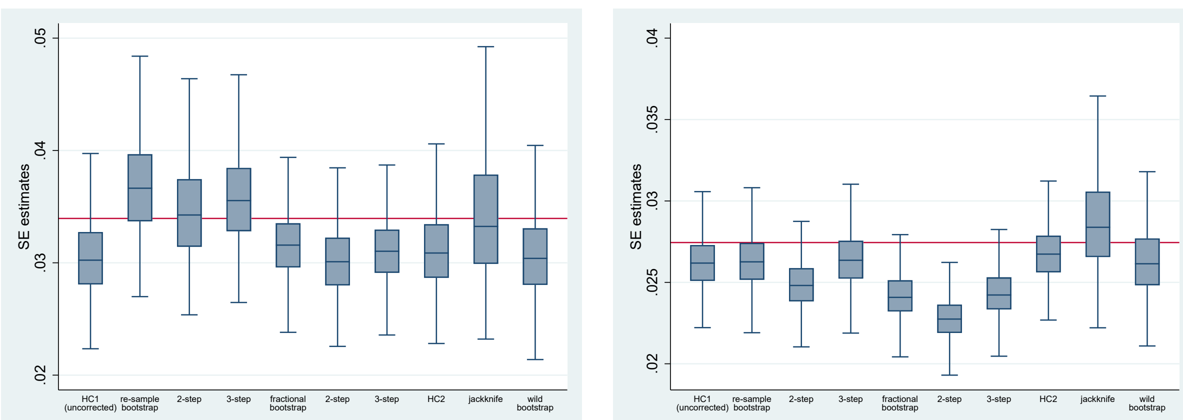

Figure 2: Summarizing different methods for computing Gamma PML standard errors for the two-way gravity model, $N = 50$. Left pane: case where PPML is correctly specified, i.e., $Var(\omega_{ij}) = \lambda_{ij}^{-1}$. Right pane: case where Gamma PML is correctly specified, i.e., $Var(\omega_{ij}) = 1$.

Table 1: Bias Corrections used in Simulations

Method	Notes
A. Bias corrections	
Analytical bias correction	Corrections for two-way Gamma PML based on Fernández-Val and Weidner (2016). Corrections for three-way PPML based on Weidner and Zylkin (2021).
Split-panel Jackknife (SPJ)	Randomly and repeatedly divides sample into 4 subsamples by splitting along country dimension like in Weidner and Zylkin (2021).
Node jackknife	Leaves out one country at a time.
Re-sample bootstrap (bootstrap)*	For the three-way gravity model I use a cluster bootstrap that re-samples pairs. Produces 2-step and 3-step bootstrap bias corrections as by-products.
Fractional Weight Bootstrap (FRWB)*	Weights are drawn from a uniform Dirichlet distribution. For the three-way gravity model, the same weight is assigned to all observations belonging to the same pair. Produces 2-step and 3-step FRWB bias corrections as by-products.
B. Standard errors	
HC1 / CR1	Refers to "default" heteroskedasticity-robust SEs (for the two-way gravity models) or cluster-robust SEs (for the three-way gravity model) that only use a degrees of freedom correction.
HC2 / CR2	HC2- and CR2-corrected SEs based on Weidner and Zylkin (2021).
Jackknife SEs	For two-way gravity models, jackknife SEs are constructed by leaving one observation out at a time. For three-way gravity models, I use a cluster jackknife that leaves out one pair at a time.
Bootstrap SEs	Uses same re-sampling as bootstrap bias correction. 2-step and 3-step bootstrap SEs obtained as by-products.
FRWB SEs	Uses same weights as FRWB bias correction. 2-step and 3-step FRWB SEs obtained as by-products.
Wild Bootstrap SEs	Based on Kline et al. (2018). Only used for the two-way gravity model.

* To limit the number of possibilities to be computed, all bootstrap and FRWB bias corrections (including their 2-step and 3-step variants) are paired only with the standard errors produced using that same bootstrap method.

Table 2: Improving coverage for two-way FE-PML gravity estimators (case 1)

	N=50				N=100			
	Bias	Bias/SD	SE/SD	95% Cov.	Bias	Bias/SD	SE/SD	95% Cov.
A. PPML (case 1)								
PPML, uncorrected	-0.0014	-0.0575	0.9350	0.936	0.0006	0.0474	0.9377	0.939
PPML with corrected SEs/CIs								
Bootstrap SEs / CIs	-0.0014	-0.0575	1.0011	0.948 [0.942]	0.0006	0.0474	0.9698	0.944 [0.937]
2-step bootstrap SEs / CIs	-0.0014	-0.0575	1.0013	0.948 [0.942]	0.0006	0.0474	0.9699	0.944 [0.937]
3-step bootstrap SEs / CIs	-0.0014	-0.0575	1.0011	0.948 [0.942]	0.0006	0.0474	0.9699	0.944 [0.937]
FRW bootstrap SEs / CIs	-0.0014	-0.0575	0.9261	0.938 [0.926]	0.0006	0.0474	0.9309	0.934 [0.924]
Jackknife SEs	-0.0014	-0.0575	1.032	0.955	0.0006	0.0474	0.9878	0.945
Analytical (HC2) SEs	-0.0014	-0.0575	0.982	0.950	0.0006	0.0474	0.9633	0.947
B. Gamma PML (case 1)								
Gamma PML, uncorrected	0.0371	1.0923	0.9112	0.722	0.0225	1.2356	0.9274	0.695
Re-centered Gamma PML								
Analytical BC	0.0100	0.2560	0.7892	0.875	0.0051	0.2479	0.8232	0.874
Bootstrap BC	0.0160	0.4372	0.8477	0.871	0.0087	0.4464	0.8710	0.870
2-step Bootstrap BC	0.0158	0.4241	0.8305	0.866	0.0080	0.4030	0.8533	0.873
3-step Bootstrap BC	0.0184	0.5127	0.8603	0.862	0.0098	0.5090	0.8793	0.859
FRW boot BC	0.0203	0.5691	0.8667	0.854	0.0107	0.5637	0.8857	0.849
Split-panel Jackknife BC	0.0109	0.2930	0.8282	0.884	0.0063	0.3224	0.8609	0.880
Node Jackknife BC	0.0080	0.2075	0.7984	0.881	0.0044	0.2167	0.8292	0.882
Fully corrected Gamma PML (top 3 + selected others)								
Node J. + bootstrap SEs	0.0080	0.2075	0.9569	0.932	0.0044	0.2167	1.0383	0.952
SPJ + bootstrap SEs	0.0109	0.2930	0.9927	0.931	0.0063	0.3224	1.0779	0.946
Analytical + bootstrap SEs	0.0100	0.2560	0.9459	0.929	0.0051	0.2479	1.0307	0.945
Bootstrap BC + SEs / CIs	0.0160	0.4372	1.0160	0.924 [0.854]	0.0087	0.4464	1.0907	0.938 [0.837]
2-step Bootstrap BC + SEs / CIs	0.0158	0.4241	0.9360	0.903 [0.805]	0.0080	0.4030	1.0507	0.933 [0.811]
3-step Bootstrap BC + SEs / CIs	0.0184	0.5127	1.0010	0.911 [0.850]	0.0098	0.5090	1.0708	0.925 [0.834]
FRWB + FRWB SEs	0.0203	0.5691	0.8906	0.865 [0.774]	0.0107	0.5637	0.9849	0.892 [0.793]
Analytical + HC2 SEs	0.0100	0.2560	0.8058	0.884	0.0051	0.2479	0.8316	0.876
Node J. + Jackknife SEs	0.0080	0.0348	0.8992	0.901	0.0044	0.2167	0.8604	0.870
SPJ + Jackknife SEs	0.0109	0.2930	0.9328	0.918	0.006	0.322	0.893	0.874

Notes: 1,000 repetitions + 250 bootstrap trials per repetition. Model: $y_{ij} = \exp(\alpha_i + \gamma_j + 1 \cdot x_{ij})\omega_{ij}$. Case 1 is the case where PPML is correctly specified. Square brackets indicate confidence intervals computed using a percentile method instead of using bootstrap standard errors.

Table 3: Improving coverage for two-way FE-PML gravity estimators (case 2)

