



## Full Length Articles

## Bias and consistency in three-way gravity models

Martin Weidner<sup>a</sup>, Thomas Zylkin<sup>b,\*</sup><sup>a</sup> Nuffield College and Department of Economics, University of Oxford, Oxford OX1 3UQ, USA<sup>b</sup> Robins School of Business, University of Richmond, Richmond, VA 23217, USA

## ARTICLE INFO

## Article history:

Received 20 January 2020

Received in revised form 22 June 2021

Accepted 25 June 2021

Available online 6 July 2021

Repository link: <https://www.dropbox.com/s/8zchjamfg2bqipd/Code%20repository%20for%20Weidner%20and%20Zylkin%20JIE%202021.zip?dl=0>

## JEL classification codes:

C13

C50

F10

## Keywords:

Structural gravity

Trade agreements

Asymptotic bias correction

## ABSTRACT

We study the incidental parameter problem for the “three-way” Poisson Pseudo-Maximum Likelihood (“PPML”) estimator recently recommended for identifying the effects of trade policies and in other panel data gravity settings. Despite the number and variety of fixed effects involved, we confirm PPML is consistent for fixed  $T$  and we show it is in fact the only estimator among a wide range of PML gravity estimators that is generally consistent in this context when  $T$  is fixed. At the same time, asymptotic confidence intervals in fixed- $T$  panels are not correctly centered at the true parameter values, and cluster-robust variance estimates used to construct standard errors are generally biased as well. We characterize each of these biases analytically and show both numerically and empirically that they are salient even for real-data settings with a large number of countries. We also offer practical remedies that can be used to obtain more reliable inferences of the effects of trade policies and other time-varying gravity variables, which we make available via an accompanying Stata package called `ppml_f_bias`.

© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Despite intense and longstanding empirical interest, the effects of bilateral trade agreements on trade are still considered highly difficult to assess. As emphasized in a recent practitioner's guide put out by the WTO (Yotov et al., 2016), many current estimates in the literature suffer from easily identifiable sources of bias (or “estimation challenges”). This is not for lack of awareness. Papers showing leading causes of bias in the gravity equation are often among the most widely celebrated and cited in the trade field, if not in all of Economics.<sup>1</sup> In particular, it is now generally accepted that trade flows across different partners are interdependent via the network structure of trade (the main contribution of Anderson and van Wincoop, 2003), that log-transforming the dependent variable is not innocuous (as argued by Santos Silva and Tenreyro, 2006), and—most relevant

\* Corresponding author.

E-mail addresses: [martin.weidner@ox.ac.uk](mailto:martin.weidner@ox.ac.uk), [martin.weidner@economics.ox.ac.uk](mailto:martin.weidner@economics.ox.ac.uk) (M. Weidner), [tzylkin@richmond.edu](mailto:tzylkin@richmond.edu), [tzylkin@richmond.edu](mailto:tzylkin@richmond.edu) (T. Zylkin).

<sup>1</sup> For some context, if we start citation counts in 2003, Anderson and van Wincoop (2003) and Santos Silva and Tenreyro (2006) are, respectively, the most cited articles in the *American Economic Review* and the *Review of Economics and Statistics*. Paling only slightly in this exclusive company, Baier and Bergstrand (2007) is the 4th most-cited article in the *Journal of International Economics*, having gathered “only” 2700 citations. Readers familiar with these other papers will also likely be familiar with Helpman et al. (2008)'s work on the selection process underlying zero trade flows, an issue we do not take up here.

to the context of trade agreements—that earlier, puzzlingly small estimates of the effects of free trade agreements were almost certainly biased downwards by treating them as exogenous (Baier and Bergstrand, 2007).

As a consequence—and aided by some recent computational developments—researchers seeking to identify the effects of trade agreements have naturally moved towards more advanced estimation strategies that take on board all of the above concerns.<sup>2</sup> In particular, a “three-way” fixed effects Poisson Pseudo-Maximum Likelihood (“FE-PPML”) estimator with time-varying exporter and importer fixed effects to account for network dependence and time-invariant exporter-importer (“pair”) fixed effects to address endogeneity has recently emerged as a logical workhorse method for empirical trade policy analysis.<sup>3</sup> It also has clear potential application to the study of network data more generally, such as data on urban commuting or migration (e.g., Brinkman and Lin, 2019; Gaduh et al., 2020; Allen et al., 2018; Beverelli and Orefice, 2019).

However, one reason why some researchers may hesitate in embracing this estimator is the current lack of clarity regarding how the three fixed effects in the model may bias estimation, especially in the standard “fixed  $T$ ” case where the number of time periods is small. Even though FE-PPML estimates can be shown to be asymptotically unbiased with a single fixed effect (a well-known result) as well as in a two-way setting where both dimensions of the panel become large (Fernández-Val and Weidner, 2016), the latter result does not come strictly as a generalization of the former one, leaving it potentially unclear whether a three-way model with a fixed time dimension should be expected to inherit the nice asymptotic properties of these other models.

Accordingly, the question we investigate in this paper is the extent to which the three-way FE-PPML estimator is affected by incidental parameter problems (IPPs). As is well known in both statistics and econometrics (Neyman and Scott, 1948; Lancaster, 2002), IPPs arise when estimation noise from estimates of fixed effects and other “incidental parameters” contaminates the scores of the main parameters of interest, inducing bias. In the worst case, this bias renders the estimates inconsistent, making estimation inadvisable. As we will show, while inconsistency is actually not a problem for three-way FE-PPML, both the estimated coefficients and standard errors are affected by meaningful biases due to IPPs that researchers should be aware of.

To state our main results more precisely, in gravity settings where the number of countries ( $N$ ) goes to infinity and  $T$  is small, we find the following:

1. Consistency of point estimates of FE-PPML: The point estimates produced by three-way FE-PPML estimator in gravity settings are asymptotically consistent.
2. Inconsistency of other FE-PML estimators: FE-PPML is the only estimator in a set of related FE-PML estimators sometimes considered in this context that is generally consistent. FE-Gamma PML, for example, should not be used because it is only consistent under strict assumptions.
3. Asymptotic bias: Point estimates of the three-way FE-PPML estimator are nonetheless *asymptotically biased*, meaning that the asymptotic distribution of the estimates is not centered at the truth as  $N \rightarrow \infty$ . In other words, it approaches the truth “at an angle” asymptotically (see Figure 1 for an illustration.)
4. Biased standard error estimates: Estimates of cluster-robust sandwich-type standard errors are likewise asymptotically biased due to an IPP.
5. Bias corrections improve inferences: Simulations show that using analytical bias corrections to address each of these biases leads to improved inferences. These corrections are available to use via the Stata package `ppml_fe_bias`.

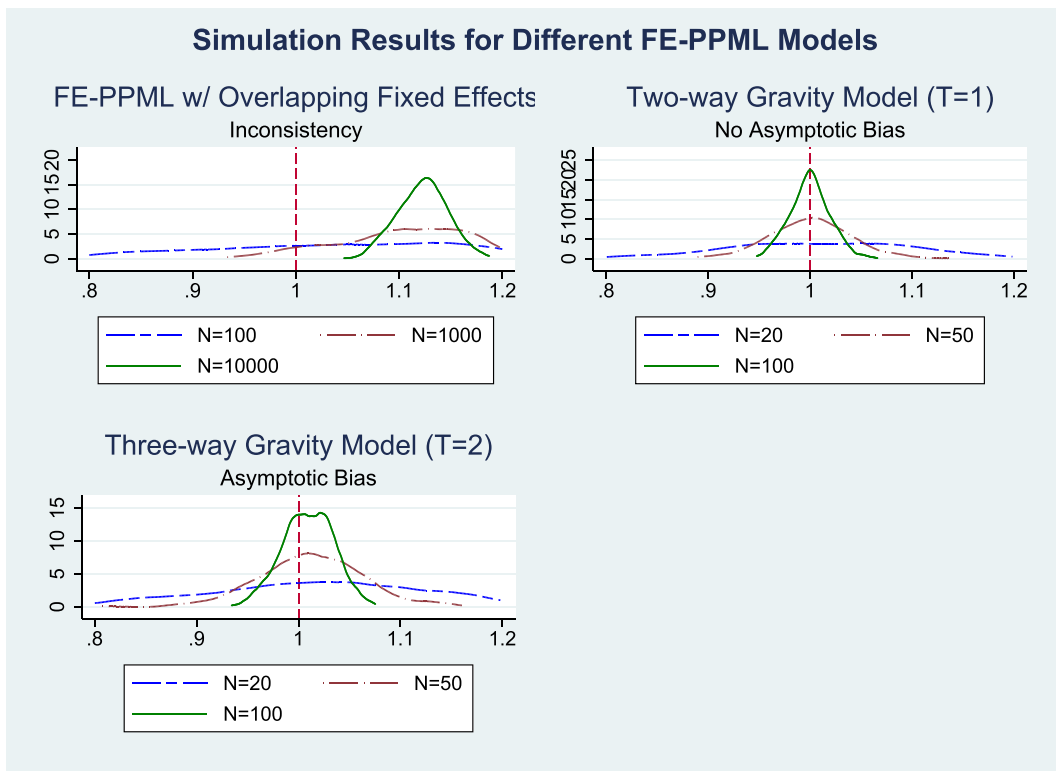
To first explain our consistency results, our basic strategy involves using the first-order conditions of FE-PPML to “profile out” (solve for) the pair fixed effect terms from the first-order conditions of the other parameters. Notably, this allows us to re-express the three-way gravity model as a two-way model in which the only remaining incidental parameters are the exporter-time and importer-time fixed effects. Three-way FE-PPML is therefore consistent in fixed- $T$  settings for largely the same reasons the two-way models considered in Fernández-Val and Weidner (2016) are consistent, and we provide suitably modified versions of the regularity conditions and consistency results established by Fernández-Val and Weidner (2016) for the simpler two-way case.

At the same time, it does not also follow that Fernández-Val and Weidner (2016)’s earlier results for the *asymptotically unbiasedness* of the two-way FE-PPML estimator similarly carry over to the three-way case when  $T$  is fixed. The key is that the resulting two-way estimator that is obtained after profiling out the pair fixed effects has its own special properties with respect to IPPs. When  $T$  is fixed, the estimation noise in the remaining exporter-time and importer-time fixed effects induces an asymptotic bias of order  $1/N$  as  $N$  grows large, a result that is broadly consistent with most of the two-way settings studied in Fernández-Val and Weidner (2016). However, when instead both  $N$  and  $T$  grow large at the same rate, the estimator turns out to be unbiased asymptotically, analogous to what Fernández-Val and Weidner (2016) found in the two-way FE-PPML case.

The reason why asymptotic bias is a concern is that the asymptotic standard deviation is itself of order  $1/(N\sqrt{T})$ . Thus, when  $T$  is fixed, the bias in each point estimate will be of comparable magnitude to its standard error as  $N \rightarrow \infty$ , causing the asymptotic distribution of estimates to be incorrectly centered as discussed above. In practice, this is a less severe problem than inconsistency, but it does mean that standard hypothesis tests for assessing statistical significance are not reliable. One of the objectives of this

<sup>2</sup> Larch et al. (2019), Correia et al. (2020), and Stammann (2018) describe algorithms that enable fast estimation of the three-way models considered here.

<sup>3</sup> Pair fixed effects are of course no substitute for good instruments. However, instruments for trade policy changes which are also exogenous to trade are understandably hard to come by. As discussed in Head and Mayer (2014)’s essential handbook chapter on gravity estimation, pair fixed effects have the advantage that the effects of trade agreements and other trade policies are identified from time-variation in trade within pairs. Causal interpretations follow if standard “parallel trend” assumptions are satisfied.



**Fig. 1.** Kernel density plots of FE PPML estimates for 3 different models, using 500 replications. Clockwise from top left, the 3 models are:  $y_{it} = \exp[\alpha_i \times 1(t \leq 2) + \gamma_i \times 1(t \geq 2) + x_{it}\beta]\omega_{it}$ , with the  $t$  dimension of the panel fixed at  $T = 3$ ; a two-way gravity model with  $y_{ij} = \exp[\alpha_i + \gamma_j + x_{ij}\beta]\omega_{ij}$ ; a three-way gravity model with  $y_{ijt} = \exp[\alpha_{it} + \gamma_{jt} + \eta_{ijt} + x_{ijt}\beta]\omega_{ijt}$  and  $T = 2$ . Respectively, these models exemplify inconsistency, no asymptotic bias, and asymptotic bias. The  $i$  and  $j$  dimensions of the panel both have size  $N$  in the latter two models. The true value of  $\beta$  is 1 (indicated by the vertical dotted lines) and the data is generated using  $\text{Var}(y|\cdot) = \mathbb{E}(y|\cdot)$ .

paper will be to adapt some of the leading remedies from the recent literature on “large  $T$ ” IPPs (see, e.g., Arellano and Hahn, 2007) in order to re-center the asymptotic distribution of estimates and thereby restore asymptotically valid inferences.<sup>4</sup>

The bias in the estimated standard errors is similar to one that has been found in two-way gravity settings by several recent studies (Egger and Staub, 2015; Jochmans, 2017; Pfaffermayr, 2019, 2021). Intuitively, because the origin-time and destination-time fixed effects in the model each converge to their true values at a rate of only  $1/\sqrt{N}$  (not  $1/N$ ), the cluster-robust sandwich estimator for the variance has a leading bias of order  $1/N$  (not  $1/N^2$ ), and standard errors in turn have a bias of order  $1/\sqrt{N}$ . This latter type of bias is related to the general result that standard “heteroskedasticity-robust” variance estimators are downward-biased in small samples (see, e.g., MacKinnon and White, 1985; Imbens and Kolesar, 2016), including for PML estimators (Kauermann and Carroll, 2001), but is more severe in this setting due to an IPP. We should therefore be concerned that estimated confidence intervals may be too narrow in addition to being off-center.

For the bias in point estimates, we construct two-way analytical and jackknife bias corrections inspired by the corrections proposed in Fernández-Val and Weidner (2016, 2018). For the bias in standard errors, we show how Kauermann and Carroll (2001)’s method for correcting the PML sandwich estimator may be adapted to the case of a conditional estimator with multi-way fixed effects and cluster-robust standard errors. Our simulations confirm that these methods are usually effective at improving inferences. The jackknife correction reduces more of the bias in point estimates than the analytical correction in smaller samples, but the analytical correction does a better job at improving coverage, especially when also paired with corrected standard errors.

For our empirical applications, we first estimate the average effects of a free trade agreement (FTA) on trade for a range of different industries using what would typically be considered a large trade data set, with 167 countries and 5 time periods. The biases we uncover vary in size across the different industries, but are generally large enough to indicate that our bias corrections should be worthwhile in most three-way gravity settings. For aggregate trade data (which yields results that are fairly representative), the estimated coefficient for FTA has an implied downward bias about 15%–22% of the estimated standard error, and the implied downward bias in the standard error itself is about 11% of the original standard error. As a means of further

<sup>4</sup> The new literature on “large  $T$ ” asymptotic bias in nonlinear FE models has emerged as a recent response to the well-known “fixed  $T$ ” consistency problem first described in Neyman and Scott (1948). Examples include Phillips and Moon (1999), Hahn and Kuersteiner (2002), Lancaster (2002), Woutersen (2002), Alvarez and Arellano (2003), Carro (2007), Arellano and Bonhomme (2009), Fernández-Val and Vella (2011), and Kato et al. (2012). Unlike in most other settings explored in this literature, the panel estimator we consider is consistent regardless of  $T$ .

demonstration, we also apply our corrections to replication data from several recent papers that have used three-way gravity models. This latter exercise reveals several instances in which our methods make a material difference for assessing statistical significance. It also highlights the possibility that the bias in standard errors can sometimes be severe, as much as 40% or more in some cases.

Aside from Fernández-Val and Weidner (2016)'s work on two-way nonlinear models, Pesaran (2006), Bai (2009), Hahn and Moon (2006), and Moon and Weidner (2017) have each conducted similar analyses for two-way linear models with interacted individual and time fixed effects. Turning to three-way models, Hinz et al. (2019) have recently developed bias corrections for dynamic three-way probit and logit models based on asymptotics suggested by Fernández-Val and Weidner (2018) where all three panel dimensions grow at the same rate. Though widely applicable, this approach is not appropriate for our setting because of the different role played by the time dimension when the estimator is FE-PPML.<sup>5</sup> In the network context, Graham (2017), Dzemski (2019), and Chen et al. (2019) have studied asymptotic bias in network models with node-specific (possibly sender- and receiver-specific) fixed effects. Chen et al. (2019)'s analysis is especially notable in that they allow these node-specific effects to be vectors rather than scalars, similar to the exporter-time and importer-time fixed effects that feature in gravity models. Our bias expansions substantially differ from those of Chen et al. (2019) because the equivalent outcome variable in our setting (trade flows observed over time for a given pair) is also a vector rather than a scalar and because we work with a conditional moment model where the distribution of the outcome may be misspecified.

In what follows, Section 2 first provides a discussion of why IPPs are a concern for gravity models and of the no-IPP properties of FE-PPML. Section 3 then establishes bias and consistency results for the three-way gravity model specifically and discusses how to implement bias corrections. Sections 4 and 5 respectively present simulation evidence and empirical applications. Section 6 concludes, and an Appendix adds further simulation results and technical details, including proofs.

## 2. Gravity models and IPPs

Gravity models are now routinely estimated using FE-PPML with multiple sets of fixed effects. As we discuss in this section, these practices follow naturally from the gravity model's theoretical microfoundations but are not without need for further scrutiny. In particular, because the underlying model is nonlinear, it is important to clarify that, while PPML is known to be free from incidental parameter bias in some special cases, it is by no means immune to IPPs in general. It will also be useful for us to provide some general discussion of IPPs and the different ways in which they may manifest.

### 2.1. Fixed effects and gravity models

As documented in Head and Mayer (2014), the emergence of rich and varied theoretical foundations for the gravity equation has fueled a "fixed effects revolution" in the gravity literature over the last two decades. As such, we find it useful to briefly describe a simple trade model and discuss how it may be used to motivate an estimating equation with either two-way or three-way fixed effects.

To establish some notation we will use throughout the paper, we will consider a world with  $N$  countries and we will let  $i$  and  $j$  respectively be indices for exporter and importer. For now, we will focus on deriving a two-way gravity model where the two fixed effects account for each country's multilateral resistance. Later, we will add a time dimension and a third fixed effect that absorbs all time-invariant components of trade costs.

To add some theoretical structure, suppose that trade flows are given by the following gravity equation:

$$y_{ij} = \frac{y_i}{\Pi_i^{-\theta}} \frac{y_j}{P_j^{-\theta}} \tau_{ij}^{-\theta}. \quad (1)$$

Here,  $y_i := \sum_j y_{ij}$  and  $y_j := \sum_i y_{ij}$  are the market sizes of the two countries,  $\tau_{ij} \geq 1$  is a bilateral trade cost,  $\theta > 0$  is the trade elasticity, and  $\Pi_i$  and  $P_j$  respectively are the outward and inward multilateral resistances from Anderson and van Wincoop (2003), which capture how bilateral trade flows depend on each country's opportunities for trade with third countries. More formally, these latter terms are derived from the following two relationships that are inherent to all general equilibrium gravity models:

$$\Pi_i^{-\theta} = \sum_{j=1}^N \frac{y_j}{P_j^{-\theta}} \tau_{ij}^{-\theta}, \quad P_j^{-\theta} = \sum_{i=1}^N \frac{y_i}{\Pi_i^{-\theta}} \tau_{ij}^{-\theta}. \quad (2)$$

As shown, these terms respectively aggregate the exporter's ability to export goods to more desirable import markets and the importer's ability to import from more capable exporters.<sup>6</sup>

<sup>5</sup> Also related are the GMM-based differencing strategies for two-way FE models proposed by Charbonneau (2017) and Jochmans (2017). These strategies rely on differencing the data in such a way that the resulting GMM moments do not depend on any of the incidental parameters. In principle, these methods could be extended to allow for differencing across a time dimension as well in a three-way panel.

<sup>6</sup> This presentation of the gravity model readily conforms to the trade models used in Eaton and Kortum (2002) or Anderson and van Wincoop (2003), though the interpretation of  $\theta$  differs across the two models. With some minor modifications, this setup can also be made compatible with any of the theoretical gravity models considered in Head and Mayer (2014) or Costinot and Rodríguez-Clare (2014). To be clear, our econometric results do not require any particular microfoundation for the gravity equation.

For estimation, it is typical to parameterize the trade cost  $\tau_{ij}$  as depending exponentially on some variables of interest, i.e.,

$$\tau_{ij}^{-\theta} = \exp(x_{ij}\beta)\omega_{ij}, \tag{3}$$

where  $x_{ij}$  are the components of trade costs whose effects we wish to estimate. Because not all trade costs are reflected in  $x_{ij}$ , we also allow for “unobserved” trade costs via the idiosyncratic trade cost term  $\omega_{ij}$ . Combining (1) with (3) then delivers the following estimating equation:

$$y_{ij} = \exp(\alpha_i + \gamma_j + x_{ij}\beta)\omega_{ij}, \tag{4}$$

where  $\alpha_i = \ln(y_i/\Pi_i^{-\theta})$  and  $\gamma_j = \ln(y_j/P_j^{-\theta})$  are origin and destination fixed effects that absorb market sizes and multilateral resistances and  $\omega_{ij}$  now provides a multiplicative error term.<sup>7</sup> When we introduce the three-way model, all of the terms shown in (4) will have a further subscript for time, and the unobserved trade cost will have a time-invariant component that will motivate the use of an added  $ij$  fixed effect.

To motivate the arc of the rest of the paper, several points stand out from the estimation suggested by (4). First, the implied moment condition for estimation is

$$\mathbb{E}(y_{ij}|x_{ij}, \alpha_i, \gamma_j) = \lambda_{ij} \exp(\alpha_i + \gamma_j + x_{ij}\beta), \tag{5}$$

which follows after imposing  $\mathbb{E}(\omega_{ij}|x_{ij}, \alpha_i, \gamma_j) = 1$ .<sup>8</sup> As discussed in Santos Silva and Tenreyro (2006), consistent estimation of the trade cost parameters in  $\beta$  therefore generally requires a nonlinear model. Second, because unobserved trade costs enter the country-specific terms  $\alpha_i$  and  $\gamma_j$  through the system of multilateral resistances, we treat  $\alpha_i$  and  $\gamma_j$  as unknown parameters that will be noisily estimated, raising concerns about a possible IPP.<sup>9</sup> As we go on to discuss, the FE-PPML estimator that is most often used in this context has some special robustness against IPPs, but this robustness does not hold for FE-PPML in general, especially once we deviate from the two-way gravity setting implied by (5).

### 2.2. The incidental parameter problem

In the context of fixed effects models, IPPs occur when the estimation noise in the fixed effects contaminates the scores of the other parameters being estimated, inducing a bias. This bias can manifest in a variety of different ways; thus it is useful to provide a generic characterization that can illustrate the different possibilities that may arise. To that end, let  $n$  be the total number of observations and let  $p$  be the total number of parameters being estimated, inclusive of any fixed effects. As described in Fernández-Val and Weidner (2018), what we need to be concerned with is the number of observations that are available to estimate each fixed effect, i.e.,  $n/p$ . More precisely, when appropriate regularity conditions are satisfied, the estimated  $\hat{\beta}$  may generally be thought of as having the following bias and standard deviation:

$$\text{bias}(\hat{\beta}) = \frac{bp}{n} + o(p/n), \quad \text{std}(\hat{\beta}) = \frac{c}{\sqrt{n}} + o(n^{-1/2}), \tag{6}$$

where  $b \in \mathbb{R}$  and  $c > 0$  are constants that depend on the model being estimated. As this presentation emphasizes, all estimators in nonlinear settings are generally biased in small samples, but this is not the same thing as saying that the bias always poses a problem for inferring statistical significance. As we will discuss, in larger samples, what matters is whether the bias disappears faster than the standard error as  $n \rightarrow \infty$ .<sup>10</sup>

To provide a simple taxonomy of the cases that can arise, consider first the standard textbook treatment of maximum likelihood estimation, where we usually have  $p$  fixed while  $n \rightarrow \infty$ . In this case, the bias in  $\hat{\beta}$  becomes asymptotically negligible as compared to the standard deviation, which crucially means that estimated confidence intervals can be expected to be centered at the truth when the data becomes sufficiently large. By contrast, in the classical IPP of Neyman and Scott (1948), the number of

<sup>7</sup> Alternatively, it is sometimes common to write trade costs as a log-linear function, i.e.,  $\ln \tau_{ij} = x_{ij}\beta + e_{ij}$ , with  $e_{ij}$  now reflecting unobserved (log) trade costs. Interestingly, these two ways of specifying the error term do not necessarily have equivalent implications for estimation. In the log-linear formulation, if the log-error term  $e_{ij}$  is assumed to be heteroskedastic with mean zero, Jensen's inequality implies that estimation in levels will be biased and log-OLS will be consistent.

<sup>8</sup> Note that consistent estimation of  $\beta$  actually does not require  $\mathbb{E}(\omega_{ij}|\cdot) = 1$  in this case but rather  $\mathbb{E}(\omega_{ij}|\cdot) = \tilde{\omega}_i \tilde{\omega}_j$ , where  $\tilde{\omega}_i$  and  $\tilde{\omega}_j$  could be country-specific components of unobserved trade costs that would be absorbed by the fixed effects. For the three-way model, one requires  $\mathbb{E}(\omega_{ijt}|\cdot) = \tilde{\omega}_{it} \tilde{\omega}_{jt} \tilde{\omega}_{ijt}$ .

<sup>9</sup> As demonstrated in Pfaffermayr (2021), if we assume (i) there are no unobserved trade costs (such that  $\omega_{ij}$  does not enter the system in (2)) and (ii) the aggregate quantities  $y_i$  and  $y_j$  are perfectly observed, then it is better to regard  $\alpha_i$  and  $\gamma_j$  as reflecting constraints rather than as incidental parameters. Since Fally (2015) shows that constrained PPML and FE-PPML produce the same estimates in this context, there is no concern about IPPs if these assumptions are met.

<sup>10</sup> In the panel data literature, these results for the bias and standard deviation are usually derived not for  $\hat{\beta}$  directly, but for the asymptotic distribution of  $\hat{\beta}$  (because there are cases where  $\hat{\beta}$  may not have a first or second moment, but nevertheless has a well-defined limiting distribution with finite moments). We ignore this distinction for our heuristic discussion here.

parameters grows at the same rate as the number of observations, implying that the bias does not converge to zero asymptotically. In that case, the fixed effect estimator is inconsistent.

The gravity model with two-way fixed effects then serves to illustrate a *third* possibility that will also be applicable to the results that follow for three-way gravity models. In the two-way gravity setting,  $p$  is on the order of  $2N$ , where  $N$  is the number of countries, and  $n$  is on the order of  $N^2$ . Consequently, the bias and standard deviation of  $\hat{\beta}$  are given by

$$\text{bias}(\hat{\beta}) = \frac{2b}{N} + o(1/N), \quad \text{std}(\hat{\beta}) = \frac{c}{N} + o(1/N). \quad (7)$$

In this case, as  $N \rightarrow \infty$ , the estimated  $\hat{\beta}$  is consistent (both the standard deviation and bias converge to zero), but we also have

$$\lim_{N \rightarrow \infty} \frac{\text{bias}(\hat{\beta})}{\text{std}(\hat{\beta})} = \frac{2b}{c}.$$

As we discuss below, the two-way PPML gravity estimator is a special case where we actually have that  $b = 0$ . However, for other two-way gravity estimators (such as two-way Gamma PML for example), we generally have that  $b \neq 0$ , meaning the bias will *not* disappear relative to the standard error as  $N \rightarrow \infty$ . Compared to Neyman and Scott (1948), the IPP these estimators suffer from is not an inconsistency problem but rather an *asymptotic bias* problem, whereby the slow convergence of the fixed effects causes the asymptotic distribution for  $\hat{\beta}$  to be incorrectly centered as it converges to the truth.<sup>11</sup> This version of the IPP is more benign, but ignoring the bias will nonetheless result in invalid inferences and test results. The “large  $T$ ” panel data literature therefore discusses various methods for bias correction of  $\hat{\beta}$  that restore asymptotically valid inference. Importantly, the degree to which inferences are biased depends on the bias constant  $b$ , which cannot be known beforehand without applying such a correction.

To provide a more visual illustration of these ideas, Figure 1 presents simulation results for the three cases we have just discussed: inconsistency (top-left), no asymptotic bias (top-right), and asymptotic bias (bottom-left). Since asymptotic bias will ultimately be our focus, it is worth noting from the figure how the estimates are consistent in this case—the distribution will collapse to the true value as  $N \rightarrow \infty$ —but confidence bounds based on these estimates will clearly be inappropriate.

### 2.3. How FE-PPML is different

Our discussion of IPPs thus far has been for the generic estimation of a nonlinear model. However, the PPML estimator that is most commonly used to estimate gravity models actually behaves very differently than other estimators in this context. In the classic panel data setting with “one way” fixed effects, for example, FE-PPML has the very special property that the IPP bias constant  $b$  turns out to be zero, meaning that it is asymptotically unbiased in situations where other estimators tend to be inconsistent. As discussed in Wooldridge (1999), the reason behind this result is that the same estimator can be obtained from a multinomial model that does not depend on the fixed effects.<sup>12</sup>

This special property of FE-PPML has important implications for estimating gravity models as well. For two-way gravity settings, the asymptotic bias of  $\hat{\beta}$  when  $N \rightarrow \infty$  was worked out in Fernández-Val and Weidner (2016). They show that the two IPP contributions from  $\alpha_i$  and  $\gamma_j$  “decouple” asymptotically, such that the overall bias can be decomposed as the sum of two bias terms that would be expected in a one-way setting, i.e.,  $b_{(\alpha)}/N + b_{(\gamma)}/N$ , where the  $N$ 's come from the number of observations associated with each fixed effect. Because  $b = 0$  for the one-way FE-PPML case, we also have  $b_{(\alpha)} = b_{(\gamma)} = 0$  in the two-way setting; that is, two-way FE-PPML gravity estimates for  $\hat{\beta}$  are asymptotically unbiased just as one-way FE-PPML estimates are.<sup>13</sup>

Taken together, these results might create the impression that FE-PPML is generally immune to IPPs, regardless of what fixed effects are included in the model. Thus, it is important to clarify that a key feature of the two-way gravity model is that both fixed effect dimensions grow only with the square root of the panel size, such that the estimation noise in the estimated  $\hat{\alpha}_i$ 's and  $\hat{\gamma}_j$ 's disappears asymptotically. As we discuss in the Appendix, if we instead consider a model where both fixed effects grow with  $n$  rather than with its square root, the IPPs associated with each fixed effect do not decouple from one another, and FE-PPML in this case is actually inconsistent. An illustration of this result is provided in Fig. 1, top-left.