	N=50				N=100			
	Bias	Bias/SD	SE/SD	95% Cov.	Bias	Bias/SD	SE/SD	95% Cov.
A. PPML (case 2)								
PPML, uncorrected	-0.0028	-0.0535	0.7703	0.874	0.0003	0.0110	0.8446	0.906
PPML with corrected SEs/CIs								
Bootstrap SEs	-0.0028	-0.0535	0.8306	0.904 [0.910]	0.0003	0.0110	0.8589	0.917 [0.917]
2-step bootstrap SEs / CIs	-0.0028	-0.0535	0.8313	0.904 [0.910]	0.0003	0.0110	0.8595	0.917 [0.917]
3-step bootstrap SEs / CIs	-0.0028	-0.0535	0.8306	0.904 [0.910]	0.0003	0.0110	0.8589	0.917 [0.917]
FRW bootstrap SEs / CIs	-0.0028	-0.0535	0.7195	0.848 [0.855]	0.0003	0.0110	0.7830	0.882 [0.894]
Jackknife SEs	-0.0028	-0.0535	0.9685	0.911	0.0003	0.0110	0.9167	0.886
Analytical (HC2) SEs	-0.0028	-0.0535	0.8658	0.911	0.0003	0.0110	0.9049	0.927
B. Gamma PML (case 2)								
Gamma PML, uncorrected	-0.0006	-0.0234	0.9575	0.943	0.0004	0.0300	0.9536	0.939
Re-centered Gamma PML								
Analytical BC	-0.0007	-0.0227	0.9147	0.926	0.0005	0.0312	0.9214	0.929
Bootstrap BC	-0.0006	-0.0209	0.9246	0.931	0.0004	0.0296	0.9249	0.928
2-step Bootstrap BC	-0.0012	-0.0435	0.9217	0.925	0.0003	0.0208	0.9245	0.913
FRW boot BC	-0.0005	-0.0195	0.9300	0.933	0.0005	0.0317	0.9293	0.931
Split-panel Jackknife BC	-0.0006	-0.0221	0.9161	0.928	0.0005	0.0312	0.9241	0.929
Node Jackknife BC	-0.0004	-0.0148	0.9005	0.925	-0.0003	-0.0192	0.9192	0.927
Fully corrected Gamma PML (top 3 + selected others)								
SPJ + Jackknife SEs	-0.0006	-0.0221	1.0025	0.950	0.0005	0.0312	0.9654	0.939
Node J. + Jackknife SEs	-0.0004	-0.0148	0.9855	0.951	-0.0003	-0.0192	0.9603	0.939
Analytical + jackknife SEs	-0.0007	-0.0227	1.0010	0.949	0.0005	0.0312	0.9625	0.938
Bootstrap BC + SEs / CIs	-0.0006	-0.0209	0.9259	0.934 [0.930]	0.0004	0.0296	0.9083	0.920 [0.921]
2-step Bootstrap BC + SEs / CIs	-0.0012	-0.0435	0.8712	0.911 [0.906]	0.0003	0.0208	0.8770	0.913 [0.910]
3-step Bootstrap BC + SEs / CIs	-0.0017	-0.0588	0.9306	0.932 [0.930]	0.0000	0.0033	0.9114	0.920 [0.923]
FRWB + FRWB SEs / CIs	-0.0005	-0.0195	0.8554	0.906 [0.897]	0.0005	0.0317	0.8672	0.909 [0.905]
Analytical + HC2 SEs	-0.0007	-0.0227	0.9340	0.935	0.0004	0.0312	0.9308	0.933
Uncorrected + HC2 SEs	-0.0006	-0.0234	0.9776	0.945	0.0004	0.0300	0.9634	0.943
Uncorrected + boot. SEs	-0.0006	-0.0234	0.9588	0.941	0.0004	0.0300	0.9365	0.933
Uncorrected + jack SEs	-0.0006	-0.0234	1.0478	0.960	0.0004	0.0300	0.9962	0.948

Notes: 1,000 repetitions + 250 bootstrap trials per repetition. Model: $y_{ij} = \exp(\alpha_i + \gamma_j + 1 \cdot x_{ij})\omega_{ij}$. Case 2 is the case where Gamma PML is correctly specified. Square brackets indicate confidence intervals computed using a percentile method instead of using bootstrap standard errors.

Table 4: Improving coverage for the three-way FE-PPML gravity estimator (N=50,T=10)

	Case 1				Case 2			
	Bias	Bias/SD	SE/SD	95% Cov.	Bias	Bias/SD	SE/SD	95% Cov.
PPML, uncorrected	0.0039	0.4675	0.9321	0.896	-0.0026	-0.1751	0.8416	0.892
Re-centered PPML (w/ uncorrected SEs)								
Analytical BC	0.0005	0.0543	0.9221	0.928	-0.0018	-0.1135	0.7804	0.877
Bootstrap BC	0.0002	0.0189	0.9205	0.929	-0.0023	-0.1371	0.7208	0.873
2-step bootstrap BC	0.0002	0.0235	0.9206	0.929	-0.0022	-0.1337	0.7581	0.879
3-step bootstrap BC	0.0002	0.0189	0.9205	0.929	-0.0025	-0.1593	0.7940	0.879
FRW bootstrap BC	0.0006	0.0702	0.9220	0.927	-0.0027	-0.1725	0.7904	0.882
Node Jackknife BC	-0.0003	-0.0365	0.9035	0.928	-0.0034	-0.0887	0.3237	0.859
Split-panel Jackknife BC	0.0001	0.0076	0.9121	0.928	-0.0012	-0.1452	0.7767	0.874
Uncorrected PPML with corrected SEs								
WZ CR2 SEs	0.0039	0.4675	0.9784	0.917	-0.0026	-0.1757	0.9336	0.918
Bootstrap SEs	0.0039	0.4675	1.0816	0.946	-0.0026	-0.1757	0.8949	0.916
2-step bootstrap SEs	0.0039	0.4675	1.0798	0.946	-0.0026	-0.1757	0.8952	0.916
FRW bootstrap SEs	0.0039	0.4675	0.9927	0.925	-0.0026	-0.1757	0.7912	0.876
Jackknife SEs	0.0039	0.4675	1.0351	0.934	-0.0026	-0.1757	1.303	0.949
Fully corrected PPML								
Analytical BC + CR2 SEs	0.0005	0.0543	0.9678	0.945	-0.0018	-0.1135	0.8616	0.905
Analytical BC + boot. SEs	0.0005	0.0543	1.0699	0.961	-0.0018	-0.1135	0.8310	0.890
Analytical BC + jack. SEs	0.0005	0.0543	1.0239	0.956	-0.0018	-0.1135	1.2085	0.935
Analytical BC + FRWB SEs	0.0005	0.0543	0.9820	0.950	-0.0018	-0.1135	0.7339	0.855
Bootstrap BC + SEs / CIs	0.0002	0.0188	1.0681	0.960 [0.938]	-0.0025	-0.1574	0.8288	0.889 [0.879]
2-step boot. BC + SEs / CIs	0.0002	0.0235	1.0665	0.960 [0.936]	-0.0022	-0.1441	0.8288	0.889 [0.879]
3-step boot. BC + SEs / CIs	0.0002	0.0189	1.0681	0.960 [0.938]	-0.0024	-0.1462	0.8076	0.890 [0.879]
FRWB BC + SEs / CIs	0.0006	0.0701	0.9819	0.946 [0.913]	-0.0027	-0.1803	0.7487	0.859 [0.843]
SPJ + CR2 SEs	0.0001	0.0076	0.9573	0.936	-0.0012	-0.0681	0.7561	0.907
SPJ + boot. SEs	0.0001	0.0076	1.0583	0.959	-0.0012	-0.0681	0.7292	0.888
SPJ + FRWB SEs	0.0001	0.0076	0.9714	0.943	-0.0012	-0.0681	0.6440	0.859
SPJ + Jackknife SEs	0.0001	0.0076	1.0128	0.948	-0.0012	-0.0681	1.0605	0.935

Notes: 1,000 repetitions + 250 bootstrap trials per repetition. Model: $y_{ijt} = \exp(\alpha_{it} + \gamma_{jt} + \eta_{ij} + 1 \cdot x_{ijt})\omega_{ijt}$. Case 1 is the case where PPML is correctly specified. Case 2 is the case where Gamma PML is correctly specified. Square brackets indicate confidence intervals computed using a percentile method instead of using bootstrap standard errors.

Table 5: Empirical application (three-way PPML)

	Estimate		Standard Error
PPML ($\widehat{\beta}$)	.0821	Cluster-Robust (CR1)	.0275
WZ analytical BC ($\widehat{\beta}_A$)	.0857	Weidner-Zylkin CR2	.0305
Avg. bootstrap estimate ($\widehat{\beta}_B$)	.0786	Bootstrap SE	.0304
Bootstrap BC ($2\widehat{\beta} - \widehat{\beta}_B$)	.0856	FRWB SE	.0261
Bootstrap the analytical BC	.0818	Jackknife SE	.0344
Avg, FRWB estimate ($\widehat{\beta}_F$)	.0794		
FRWB BC ($2\widehat{\beta} - \widehat{\beta}_F$)	.0848		
Split-panel Jackknife BC ($\widehat{\beta}_S$)	.0877		

Notes; Original data from Weidner and Zylkin 2021. Trade between 167 countries observed over 5 years (1995, 2000, 2005, 2010, 2015). The model being estimated is $y_{ijt} = \exp(\alpha_{it} + \gamma_{jt} + \eta_{ij} + \beta FTA_{ijt})\omega_{ijt}$. Bootstrap results use 5,000 bootstrap replications.

Appendix (not for publication)

This Appendix elaborates on the formal results stated in the main text and also provides some further technical details and examples. To aid in coherence, it also provides some restatements of the basic environment and notation.

Basic Model

Let $y_{ij} \geq 0$ be generated from a two-way fixed effects model. Its conditional mean is given by

$$\mathbb{E}(y_{ij}|x_{ij}, \alpha_i, \gamma_j) = \exp\left(x_{ij}\beta + \alpha_i + \gamma_j\right),$$

though the exact functional relationship is not crucial. As in a trade model, both i and j run from $1, ..., N$. Only the conditional mean is specified, like in Remark 3 of Fernández-Val and Weidner (2016) (henceforth, FW) and in Weidner and Zylkin (2021) (henceforth, WZ). The data does not include observations for which $i = j$, such that the sample size is $N(N-1)$. Since i and j are both indices for locations rather than time, there is cross-sectional independence across both i and j. All the other assumptions are presumed to be the same as in FW, as re-stated in WZ's Theorem 1 in their appendix.

Estimation

Before bootstrapping, the original estimate $\widehat{\beta}$ is obtained using

$$(\widehat{\beta}, \widehat{\alpha}, \widehat{\gamma}) := \arg\max_{\beta,\alpha\gamma} \; \mathcal{L}(\beta, \alpha, \gamma) := \sum_{i,j} \ell_{ij}(\beta, \alpha_i, \gamma_j),$$

where ℓ_{ij} is the (log-)pseudolikelihood of observation ij. Examples of applicable PML estimators include Poisson PML (which is asymptotically unbiased) and Gamma PML (which is not).

As in the main text, let ℓ^*_{ij} be the information-orthogonalized version of ℓ_{ij} satisfying $\mathbb{E}(\ell_{ij}^{*\beta\alpha_i}) = \mathbb{E}(\ell_{ij}^{*\beta\gamma_j}) = 0$. This can be obtained by re-parameterizing the original model so that x_{ij} is orthogonalized with respect to the fixed effects. Like in WZ, this orthogonalized version of x_{ij} is given by $\widetilde{x}_{ij}$. Let a "bar" denote an expectation, e.g., $\overline{\ell}_{ij}^{*\beta\beta} = \mathbb{E}(\ell_{ij}^{*\beta\beta})$.