To synthesize these points, FE-PPML has a very special property—one can condition out one of the fixed effects—but this property has an important limitation—the resulting multinomial model does not inherit the same no-bias properties as the original PPML estimator with respect to any further fixed effects. Both of these results will be fleshed out in more detail in the following

<sup>11</sup> Consistency here follows from how the number of fixed effects grows only with the square root of the sample size, as discussed in Egger et al. (2011). The bias in the asymptotic distribution for two-way models was proven by Fernández-Val and Weidner (2016), discussed below.

<sup>12</sup> The earliest references to present versions of this result include Andersen (1970), Palmgren (1981), and Hausman et al. (1984). Wooldridge (1999)'s contribution is to show that FE-PPML is consistent even when the assumed distribution of the data is misspecified. Our Lemma 2 in the Appendix clarifies that FE-PPML is relatively unique in this regard versus similar estimators.

<sup>13</sup> Note that Theorem 4.1 in Fernández-Val and Weidner (2016) is written for the correctly specified case, where  $y_{ij}$  is actually Poisson distributed. However, Remark 3 in the paper gives the extension to conditional moment models, where for the FE-PPML case only the moment condition in (5) needs to hold. Their paper considers standard panel models, as opposed to trade models, but the only technical difference is that  $y_{ij}$  is often not observed for the trade model when  $i \neq j$ . This missing diagonal has no meaningful effect on any of the results we discuss.

section when we recast the three-way gravity model as a two-way multinomial model in order to obtain an appropriate expression for the bias. Doing so will also allow us to highlight another reason why FE-PPML is not immune to IPPs: even for the two-way gravity model, while the  $\alpha_i$  and  $\gamma_j$  parameters do not induce an IPP bias in  $\hat{\beta}$ , they nonetheless have implications for the estimated variance that are not innocuous; we thus will devote attention to this issue as well.

### 3. Results for the three-way gravity model

To recap the sequence of results just described, we know that FE-PPML estimates with one fixed effect do not suffer from an IPP. We also know that FE-PPML may have an IPP in models with more than one fixed effect, but it is both consistent and asymptotically unbiased in two-way gravity settings where neither fixed effect dimension grows at the same rate as the size of the panel. As we will now show, each of these earlier results will be useful for understanding the more complex case of a three-way gravity model that adds a time dimension and a third set of fixed effects to the above two-way model. We also describe a series of bias corrections for the three-way model, including for the possible downward bias of the estimated standard errors.

#### 3.1. Consistency

To formally introduce the three-way model, we add an explicit time subscript  $t \in \{1, \dots, T\}$  to  $y_{ij}$ ,  $x_{ij}$ , and  $\omega_{ij}$  from the prior model and add a “country-pair”-specific fixed effect  $\eta_{ij}$ , such that trade costs are now given by  $\tau_{ijt}^{-\theta} = e^{\alpha_{it}\beta + \eta_{ij}}\omega_{ijt}$ . All other elements in the original trade model likewise acquire a time subscript, meaning that  $\alpha_{it} = \ln y_{it}/\Pi_{it}^{-\theta}$  and  $\gamma_{jt} = \ln y_{jt}/P_{jt}^{-\theta}$  also must be indexed by  $t$ . The model now reads

$$\mathbb{E}(y_{ijt}|x_{ijt}, \alpha_{it}, \gamma_{jt}, \eta_{ij}) = \lambda_{ijt} := \exp(x'_{ijt}\beta + \alpha_{it} + \gamma_{jt} + \eta_{ij}), \tag{8}$$

where the three fixed effects now respectively index exporter-time, importer-time, and country-pair.<sup>14</sup> The unobserved trade cost  $\omega_{ijt} \geq 0$  continues to serve as an error term, such that  $y_{ijt} = \lambda_{ijt}\omega_{ijt} \geq 0$ . We thus allow for zero trade flows. For the asymptotics using the three-way model, we consider  $T$  fixed, while  $N \rightarrow \infty$ . The FE-PPML estimator maximizes

$$\mathcal{L}(\beta, \alpha, \gamma, \eta) := \sum_{i=1}^N \sum_{j \neq i}^N \sum_{t=1}^T (y_{ijt} \log \lambda_{ijt} - \lambda_{ijt})$$

over  $\beta, \alpha, \gamma$  and  $\eta$ .

Our strategy for showing the consistency of this estimator will capitalize on the special properties of FE-PPML discussed in the previous section. In particular, we will exploit the fact that not all of the fixed effect dimensions grow at the same rate as  $N$  increases. The numbers of exporter-time and importer-time fixed effects each increase with  $N$  (as before), but the dimension of the pair fixed effect  $\eta$  increases with  $N^2$ , since adding another country to the data adds another  $N - 1$  pairs to the estimation. It therefore makes sense to first “profile out” (i.e., solve for)  $\eta$  so that we may deal with the remaining two fixed effects in turn. For given values of  $\beta, \alpha, \gamma$ , maximizing over  $\eta$  gives us

$$\exp[\hat{\eta}_{ij}(\beta, \alpha, \gamma)] = \frac{\sum_{t=1}^T y_{ijt}}{\sum_{t=1}^T \mu_{ijt}}, \quad \mu_{ijt} := \exp(x'_{ijt}\beta + \alpha_{it} + \gamma_{jt}). \tag{9}$$

We therefore have

$$\mathcal{L}(\beta, \alpha, \gamma) := \max_{\eta} \mathcal{L}(\beta, \alpha, \gamma, \eta) = \sum_{i=1}^N \sum_{j \neq i}^N \ell_{ij}(\beta, \alpha_{it}, \gamma_{jt}), \tag{10}$$

with

$$\ell_{ij}(\beta, \alpha_{it}, \gamma_{jt}) := \sum_{t=1}^T y_{ijt} \log \left( \frac{\mu_{ijt}}{\sum_{s=1}^T \mu_{ijs}} \right) + \text{terms not depending on parameters}, \tag{11}$$

thus leaving us with the likelihood of a multinomial model where the only incidental parameters are  $\alpha_{it}$  and  $\gamma_{jt}$ . Using (8), one can easily verify that there is no bias in the score of the profile log-likelihood  $\ell_{ij}(\beta, \alpha_{it}, \gamma_{jt})$  when evaluated at the true parameters  $\beta^0$ ,

<sup>14</sup> Note that a multiplicative error term is not necessary to deliver the moment condition in (8). We could instead have an additive error term  $\varepsilon_{ijt} = y_{ijt} - \lambda_{ijt}$ . In this case, it would be more natural to think of it as coming from measurement error. In addition, note that we assume the true model is as written in (8) and assume away, e.g., any unobserved heterogeneity in  $\beta$ . Allowing for this type of heterogeneity is an important extension for future work to address.

$\alpha_{it}^0$ , and  $\gamma_{jt}^0$ . The reason for this is exactly the same as for the classic panel data setting discussed above. Furthermore, the remaining fixed effects  $\alpha_{it}$  and  $\gamma_{jt}$  grow only with the square root of the sample size as  $N \rightarrow \infty$ , implying that they are consistently estimated. This in turn leads us to the following result:

**Proposition 1.** *So long as the set of non-fixed effect regressors  $x_{ijt}$  is exogenous to the disturbance  $\omega_{ijt}$  after conditioning on the fixed effects  $\alpha_{it}$ ,  $\gamma_{jt}$ , and  $\eta_{ij}$ , FE-PPML estimates of  $\beta$  from the three-way gravity model are consistent for  $N \rightarrow \infty$ .<sup>15</sup>*

Intuitively, this result follows because of how the special properties of FE-PPML allow us to rewrite the three-way gravity model as a two-way model without introducing a  $1/T$  bias. The form of the bias in  $\hat{\beta}$  is therefore the same as in (7), such that three-way FE-PPML is consistent as  $N \rightarrow \infty$  largely for the same reason two-way FE-PPML and other two-way PML gravity estimators are generally consistent. However, in the context of *three-way* estimators, we can also state a stronger result that applies more narrowly to FE-PPML in particular:

**Proposition 2.** *Assume the conditional mean is given by  $\lambda_{ijt} = \exp(x_{ijt}'\beta + \alpha_{it} + \gamma_{jt} + \eta_{ij})$  and consider the class of “three-way” FE-PML gravity estimators with FOC's given by*

$$\hat{\beta} : \sum_{i=1}^N \sum_{j \neq i}^N \sum_{t=1}^T x_{ijt} (y_{ijt} - \hat{\lambda}_{ijt}) g(\hat{\lambda}_{ijt}) = 0, \quad \hat{\alpha}_{it} : \sum_{j=1}^N (y_{ijt} - \hat{\lambda}_{ijt}) g(\hat{\lambda}_{ijt}) = 0,$$

$$\hat{\gamma}_{jt} : \sum_{i=1}^N (y_{ijt} - \hat{\lambda}_{ijt}) g(\hat{\lambda}_{ijt}) = 0, \quad \hat{\eta}_{ij} : \sum_{t=1}^T (y_{ijt} - \hat{\lambda}_{ijt}) g(\hat{\lambda}_{ijt}) = 0,$$

where  $i, j = 1, \dots, N, t = 1, \dots, T$ , and  $g(\hat{\lambda}_{ijt})$  is an arbitrary function of  $\hat{\lambda}_{ijt}$  that can be specialized to construct various PML estimators. For example,  $g(\hat{\lambda}_{ijt}) = 1$  delivers PPML,  $g(\hat{\lambda}_{ijt}) = \hat{\lambda}_{ijt}^{-1}$  delivers Gamma PML, etc. If  $T$  is fixed, then for  $\hat{\beta}$  to be consistent under general assumptions about  $\text{Var}(y|x, \alpha, \gamma, \eta)$ , we must have that  $g(\lambda_{ijt})$  is constant over the range of  $\lambda$ 's that are realized in the data-generating process. That is, the estimator must be equivalent to FE-PPML.

In other words, three-way FE-PPML is unique among three-way PML estimators in that its consistency does not require strong assumptions about the conditional variance of  $y_{ijt}$ . To draw an appropriate contrast, it is possible to obtain a closed form solution for the pair fixed effect  $\hat{\eta}_{ij}$  so long as  $g(\hat{\lambda}_{ijt})$  is of the form  $g(\hat{\lambda}_{ijt}) = \hat{\lambda}_{ijt}^q$ , where  $q$  can be any real number. Notably, this latter class of estimators not only includes FE-PPML (for which  $q = 0$ ), but also includes other popular gravity estimators such as Gamma PML ( $q = -1$ ) and Gaussian PML ( $q = 1$ ). However, as we discuss in the Appendix, these other estimators are only consistent if the conditional variance is proportional to  $\lambda_{ijt}^{1-q}$ , in which case they inherit the properties of their associated MLE estimators.

### 3.2. Asymptotic bias

Because three-way FE-PPML inherits the consistency properties of the two-way estimator, one might expect that it also inherits its “no asymptotic bias” properties as well. However, this is where the limitations of PPML's no-IPP properties become apparent. While the profile log-likelihood in (11) is now of a similar form to the two-way models considered in Fernández-Val and Weidner (2016), notice that it no longer resembles the original FE-PPML log-likelihood. Their no-bias result for two-way FE-PPML therefore does not carry over to the three-way model, and it is possible to show that FE-PPML estimates have an asymptotic bias in this setting.

Before proceeding, it is helpful to first revisit the intuition established in Section 2.2 that shapes how we expect the bias to behave. As we have discussed, in a model with  $p$  parameters and  $n$  observations, the bias should be proportional to  $p/n$ , whereas the standard error should vary with  $1/\sqrt{n}$ . After profiling out  $\eta$ , the resulting two-way model has  $\sim 2NT$  parameters vs.  $\sim N^2T$  observations. If  $T$  is held fixed, we would expect the bias and the standard error to decrease at the same rate ( $1/N$ ), raising concerns about a possible asymptotic bias problem and guiding us as to its form. Readers should keep this intuition in mind in reading through the technical details that follow.

To illustrate more precisely where the bias comes from, it is necessary to examine how the estimated fixed effects enter the score for  $\beta$  using a Taylor expansion. To that end, first let  $\phi = \text{vec}(\alpha, \gamma)$  be a vector that collects all of the exporter-time and importer-time fixed effects, such that we can rewrite  $\ell_{ij}$  slightly as  $\ell_{ij} = \ell_{ij}(\beta, \phi)$ . We can then similarly define the function

<sup>15</sup> This consistency result can be seen as a corollary of the asymptotic normality result in Proposition 3 below, for which formal regularity conditions are stated in Assumption 1 of the Appendix.

$\hat{\phi}(\beta)$  as collecting the estimated fixed effects  $\hat{\alpha}$  and  $\hat{\gamma}$  as functions of  $\beta$ . Next, we construct a second-order expansion of the score for  $\beta$  around the true set of fixed effects  $\phi^0$  and evaluated at the true parameter  $\beta^0$ :

$$\begin{aligned} \mathbb{E} \left[ \frac{\partial \ell_{ij}(\beta^0 \hat{\phi}(\beta^0))}{\partial \beta} \right] &\approx \underbrace{\mathbb{E} \left[ \frac{\partial \ell_{ij}(\beta^0 \phi^0)}{\partial \beta} \right]}_{=0} + \underbrace{\mathbb{E} \left[ \frac{\partial^2 \ell_{ij}(\beta^0 \phi^0)}{\partial \beta \partial \phi'} (\hat{\phi}(\beta^0) - \phi^0) \right]}_{\neq 0 \text{ (reflects estimation error in FEs)}} \\ &+ \underbrace{\frac{1}{2} \sum_{f,g}^{\dim \phi} \mathbb{E} \left[ \frac{\partial^3 \ell_{ij}(\beta^0 \phi^0)}{\partial \beta \partial \phi_f \partial \phi_g} (\hat{\phi}_f(\beta^0) - \phi_f^0) (\hat{\phi}_g(\beta^0) - \phi_g^0) \right]}_{\neq 0 \text{ (reflects variances and covariances of FEs)}}. \end{aligned} \tag{12}$$

This expression is near-identical to a similar expansion that appears in Fernández-Val and Weidner (2016)—differing mainly in that  $\ell_{ij}$  is a vector rather than a scalar—and communicates the same essential insights: because the latter two terms in (12) are generally not equal to zero, the score for  $\beta$  is biased, with the bias depending on the interaction between the higher-order partial derivatives of  $\ell_{ij}$  and the estimation errors in  $\hat{\alpha}_i$  and  $\hat{\gamma}_j$  as well as their variances and covariances.

Demonstrating the bias in this particular setting then requires that we introduce some additional notation, mainly to provide some shorthand for the higher-order partial derivatives of  $\ell_{ij}$  that appear in (12). To do so, we first find it convenient to let  $\vartheta_{ijt} = \lambda_{ijt} / \sum_r \lambda_{ijr}$ . We then define the  $T \times 1$  “score” vector  $S_{ij}$ , the  $T \times T$  “Hessian” matrix  $H_{ij}$  and the  $T \times T \times T$  cubic tensor  $G_{ij}$  (the “third partial”), with their respective elements given by

$$\begin{aligned} S_{ij,t} &= \frac{\partial \ell_{ij}}{\partial \alpha_{it}} = y_{ijt} - \vartheta_{ijt} \sum_r y_{ijr}, \\ H_{ij,ts} &= - \frac{\partial^2 \ell_{ij}}{\partial \alpha_{it} \partial \alpha_{is}} = \begin{cases} \vartheta_{ijt} (1 - \vartheta_{ijt}) \sum_r y_{ijr} & \text{if } t = s, \\ -\vartheta_{ijs} \vartheta_{ijt} \sum_r y_{ijr} & \text{if } t \neq s, \end{cases} \\ G_{ij,tsr} &= \frac{\partial^3 \ell_{ij}}{\partial \alpha_{it} \partial \alpha_{is} \partial \alpha_{ir}}, \end{aligned}$$

where it should be understood that all of these terms are evaluated at the true values for all parameters. Explicit formulas for  $G_{ij,tsr}$  are provided in the Appendix.