Bootstrap estimation

For each bootstrap replication b, we obtain the bootstrap estimate $\widehat{\beta}^{(b)}$using

$$(\widehat{\beta}^{(b)}, \widehat{\alpha}^{(b)}, \widehat{\gamma}^{(b)}) := \arg\max_{\beta,\alpha\gamma} \; \mathcal{L}^{(b)}(\beta, \alpha, \gamma) = \sum_{i,j} W_{ij}^{(b)} \ell_{ij}(\beta, \alpha_i, \gamma_j), \tag{10}$$

where each $W_{ij}^{(b)}$ is a randomly and independently generated weighting parameter satisfying $\mathbb{E}(W_{ij}^{(b)}) = \text{Var}(W_{ij}^{(b)}) = 1$. In the case of the traditional re-sampling bootstrap, each $W_{ij}^{(b)}$ is a randomly and independently generated integer ≥ 0 such that $\sum_{i,j} W_{ij}^{(b)} = N(N-1)$ for all b. If $W_{ij}^{(b)} = 0$, pair ij is not included in the estimation because it receives no weight; if $W_{ij}^{(b)} \geq 2$, it is "as though" it has been sampled more than once, but one could equivalently infer that it is being given a larger weight than in the original estimation. Either perspective gives identical estimates.

Instead of restricting $W_{ij}^{(b)}$ to be an integer, it can also be drawn from a continuous distribution. In this case, (10) becomes equivalent to the "fractional weight bootstrap" (Gotwalt et al. 2018), also known as the Bayesian Bootstrap (Rubin 1981). A popular choice for generating $W_{ij}^{(b)}$ is a uniform Dirichlet distribution.

The bootstrap estimation is performed B times, drawing a new set of weights each time. The average bootstrap estimate is $\overline{\beta}^B = B^{-1}\sum_{b=1}^{B}\widehat{\beta}^{(b)}$. The bootstrap bias-corrected estimate is then given by $\widehat{\beta}^{BBC} = 2\widehat{\beta} - \overline{\beta}^B$. As discussed in the main text, the standard intuition behind why this correction works is that bootstrap sampling effectively treats $\widehat{\beta}$, not β^0, as the true parameter (see Efron 1982, Hahn et al. 2024). In the following, I will prove this result holds for two-way gravity models estimated by PML and for three-way gravity models estimated by PPML. So as not to give the impression that these results follow trivially from the known properties of the bootstrap, I will also include some examples of similar applications where the bootstrap does not double the bias.

Bias of the bootstrap estimates

The strategy here is to apply the results of FW to the weighted pseudolikelihood in (10). What we are looking to verify is whether the bias of the average bootstrap estimate equates to 2 times that of the original estimate as $N \to \infty$ and $B \to \infty$. It will also suffice to show that any individual bootstrap estimate has twice the bias as $N \to \infty$, as stated in Proposition 1.

As above let $\mathcal{L}^{(b)} := \sum_{i,j} W_{ij}^{(b)} \ell_{ij}$ be the bootstrap-weighted likelihood. For a given draw of bootstrap weights $W^{(b)} = \{W_{ij}^{(b)}\}_{i\neq j}$,we have

$$\frac{\partial \mathcal{L}^{*(b)}}{\partial \beta} = \sum_{i,j} W_{ij}^{(b)} \frac{\partial \ell_{ij}^{*}}{\partial \beta}, \qquad \frac{\partial \mathcal{L}^{*(b)}}{\partial \alpha_i} = \sum_{j} W_{ij}^{(b)} \frac{\partial \ell_{ij}^{*}}{\alpha_i}, \qquad \frac{\partial \mathcal{L}^{*(b)}}{\partial \gamma_j} = \sum_{i} W_{ij}^{(b)} \frac{\partial \ell_{ij}^{*}}{\gamma_j},$$

and so on. Also suppose that the distribution of bootstrap weights satisfies the following:

$$\mathbb{E}\left(W_{ij}^{(b)}\right) = 1, \qquad \mathrm{Var}\left(W_{ij}^{(b)}\right) = 1, \qquad \mathbb{E}\left(\left(W_{ij}^{(b)}\right)^2\right) = 2,$$

As noted in the main text, assuming that $\mathrm{Var}\left(W_{ij}^{(b)}\right) = 1$ is a simplification. Using the actual multinomial or Dirichlet variance does not make a difference to the asymptotic bias results because they both approach 1 as $N \to \infty$. Furthermore, $W_{ij}^{(b)}$ is independent of ℓ_{ij} and all its partial derivatives when evaluated at the true parameters.

For a single bootstrap draw b, the results of FW Remark 3 and WZ Theorem 1 imply that the bias of the bootstrap-weighted estimator's score is

$$\frac{1}{N-1} H_N^{(b)-1} \left(B_N^{(b)} + D_N^{(b)}\right),$$

where

$$H_N^{(b)} = -\frac{1}{N(N-1)} \sum_{i,j=1}^{N} W_{ij}^{(b)} \overline{\ell}_{ij}^{*\beta\beta}$$

and where each element of $B_N^{(b)}$ is given by

$$\begin{aligned} B_N^{(b),m} = & -\frac{1}{N} \sum_{i=1}^{N} \frac{\frac{1}{N-1}\sum_{j\neq i} \left(W_{ij}^{(b)}\right)^2 \mathbb{E}\left(\ell_{ij}^{*\beta_m\alpha_i}\, \ell_{ij}^{\alpha_i}\right)}{\frac{1}{N-1}\sum_{j\neq i}^{N} W_{ij}^{(b)} \overline{\ell}_{ij}^{\alpha_i\alpha_i}} \\ & + \frac{1}{2N} \sum_{i=1}^{N} \frac{\left(\frac{1}{N-1}\sum_{j\neq i} W_{ij}^{(b)} \overline{\ell}_{ij}^{*\beta_m\alpha_i\alpha_i}\right)\left[\frac{1}{N-1}\sum_{j\neq i}\left(W_{ij}^{(b)}\right)^2 \left[\mathbb{E}\left(\ell_{ij}^{\alpha_i}\, \ell_{ij}^{\alpha_i}\right)\right]\right]}{\left(\frac{1}{N-1}\sum_{j}^{N} W_{ij}^{(b)} \overline{\ell}_{ij}^{\alpha_i\alpha_i}\right)^2}, \end{aligned}$$

with an analogous expression following for the elements of $D_N^{(b),m}$. The terms depending on $W_{ij}^{(b)}$ are pulled outside the expectations because of their independence and because we are treating them as fixed for a given bootstrap draw. The $\frac{1}{N-1}$ normalization terms are added to aid in the next step.

As $N \to \infty$, we have

$$\lim_{N\to\infty} \frac{1}{N-1} \sum_{j\neq i} \left(W_{ij}^{(b)}\right)^2 \mathbb{E}\left(\ell_{ij}^{*\beta_m \alpha_i}\, \ell_{ij}^{\alpha_i}\right) = \lim_{N\to\infty} \frac{2}{N-1} \sum_{j\neq i} \mathbb{E}\left(\ell_{ij}^{*\beta_m \alpha_i}\, \ell_{ij}^{\alpha_i}\right),$$

$$\lim_{N\to\infty} \frac{1}{N-1} \sum_{j\neq i} W_{ij}^{(b)} \overline{\ell}_{ij}^{*\beta_m \alpha_i \alpha_i} = \lim_{N\to\infty} \frac{1}{N-1} \sum_{j\neq i} \mathbb{E}\left(\overline{\ell}_{ij}^{*\beta_m \alpha_i \alpha_i}\right),$$

$$\lim_{N\to\infty} \frac{1}{N-1} \sum_{j\neq i} \left(W_{ij}^{(b)}\right)^2 \mathbb{E}\left(\ell_{ij}^{\alpha_i}\, \ell_{ij}^{\alpha_i}\right) = \lim_{N\to\infty} \frac{2}{N-1} \sum_{j\neq i} \mathbb{E}\left(\ell_{ij}^{\alpha_i}\, \ell_{ij}^{\alpha_i}\right),$$

$$\lim_{N\to\infty} \frac{1}{N-1} \sum_{j\neq i}^{N} W_{ij}^{(b)} \overline{\ell}_{ij}^{\alpha_i \alpha_i} = \lim_{N\to\infty} \frac{1}{N-1} \sum_{j\neq i}^{N} \overline{\ell}_{ij}^{\alpha_i \alpha_i},$$

and the same simplifications also applying for the corresponding elements of D_N. The key steps here are to use $\mathbb{E}\left(W_{ij}^{(b)}\right) = 1$ and $\mathbb{E}\left(\left(W_{ij}^{(b)}\right)^2\right) = 2$, which are applicable because the averages taken over $N-1$ observations become equivalent to population expectations as $N \to \infty$. Moreover, we also have that

$$\lim_{N\to\infty} -\frac{1}{N(N-1)} \sum_{i,j=1}^{N} W_{ij}^{(b)} \overline{\ell}_{ij}^{*\beta\beta} = \lim_{N\to\infty} -\frac{1}{N(N-1)} \sum_{i,j=1}^{N} \overline{\ell}_{ij}^{*\beta\beta},$$

implying that $H_N^{(b)}$ converges to H_N.

Thus, as $N \to \infty$, $B_N^{(b)} \to 2B_N$ and $B_N^{(b)} \to 2D_N$ while $H_N^{(b)} \to H_N$ for all b, guaranteeing the asymptotic bias of each bootstrap estimate is twice that of the original estimate. Interestingly, this argument does not require $B \to \infty$. We only need $B \to \infty$ for the resulting bias correction not to contribute to the asymptotic variance of the bias-corrected estimator.