The value of defining these objects is that they allow us to easily form terms identified by (12) as being important for the bias of the score. For example,  $S_{ij}$  allows us to obtain  $\partial \ell_{ij} / \partial \beta^k = x_{ij,k} S_{ij}$ . Likewise, we also have that  $\partial^2 \ell_{ij} / \partial \alpha_i \partial \beta^k = \partial^2 \ell_{ij} / \partial \gamma_j \partial \beta^k = -H_{ij} x_{ij,k}$  and that

$$\frac{\partial^3 \ell_{ij}}{\partial \alpha_i \partial \alpha'_i \partial \beta^k} = \frac{\partial^3 \ell_{ij}}{\partial \alpha_i \partial \gamma'_j \partial \beta^k} = \frac{\partial^3 \ell_{ij}}{\partial \gamma_j \partial \alpha'_i \partial \beta^k} = \frac{\partial^3 \ell_{ij}}{\partial \gamma_j \partial \gamma'_j \partial \beta^k} = G_{ij} x_{ij,k},$$

where we use the convention that  $G_{ij} x_{ij,k}$  is a  $T \times T$  matrix with elements  $[G_{ij} x_{ij,k}]_{st} = \sum_r G_{ijrst} x_{ijr,k}$ . We also find it useful to define the expected Hessian  $\bar{H}_{ij} = \mathbb{E}(H_{ij} | x_{ij})$  and, similarly, the expected third partial  $\bar{G}_{ij} = \mathbb{E}(G_{ij} | x_{ij})$ .<sup>16</sup> Finally, we define the  $K$ -vector  $\tilde{x}_{ij}$  as an appropriate two-way within-transformation of  $x_{ij}$  that purges it of the fixed effects; see the Appendix for details.

Obtaining a tractable expression for the bias then involves following the logic of (12) and plugging in the just-defined objects  $S_{ij}$ ,  $H_{ij}$ ,  $G_{ij}$ , and  $\tilde{x}_{ij}$  where appropriate. Before doing so, we invoke the assumption that observations are serially correlated within pairs but independent across pairs, as is commonly assumed in the literature (see Yotov et al., 2016.) This assumption turns out to cause the IPPs associated with  $\alpha_i$  and  $\gamma_j$  to “decouple”,<sup>17</sup> leading to the following proposition:

**Proposition 3.** Under appropriate regularity conditions (Assumption A in the Appendix), for  $T$  fixed and  $N \rightarrow \infty$  we have

$$\sqrt{N(N-1)} \left( \hat{\beta} - \beta^0 - \frac{W_N^{-1}(B_N + D_N)}{N-1} \right) \rightarrow_d \mathcal{N} \left( 0, W_N^{-1} \Omega_N W_N^{-1} \right),$$

<sup>16</sup> Because  $\bar{H}_{ij}$  is only positive semi-definite (not positive definite), we use a Moore-Penrose pseudoinverse whenever the analysis requires we work with an inverse of  $\bar{H}_{ij}$ . Specifically, we have that  $\bar{H}_{ij} \nu_T = 0$ , where  $\nu_T = (1, \dots, 1)$  is a  $T$ -vector of ones. Thus,  $\bar{H}_{ij}$  is only of rank  $T - 1$  rather than of rank  $T$ .

<sup>17</sup> In particular, all elements of the cross-partial objects  $\mathbb{E}[\partial^2 \ell_{ij} / \partial \alpha_i \partial \gamma_j]$ ,  $\mathbb{E}[\partial^3 \ell_{ij} / \partial \alpha_i \alpha'_i \partial \gamma_j]$ , etc. can be shown to be asymptotically small for  $N \rightarrow \infty$ . Thus, in what follows,  $B_N$  reflects the contribution of the  $\alpha_i$  parameters to the bias and  $D_N$  reflects the contribution of the  $\gamma_j$  parameters. As we discuss in the Appendix, relaxing this assumption can change the expression of the bias.

where  $W_N$  and  $\Omega_N$  are  $K \times K$  matrices given by

$$W_N = \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{j \neq i} \tilde{x}'_{ij} \bar{H}_{ij} \tilde{x}_{ij},$$

$$\Omega_N = \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{j \neq i} \tilde{x}'_{ij} \left[ \text{Var}(S_{ij} | x_{ij}) \right] \tilde{x}_{ij},$$

and  $B_N$  and  $D_N$  are  $K$ -vectors with elements given by

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

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

where a  $\dagger$  denotes a Moore-Penrose pseudoinverse.

The above proposition establishes the asymptotic distribution of the three-way gravity estimator as  $N \rightarrow \infty$ , including the asymptotic bias  $(N - 1)^{-1} W_N^{-1} (B_N + D_N)$ . Intuitively, this bias can be decomposed as the product of the inverse expected Hessian with respect to  $\beta$  (i.e.  $W_N^{-1}$ ), the rate of asymptotic convergence (essentially  $1/N$ ), and the bias of the score from (12), which here is given by the combined term  $B_N + D_N$ .  $B_N$  reflects the contribution to the bias from the noise in  $\hat{\alpha}_i$ , whereas  $D_N$  reflects the contribution from the noise in  $\hat{\gamma}_j$ . The first terms in both  $B_N$  and  $D_N$  come from the second term in (12), reflecting the estimation error in the estimated fixed effects, and the second terms in  $B_N$  and  $D_N$  echo the third term in (12), reflecting their variance.

Thus, in the end, the bias reduces to essentially the same simple formula we gave in Section 2.2 for the bias of the two-way gravity model, i.e.,

$$\frac{1}{N-1} b_{(\alpha)} + \frac{1}{N-1} b_{(\gamma)},$$

where  $b_{(\alpha)} = W_N^{-1} B_N$  and  $b_{(\gamma)} = W_N^{-1} D_N$  are constants that do not vary with  $N$ . Importantly, and unlike in the two-way FE-PPML setting, the three-way model does not give us the no-bias result that  $B_N = D_N = 0$ , as the following discussion helps to illustrate.

### 3.2.1. Illustrating the bias using the $T = 2$ case

Admittedly, the complexity of the objects that appear in Proposition 3 may make it difficult to appreciate the general point that the three-way estimator is not asymptotically unbiased. One way to make these details more transparent is to focus our attention on the simplest possible panel model where  $T = 2$ . The convenient thing about this simplified setting is that the likelihood function  $\ell_{ij}$  can be reduced to just a scalar:  $\ell_{ij} = y_{ij1} \log \vartheta_{ij1} + y_{ij2} \log (1 - \vartheta_{ij1})$ , where now

$$\vartheta_{ij1} = \frac{\exp(\Delta x_{ij} \beta + \Delta \alpha_i + \Delta \gamma_j)}{\exp(\Delta x_{ij} \beta + \Delta \alpha_i + \Delta \gamma_j) + 1},$$

with  $\Delta x_{ij} = x_{ij1} - x_{ij2}$ ,  $\Delta \alpha_i = \alpha_{i1} - \alpha_{i2}$ , and  $\Delta \gamma_j = \gamma_{j1} - \gamma_{j2}$ . These normalizations allow us to express all of the objects that appear in Proposition 3 as also just scalars, and we can therefore easily derive the following result:

**Remark 1.** For  $T = 2$ , we calculate  $S_{ij} = \vartheta_{ij2} y_{ij1} - \vartheta_{ij1} y_{ij2}$ ,  $H_{ij} = \vartheta_{ij1} \vartheta_{ij2} (y_{ij1} + y_{ij2})$ ,  $\bar{H}_{ij} = \vartheta_{ij1} \lambda_{ij2}$ ,  $G_{ij} = \vartheta_{ij1} \vartheta_{ij2} (\vartheta_{ij1} - \vartheta_{ij2}) (y_{ij1} + y_{ij2})$ ,  $\bar{G}_{ij} = \vartheta_{ij1} (\vartheta_{ij1} - \vartheta_{ij2}) \lambda_{ij2}$ , and  $\Delta \tilde{x}_{ij} = \tilde{x}_{ij1} - \tilde{x}_{ij2}$ . The bias term  $B_N^k$  in Proposition 3 can then be written as

$$B_N^k = \text{p lim}_{N \rightarrow \infty} \left[ -\frac{\frac{1}{N} \sum_{i=1}^N \sum_{j \neq i} \Delta \tilde{x}_{ij} \vartheta_{ij1} \vartheta_{ij2} \left[ \vartheta_{ij2} \mathbb{E}(y_{ij1}^2) - \vartheta_{ij1} \mathbb{E}(y_{ij2}^2) + (\vartheta_{ij2} - \vartheta_{ij1}) \mathbb{E}(y_{ij1} y_{ij2}) \right]}{\sum_{j \neq i} \vartheta_{ij1} \lambda_{ij2}} \right. \\ \left. + \frac{\frac{1}{2N} \sum_{i=1}^N \left\{ \sum_{j \neq i} \Delta \tilde{x}_{ij} \vartheta_{ij1} (\vartheta_{ij1} - \vartheta_{ij2}) \lambda_{ij2} \right\} \left\{ \sum_{j \neq i} \vartheta_{ij2}^2 \mathbb{E}(y_{ij1}^2) + \vartheta_{ij1}^2 \mathbb{E}(y_{ij2}^2) - 2 \vartheta_{ij1} \vartheta_{ij2} \mathbb{E}(y_{ij1} y_{ij2}) \right\}}{\left[ \sum_{j \neq i} \vartheta_{ij1} \lambda_{ij2} \right]^2} \right],$$

with an analogous expression also following for  $D_N^k$ .

Two points then stand out based on the above expression: (i) all bias terms in  $B_N^k$  and  $D_N^k$  generally depend on the distribution of  $y_{ij}$  and do not depend on it in the same way; (ii) none of these terms generally equals 0. The first of these two observations can be seen from how the bias depends on the expected second moments of  $y_{ij}$  (e.g.,  $\mathbb{E}(y_{ij1}^2)$ ,  $\mathbb{E}(y_{ij1}y_{ij2})$ , etc.), marking an important difference from the models that were considered in Fernández-Val and Weidner (2016).<sup>18</sup> Among other things, the difficulty associated with estimating these second moments means that analytical bias corrections may not necessarily offer superior performance to distribution-free method such as the jackknife. The second observation mainly follows from the first. It can also be shown that  $\sum_{j \neq i} \bar{C}_{ij} \Delta \bar{x}_{ij} \neq 0$ , which ensures that the second term can never be zero.

3.2.2. What if  $T$  is large?

While Proposition 3 only focuses on asymptotics where  $N \rightarrow \infty$ , the three-way gravity panel also features a time dimension ( $T$ ), and it is interesting to wonder how the above results may depend on changes in  $T$ . As we show in the Appendix, large  $T$  only makes a difference for the asymptotic order of the bias of  $\hat{\beta}$  if there is only weak time dependence between observations belonging to the same country pair, in the sense described by Hansen, 2007.<sup>19</sup> We will henceforth assume any time dependence is weak. The following remark then describes some additional asymptotic results for when  $T$  is large.

**Remark 2.** Under asymptotics where  $T \rightarrow \infty$ , we have the following:

- (i) If  $N$  is fixed and  $T \rightarrow \infty$ , then  $\hat{\beta}$  is generally inconsistent.
- (ii) As  $N, T \rightarrow \infty$ , the combined bias term  $(N - 1)^{-1} W_N^{-1} (B_N + D_N)$  goes to zero at a rate of  $1/(NT)$ . Therefore, because the standard error is of order  $1/(N\sqrt{T})$ , there is no bias in the asymptotic distribution of  $\hat{\beta}$  as  $N$  and  $T$  both  $\rightarrow \infty$ .

To elaborate further, letting  $T \rightarrow \infty$  is obviously not sufficient for either  $\alpha$  or  $\gamma$  to be consistently estimated and does not solve the IPP, as stated in part (i). However, as part (ii) tells us,  $T$  still plays an interesting role in conditioning the bias when both  $N$  and  $T$  jointly become large. Intuitively, because  $W_N^{-1}$  is of order  $1/T$  as  $T \rightarrow \infty$ , whereas  $B_N$  and  $D_N$  are both of order 1, the bias in  $\hat{\beta}$  effectively vanishes at a rate of  $1/(NT)$  as both  $N, T \rightarrow \infty$ , such that it disappears asymptotically in relation to the order- $1/(N\sqrt{T})$  standard error. However, since  $T$  is usually small relative to  $N$  in this context, it remains to be seen whether these asymptotic results carry over to practical settings.

3.3. Downward bias in robust standard errors

Of course, even if the point estimates are correctly centered, inferences will still be unreliable if the estimates of the variance used to construct confidence intervals are not themselves unbiased. For PPML, confidence intervals are typically obtained using a “sandwich” estimator for the variance that accounts for the possible misspecification of the model. However, as shown by Kauermann and Carroll (2001), the PPML sandwich estimator is generally downward-biased in finite samples. Furthermore, for gravity models (both two-way and three-way), the bias in the sandwich estimator can itself be formalized as a kind of IPP.<sup>20</sup>

To illustrate the bias of the sandwich estimator in our three-way setting, recall that we can express the variance of  $\hat{\beta}$  as  $\text{Var}(\hat{\beta} - \beta) = N^{-1} (N - 1)^{-1} W_N^{-1} \Omega_N W_N^{-1}$ . As is also true for the linear model (cf., MacKinnon and White, 1985; Imbens and Kolesar, 2016), the bias arises because plugin estimates for the “meat” of the sandwich  $\Omega_N$  depend on the estimated score variance  $\mathbb{E}(\hat{S}_{ij} \hat{S}_{ij}')$  rather than on the true variance  $\mathbb{E}(S_{ij} S_{ij}')$ . Even though  $\mathbb{E}(\hat{S}_{ij} \hat{S}_{ij}')$  is a consistent estimate for  $\mathbb{E}(S_{ij} S_{ij}')$ , it will generally be downward-biased in finite samples. Notably, this bias may be especially slow to vanish for models with gravity-like fixed effects.

To see this, we follow the same approach as Kauermann and Carroll (2001). Specifically, we use the special case where  $\mathbb{E}(S_{ij} S_{ij}') = \kappa \bar{H}_{ij}$  (such that  $\Omega_N = \kappa W_N$ , meaning PPML is correctly specified) to demonstrate that  $\mathbb{E}(\hat{S}_{ij} \hat{S}_{ij}')$  generally has a downward bias. Under this assumption, it is possible to show that the expected outer product of the fitted score  $\mathbb{E}(\hat{S}_{ij} \hat{S}_{ij}')$  has a first-order bias of

$$\mathbb{E}(\hat{S}_{ij} \hat{S}_{ij}' - S_{ij} S_{ij}') \approx - \underbrace{\frac{\kappa}{N(N-1)} \bar{H}_{ij} \bar{x}_{ij}' W_N^{-1} \bar{x}_{ij}' \bar{H}_{ij}}_{\text{order } 1/N^2 \text{ (from noise in } \hat{\beta})}} - \underbrace{\frac{\kappa}{N(N-1)} \bar{H}_{ij} d_{ij}' W_N^{(\phi)-1} d_{ij}' \bar{H}_{ij}}_{\text{order } 1/N \text{ (from noise in FEs)}} \tag{13}$$

<sup>18</sup> The specific examples they use are the Poisson model, which is unbiased, and the probit model, which requires the distribution of  $y_{ij}$  to be correctly specified. They also provide a bias expansion for “conditional moment” models that allow the distribution of  $y_{ij}$  to be misspecified. Beyond this theoretical discussion, bias corrections for misspecified models have yet to receive much attention, however.

<sup>19</sup> By “weak” time dependence, we mean that any such dependence dissipates as the temporal distance between observations increases. Alternatively, if observations are correlated regardless of how far apart they are in time, the standard error is always of order  $1/N$  (see Hansen, 2007), and the same will also be true for the asymptotic bias. The latter is arguably a less natural assumption in this context, however.

<sup>20</sup> This type of IPP has similar origins to the one described in Verdier (2020), who considers a dyadic linear model with two-way FEs and sparse matching between the two panel dimensions.

<sup>21</sup> A detailed derivation of (13) is provided in the Appendix.

where  $W_N^{(\phi)} := \mathbb{E}_N[-\partial^2 \ell_{ij} / \partial \phi \partial \phi']$  captures the expected Hessian of the concentrated likelihood with respect to  $\alpha$  and  $\gamma$  and where  $d_{ij}$  is a  $T \times \dim(\phi)$  matrix of dummies such that each row satisfies  $d_{ijt}\phi = \alpha_{it} + \gamma_{jt}$ .<sup>21</sup>

The two terms on the right-hand side of (13) are both negative definite, implying that the sandwich estimator is generally downward-biased—and definitively so if the model is correctly specified. Since we work with cluster-robust standard errors, a relevant comparison to draw here is with Cameron et al. (2008), who have previously shown that the cluster-robust sandwich estimator has a downward bias that depends on the number of clusters. In our setting, the standard Cameron et al. (2008) bias is reflected in the first term on the righthand-side of (13), which captures how the bias depends on the variance of  $\hat{\beta}$ . The second term, which arises because of an IPP, captures how much of the bias is due to the variance in the estimated origin-time and destination-time fixed effects in  $\hat{\phi}$ . The former term decreases with  $1/N^2$ —i.e., with the number of pairs/clusters—but the latter term only decreases with  $1/N$ , since increasing  $N$  by 1 only adds 1 additional observation of each origin-time and destination-time fixed effect.<sup>22</sup>

All together, this analysis implies that the estimated standard error for  $\hat{\beta}$  will exhibit a bias that only disappears at the relatively slow rate of  $1/\sqrt{N}$ . We should therefore be concerned that asymptotic confidence intervals for  $\hat{\beta}$  may exhibit inadequate coverage even in moderately large samples, similar to what has been found for the two-way FE-PPML estimator in recent simulation studies by Egger and Staub (2015), Jochmans (2017), and Pfaffermayr (2019). Indeed, the bias approximation we have derived in (13) can be readily adapted to the two-way setting or even to more general settings with  $k$ -way fixed effects.

### 3.4. Bias corrections for the three-way gravity model

We now present two methods for correcting the bias in estimates: a jackknife method based on the split-panel jackknife of Dhaene and Jochmans (2015) and an analytical correction based on the expansion shown in Proposition 3. We also provide an analytical correction for the downward bias in standard errors.

#### 3.4.1. Jackknife bias correction

The advantage of the jackknife correction is that it does not require explicit estimation of the bias yet still has a simple and powerful applicability. To see this, note first that the asymptotic bias we characterize can be written as

$$\frac{1}{N} B^\beta + o_p(N^{-1}),$$

where  $B^\beta$  is a combined term that captures any suspected asymptotic bias contributions of order  $1/N$ . The specific jackknife we will apply for our current purposes is a split-panel jackknife based on Dhaene and Jochmans (2015). As in Dhaene and Jochmans (2015), we want to divide the overall data set into subpanels of roughly even size and then estimate  $\hat{\beta}_{(p)}$  for each subpanel  $p$ . Given the gravity structure of the model, we first divide the set of countries into evenly-sized groups  $a$  and  $b$ . We then consider 4 subpanels of the form “(a,b)”, where “(a,b)” denotes a subpanel where exporters from group  $a$  are matched with importers from group  $b$ . The other three subpanels are (a,a), (b,a), and (b,b). For randomly-generated data, we can define  $a$  and  $b$  based on their ordering in the data (i.e.,  $a = i : i \leq N/2; b = i : i > N/2$ ). For actual data, it would be more sensible to draw these subpanels randomly and repeatedly.<sup>23</sup>

The split-panel jackknife estimator for  $\beta$ ,  $\tilde{\beta}_N^J$ , is then defined as

$$\tilde{\beta}_N^J := 2\hat{\beta} - \sum_p \frac{\hat{\beta}_{(p)}}{4}. \tag{14}$$

This correction works to reduce the bias because, so long as the distribution of  $y_{ij}$  and  $x_{ij}$  is homogeneous across both the  $i$  and  $j$  dimensions of the panel,<sup>24</sup> each  $\hat{\beta}_{(p)}$  has a leading bias term equal to  $2B^\beta/N$ . The average  $\hat{\beta}_{(p)}$  across these four subpanels thus also has a leading bias of  $2B^\beta/N$  and any terms depending on  $B^\beta/N$  cancel out of (14). Thus, the bias-corrected estimate  $\tilde{\beta}_N^J$  only has a bias of order  $o_p(N^{-1})$ , which is obtained by combining the second-order bias from  $\hat{\beta}$  with that of the average subpanel estimate. This latter bias can be shown to be larger than the original second-order bias in  $\hat{\beta}$ , but the overall bias should still be smaller because of the elimination of the leading bias term.

<sup>21</sup> A detailed derivation of (13) is provided in the Appendix.

<sup>22</sup> The  $[N(N-1)]^{-1} W_N^{(\phi)-1}$  matrix that appears in the second term is the inverse Hessian with respect to the fixed effects and thus reflects their variance. Because adding a new country only adds one new observation of each fixed effect, the diagonal elements of this matrix decrease with only  $1/N$  as  $N$  increases, despite how the formula is written. Pfaffermayr (2019) makes a similar point about the order of the bias of the standard errors for the two-way FE-PPML estimator, albeit using a slightly different analysis.

<sup>23</sup> This is just one possible way to construct a jackknife correction for two-way panels. We have also experimented with splitting the panel one dimension at a time as in Fernández-Val and Weidner (2016), but we find the present method performs significantly better at reducing the bias.

<sup>24</sup> By “homogeneity” we mean that the vector  $(y_{ij}, x_{ij}, \alpha_i, \gamma_j)$  is identically distributed across both  $i$  and  $j$ , which is the appropriate translation of Assumption 4.3 in Fernández-Val and Weidner (2016) to our setting. This does allow  $(y_{ij}, x_{ij})$  to be heterogeneously distributed conditional on the fixed effects in the sense that the fixed effects themselves introduce heterogeneity into the model. Nonetheless, this is a strong assumption. One of the main advantages of the analytical bias correction is that it does not require such assumptions.

### 3.4.2. Analytical bias correction

Our analytical correction for the bias is based on the bias expression in Proposition 3. In the Appendix, we show how appropriate sample analogs  $\hat{W}_N, \hat{B}_N, \hat{D}_N$  of the expressions for  $W_N, B_N, D_N$  can be formed. The resulting bias-corrected estimate is then given by

$$\hat{\beta} - (N-1)^{-1} \hat{W}_N^{-1} (\hat{B}_N + \hat{D}_N).$$

It is possible to show that these plug-in corrections lead to estimates that are asymptotically unbiased as  $N \rightarrow \infty$ . Still, for finite samples, it is evident that the bias in some of these plug-in objects could cause the analytical bias correction to itself exhibit some bias. For this reason, it is not obvious a priori whether the analytical correction will outperform the jackknife at reducing the bias in  $\hat{\beta}$ . One clear advantage the analytical correction has over the jackknife is that it does not require any homogeneity restrictions on the distribution of  $y_{ij}$  and  $x_{ij}$  in order to be valid.

### 3.4.3. Bias-corrected standard errors

Under the assumption of clustered errors within pairs, a natural correction for the variance estimate is available based on (13). Specifically, let

$$\hat{\Omega}^U = \frac{1}{N(N-1)} \sum_{i,j} \hat{x}_{ij} \left[ \mathbf{I}_T - \frac{1}{N(N-1)} \bar{H}_{ij} \hat{x}_{ij} \hat{W}_N^{-1} \hat{x}' - \frac{1}{N(N-1)} \bar{H}_{ij} d_{ij} \hat{W}_N^{(\phi)-1} d'_{ij} \right]^{-1} \hat{S}_{ij} \hat{S}'_{ij} \hat{x}_{ij},$$

where  $\mathbf{I}_T$  is a  $T \times T$  identity matrix and  $\hat{W}_N^{(\phi)}$  is a plugin estimate for  $W_N^{(\phi)}$ . The corrected variance estimate is then given by

$$\hat{V}^U = \frac{1}{N(N-1)-1} \hat{W}^{-1} \hat{\Omega}^U \hat{W}^{-1}.$$

The logic of this adjusted variance estimate follows directly from Kauermann and Carroll (2001): if the PPML estimator is correctly specified (such that  $E(S_{ij} S'_{ij}) = \kappa \bar{H}_{ij}$ ), then  $\hat{V}^U$  can be shown to eliminate the first-order bias in  $\hat{V}(\hat{\beta} - \beta^0)$  shown in (13). It is not generally unbiased otherwise, but it is plausible that it should eliminate a significant portion of any downward bias under other variance assumptions as well.

### 3.4.4. Other practicalities

As we have noted, implementations of our analytical corrections are available via our Stata command `ppml_fe_bias`. Since this command can be applied to data sets that do not conform exactly to our theoretical framework, we provide here some brief comments on its applicability in general cases that may not be explicitly covered in the notation and formulas used above. For example, it is worth pointing out that our corrections may be used with data sets that have missing values. They also continue to apply if the data includes the diagonal ( $i = j$ ) terms versus treating them as missing or inapplicable. Similarly, we can allow for the data to have unequal numbers of exporters and importers. In the latter case, we need the numbers of exporters and importers to grow at the same rate asymptotically for our results to apply.

One practicality that is not covered in our current framework is the case of “four way” gravity models that have an added index for industry. Though a full analytical characterization of this type of model is left for future work, our Appendix describes how a modified version of the heuristic from Fernández-Val and Weidner (2018) may be used to assess the order of the bias and derive a suitable jackknife correction. Interestingly, this discussion reveals that the asymptotic bias problem may be more severe for four-way models than for three-way models. The heuristic we propose may be used to assess IPPs in other, non-trade settings as well.

## 4. Simulation evidence

For our simulation analysis, we assume the following: (i) the data generating process (DGP) for the dependent variable is of the form  $y_{ijt} = \lambda_{ijt} \omega_{ijt}$ , where  $\omega_{ijt}$  is a log-normal disturbance with mean 1 and variance  $\sigma_{ijt}^2$ . (ii)  $\beta = 1$ . (iii) The model-relevant fixed effects  $\alpha, \gamma$ , and  $\eta$  are each  $\sim \mathcal{N}(0, 1/16)$ . (iv)  $x_{ijt} = x_{ijt-1}/2 + \alpha + \gamma + \nu_{ijt}$ , where  $\nu_{ijt} \sim \mathcal{N}(0, 1/16)$ .<sup>25</sup> (v) Taking our cue

<sup>25</sup> These assumptions on  $\alpha, \gamma, \eta, x_{ijt}$ , and  $\nu_{ijt}$  are taken from Fernández-Val and Weidner (2016). Notice that  $x_{ijt}$  is strictly exogenous with respect to  $\omega_{ijt}$  conditional on  $\alpha, \gamma$ , and  $\eta$ .

from Santos Silva and Tenreyro (2006), we consider 4 different assumptions about the disturbance  $\omega_{ijt}$ :

- DGP I** :  $\sigma_{ijt}^2 = \lambda_{ijt}^{-2}$ ;  $\text{Var}(y_{ijt}|x_{it}, \alpha, \gamma, \eta) = 1$ .
- DGP II** :  $\sigma_{ijt}^2 = \lambda_{ijt}^{-1}$ ;  $\text{Var}(y_{ijt}|x_{it}, \alpha, \gamma, \eta) = \lambda_{ijt}$ .
- DGP III** :  $\sigma_{ijt}^2 = 1$ ;  $\text{Var}(y_{ijt}|x_{it}, \alpha, \gamma, \eta) = \lambda_{ijt}^2$ .
- DGP IV** :  $\sigma_{ijt}^2 = 0.5\lambda_{ijt}^{-1} + 0.5e^{2x_{ijt}}$ ;  $\text{Var}(y_{ijt}|x_{it}, \alpha, \gamma, \eta) = 0.5\lambda_{ijt} + 0.5e^{2x_{ijt}}\lambda_{ijt}^2$ ,

where we also allow for serial correlation within pairs by imposing

$$\text{Cov}[\omega_{ijs}, \omega_{ijt}] = \exp\left[0.3^{|s-t|} \times \sqrt{\ln(1 + \sigma_{ijs}^2)} \sqrt{\ln(1 + \sigma_{ijt}^2)}\right] - 1,$$

such that the degree of correlation weakens for observations further apart in time.<sup>26</sup>

The relevance of these various assumptions to commonly used error distributions is best described by considering the conditional variance  $\text{Var}(y_{ijt}|x_{it}, \alpha, \gamma, \eta)$ . For example, DGP I assumes that the conditional variance is constant, as in a Gaussian process with i.i.d disturbances. In DGP II, the conditional variance equals the conditional mean, as in a Poisson distribution. DGP III—which we will also refer to as “log-homoskedastic”—is the unique case highlighted in Santos Silva and Tenreyro (2006) where the assumption that the conditional variance is proportional to the square of the conditional mean leads to a homoskedastic error when the model is estimated in logs using a linear model. Finally, DGP IV provides a “quadratic” error distribution that mixes DGP II and DGP III and also allows for overdispersion that depends on  $x_{ijt}$ . As shown by Santos Silva and Tenreyro (2006), this type of DGP tends to induce a relatively large degree of bias.

Tables 1 and 2 present simulation evidence comparing the uncorrected three-way FE-PPML estimator with results computed using the analytical and jackknife corrections described in Section 3.4. As in the prior simulations, we again compute results for a variety of different panel sizes—in this case for  $N = 20, 50, 100$  and  $T = 2, 5, 10$ .<sup>27</sup> In order to validate our analytical predictions regarding these estimates, we compute the average bias, the ratios of the average bias to the average standard error and of the average standard error to the standard deviation of the simulated estimates, and the probability that the estimated 95% confidence interval covers the true estimate of  $\beta = 1$ . In particular, we expect that the bias in  $\hat{\beta}$  should be decreasing in either  $N$  or  $T$  but should remain large relative to the estimated standard error and induce inadequate coverage for small  $T$ . We are also interested in whether the usual cluster-robust standard errors accurately reflect the true dispersion of estimates. Results for DGPs I and II are shown in Table 1, whereas Table 2 shows results for DGPs III and IV.