Remarks

1. (Bias expansion for the unweighted estimator) The bias of two-way *unweighted* fixed effects estimators derived in FW may be described as follows. Let $\phi = \text{vec}(\alpha, \gamma)$. ϕ^0 is the vector of true fixed effect parameters, and $\widehat{\phi}(\beta)$ are the fixed effects estimates that correspond with a particular value of β (i.e., by solving the first-order conditions for each element of ϕ given a value for β). The first step is to obtain a second-order expansion of the score for β around the true set of fixed effects ϕ^0 and evaluated at the

true target parameter β^0:

$$\frac{\partial \mathcal{L}^*\left(\beta^0, \widehat{\phi}\left(\beta^0\right)\right)}{\partial \beta} \approx \sum_{i,j} \frac{\partial \ell^*_{ij}\left(\beta^0, \phi^0\right)}{\partial \beta} + \sum_{i,j} \frac{\partial^2 \ell^*_{ij}\left(\beta^0, \phi^0\right)}{\partial \beta \partial \phi'} \left(\widehat{\phi}(\beta) - \phi^0\right)$$
$$+ \frac{1}{2} \sum_{i,j} \left\{ \sum_{f,g}^{\dim \phi} \frac{\partial^3 \ell^*_{ij}\left(\beta^0, \phi^0\right)}{\partial \beta \partial \phi_f \partial \phi_g} \left(\widehat{\phi}_f(\beta^0) - \phi^0_f\right) \left(\widehat{\phi}_g(\beta^0) - \phi^0_g\right) \right\}. \quad (11)$$

As $N \to \infty$, the first term in this expression, which contributes only to the variance, becomes 0 but the latter two, which contribute only to the bias, do not. To say more, these latter two terms may be simplified using the following:

- $\left(\widehat{\phi}(\beta) - \phi^0\right)$ in the first term may be replaced with the following first-order approximation:

$$\widehat{\phi}(\beta) - \phi^0 \approx -\left(\frac{\partial^2 \overline{\mathcal{L}}}{\partial \phi \partial \phi'}\right)^{-1} \frac{\partial \mathcal{L}}{\partial \phi}.$$

- For all $f \neq g$, $\frac{\partial^2 \ell_{ij}(\beta_0, \phi_0)}{\partial \phi_f \partial \phi_g}$ either equals 0 or converges in probability to 0 as $N \to \infty$. For example, for $i' \neq i$, $j' \neq j$, we always have that, e.g., $\frac{\partial^2 \ell_{ij}}{\partial \alpha_i \partial \alpha_{i'}} = \frac{\partial^2 \ell_{ij}}{\partial \alpha_i \partial \alpha_{j'}} = 0$. More subtly, FW also show that

$$\left\| \overline{\mathcal{H}}^{-1} - \text{diag}\left(\overline{\mathcal{H}}_{(\alpha\alpha),} \overline{\mathcal{H}}_{(\gamma\gamma)}\right)^{-1} \right\|_{\max} = O_P(N^{-1}),$$

  where $\overline{\mathcal{H}} = \mathbb{E}\left[-\partial_{\phi\phi'}\mathcal{L}\right]$, $\overline{\mathcal{H}}_{(\alpha\alpha)} = \mathbb{E}\left[-\partial_{\alpha\alpha'}\mathcal{L}\right]$, $\overline{\mathcal{H}}_{(\gamma\gamma)} = \mathbb{E}\left[-\partial_{\gamma\gamma'}\mathcal{L}\right]$. This is an important result from FW that ensures all terms depending on, e.g., $\partial^2_{\alpha_i\gamma_j}\ell_{ij}$, $\partial^3_{\alpha_i\alpha_i\gamma_j}\ell_{ij}$, etc. disappear asymptotically from the expansion in (11). In addition, it implies we can use the approximations $\widehat{\alpha}_i - \alpha^0_i \approx (\partial^2_{\alpha_i\alpha_i}\overline{\ell}_{ij})^{-1}\partial_{\alpha_i}\ell_{ij}$ and $\widehat{\gamma}_j - \gamma^0_j \approx (\partial^2_{\gamma_j\gamma_j}\overline{\ell}_{ij})^{-1}\partial_{\gamma_j}\ell_{ij}$, which do not require inverting the full Hessian.

- FW also show that, after decomposing terms such as $N^{-1}\sum_j \left(\partial^3_{\beta\alpha_i\alpha_i}\ell^*_{ij}\right)(\widehat{\alpha}_i - \alpha^0_i)^2$ into, e.g., $N^{-1}\sum_j \left(\partial^3_{\beta\alpha_i\alpha_i}\overline{\ell}^*_{ij}\right)(\widehat{\alpha}_i - \alpha^0_i)^2 + N^{-1}\sum_j \left(\partial^3_{\beta\alpha_i\alpha_i}\ell^*_{ij} - \partial^3_{\beta\alpha_i\alpha_i}\overline{\ell}^*_{ij}\right)(\widehat{\alpha}_i - \alpha^0_i)^2$, the latter component $N^{-1}\sum_j \left(\partial^3_{\beta\alpha_i\alpha_i}\ell^*_{ij} - \partial^3_{\beta\alpha_i\alpha_i}\overline{\ell}^*_{ij}\right)(\widehat{\alpha}_i - \alpha^0_i)^2$is $o_P(N^{-1})$ as $N \to \infty$.

Consequently, ignoring terms that equal zero or vanish asymptotically as $N \to 0$, the above expansion in (11) may be re-written as:

$$
\begin{aligned}
\frac{\partial \mathcal{L}^*\left(\beta^0, \widehat{\phi}\left(\beta^0\right)\right)}{\partial \beta} \approx & \sum_{i,j} \frac{\ell_{ij}^{*\beta}}{\partial \beta} + \sum_{i,j} \frac{\partial^2 \ell_{ij}^*}{\partial \beta \partial \alpha_i}\left(\widehat{\alpha}_i - \alpha_i^0\right) + \sum_{i,j} \frac{\partial^2 \ell_{ij}^*}{\partial \beta \partial \gamma_j}\left(\widehat{\gamma}_j - \gamma_j^0\right) \\
& + \frac{1}{2}\sum_{i,j}\left\{\frac{\partial^3 \overline{\ell}_{ij}^*}{\partial \beta \partial \alpha_i \partial \alpha_i}\left(\widehat{\alpha}_i - \alpha_i^0\right)^2 + \frac{\partial^3 \overline{\ell}_{ij}^*}{\partial \beta \partial \gamma_j \partial \gamma_j}\left(\widehat{\gamma}_j - \gamma_j^0\right)^2\right\} \\
= & \sum_{i,j} \frac{\partial \ell_{ij}^*}{\partial \beta} - \sum_{i,j} \frac{\partial^2 \ell_{ij}^*}{\partial \beta \partial \alpha_i} \sum_{i,j}\left(\sum_j \frac{\partial^2 \overline{\ell}_{ij}^*}{\partial \alpha_i \partial \alpha_i}\right)^{-1} \frac{\partial \ell_{ij}}{\partial \alpha_i} \\
& - \sum_{i,j} \frac{\partial^2 \ell_{ij}^*}{\partial \beta \partial \gamma_j} \sum_{i,j}\left(\sum_i \frac{\partial^2 \overline{\ell}_{ij}^*}{\partial \gamma_j \partial \gamma_j}\right)^{-1} \frac{\partial \ell_{ij}}{\partial \gamma_j} \\
& + \frac{1}{2}\sum_{i,j} \frac{\partial^3 \overline{\ell}_{ij}^*}{\partial \beta \partial \alpha_i \partial \alpha_i} \sum_{i,j}\left(\sum_j \frac{\partial^2 \overline{\ell}_{ij}^*}{\partial \alpha_i \partial \alpha_i}\right)^{-2}\left(\frac{\partial \ell_{ij}}{\partial \alpha_i}\right)^2 \\
& + \frac{1}{2}\sum_{i,j} \frac{\partial^3 \overline{\ell}_{ij}^*}{\partial \beta \partial \gamma_j \partial \gamma_j} \sum_{i,j}\left(\sum_i \frac{\partial^2 \overline{\ell}_{ij}^*}{\partial \gamma_j \partial \gamma_j}\right)^{-2}\left(\frac{\partial \ell_{ij}}{\partial \gamma_j}\right)^2,
\end{aligned}
$$

where it should henceforth be understood that all partial derivatives of the pseudo-likelihood and other functions of the parameters are evaluated at the true parameter values unless otherwise indicated. Re-arranging sums, this last expression boils down to:

$$
\begin{aligned}
\frac{\partial \mathcal{L}^*\left(\beta^0, \widehat{\phi}\left(\beta^0\right)\right)}{\partial \beta} \approx & \sum_{i,j} \frac{\partial \ell_{ij}^*}{\partial \beta} \\
& + \sum_i \left\{-\left(\sum_j \frac{\partial^2 \overline{\ell}_{ij}^*}{\partial \alpha_i \partial \alpha_i}\right)^{-1} \sum_j \mathbb{E} \frac{\partial^2 \ell_{ij}^*}{\partial \beta \partial \alpha_i} \frac{\partial \ell_{ij}}{\partial \alpha_i} + \frac{1}{2}\left(\sum_j \frac{\partial^2 \overline{\ell}_{ij}^*}{\partial \alpha_i \partial \alpha_i}\right)^{-2} \sum_j \frac{\partial^3 \overline{\ell}_{ij}^*}{\partial \beta \partial \alpha_i \partial \alpha_i} \sum_j \left(\frac{\partial \ell_{ij}}{\partial \alpha_i}\right)^2\right\} \\
& + \sum_j \left\{-\left(\sum_i \frac{\partial^2 \overline{\ell}_{ij}^*}{\partial \gamma_j \partial \gamma_j}\right)^{-1} \sum_i \mathbb{E} \frac{\partial^2 \ell_{ij}^*}{\partial \beta \partial \gamma_j} \frac{\partial \ell_{ij}}{\partial \gamma_j} + \frac{1}{2}\left(\sum_i \frac{\partial^2 \overline{\ell}_{ij}^*}{\partial \gamma_j \partial \gamma_j}\right)^{-2} \sum_i \frac{\partial^3 \overline{\ell}_{ij}^*}{\partial \beta \partial \gamma_j \partial \gamma_j} \sum_i \left(\frac{\partial \ell_{ij}}{\partial \gamma_j}\right)^2\right\}
\end{aligned}
\tag{12}
$$