The results in both tables collectively confirm the presence of bias and the viability of the analytical and jackknife bias corrections. The average bias is generally larger for DGPs I and IV than II and III. As expected, it generally falls with both  $N$  and  $T$  across all the different DGPs, though only weakly so for DGP III (the log-homoskedastic case), which generally only has a small bias.<sup>28</sup> To use DGP II—the Poisson case, where PPML should otherwise be an optimal estimator—as a representative example, we see that the average bias falls from 3.840% for the smallest sample where  $N = 20, T = 2$  to a low of 0.249% at the other extreme where  $N = 100, T = 10$ . For DGP IV, the least favorable of these cases, the average bias ranges from  $-6.1364\%$  down to  $-1.878\%$ . On the whole, these results support our main theoretical findings that  $\beta$  should be consistently estimated even for fixed  $T$  but has an asymptotic bias that depends on the number of countries and on the number of time periods.

Interestingly, while the average bias almost always decreases with  $T$ , the ratio of the bias to standard error usually does not, seemingly contrary to the expectations laid out in Remark 2. Evidently, increasing  $T$  does not automatically reduce the bias at a rate of  $1/T$ . As we discuss in more detail below, researchers should thus be careful to note that the implications of Remark 2 do not necessarily apply to settings with small  $T$  or even moderately large  $T$ . Furthermore, the estimated cluster-robust standard errors themselves clearly exhibit a bias in all cases as well. Even when  $N = 100$ , SE/SD ratios are uniformly below 1; generally they are closer to 0.9 or 0.95, and for DGPs I and IV, they are often closer to 0.85 or even 0.8. Because of these biases, the simulated FE-PPML coverage ratios are unsurprisingly below the 0.95 we would expect for an unbiased estimator.

Bias corrections to the point estimates do help with addressing some, but not all, of these issues. The jackknife generally performs more reliably than the analytical correction at reducing the average bias when compared across all values of  $N$  and  $T$ ; notice how, for the Poisson case, for example, the average bias left by the jackknife correction is never greater than 0.25% in absolute magnitude, whereas the analytical-corrected estimates still have average biases ranging between 0.01% and 1.12%. However, when  $N = 100$ , the analytical correction begins to closely match the jackknife, especially when  $T = 10$ . All the same, both

<sup>26</sup> The 0.3 that appears here serves as a quasi-correlation parameter. Replacing 0.3 with 1 would be analogous to assuming disturbances are perfectly correlated within pairs. Replacing it with 0 removes any serial correlation. Choosing other values for this parameter produces similar results.

<sup>27</sup> Note that the trade literature currently recommends using wide intervals of 4–5 years between time periods so as to allow trade flows time to adjust to changes in trade costs (see Cheng and Wall, 2005.) Thus, for practical purposes,  $T = 10$  may be thought of as a relatively “long” panel in this context that might span 40+ years. IPPS do not necessarily vanish for larger values of  $T$ , as discussed further below.

<sup>28</sup> Numerically, we have found that the two terms that appear in both  $B_N$  and  $D_N$  in Proposition 3 tend to have opposite signs when the DGP is log-homoskedastic, thereby mitigating one another.

**Table 1**  
Finite-sample Properties of Three-way FE-PPML Estimates.

	N = 20			N = 50			N = 100		
	T = 2	T = 5	T = 10	T = 2	T = 5	T = 10	T = 2	T = 5	T = 10
<b>A. Gaussian DGP ("DGP I")</b>									
<i>Average bias (×100)</i>									
FE-PPML	6.647	3.764	2.364	2.854	1.497	0.913	1.463	0.755	0.428
Analytical	2.261	1.005	0.543	0.529	0.165	0.104	0.152	0.043	0.011
Jackknife	0.558	0.024	-0.304	0.247	0.014	-0.035	0.074	0.006	-0.025
<i>Bias / SE ratio</i>									
FE-PPML	0.690	0.796	0.753	0.668	0.732	0.687	0.644	0.708	0.626
Analytical	0.235	0.213	0.173	0.124	0.081	0.078	0.067	0.040	0.016
Jackknife	0.058	0.005	-0.097	0.058	0.007	-0.026	0.033	0.006	-0.037
<i>SE / SD ratio</i>									
FE-PPML	0.836	0.846	0.868	0.901	0.916	0.936	0.934	0.949	0.968
Analytical	0.792	0.813	0.845	0.859	0.887	0.921	0.899	0.929	0.958
Jackknife	0.718	0.752	0.783	0.831	0.863	0.900	0.887	0.918	0.949
<i>Coverage probability (should be 0.95 for an unbiased estimator)</i>									
FE-PPML	0.836	0.804	0.823	0.856	0.846	0.870	0.874	0.870	0.889
Analytical	0.871	0.883	0.895	0.907	0.921	0.930	0.922	0.933	0.941
Jackknife	0.841	0.866	0.876	0.901	0.911	0.926	0.916	0.929	0.939
<b>B. Poisson DGP ("DGP II")</b>									
<i>Average bias (×100)</i>									
FE-PPML	3.840	2.161	1.358	1.621	0.857	0.541	0.825	0.435	0.249
Analytical	1.193	0.621	0.421	0.237	0.095	0.094	0.063	0.024	0.010
Jackknife	0.229	0.002	-0.153	0.079	0.007	0.001	0.018	0.004	-0.013
<i>Bias / SE ratio</i>									
FE-PPML	0.403	0.466	0.446	0.394	0.439	0.430	0.386	0.436	0.390
Analytical	0.125	0.134	0.138	0.058	0.049	0.075	0.029	0.024	0.016
Jackknife	0.024	0.000	-0.050	0.019	0.004	0.001	0.008	0.004	-0.020
<i>SE / SD ratio</i>									
FE-PPML	0.865	0.869	0.891	0.931	0.940	0.957	0.957	0.965	0.982
Analytical	0.824	0.838	0.869	0.903	0.924	0.947	0.939	0.957	0.977
Jackknife	0.748	0.774	0.804	0.875	0.898	0.924	0.926	0.942	0.966
<i>Coverage probability (should be 0.95 for an unbiased estimator)</i>									
FE-PPML	0.887	0.880	0.892	0.912	0.905	0.919	0.918	0.919	0.925
Analytical	0.888	0.897	0.902	0.920	0.931	0.938	0.934	0.939	0.948
Jackknife	0.857	0.870	0.884	0.916	0.922	0.934	0.928	0.936	0.945

**Notes:** Results computed using 5000 repetitions. The model being estimated is  $y_{ijt} = \lambda_{ijt}\omega_{ijt}$ , where  $\lambda_{ijt} = \exp(\alpha_{it} + \gamma_{jt} + \eta_{ij} + \beta x_{ijt})$ . The data is generated using  $\alpha_{it} \sim \mathcal{N}(0, 1/16)$ ,  $\gamma_{jt} \sim \mathcal{N}(0, 1/16)$ ,  $\eta_{ij} \sim \mathcal{N}(0, 1/16)$  and  $\beta = 1$ .  $x_{ijt} = x_{ijt-1}/2 + \alpha_{it} + \gamma_{jt} + \eta_{ij} + \nu_{ijt}$ , with  $x_{ij0} = \eta_{ij} + \nu_{ij0}$  and  $\nu_{ijt} \sim \mathcal{N}(0, 1/2)$ . Results are shown for two different assumptions about  $V(y_{ijt})$ . The "Gaussian" DGP (panel A) assumes  $V(\omega_{ijt}) = \lambda_{ijt}^2$ . The "Poisson" DGP (panel B) assumes  $V(\omega_{ijt}) = \lambda_{ijt}^{-1}$ . SE/SD refers to the ratio of the average standard error of  $\hat{\beta}$  relative to the standard deviation of  $\hat{\beta}$  across simulations. Coverage probability refers to the probability  $\beta^0$  is covered in the 95% confidence interval for  $\hat{\beta}_{NT}$ . "Analytical" and "Jackknife" respectively indicate Analytical and Jackknife bias-corrected FE-PPML estimates. "FE-PPML" indicates uncorrected estimates. SEs allow for within- $ij$  clustering.

corrections generally have a positive effect, and the better across-the-board bias-reduction performance of the jackknife comes at the important cost of a relatively large increase in the variance. Thus, the analytical correction generally performs as well as or better than the jackknife in terms of improving coverage even in the smaller samples. Neither correction is sufficient to bring coverage ratios to 0.95, however, though corrected Gaussian-DGP estimates and Poisson-DGP estimates both exceed 0.94 using the analytical correction when  $N = 100$  and  $T = 10$ , with the latter reaching 0.948.

Table 3 then evaluates the efficacy of our bias correction for the estimated variance. Keeping in mind that this correction is calibrated for the case of a correctly specified variance (which corresponds to DGP II), we would naturally expect that the effect of this correction should vary depending on the conditional distribution of the data. In that light, it is encouraging that we observe positive effects across all cases. The best results by far are for the DGPs I, II, and III, where combining the analytical bias correction for the point estimates with the correction for the variance yields coverage ratios that range between 0.925 and 0.952 when  $N$  is either 50 or 100 and generally get closer to the target value of 0.95 as either  $N$  or  $T$  increases. These corrections lead to dramatic improvements in coverage for DGP IV as well, but there the remaining biases in both the point estimate and the standard error remain large even for  $N = 100$  and  $T = 10$ .

To summarize, these simulations suggest that combining an analytical bias correction for  $\hat{\beta}$  with a further correction for the variance based on (13) should be a reliable way of reducing bias and improving coverage. At the same time, it should be noted that neither should be expected to offer a complete bias removal. For smaller samples, if reducing bias on average is heavily favored, and if the distribution of  $y_{ij}$  and  $x_{ij}$  can be reasonably assumed to be homogeneous across  $i$  and  $j$ , then the split-panel jackknife method might be preferable to the analytical correction method.

**Table 2**  
Finite-sample Properties of Three-way FE-PPML Estimates.

	N = 20			N = 50			N = 100		
	T = 2	T = 5	T = 10	T = 2	T = 5	T = 10	T = 2	T = 5	T = 10
<b>A. Log-homoskedastic DGP ("DGP III")</b>									
Average bias ( $\times 100$ )									
FE-PPML	0.368	-0.086	-0.248	0.143	-0.092	-0.119	0.083	-0.051	-0.088
Analytical	-0.103	-0.108	-0.048	-0.065	-0.073	-0.004	-0.012	-0.024	-0.016
Jackknife	-0.189	-0.286	-0.302	-0.064	-0.077	-0.038	-0.012	-0.021	-0.023
Bias / SE ratio									
FE-PPML	0.036	-0.017	-0.074	0.032	-0.042	-0.081	0.036	-0.044	-0.114
Analytical	-0.010	-0.021	-0.014	-0.015	-0.033	-0.003	-0.005	-0.021	-0.021
Jackknife	-0.019	-0.057	-0.090	-0.014	-0.035	-0.026	-0.005	-0.018	-0.030
SE / SD ratio									
FE-PPML	0.853	0.838	0.848	0.920	0.910	0.923	0.949	0.943	0.956
Analytical	0.801	0.787	0.799	0.881	0.874	0.885	0.922	0.918	0.930
Jackknife	0.727	0.726	0.744	0.853	0.850	0.863	0.908	0.902	0.920
Coverage probability (should be 0.95 for an unbiased estimator)									
FE-PPML	0.903	0.898	0.899	0.933	0.927	0.923	0.937	0.936	0.940
Analytical	0.884	0.874	0.882	0.919	0.917	0.914	0.931	0.929	0.934
Jackknife	0.846	0.843	0.852	0.912	0.907	0.902	0.924	0.923	0.935
<b>B. Quadratic DGP ("DGP IV")</b>									
Average bias ( $\times 100$ )									
FE-PPML	-6.136	-5.658	-5.098	-3.560	-3.348	-2.935	-2.207	-2.114	-1.878
Analytical	-4.592	-3.983	-3.415	-2.088	-1.907	-1.588	-1.021	-0.990	-0.857
Jackknife	-3.752	-3.517	-3.116	-1.794	-1.674	-1.414	-0.922	-0.893	-0.774
Bias / SE ratio									
FE-PPML	-0.510	-0.898	-1.171	-0.640	-1.114	-1.384	-0.718	-1.242	-1.544
Analytical	-0.382	-0.632	-0.784	-0.375	-0.634	-0.749	-0.332	-0.582	-0.704
Jackknife	-0.312	-0.558	-0.716	-0.323	-0.557	-0.667	-0.300	-0.525	-0.636
SE / SD ratio									
FE-PPML	0.799	0.759	0.749	0.847	0.809	0.809	0.877	0.844	0.844
Analytical	0.734	0.693	0.682	0.781	0.738	0.734	0.811	0.773	0.770
Jackknife	0.662	0.635	0.633	0.747	0.710	0.707	0.795	0.756	0.755
Coverage probability (should be 0.95 for an unbiased estimator)									
FE-PPML	0.849	0.760	0.700	0.844	0.734	0.673	0.848	0.705	0.619
Analytical	0.830	0.775	0.749	0.849	0.799	0.780	0.871	0.823	0.801
Jackknife	0.795	0.749	0.729	0.840	0.795	0.786	0.867	0.818	0.803

**Notes:** Results computed using 5000 repetitions. The model being estimated is  $y_{ijt} = \lambda_{ijt}\omega_{ijt}$ , where  $\lambda_{ijt} = \exp(\alpha_{it} + \gamma_{jt} + \eta_{ij} + \beta x_{ijt})$ . The data is generated using  $\alpha_{it} \sim \mathcal{N}(0, 1/16)$ ,  $\gamma_{jt} \sim \mathcal{N}(0, 1/16)$ ,  $\eta_{ij} \sim \mathcal{N}(0, 1/16)$  and  $\beta = 1$ .  $x_{ijt} = x_{ijt-1}/2 + \alpha_{it} + \gamma_{jt} + \eta_{ij} + v_{ijt}$ , with  $x_{ij0} = \eta_{ij} + v_{ij0}$  and  $v_{ijt} \sim \mathcal{N}(0, 1/2)$ . Results are shown for two different assumptions about  $V(y_{ijt})$ . The "Log-homoskedastic" DGP (panel A) assumes  $V(y_{ijt}) = \lambda_{ijt}^2$ . The "Quadratic" DGP (Panel D) assumes  $\omega_{ijt}$  is log-normal with variance equal to  $\lambda_{ijt}^{-1} + \exp(2x_{ijt})$ . SE/SD refers to the ratio of the average standard error of of  $\hat{\beta}$  relative to the standard deviation of  $\hat{\beta}$  across simulations. Coverage probability refers to the probability  $\beta^0$  is covered in the 95% confidence interval for  $\hat{\beta}_{NT}$ . "Analytical" and "Jackknife" respectively indicate Analytical and Jackknife bias-corrected FE-PPML estimates. "FE-PPML" indicates uncorrected estimates. SEs allow for within- $ij$  clustering.

#### 4.1. What happens for larger values of T?

Based on our Remark 2, one might expect that increasing the size of the time dimension should reduce the asymptotic bias in relation to the standard error. However, in the range of  $T$  values we used in Tables 1–3, this is not what we observe. The question thus arises: can we say if there exists a "large enough" value of  $T$  beyond which researchers may feel relatively secure about IPPs?

Fig. 2 addresses this question by presenting simulated bias/SE and coverage ratios for a wider range of  $T$  values, spanning from 2 to 100.  $N$  is fixed at 100, and the data is otherwise generated the same way as before. If we focus just on the first two DGPs, the Gaussian and Poisson cases, we do indeed observe steady improvements in both ratios as  $T$  increases, though coverage fails to hit 0.95 in either case. However, for both DGP III (log-homoskedastic) and DGP IV (quadratic), we actually observe bias/SE ratios getting worse as  $T$  approaches 100. In the case of DGP IV, coverage gets worse as well.<sup>29</sup> The main takeaway is that gravity panels with seemingly large time spans are not necessarily immune to IPPs. To reconcile these findings with our theory, note that Remark 2 only says that three-way PPML estimates become asymptotically unbiased as both  $N$  and  $T$  become large simultaneously. In further simulations, we have confirmed that both the bias/SE ratio and coverage improve across all DGPs when we compare, e.g.,  $N = T = 200$  with  $N = T = 100$ .

<sup>29</sup> For scale reasons, coverage results for DGP IV are not shown. For  $T = 100$ , we find that coverage is only 0.52 in this case. As shown in the left-hand panel of Figure 2, the reason is because the bias tends to decrease more slowly than the standard error as  $T$  increases while  $N$  is fixed under this DGP.

**Table 3**  
Improving Coverage in Three-way FE-PPML Gravity Estimates.

	N = 20			N = 50			N = 100		
	T = 2	T = 5	T = 10	T = 2	T = 5	T = 10	T = 2	T = 5	T = 10
<b>A. Gaussian DGP ("DGP I")</b>									
<i>SE / SD ratio with corrected SEs</i>									
FE-PPML	0.936	0.930	0.946	0.953	0.957	0.971	0.965	0.972	0.988
Analytical	0.887	0.894	0.921	0.908	0.927	0.956	0.929	0.952	0.978
Jackknife	0.803	0.827	0.854	0.879	0.902	0.935	0.917	0.941	0.968
<i>Coverage probability with corrected SEs (should be 0.95 for an unbiased estimator)</i>									
(uncorrected)	0.836	0.804	0.823	0.856	0.846	0.870	0.874	0.870	0.889
FE-PPML	0.877	0.846	0.856	0.881	0.866	0.885	0.889	0.877	0.895
Analytical	0.909	0.912	0.918	0.925	0.930	0.941	0.931	0.940	0.948
Jackknife	0.884	0.901	0.908	0.920	0.923	0.938	0.926	0.935	0.944
<b>B. Poisson DGP ("DGP II")</b>									
<i>SE / SD ratio with corrected SEs</i>									
FE-PPML	0.966	0.956	0.974	0.979	0.980	0.993	0.983	0.986	1.001
Analytical	0.920	0.922	0.950	0.949	0.963	0.983	0.964	0.978	0.996
Jackknife	0.834	0.851	0.879	0.920	0.936	0.960	0.951	0.962	0.985
<i>Coverage probability with corrected SEs (should be 0.95 for an unbiased estimator)</i>									
(uncorrected)	0.887	0.880	0.892	0.912	0.905	0.919	0.918	0.919	0.925
FE-PPML	0.923	0.915	0.916	0.927	0.921	0.930	0.925	0.927	0.931
Analytical	0.923	0.929	0.930	0.938	0.942	0.949	0.942	0.945	0.952
Jackknife	0.900	0.903	0.915	0.932	0.935	0.942	0.936	0.941	0.949
<b>C. Log-homoskedastic DGP ("DGP III")</b>									
<i>SE / SD ratio with corrected SEs</i>									
FE-PPML	0.966	0.943	0.954	0.979	0.967	0.979	0.984	0.976	0.989
Analytical	0.906	0.886	0.899	0.938	0.928	0.940	0.956	0.951	0.962
Jackknife	0.823	0.817	0.837	0.908	0.902	0.916	0.942	0.934	0.952
<i>Coverage probability with corrected SEs (should be 0.95 for an unbiased estimator)</i>									
(uncorrected)	0.903	0.898	0.899	0.933	0.927	0.923	0.937	0.936	0.940
FE-PPML	0.938	0.934	0.932	0.947	0.942	0.940	0.945	0.943	0.949
Analytical	0.922	0.914	0.915	0.938	0.932	0.929	0.940	0.939	0.943
Jackknife	0.892	0.889	0.899	0.930	0.925	0.924	0.935	0.934	0.942
<b>D. Quadratic DGP ("DGP IV")</b>									
<i>SE / SD ratio with corrected SEs</i>									
FE-PPML	0.952	0.861	0.929	0.947	0.948	0.949	0.970	0.924	0.882
Analytical	0.877	0.793	0.845	0.873	0.865	0.865	0.897	0.853	0.810
Jackknife	0.781	0.729	0.786	0.833	0.824	0.837	0.889	0.828	0.789
<i>Coverage probability with corrected SEs (should be 0.95 for an unbiased estimator)</i>									
(uncorrected)	0.849	0.760	0.700	0.844	0.734	0.673	0.848	0.705	0.619
FE-PPML	0.900	0.820	0.808	0.894	0.806	0.762	0.892	0.752	0.652
Analytical	0.894	0.830	0.838	0.908	0.864	0.862	0.918	0.856	0.818
Jackknife	0.844	0.808	0.814	0.894	0.860	0.866	0.908	0.848	0.822

**Notes:** Results computed using 5000 repetitions. The data is generated in the same manner as Tables 1 and 2. SE/SD refers to the ratio of the average standard error of  $\hat{\beta}$  relative to the standard deviation of  $\hat{\beta}$  across simulations. Coverage probability refers to the probability  $\beta^0$  is covered in the 95% confidence interval for  $\hat{\beta}$ . "Analytical" and "Jackknife" respectively indicate Analytical and Jackknife bias-corrected FE-PPML estimates. "FE-PPML" estimates use corrections for the SEs only. SEs allow for within- $ij$  clustering. The corrected SEs correct for first-order finite sample bias in the estimated variance.

## 5. Empirical applications

For our main empirical application, we estimate the average effects of an FTA for a variety of different industries using a panel with a relatively large number of countries. The value of this exercise is that we expect that trade flows could be distributed very differently across different industries. Based on our results so far, this should lead to a range of different bias behaviors in the data.

Our trade data is from the BACI database of Gaulier and Zignago (2010), from which we extract data on trade flows between 167 countries for the years 1995, 2000, 2005, 2010, and 2015. Countries are chosen so that the same 167 countries always appear as both exporters and importers in every period; hence, the data readily maps to the setting just described with  $N = 167$  and  $T = 5$ . We combine this trade data with data on FTAs from the NSF-Kellogg database maintained by Scott Baier and Jeff

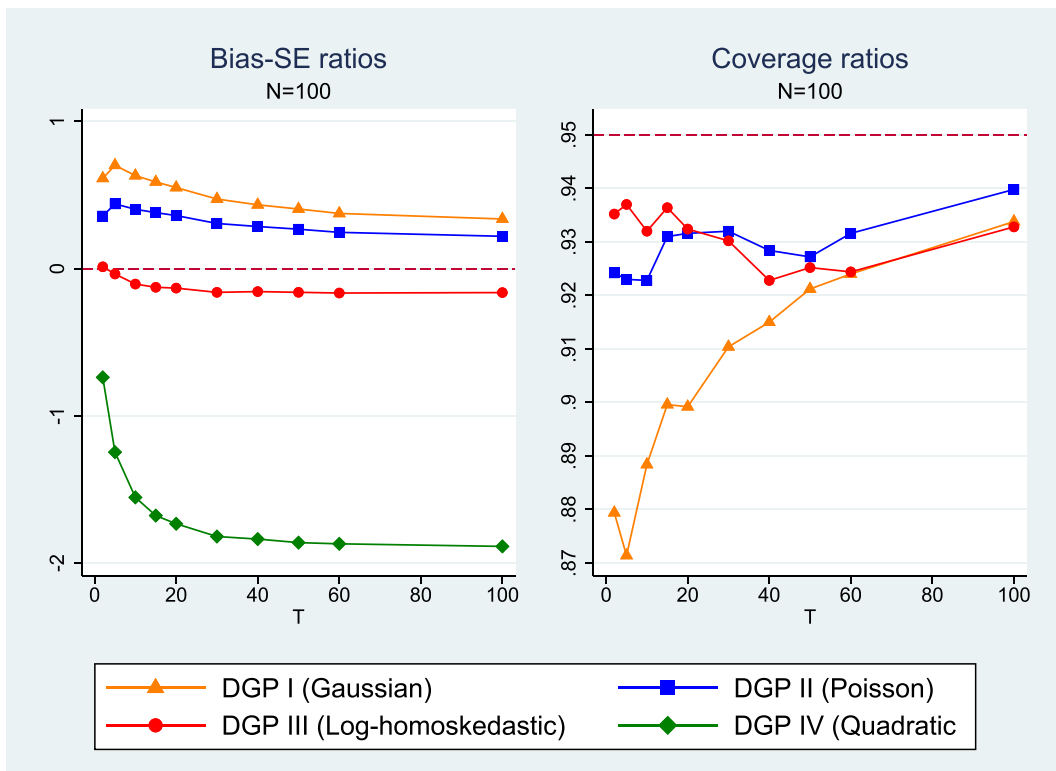


Fig. 2. Simulated bias/SE and coverage ratios for three-way FE-PPML estimates as the time dimension ( $T$ ) becomes increasingly large. The data is generated in the same way as in Tables 1–3. The  $i$  and  $j$  dimensions are fixed at  $N = 100$ . The values for  $T$  are 2, 5, 10, 15, 20, 30, 40, 50, 60, and 100. Coverage for DGP IV (not shown) ranges between 0.52 and 0.84 and generally gets progressively worse at higher values of  $T$ .

Bergstrand, which we crosscheck against data from the WTO in order to incorporate agreements from more recent years.<sup>30</sup> The specification we estimate is

$$y_{ijt} = \exp [\alpha_{it} + \gamma_{jt} + \eta_{ij} + \beta FTA_{ijt}] \omega_{ijt}, \tag{15}$$

where  $y_{ijt}$  is trade flows (measured in current USD),  $FTA_{ijt}$  is a 0/1 dummy for whether or not  $i$  and  $j$  have an FTA at time  $t$ , and  $\omega_{ijt}$  is an error term. As we have noted, estimation of specifications such as (15) via PPML has become an increasingly standard method for estimating the effects of FTAs and other trade policies and is currently recommended as such by the WTO (see Yotov et al., 2016.)

Table 4 presents results from FE-PPML estimation of (15), including results obtained using our bias corrections. The estimation is applied separately to 28 2 digit ISIC (rev. 3) industries as well as to aggregate trade. While we do indeed see a range of different biases in the industry-level estimates, the results for aggregate trade flows, shown in the bottom row of Table 4, are fairly representative. To provide some basic interpretation, the coefficient on  $FTA_{ijt}$  for aggregate trade is initially estimated to be 0.082, which equates to an  $e^{0.082} - 1 = 8.5\%$  average “partial” effect of an FTA on trade.<sup>31</sup> The estimated standard error is 0.027, implying that this effect is statistically different from zero at the  $p < 0.01$  significance level. Our bias-corrected estimates do not paint an altogether different picture, but do highlight the potential for meaningful refinement. Both the analytical and jackknife bias corrections for  $\beta$  suggest a downward bias of 0.04–0.06, or about 15%–22% of the estimated standard error. As our bias-corrected standard errors show (in the last column of Table 4), the initially estimated standard error itself has an implied downward bias of 11% (i.e., 0.027 versus 0.030).

Turning to the industry-level estimates, the analytical bias correction more often than not indicates a downward bias ranging between 5%–20% of the estimated standard error, though exceptions are present on both sides of this range. Estimates for the Chemical and Furniture industries appear to be unbiased, for example, and some (such as Tobacco) are associated with an upward

<sup>30</sup> This database is available for download on Jeff Bergstrand’s website: <https://www3.nd.edu/~jbergstr/>. The most recent version runs from 1950 to 2012. The additional data from the WTO is needed to capture agreements that entered into force between 2012 and 2015.

<sup>31</sup> The term “partial effect” is conventionally used to distinguish this type of estimate from the “general equilibrium” effects of an FTA, which would typically be calculated by solving a general equilibrium trade model where prices, incomes, and output levels (which are otherwise absorbed by the  $\alpha_{it}$  and  $\gamma_{jt}$  fixed effects) are allowed to evolve endogenously in response to the FTA. In the context of such models,  $\beta$  can usually be interpreted as capturing the average effect of an FTA on bilateral trade frictions specifically, holding fixed all other determinants of trade.

**Table 4**  
Bias Correction Results Using BACI Trade Data ( $N = 167$ ).

Industry	Code	Original estimates		Bias-corrected estimates		
		$\hat{\beta}$	SE	Analytical	Jackknife	SE
Agriculture	1	0.097	(0.046)	0.107	0.109	(0.051)
Forestry	2	-0.203	(0.125)	-0.197	-0.185	(0.156)
Fishing	5	0.127	(0.141)	0.140	0.189	(0.164)
Coal	10	0.025	(0.131)	0.011	-0.079	(0.163)
Metal Ores	13	0.040	(0.100)	0.032	-0.020	(0.124)
Other Mining & Quarrying n.e.c.	14	0.048	(0.097)	0.079	0.098	(0.108)
Food & Beverages	15	0.020	(0.043)	0.027	0.029	(0.048)
Tobacco	16	0.535	(0.139)	0.525	0.598	(0.163)
Textiles	17	0.229	(0.045)	0.227	0.238	(0.055)
Apparel	18	0.092	(0.092)	0.094	0.132	(0.120)
Leather Products	19	0.222	(0.067)	0.218	0.242	(0.079)
Wood & Cork Products	20	0.077	(0.109)	0.098	0.102	(0.127)
Paper & Paper Products	21	-0.002	(0.062)	-0.004	-0.018	(0.071)
Printed & Recorded Media	22	-0.114	(0.065)	-0.144	-0.177	(0.076)
Coke & Refined Petroleum	23	0.254	(0.076)	0.290	0.340	(0.090)
Chemicals & Chemical Products	24	0.072	(0.035)	0.072	0.077	(0.040)
Rubber & Plastic Products	25	0.141	(0.030)	0.146	0.154	(0.035)
Non-metallic Mineral Products	26	0.218	(0.049)	0.225	0.232	(0.058)
Basic Metal Products	27	0.267	(0.102)	0.272	0.302	(0.115)
Fabricated Metal Products (excl. Machinery)	28	0.196	(0.036)	0.207	0.226	(0.041)
Machinery & Equipment n.e.c.	29	0.049	(0.036)	0.052	0.055	(0.041)
Office, Accounting, and Computer Equipment	30	-0.036	(0.062)	-0.044	-0.038	(0.074)
Electrical Equipment	31	0.214	(0.045)	0.226	0.240	(0.052)
Communications Equipment	32	-0.127	(0.067)	-0.143	-0.174	(0.081)
Medical & Scientific Equipment	33	0.062	(0.039)	0.069	0.087	(0.044)
Motor Vehicles, Trailers & Semi-trailers	34	0.158	(0.064)	0.170	0.196	(0.077)
Other Transport Equipment	35	0.207	(0.124)	0.230	0.251	(0.137)
Furniture & Other Manufacturing n.e.c.	36	0.225	(0.073)	0.225	0.228	(0.082)
Total	All	0.082	(0.027)	0.086	0.088	(0.030)

**Notes:** These results are computed using ISIC Rev. 3 industry-level trade data for trade between 167 countries during years 1995, 2000, 2005, 2010, & 2015. The original data is from BACI. The model being estimated is  $y_{ijt} = \exp(\alpha_{it} + \gamma_{jt} + \eta_{ij} + \beta FTA_{ijt})\omega_{ijt}$ , where  $y_{ijt}$  is the trade volume and  $FTA_{ijt}$  is a dummy for the presence of an FTA.  $\alpha_{it}$ ,  $\gamma_{jt}$ , &  $\eta_{ij}$  respectively denote exporter-time, importer-time, & exporter-importer fixed effects. We estimate each industry separately. The jackknife corrections use the average of 200 randomly-assigned split-panel partitions. SEs are clustered by exporter-importer.

bias. On the other end of the spectrum, implied downward biases can also be larger than 20% of the standard error, as is seen for Petroleum (47%), Fabricated Metal Products (31%), Electrical Equipment (27%), and Agriculture (22%). The biases implied by the jackknife are often even larger (see Fabricated Metal Products, for example), consistent with what we found in our simulations for smaller panel sizes. One possible interpretation is that the jackknife-corrected estimates are giving us a less conservative alternative to the analytical corrections in these cases. Indeed, the general correspondence between the two sets of results adds validity to both methods. However, as we have noted, these jackknife estimates could also be reflecting non-homogeneity in the data and/or the higher variance introduced by the jackknife. Implied biases in the standard error, meanwhile, tend to range between 10%–20% of the original standard error, again with some exceptions.