where I also use the result from FW that, e.g., $\sum_j \left(\partial^2_{\beta\alpha_i}\ell^*_{ij} - \partial^2_{\beta\alpha_i\alpha_i}\overline{\ell}^*_{ij}\right)\sum_j \partial_{\alpha_i}\ell_{ij}$ is the product of sums of order $\sqrt{N}$ whose constituent terms are mean zero but weakly correlated. By the weak law of large numbers, I therefore replace $\sum_j \partial^2_{\beta\alpha_i}\ell^*_{ij}\sum_j \partial_{\alpha_i}\ell_{ij}$ and $\sum_i \partial^2_{\beta\gamma_j}\ell^*_{ij}\sum_i \partial_{\gamma_j}\ell_{ij}$ respectively with $\sum_j \mathbb{E}(\partial^2_{\beta\alpha_i}\ell^*_{ij}\partial_{\alpha_i}\ell_{ij})$ and $\sum_i \mathbb{E}(\partial^2_{\beta\gamma_j}\ell^*_{ij}\partial_{\gamma_j}\ell_{ij})$.

Comparing (12) with the bias formulas given earlier shows that $\mathbb{E}\left[N^{-1}\partial_\beta \mathcal{L}^*\left(\beta^0, \widehat{\phi}\left(\beta^0\right)\right)\right] \approx B_N + D_N$. Further, as $N \to \infty$, the loss of asymptotically small terms ensures that the approximation becomes exact, i.e., $\lim_{N\to\infty} N^{-1}\partial_\beta \mathcal{L}^*\left(\beta^0, \widehat{\phi}\left(\beta^0\right)\right) = \lim_{N\to\infty} B_N + D_N = O_P(1)$. Finally, using a first-order expansion for $\widehat{\beta}$, we have

$$\begin{aligned}\left(\widehat{\beta} - \beta^0\right) &\approx \left(-n^{-1}\frac{\partial^2 \overline{\mathcal{L}}^*}{\partial\beta\partial\beta}\right)^{-1}\left(n^{-1}\frac{\partial \mathcal{L}^*\left(\beta^0, \widehat{\phi}\left(\beta^0\right)\right)}{\partial\beta}\right)\\ &= H_N^{-1}\left(n^{-1}\sum_{i,j}\frac{\partial \ell^*_{ij}(\beta^0, \phi^0)}{\partial\beta} + \frac{H_N^{-1}\left(B_N + D_N\right)}{N-1}\right),\end{aligned}$$

where $n = N(N-1)$ is used as shorthand for the full sample size. As $N \to \infty$, the product $N\left(\widehat{\beta} - \beta^0\right)$ converges to $H_N^{-1}\left(B_N + D_N\right)$, confirming $\widehat{\beta}$ has a bias of order $1/N$. Tracing this result back to the beginning of the derivation, this bias appears due to how the estimation noise in the fixed effects $\widehat{\phi}(\beta) - \phi^0$ enters the score for β.

2. (Corresponding expansion for the weighted estimator) As noted above, the bootstrap-weighted pseudolikelihood has a weighted score, Hessian, and other partial derivatives. For bootstrap draw b, the first-order expansion for $\widehat{\beta}^{(b)}$ can be obtained using

$$\begin{aligned}0 &\approx n^{-1}\frac{\partial \mathcal{L}^{*(b)}\left(\beta^0, \widehat{\phi}^{(b)}\left(\beta^0\right)\right)}{\partial\beta} + n^{-1}\frac{\partial^2 \mathcal{L}^{*(b)}}{\partial\beta\partial\beta}\left(\widehat{\beta}^{(b)} - \beta^0\right)\\ &\approx n^{-1}\frac{\partial \mathcal{L}^{*(b)}\left(\beta^0, \widehat{\phi}^{(b)}\left(\beta^0\right)\right)}{\partial\beta} + H_N\left(\widehat{\beta}^{(b)} - \beta^0\right),\end{aligned}$$

where in the second line I use $n^{-1}\partial^2_{\beta\beta}\mathcal{L}^{*(b)} = H_N + n^{-1}(\partial^2_{\beta\beta}\mathcal{L}^{*(b)} - \partial^2_{\beta\beta}\overline{\mathcal{L}}^*)$ and leave out the second component $n^{-1}(\partial^2_{\beta\beta}\mathcal{L}^{*(b)} - \partial^2_{\beta\beta}\overline{\mathcal{L}}^*)$ because of it is of small enough order. Therefore, we can approximate $\widehat{\beta}^{(b)} - \beta^0$ using

$$\left(\widehat{\beta}^{(b)} - \beta^0\right) \approx H_N^{-1}\left(n^{-1}\frac{\partial \mathcal{L}^{*(b)}\left(\beta^0, \widehat{\phi}^{(b)}\left(\beta^0\right)\right)}{\partial\beta}\right).$$

Next, like in the steps laid out in the previous remark, I expand of $\partial_\beta \mathcal{L}^{*(b)}\left(\beta_0, \widehat{\phi}^{(b)}\left(\beta_0\right)\right)$

around ϕ_0:

$$\frac{\partial \mathcal{L}^{*(b)}\left(\beta^0, \widehat{\phi}^{(b)}\left(\beta^0\right)\right)}{\partial \beta} = \sum_{i,j} W_{ij}^{(b)} \frac{\partial \ell^{*}\left(\beta^0, \widehat{\phi}^{(b)}\left(\beta^0\right)\right)}{\partial \beta}$$

$$\approx \sum_{i,j} W_{ij}^{(b)} \frac{\partial \ell_{ij}^{*}}{\partial \beta} + \sum_{i,j} W_{ij}^{(b)} \frac{\partial^2 \ell_{ij}^{*}}{\partial \beta \partial \phi'} \left(\widehat{\phi}^{(b)}(\beta) - \phi^0\right)$$

$$+ \frac{1}{2} \sum_{i,j} \left\{ \sum_{f,g}^{\dim \phi} W_{ij}^{(b)} \frac{\partial^3 \ell_{ij}^{*}}{\partial \beta \partial \phi_f \partial \phi_g} \left(\widehat{\phi}_f^{(b)}(\beta^0) - \phi_f^0\right) \left(\widehat{\phi}_g^{(b)}(\beta^0) - \phi_g^0\right) \right\}. \tag{13}$$

The estimation noise term $\widehat{\phi}^{(b)}(\beta) - \phi^0$ in this case may be replaced with

$$\widehat{\phi}^{(b)}(\beta) - \phi^0 \approx -\left(\frac{\partial^2 \overline{\mathcal{L}}^{*}}{\partial \phi \partial \phi'}\right)^{-1} \frac{\partial \mathcal{L}^{*(b)}}{\partial \phi} = \left(\frac{\partial^2 \overline{\mathcal{L}}^{*}}{\partial \phi \partial \phi'}\right)^{-1} \frac{\partial \left(\sum_{i,j} W_{i,j}^{(b)} \ell_{ij}^{*}\right)}{\partial \phi}.$$

Therefore, again applying Lemma D.1 from FW, terms such as $\widehat{\alpha}_i^{(b)} - \alpha_i^0$ are approximated with

$$\widehat{\alpha}_i^{(b)} - \alpha_i^0 \approx \left(-\sum_j \frac{\partial^2 \overline{\ell}_{ij}^{*}}{\partial \alpha_i \partial \alpha_i}\right)^{-1} \sum_j W_{ij}^{(b)} \frac{\partial \ell_{ij}^{*}}{\partial \alpha_i}.$$

Following the same steps as before, but now keeping track of the bootstrap weights, the expansion of $\partial_\beta \mathcal{L}^{*(b)}\left(\beta_0, \widehat{\phi}^{(b)}\left(\beta_0\right)\right)$ in (13) simplifies to:

$$\sum_{i,j} W_{ij}^{(b)} \frac{\partial \ell_{ij}^{*}}{\partial \beta} + \sum_i \left\{ -\left(\sum_j \frac{\partial^2 \overline{\ell}_{ij}^{*}}{\partial \alpha_i \partial \alpha_i}\right)^{-1} \sum_j \left(W_{ij}^{(b)}\right)^2 \mathbb{E} \frac{\partial^2 \ell_{ij}^{*}}{\partial \beta \partial \alpha_i} \frac{\partial \ell_{ij}^{*}}{\partial \alpha_i} + \frac{1}{2} \left(\sum_j \frac{\partial^2 \overline{\ell}_{ij}^{*}}{\partial \alpha_i \partial \alpha_i}\right)^{-2} \sum_j \left(W_{ij}^{(b)}\right)^2 \frac{\partial^3 \overline{\ell}_{ij}^{*}}{\partial \beta \partial \alpha_i \partial \alpha_i} \left(\frac{\partial}{\partial}\right.\right.$$

$$+ \sum_j \left\{ -\left(\sum_i \frac{\partial^2 \overline{\ell}_{ij}^{*}}{\partial \gamma_j \partial \gamma_j}\right)^{-1} \sum_i \left(W_{ij}^{(b)}\right)^2 \mathbb{E} \frac{\partial^2 \ell_{ij}^{*}}{\partial \beta \partial \gamma_j} \frac{\partial \ell_{ij}^{*}}{\partial \gamma_j} + \frac{1}{2} \left(\sum_i \frac{\partial^2 \overline{\ell}_{ij}^{*}}{\partial \gamma_j \partial \gamma_j}\right)^{-2} \sum_i \left(W_{ij}^{(b)}\right)^2 \frac{\partial^3 \overline{\ell}_{ij}^{*}}{\partial \beta \partial \gamma_j \partial \gamma_j} \left(\frac{\partial}{\partial}\right.\right.$$

As explained above, the presence of the $\left(W_{ij}^{(b)}\right)^2$ terms is the mechanism that amplifies the bias.

3. (Correctly specified likelihood) When the likelihood is correctly specified, the information equality $\text{Var}\left(\partial_{\alpha_i} \ell_{ij}^{*}\right) = -\partial_{\alpha_i \alpha_i} \overline{\ell}_{ij}^{*}$ may be used to simplify the elements of $B_N^{(b)}$ and

$D_N^{(b)}$ to

$$B_N^{(b),m} = \frac{1}{N}\sum_{i=1}^{N} \frac{\sum_{j\neq i} -\left(W_{ij}^{(b)}\right)^2 \mathbb{E}\left(\ell_{ij}^{*\beta_m\alpha_i}\ \ell_{ij}^{\alpha_i}\right) + \frac{1}{2}\sum_{j\neq i}\left(W_{ij}^{(b)}\right)^2 \overline{\ell}_{ij}^{*\beta_m\alpha_i\alpha_i}}{\sum_{j\neq i} W_{ij}^{(b)}\overline{\ell}_{ij}^{\alpha_i\alpha_i}},$$

$$D_N^{(b),m} = \frac{1}{N}\sum_{j=1}^{N} \frac{\sum_{i\neq j} -\left(W_{ij}^{(b)}\right)^2 \mathbb{E}\left(\ell_{ij}^{*\beta_m\gamma_j}\ \ell_{ij}^{\gamma_j}\right) + \frac{1}{2}\sum_{j\neq i}\left(W_{ij}^{(b)}\right)^2 \overline{\ell}_{ij}^{*\beta_m\gamma_j\gamma_j}}{\sum_{i\neq j} W_{ij}^{(b)}\overline{\ell}_{ij}^{\gamma_j\gamma_j}}.$$

4. (Three-way gravity models) In WZ, the bias of PPML for three-way gravity models in fixed T settings is shown to be very similar to that of a two-way estimator, except for the fact that y_{ij}, α_i, γ_j, and $x_{ij,m}$ are all T-vectors. For example, $y_{ij} = (y_{ij1}, ..., y_{ijT})$, $\alpha_i = (\alpha_{i1}, ..., \alpha_{iT})$, and so on. In the appropriate analogue of B_N^m, $\ell_{ij}^{*\beta_m\alpha_i}$ and $\ell_{ij}^{\alpha_i}$ become T-vectors, $\ell_{ij}^{\alpha_i\alpha_i}$ and $\ell_{ij}^{*\beta_m\alpha_i\alpha_i}$ become $T\times T$ matrices, and matrix inverses are used in place of dividing, with similar generalizations applying to the components of D_N^k. In addition, the trace operator is used to reduce the resulting matrix computations to scalars when computing the bias.

   Despite this added complexity, applying a cluster bootstrap that clusters by ij continues to produce the same bias magnification as in the two-way case. For fixed T while $N\to\infty$, the α-specific bias term $B_N^{(b)}$ has elements that look like this:

$$\begin{aligned} B_N^{(b),m} &= -\frac{1}{N}\sum_{i=1}^{N} \mathrm{Tr}\left[\left(\sum_{j}^{N}\overline{\ell}_{ij}^{*\alpha_i\alpha_i}\right)^{-1}\sum_{j}^{N}\left(W_{ij}^{(b)}\right)^2\mathbb{E}\left(\ell_{ij}^{*\beta_m\alpha_i}\ell_{ij}^{\alpha_i\prime}\middle|x_{ij,m}\right)\right] \\ &+ \frac{1}{2\,N}\sum_{i=1}^{N}\mathrm{Tr}\left[\left(\sum_{j}^{N}\overline{\ell}_{ij}^{*\alpha_i\alpha_i}\right)^{-1}\left(\sum_{j}^{N}\overline{\ell}_{ij}^{*\beta_m\alpha_i\alpha_i}\right)\left[\sum_{j}^{N}\left(W_{ij}^{(b)}\right)^2\mathbb{E}\left(\ell_{ij}^{\alpha_i}\ \ell_{ij}^{\alpha_i\prime}\middle|x_{ij,m}\right)\right]\left(\sum_{j}^{N}\overline{\ell}_{ij}^{*\alpha_i\alpha_i}\right)^{-1}\right]. \end{aligned}$$

   Once again, I omit the analogous expression for $D_N^{(b),m}$, which requires exchanging i with j and α_i with γ_j where appropriate. We can see again that the inclusion of the $\left(W_{ij}^{(b)}\right)^2$ terms will inflate the asymptotic bias by a factor of 2, just like in the two-way case.

Proof of Proposition 2

Drawing on the expansions derived for $\widehat{\beta}$ and $\widehat{\beta}^{(b)}$ derived in Remarks 1 and 2 above, the bootstrap bias-corrected estimate can be analyzed using

$$\widehat{\beta}^{BBC} - \beta^0 = 2\left(\widehat{\beta} - \beta^0\right) - \left(\overline{\beta}^B - \beta^0\right).$$
$$\approx \frac{H_N^{-1}}{N(N-1)} \sum_{i,j} \left(2\frac{\partial \ell_{ij}^*}{\partial \beta} - \mathbb{E}_W\left(W_{ij}^{(b)} \frac{\partial \ell_{ij}^*}{\partial \beta}\right)\right),$$

where $\mathbb{E}_W$ denotes the expectation over the distribution of the bootstrap weights across bootstrap trials. Therefore,

$$\widehat{\beta}^{BBC} - \beta^0 \approx \frac{H_N^{-1}}{N(N-1)} \sum_{i,j} \frac{\partial \ell_{ij}^*}{\partial \beta}$$
$$\Longrightarrow \text{Var}\left(\widehat{\beta}^{BBC} - \beta^0\right) = \text{Var}\left(\widehat{\beta} - \beta^0\right) = \frac{1}{N(N-1)} H_N^{-1} \Omega_N H_N^{-1}.$$

This result confirms that the bootstrap bias correction does not inflate the asymptotic variance, as stated in Proposition 2. Moreover, $\widehat{\beta}^{BBC}$ is unbiased.

Examples where the bootstrap does not double the bias

Here, I focus on two examples where dependence in the error term causes the bootstrap bias correction to fail: (i) a standard panel data model with weak time dependence, (ii) a two-way gravity model where errors are correlated for pairs sharing the same exporter or the same importer. I will also show that using a cluster bootstrap will not be effective for correcting the bias either.

First, the standard panel data model is given by

$$y_{it} = \exp(\alpha_i + x_{it}\beta)\omega_{it}, \qquad E(y_{it}|x_{it}, \alpha_i) = \exp(\alpha_i + x_{it}\beta),$$

where $i \in 1, ..., N$ in this example indexes the units in the panel and t indexes time and runs from $1, ..., T$. Cross-sectional independence in this case means $E(\omega_{it}\omega_{jt}) = 0$ if $i \neq j$. To introduce weak time dependence, let $E(\omega_{it}\omega_{is})$ generally be positive for $s \neq t$ but decreasing in $|s - t|$ and vanishing to 0 for observations that are sufficiently distant from one another in time. As $N, T \to \infty$, a generic fixed effects PML estimator for this model has a $1/T$-

asymptotic bias given by $(1/T)H^{-1}\psi$, where each element of ψ_i is given by:

$$\begin{aligned}
\psi^m &= \frac{1}{N}\sum_i \mathbb{E}\left[\sum_t \ell_{it}^{*\beta_m\alpha_i}\left(\widehat{\alpha}_i-\alpha_i^0\right)+\frac{1}{2}\left(\sum_{t=1}^T \overline{\ell}_{it}^{*\beta_k\alpha_i\alpha_i}\right)\left(\widehat{\alpha}_i-\alpha_i^0\right)^2\right]\\
&\approx -\frac{1}{N}\sum_i \frac{\sum_{t=1}^T \mathbb{E}\left(\sum_t \ell_{it}^{*\beta_k\alpha_i}\sum_s \ell_{is}^{\alpha_i}\right)}{\sum_{t=1}^T \overline{\ell}_{it}^{\alpha_i\alpha_i}}+\frac{1}{2N}\sum_i \frac{\left(\sum_{t=1}^T \overline{\ell}_{it}^{*\beta_k\alpha_i\alpha_i}\right)\mathbb{E}\left[\sum_{t=1}^T \left(\ell_{it}^{\alpha_i}\right)\right]^2}{\left(\sum_{t=1}^T \overline{\ell}_{it}^{\alpha_i\alpha_i}\right)^2}\\
&= -\frac{1}{N}\sum_i \frac{\sum_{t=1}^T\sum_{s=1}^T \mathbb{E}\left(\ell_{it}^{*\beta_k\alpha_i}\ell_{is}^{\alpha_i}\right)}{\sum_{t=1}^T \overline{\ell}_{it}^{\alpha_i\alpha_i}}+\frac{1}{2N}\sum_i \frac{\left(\sum_{t=1}^T \overline{\ell}_{it}^{*\beta_k\alpha_i\alpha_i}\right)\left[\sum_{t=1}^T\sum_{s=1}^T \mathbb{E}\left(\ell_{it}^{\alpha_i}\ell_{is}^{\alpha_i}\right)\right]}{\left(\sum_{t=1}^T \overline{\ell}_{it}^{\alpha_i\alpha_i}\right)^2},
\end{aligned} \tag{14}$$

and H, ℓ_{it},ℓ_{it}^*, etc. refer to the appropriate analogues of the terms depending on the Hessian and likelihood that appear in the bias formulas used throughout the rest of the paper. The spotlight is on the score bias term ψ, which is where the weights used by the bootstrap meaningfully enter. The second line uses the approximation $\widehat{\alpha}_i-\alpha_i^0 \approx (\sum_{t=1}^T \overline{\ell}_{it}^{\alpha_i\alpha_i})^{-1}\ell_{it}^{\alpha_i}$. In the 3rd line, dependence across time causes $\mathbb{E}\left(\sum_t \ell_{it}^{*\beta_k\alpha_i}\sum_s \ell_{is}^{\alpha_i}\right)$ to become $\sum_{t=1}^T\sum_{s=1}^T \mathbb{E}\left(\ell_{it}^{*\beta_k\alpha_i}\ell_{is}^{\alpha_i}\right)$ and $\mathbb{E}\left[\sum_{t=1}^T \left(\ell_{it}^{\alpha_i}\right)\right]^2$to become $\sum_{s=1}^T\sum_{t=1}^T \mathbb{E}\left(\ell_{is}^{\alpha_i}\ell_{it}^{\alpha_i}\right)$.

For given set of bootstrap replications $b=1,...,B$, each bootstrap estimate is obtained by maximizing $\sum_{i,t}\ell_{it}^{(b)}=\sum_{i,t}W_{it}^{(b)}\ell_{it}$, where each $W_{it}^{(b)}$ is sampled randomly and independently and satisfies $\mathbb{E}_W(W_{it}^{(b)})=1$, $\mathbb{E}_W(W_{it}^{(b)2})=2$. Importantly, because of independence, we also have that $\mathbb{E}_W(W_{is}^{(b)}W_{it}^{(b)})=1$ for $s\neq t$. Replacing ℓ_{it} with $\ell_{it}^{(b)}$, and using the same arguments as Section 3.2, the bias of each bootstrap estimate as $N,T\to\infty$ can be expressed as $(1/T)H^{-1}\psi^{(b)}$, where each element of $\psi^{(b)}$ has the following limit:

$$\begin{aligned}
\lim_{N,T\to\infty}\psi^{(b),m} = \lim_{N,T\to\infty} &-\frac{1}{N}\sum_i \frac{\sum_{t=1}^T 2\mathbb{E}\left(\ell_{is}^{*\beta_k\alpha_i}\ell_{it}^{\alpha_i}\right)}{\sum_{t=1}^T \overline{\ell}_{it}^{\alpha_i\alpha_i}}+\frac{1}{N}\sum_i \frac{\sum_{t=1}^T\sum_{s\neq t} \mathbb{E}\left(\ell_{it}^{*\beta_k\alpha_i}\ell_{is}^{\alpha_i}\right)}{\sum_{t=1}^T \overline{\ell}_{it}^{\alpha_i\alpha_i}}\\
&+\frac{1}{2N}\sum_i \frac{\left(\sum_{t=1}^T \overline{\ell}_{it}^{*\beta_k\alpha_i\alpha_i}\right)\left[\sum_{t=1}^T 2\mathbb{E}\left(\ell_{it}^{\alpha_i}\ell_{it}^{\alpha_i}\right)\right]}{\left(\sum_{t=1}^T \overline{\ell}_{it}^{\alpha_i\alpha_i}\right)^2}\\
&+\frac{1}{2N}\sum_i \frac{\left(\sum_{t=1}^T \overline{\ell}_{it}^{*\beta_k\alpha_i\alpha_i}\right)\left[\sum_{t=1}^T\sum_{s\neq t} \mathbb{E}\left(\ell_{is}^{\alpha_i}\ell_{it}^{\alpha_i}\right)\right]}{\left(\sum_{t=1}^T \overline{\ell}_{it}^{\alpha_i\alpha_i}\right)^2}.
\end{aligned} \tag{15}$$

Unlike in the case of gravity models with cross-sectional independence, the bootstrap does not double the bias. The reason is that, for $s\neq t$, $\ell_{it}^{(b)*\beta_k\alpha_i}\ell_{is}^{(b)\alpha_i}=W_{is}^{(b)}W_{it}^{(b)}\ell_{it}^{*\beta_k\alpha_i}\ell_{is}^{\alpha_i}$ and $\ell_{is}^{(b)\alpha_i}\ell_{it}^{(b)\alpha_i}=W_{is}^{(b)}W_{it}^{(b)}\ell_{it}^{\alpha_i}\ell_{it}^{\alpha_i}$. The expectations of these terms do not produce a 2 as $T\to\infty$ because of the independence of $W_{is}^{(b)}$ and $W_{it}^{(b)}$. As (15) shows, the larger the degree of time dependence, the less accurate the bootstrap bias correction (which assumes a doubling of the bias) becomes.

The characterization of the bootstrap for this example thus far has assumed it samples observations randomly and ignores potential dependence in the data. A possible alternative is a cluster bootstrap that, like the cluster bootstrap used for the three-way gravity model, samples cross-sectional units instead of individual observations. Let the weight the bootstrap applies to each unit in this case be given by $W_i^{(b)}$. Replacing ℓ_{it} with $W_i^{(b)}\ell_{it}$ in (14), and then simplifying, leads to

$$-\frac{1}{N}\sum_i \frac{W_i^{(b)}\sum_{t=1}^T\sum_{s=1}^T \mathbb{E}\left(\ell_{it}^{*\beta_k\alpha_i}\,\ell_{is}^{\alpha_i}\right)}{\sum_{t=1}^T \overline{\ell}_{it}^{\alpha_i\alpha_i}} + \frac{1}{2N}\sum_i \frac{W_i^{(b)}\left(\sum_{t=1}^T \overline{\ell}_{it}^{*\beta_k\alpha_i\alpha_i}\right)\left[\sum_{t=1}^T\sum_{s=1}^T \mathbb{E}\left(\ell_{it}^{\alpha_i}\,\ell_{is}^{\alpha_i}\right)\right]}{\left(\sum_{t=1}^T \overline{\ell}_{it}^{\alpha_i\alpha_i}\right)^2}.$$

As $N, T \to \infty$, the cluster bootstrap does not amplify the bias of the original estimator and therefore cannot be used for bias correction. This echoes prior results for the cluster bootstrap for panel data models by Gonçalves and Kaffo (2015) and Kaffo (2015), ch. 2.

Turning to another example, consider the same gravity model introduced in Section 2.1, only with errors for observations sharing the same exporter or the same importer. That is, suppose that everything is the same as in Section 2.1 except $Cov(\omega_{ij}\omega_{ij'}) \neq 0$ for $j \neq j'$, $Cov(\omega_{ij}\omega_{i'j}) \neq 0$ for $i \neq i'$, such that cross-sectional independence no longer holds. A possible reason for this type of dependence could be if trade flows from similar locations are spatially correlated. Drawing on the previous example, the bias constants originally described in in Section 2.1 now become

$$B_N^m = -\frac{1}{N}\sum_{i=1}^N \frac{\sum_{j\neq i}\sum_{k\neq i}\mathbb{E}\left(\ell_{ij}^{*\beta_m\alpha_i}\,\ell_{ik}^{\alpha_i}\right)}{\sum_{j\neq i}^N \overline{\ell}_{ij}^{\alpha_i\alpha_i}} + \frac{1}{2\,N}\sum_{i=1}^N \frac{\left(\sum_{j\neq i}\overline{\ell}_{ij}^{*\beta_m\alpha_i\alpha_i}\right)\left[\sum_{j\neq i}\sum_{k\neq i}\mathbb{E}\left(\ell_{ij}^{\alpha_i}\,\ell_{ik}^{\alpha_i}\right)\right]}{\left(\sum_j^N \overline{\ell}_{ij}^{\alpha_i\alpha_i}\right)^2},$$

$$D_N^m = -\frac{1}{N}\sum_{j=1}^N \frac{\sum_{i\neq j}\sum_{k\neq j}\mathbb{E}\left(\ell_{ij}^{*\beta_m\gamma_j}\,\ell_{kj}^{\gamma_j}\right)}{\sum_{i\neq j}\overline{\ell}_{ij}^{\gamma_j\gamma_j}} + \frac{1}{2\,N}\sum_{j=1}^N \frac{\left(\sum_{i\neq j}\overline{\ell}_{ij}^{*\beta_m\gamma_j\gamma_j}\right)\left[\sum_{i\neq j}\sum_{k\neq j}\mathbb{E}\left(\ell_{ij}^{\gamma_j}\,\ell_{kj}^{\gamma_j}\right)\right]}{\left(\sum_{i\neq j}\overline{\ell}_{ij}^{\gamma_j\gamma_j}\right)^2}$$

Exactly like in the previous example, the replacement of terms such as $\mathbb{E}\left(\ell_{ij}^{*\beta_k\alpha_i}\,\ell_{ik}^{\alpha_i}\right)$ and $\mathbb{E}\left(\ell_{ij}^{\alpha_i}\,\ell_{ik}^{\alpha_i}\right)$ with $\mathbb{E}\left(W_{ij}^{(b)}W_{ik}^{(b)}\ell_{ij}^{*\beta_k\alpha_i}\,\ell_{ik}^{\alpha_i}\right)$ and $\mathbb{E}\left(W_{ij}^{(b)}W_{ik}^{(b)}\ell_{ij}^{\alpha_i}\,\ell_{ik}^{\alpha_i}\right)$ when the data is bootstrapped will not double the contribution of these terms to the bias. Using a cluster bootstrap that clusters by exporter and/or importer will not correct the bias either.

More Details on Implementation of k-step Bootstrap

To economize on the amount of notation needed, I will first describe an iteratively re-weighted least squares (IRLS) algorithm for unweighted log-link GLM estimation with two-way fixed effects. The point will be to show that the estimate values produced by the kth iteration from the IRLS algorithm is equivalent to the k-step estimator from Kim and Sun (2016) when the expected Hessian is used in place of the observed Hessian. As in Correia et al. (2020), the advantage of the IRLS approach is how it facilitates computation of the fixed effects estimates. I will then explain how the k-step bootstrap uses a similar algorithm that incorporates bootstrap weights for each bootstrap repetition.

IRLS algorithm. Consider the following (unweighted) least squares minimization problem:

$$(\widehat{\beta}_{[k]}, \widehat{\alpha}_{[k]}, \widehat{\gamma}_{[k]}) = \arg\min_{\beta,\alpha,\gamma} \sum_{i,j} \varphi_{ij[k-1]} \left(z_{ij[k-1]} - \alpha_i - \gamma_j - \widetilde{x}_{ij[k-1]}\beta \right)^2, \tag{16}$$

where

$$z_{ij[k-1]} = \frac{y_{ij} - \widehat{\mu}_{ij[k-1]}}{\widehat{\mu}_{ij[k-1]}} + \ln \widehat{\mu}_{ij[k-1]}, \tag{17}$$

$$\varphi_{ij[k-1]} = -\overline{\ell}^{\alpha_i\alpha_i}_{ij[k-1]} = -\overline{\ell}^{\gamma_j\gamma_j}_{ij[k-1]}, \tag{18}$$

$$\widehat{\mu}_{ij[k-1]} = e^{\widehat{\alpha}_{i[k-1]} + \widehat{\gamma}_{j[k-1]} + \widetilde{x}_{ij[k-1]}\widehat{\beta}_{[k-1]}}. \tag{19}$$

and $\widetilde{x}_{ij[k-1]}$ is a residualized version of x_{ij} that satisfies $\sum_i \varphi_{ij[k-1]}\widetilde{x}_{ij[k-1]} = 0\ \forall j$, $\sum_j \varphi_{ij[k-1]}\widetilde{x}_{ij[k-1]} = 0\ \forall i$. It is obtained by partialing out x_{ij} with respect to the fixed effects and weighting by $\varphi_{ij[k-1]}$. Importantly, replacing $\widetilde{x}^{(b)}_{ij[k-1]}$ with x_{ij} in (16) yields the same value for $\widehat{\beta}^{(b)}_{[k]}$.

Together, equations (16)-(19) define a recursive updating step that can be used to iteratively solve for the PML estimates $(\widehat{\beta}, \widehat{\alpha}, \widehat{\gamma})$ using IRLS. To see this, note that the properties of log-link generalized linear models imply that the solution to (16) satisfies

$$\widehat{\beta}_{[k]} = \widehat{\beta}_{[k-1]} - \left(\partial_{\beta\beta}\overline{\mathcal{L}}^*_{[k-1]}\right)^{-1} \left(\partial_{\beta}\mathcal{L}^*{}_{[k-1]}\right) \tag{20}$$

$$\begin{pmatrix} \widehat{\alpha}_{[k]} \\ \widehat{\gamma}_{[k]} \end{pmatrix} = \begin{pmatrix} \widehat{\alpha}_{[k-1]} \\ \widehat{\gamma}_{[k-1]} \end{pmatrix} - \begin{pmatrix} \partial_{\alpha\alpha}\overline{\mathcal{L}}_{[k-1]} & \partial_{\alpha\gamma}\overline{\mathcal{L}}_{[k-1]} \\ \partial_{\gamma\alpha}\overline{\mathcal{L}}_{[k-1]} & \partial_{\gamma\gamma}\overline{\mathcal{L}}_{[k-1]} \end{pmatrix}^{-1} \begin{pmatrix} \partial_{\alpha}\mathcal{L}^{(b)}_{[k-1]} \\ \partial_{\gamma}\mathcal{L}^{(b)}_{[k-1]} \end{pmatrix}, \tag{21}$$

which differs from the updating step presented in the main text in that it uses the expected

Hessian terms $\partial_{\beta\beta}\overline{\mathcal{L}}^*_{[k-1]}$, $\partial_{\alpha\alpha}\overline{\mathcal{L}}_{[k-1]}$, $\partial_{\alpha\gamma}\overline{\mathcal{L}}_{[k-1]}$ in place of their observed equivalents.[17] As discussed in Kim and Sun (2016), using the expected Hessian in place of the observed one makes no difference for asymptotic results regarding the k-step bootstrap. It also differs in that it does not yet incorporate bootstrap weights, discussed below.

In practice, we can use the fact that (16) is a least squares minimization problem for a linear model to greatly simplify computation. Like in Correia et al. (2020), I use within-transformations based on the Frisch-Waugh-Lovell (FWL) theorem to sweep out the fixed effects from the estimation of (16) so that no inversion of the Hessian matrix in (21) is required. Given values for $\widehat{\mu}_{ij[k-1]}$ from the previous updating step, the procedure for obtaining the k-step estimate $\widehat{\beta}_{[k]}$ is as follows:

1. Obtain values for $z_{ij[k-1]}$ and $\varphi_{ij[k-1]}$ using (17) and (18).
2. Partial out x_{ij} using $\min_{a_i^x,\gamma_j^x}\sum_{i,j}\varphi_{ij[k-1]}\left(x_{ij}-a_i^x-\gamma_j^x\right)^2$. The residuals from this step give us $\widetilde{x}_{ij[k-1]}$.
3. Similarly, partial out $z_{ij[k-1]}$ using $\min_{a_i^z,\gamma_j^z}\sum_{i,j}\varphi_{ij[k-1]}\left(z_{ij}-a_i^z-\gamma_j^z\right)^2$ and call the residuals from this step $\widetilde{z}_{ij[k-1]}$.
4. Update the k-step $\widehat{\beta}_{[k]}$ using

$$\min_{\beta}\sum_{i,j}\varphi_{ij[k-1]}\left(\widetilde{z}_{ij[k-1]}-\widetilde{x}_{ij[k-1]}\beta\right)^2 \tag{22}$$

5. (if needed) Update $\widehat{\mu}_{ij[k]}$ using $\widehat{\mu}_{ij[k]}=\exp(z_{ij[k-1]}-e_{ij[k]})$, where $e_{ij[k]}=\widetilde{z}_{ij[k-1]}-\widetilde{x}_{ij[k-1]}\widehat{\beta}_{[k]}$.

The reason why this works is a simple application of the FWL theorem. Since steps 2 and 3 project out the fixed effects from the working dependent variable z and principal regressor x, regressing $\widetilde{z}_{ij[k-1]}$ on $\widetilde{x}_{ij[k-1]}$ in step 4 produces the same result as though we had regressed z on x and also included all of the fixed effects. The computational benefit lies in steps 2 and 3, which can be quickly carried out using within transformations instead of computing explicit least squares estimates for the fixed effects. The updating of the $\widehat{\mu}_{ij[k]}$ is taken from Correia et al. (2020) and also follows from the FWL theorem, which holds that the residuals from

[17] To add some illustration: for a generic generalized linear model with $\mu_i=e^{a+bx_i}$, we have that $\ell_i^b=\varphi_i(y_i-\mu_i)\times\partial_b\mu_i=\varphi_i(y_i-\mu_i)\mu_i^{-1}x_i$, $\ell_i^{bb}=-\varphi_i x_i x_i'+(y_i-\mu_i)\times\partial_b(\varphi_i/\mu_i)$, $\overline{\ell}_i^{bb}=-\varphi_i x_i x_i'$, where $\varphi_i=-\overline{\ell}_i^{aa}$. For more details, see Hardin and Hilbe (2018).

(22) are the same as the residuals that would be obtained from estimating the full model that includes all of the fixed effects. Relative to Kim and Sun (2016), who use a solution method that requires inverting the full Hessian, this approach has the advantage that it can handle larger numbers of fixed effects much more easily. It should also generally be faster to compute.

***k*-step bootstrap algorithm**. For each bootstrap replication b, the k-step bootstrap estimates can be obtained using

$$(\widehat{\beta}^{(b)}_{[k]}, \widehat{\alpha}^{(b)}_{[k]}, \widehat{\gamma}^{(b)}_{[k]}) = \arg\min_{\beta,\alpha,\gamma} \sum_{i,j} W^{(b)}_{ij} \varphi_{ij[k-1]} \left(z^{(b)}_{ij[k-1]} - \alpha_i - \gamma_j - \widetilde{x}^{(b)}_{ij[k-1]}\beta \right)^2, \tag{23}$$

which in turn can be reduced via the FWL theorem to

$$\widehat{\beta}^{(b)}_{[k]} = \arg\min_{\beta} \sum_{i,j} W^{(b)}_{ij} \varphi_{ij[k-1]} \left(\widetilde{z}^{(b)}_{ij[k-1]} - \widetilde{x}^{(b)}_{ij[k-1]}\beta \right)^2, \tag{24}$$

where $\widetilde{x}^{(b)}_{ij[k-1]}$ and $\widetilde{z}^{(b)}_{ij[k-1]}$ can be obtained by adding the $W^{(b)}_{ij}$ bootstrap weights as additional weighting terms to steps 2 and 3 in the unweighted IRLS algorithm described above. The working dependent variable $z^{(b)}_{ij[k-1]}$ is obtained using

$$z^{(b)}_{ij[k-1]} = \frac{y_{ij} - \widehat{\mu}^{(b)}_{ij[k-1]}}{\widehat{\mu}^{(b)}_{ij[k-1]}} + \ln \widehat{\mu}^{(b)}_{ij[k-1]},$$

where $\widehat{\mu}^{(b)}_{ij[k-1]}$ is obtained using $\widehat{\mu}^{(b)}_{ij[k]} = \exp(z^{(b)}_{ij[k-1]} - \widetilde{z}^{(b)}_{ij[k-1]} + \widetilde{x}^{(b)}_{ij[k-1]}\widehat{\beta}^{(b)}_{[k]})$ like in step 5. As explained in the main text, the IRLS algorithm for implementing the k-step bootstrap is initialized using $\widehat{\mu}^{(b)}_{ij[0]} = \widehat{\mu}_{ij}$. Setting k to a fixed integer produces a $k-$step bootstrap estimate. Letting the algorithm iterate until convergence produces a regular bootstrap estimate.