To further illustrate the types of results that can occur, we also obtain replication data for several recent articles that have used three-way gravity models and re-examine their findings using our bias corrections. The results, reported in Table 5, help to demonstrate how these corrections can matter for assessing statistical significance. Instances where conventional significance levels are affected include the coefficients for  $EIA \times CONTIG$  and  $EIA \times LEGAL$  from Baier et al. (2018) and the coefficient for log approval rating from Rose (2019).<sup>32</sup> The implied bias to standard error ratios are sometimes as large as 40%–45%, as occurs for the  $EIA \times LANG$  coefficient from Baier et al. (2018) and the Total EIA Effect from Bergstrand et al. (2015). However, the largest effects are actually for the standard error, which is downward-biased by more than 40% in several cases (the Total FTA effect from Baier et al. (2019), for example). Notably, there are meaningful differences found even for the data used in Larch et al. (2019), a large data set with 213 countries and 66 time periods. These results reinforce our earlier finding that bias corrections may be useful even in settings with seemingly large  $N$  and  $T$ .

<sup>32</sup> Note that Baier et al. (2018) and Baier et al. (2014) use first-differenced OLS with added pair time trends. We estimate the three-way FE-PPML equivalents.

**Table 5**  
Further Applications.

Article	Covariate	original PPML coefficient and standard error	bias-corrected estimate	Notes
Baier et al. (2018)	EIA	0.062 (0.039)	0.073 (0.044)*	EIA is a dummy for the presence of an Economic Integration Agreement. The interaction terms are log distance and dummies for adjacency, common legal system, common religion, common language, and colonial history. Our estimation differs from BBF's in that we use PPML, whereas BBF use OLS with first-differences and pair-specific time trends. The specification we replicate is their Table 4, column 7. 183 countries, 1965–2010 every 5 years ( $N = 183, T = 10$ ).
	EIA $\times$ log DIST	-0.102 (0.031)***	-0.094 (0.035)***	
	EIA $\times$ CONTIG	-0.250 (0.096)***	-0.266 (0.140)*	
	EIA $\times$ LEGAL	0.160 (0.078)**	0.189 (0.111)*	
	EIA $\times$ RELG	0.418 (0.125)***	0.456 (0.166)***	
	EIA $\times$ LANG	-0.414 (0.103)***	-0.457 (0.147)***	
Baier et al. (2014)	EIA $\times$ CLNY	-0.082 (0.116)	-0.051 (0.147)	NRPTA is a non-reciprocal trade preference. PTA is a reciprocal preferential trade agreement. FTA is a free trade agreement with additional provisions beyond tariff reductions. CUCMECU refers to customs unions, monetary unions, and economic unions. "Total" effects are inclusive of 5 year lags. Our estimation differs from BBF's in that we use PPML, whereas BBF use OLS with first-differences and pair-specific time trends. The specification we replicate is their Table 1, Column 2a. 149 countries, 1965–2000 ( $N = 149, T = 8$ ). Reported coefficient is for the "Total FTA effect" inclusive of 5-year lags. The specification we estimate is their Eq. (4). Our estimate differs slightly because we use asymmetric pair fixed effects instead of symmetric pair fixed effects. Our standard errors differ because we cluster by pair. 69 countries, 1986–2006 ( $N = 69, T = 21$ ).
	Total NRPTA Effect	0.020 (0.091)	0.035 (0.110)	
	Total PTA Effect	0.149 (0.079)*	0.160 (0.089)*	
	Total FTA Effect	0.365 (0.045)***	0.352 (0.061)***	
Baier et al. (2019)	Total CUCMECU Effect	0.670 (0.060)***	0.669 (0.070)***	Reported coefficient is for the "Total EIA effect" inclusive of 4- and 8-year lags for the same specification as in their Table 2, column 4. Baseline estimates differ slightly because BLY originally used symmetric pair fixed effects whereas we use asymmetric pair fixed effects. 41 countries, 1990–2002, every 4 years ( $N = 41, T = 4$ ).
	Total FTA Effect	0.292 (0.051)***	0.299 (0.074)***	
Bergstrand et al. (2015)	Total EIA Effect	0.521 (0.060)***	0.546 (0.098)***	The specification we replicate is their Table 1, column 5, which in turn replicates the main specification from <a href="#">Glick and Rose (2016)</a> . Our standard errors differ because we cluster by pair. 213 countries, 1948–2013 ( $N = 213, T = 66$ ).
Larch et al. (2019)	Euro	0.030 (0.042)	0.021 (0.046)	These reported coefficient is from Table 4 of <a href="#">Rose (2019)</a> . Log approval rating is a measure of the importing country's approval of the exporting country's leader. There are 5 exporters, 149 importers, and 12 years (2006–2017).
	Other CUs	0.700 (0.107)***	0.708 (0.142)***	
	RTA	0.169 (0.040)***	0.159 (0.050)***	
Rose (2019)	Curr. Colony	0.545 (0.220)**	0.577 (0.262)**	
	log approval rating	0.053 (0.022)***	0.058 (0.028)**	

This table shows replications of several recent papers that use three-way gravity models with exporter-time, importer-time, and country-pair fixed effects. Standard errors shown in parentheses. Bias-corrected estimates use our analytical corrections for the point estimates and standard errors. \*  $p < 0.10$ , \*\*  $p < .05$ , \*\*\*  $p < .01$ .

## 6. Conclusion

Thanks to recent methodological and computational advances, nonlinear models with three-way fixed effects have become increasingly popular for investigating the effects of trade policies on trade flows. However, the asymptotic and finite-sample properties of three-way fixed effects estimators have not been rigorously studied, especially with regards to potential IPPs. The performance of the FE-PPML estimator in particular is of natural interest in this context, both because FE-PPML is known to be relatively robust to IPPs as well as because it is likely to be a researcher's first choice for estimating three-way gravity models. Our results regarding the consistency of PPML in this setting reflect these unique properties of PPML and support its current status as a workhorse estimator for estimating the effects of trade policies.

Given the consistency of PPML in this setting, and given the nice IPP-robustness properties of PPML in general, it may come as a surprise that three-way PPML estimates nonetheless suffer from an asymptotic bias that affects the validity of inferences. In theory, the bias should become less of a problem when the country and time dimensions are both large, but our experiments with the time dimension indicate the bias can be of comparable magnitude to the standard error even in ostensibly large trade data sets. Typical cluster-robust estimates of the standard error are also biased, implying estimated confidence intervals not only off-center but also too narrow.

These issues are not so severe that they leave researchers in the wilderness, but we do recommend taking advantage of the corrective measures we have described. In particular, we find that analytical bias corrections based on Taylor expansions to both the point estimates and standard errors generally lead to improved inferences when applied simultaneously. We caution that we have not found these corrections to be a panacea, however, and several avenues remain open for future work. For example, confidence interval estimates could be adjusted further to account for the uncertainty in the estimated variance—[Kauermann and Carroll \(2001\)](#) describe such a correction for the PPML case. A quasi-differencing approach similar to [Jochmans \(2017\)](#) could provide another angle of attack, and a recent contribution by [Pfaffermayr \(2021\)](#) suggests that jackknife and bootstrap confidence interval methods hold promise as well. Turning to broader applications, the essential dyadic structure of our bias corrections could be easily adapted to network models that study changes in network behavior over time, including settings that involve studying the number of interactions between network members.

## Acknowledgements

Thomas Zylkin is grateful for support from NUS Strategic Research Grant WBS: R-109-000-183-646 awarded to the Global Production Networks Centre at National University of Singapore and from the Economic and Social Research Council through ESRC grant EST013567/1. Martin Weidner acknowledges support from the Economic and Social Research Council through the ESRC Centre for Microdata Methods and Practice grant RES-589-28-0001 and from the European Research Council grants ERC-2014-CoG-646917-ROMIA and ERC-2018-CoG-819086-PANEDA. We also thank Treb Allen, two anonymous referees, Valentina Corradi, Riccardo D'adamo, Ivan Fernández-Val, Koen Jochmans, Maia Linask, Michael Pfaffermayr, Raymond Robertson, Ben Shepherd, Amrei Stammann, Yoto Yotov, and participants at seminars and conferences organized by the College of William and Mary, the FREIT Empirical Trade Online Seminar, the Econometric Society, the Midwest International Economics Group, the Southern Economic Association, and the Western Economic Association. Kelley Ranager provided excellent assistance with the creation of our graphical abstract.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jinteco.2021.103513>.

## References

- Allen, T., Dobbin, C.D.C., Morten, M., 2018. Border Walls. Discussion paper, National Bureau of Economic Research.
- Alvarez, J., Arellano, M., 2003. The time series and cross-section Asymptotics of dynamic panel data estimators. *Econometrica* 71 (4), 1121–1159.
- Andersen, E.B., 1970. Asymptotic properties of conditional maximum-likelihood estimators. *J. R. Stat. Soc. Ser. B Methodol.* 32 (2), 283–301.
- Anderson, J.E., van Wincoop, E., 2003. Gravity with gravitas: a solution to the border puzzle. *Am. Econ. Rev.* 93 (1), 170–192.
- Arellano, M., Bonhomme, S., 2009. Robust priors in nonlinear panel data models. *Econometrica* 77 (2), 489–536.
- Arellano, M., Hahn, J., 2007. Understanding bias in nonlinear panel models: some recent developments. *Econ. Soc. Monogr.* 43, 381.
- Bai, J., 2009. Panel data models with interactive fixed effects. *Econometrica* 77 (4), 1229–1279.
- Baier, S.L., Bergstrand, J.H., 2007. Do free trade agreements actually increase Members' International trade? *J. Int. Econ.* 71 (1), 72–95.
- Baier, S.L., Bergstrand, J.H., Feng, M., 2014. Economic integration agreements and the margins of international trade. *J. Int. Econ.* 93 (2), 339–350.
- Baier, S.L., Bergstrand, J.H., Clance, M.W., 2018. Heterogeneous effects of economic integration agreements. *J. Dev. Econ.* 135, 587–608.
- Baier, S.L., Yotov, Y.V., Zylkin, T., 2019. On the widely differing effects of free trade agreements: lessons from twenty years of trade integration. *J. Int. Econ.* 116, 206–226.
- Bergstrand, J.H., Larch, M., Yotov, Y.V., 2015. Economic integration agreements, border effects, and distance Elasticities in the gravity equation. *Eur. Econ. Rev.* 78.
- Beverelli, C., Orefice, G., 2019. Migration deflection: the role of preferential trade agreements. *Reg. Sci. Urban Econ.* 79, 103469.
- Brinkman, J., Lin, J., 2019. Freeway Revolts! Unpublished manuscript.
- Cameron, A.C., Gelbach, J.B., Miller, D.L., 2008. Bootstrap-based improvements for inference with clustered errors. *Rev. Econ. Stat.* 90 (3), 414–427.
- Carro, J.M., 2007. Estimating dynamic panel data discrete choice models with fixed effects. *J. Econ.* 140 (2), 503–528.
- Charbonneau, K.B., 2017. Multiple fixed effects in binary response panel data models. *Econ. J.* 20 (3), S1–S13.
- Chen, M., Fernández-Val, I., Weidner, M., 2019. Nonlinear factor models for network and panel data. *arXiv preprint. arXiv:1412.5647*.
- Cheng, I.H., Wall, H.J., 2005. Controlling for Heterogeneity in gravity models of trade. *Fed. Reserve Bank St. Louis Rev.* 87 (1), 49–63.
- Correia, S., Guimarães, P., Zylkin, T., 2020. Fast Poisson estimation with high-dimensional fixed effects. *Stata J.* 20 (1), 95–115.
- Costinot, A., Rodríguez-Clare, A., 2014. Trade theory with numbers: quantifying the consequences of globalization. *Handb. Int. Econ.* 4, 197–261.
- Dhaene, G., Jochmans, K., 2015. Split-panel jackknife estimation of fixed-effect models. *Rev. Econ. Stud.* 82 (3), 991–1030.
- Dzemska, A., 2019. An empirical model of dyadic link formation in a network with unobserved heterogeneity. *Rev. Econ. Stat.* 101 (5), 763–776.
- Eaton, J., Kortum, S., 2002. Technology, geography, and trade. *Econometrica* 70 (5), 1741–1779.
- egger, P.H., Staub, K.E., 2015. GLM estimation of trade gravity models with fixed effects. *Empir. Econ.* 50 (1), 137–175.
- egger, P., Larch, M., Staub, K.E., Winkelmann, R., 2011. The trade effects of endogenous preferential trade agreements. *Am. Econ. J. Econ. Pol.* 3 (3), 113–143.
- Fally, T., 2015. Structural gravity and fixed effects. *J. Int. Econ.* 97 (1), 76–85.
- Fernández-Val, I., Vella, F., 2011. Bias corrections for two-step fixed effects panel data estimators. *J. Econ.* 163 (2), 144–162.
- Fernández-Val, I., Weidner, M., 2016. Individual and time effects in nonlinear panel models with large N, T. *J. Econ.* 192 (1), 291–312.
- Fernández-Val, I., Weidner, M., 2018. Fixed Effect Estimation of Large T Panel Data Models. *Economics, Annual Review of*.
- Gaduh, A., Gračner, T., Rothenberg, A.D., 2020. Life in the Slow Lane: Unintended Consequences of Public Transit in Jakarta Unpublished manuscript.
- Gaulier, G., Zignago, S., 2010. "BACI: International Trade Database at the Product-Level. The 1994–2007 Version," Working Paper 2010–23. CEPII research center.
- Glick, R., Rose, A.K., 2016. Currency unions and trade: a post-EMU reassessment. *Eur. Econ. Rev.* 87, 78–91.
- Graham, B.S., 2017. An econometric model of network formation with degree heterogeneity. *Econometrica* 85 (4), 1033–1063.
- Hahn, J., Kuersteiner, G., 2002. Asymptotically unbiased inference for a dynamic panel model with fixed effects when both N and T are large. *Econometrica* 70 (4), 1639–1657.
- Hahn, J., Moon, H.R., 2006. Reducing bias of MLE in a dynamic panel model. *Econom. Theory* 22 (3), 499–512.
- Hansen, C.B., 2007. Asymptotic properties of a robust variance matrix estimator for panel data when T is large. *J. Econ.* 141 (2), 597–620.
- Hausman, J., Hall, B.H., Griliches, Z., 1984. Econometric models for count data with an application to the Patents-R&D Relationship. *Econometrica* 52 (4), 909–938.
- Head, K., Mayer, T., 2014. Gravity equations: workhorse, toolkit, and cookbook. *Handb. Int. Econ.* 4, 131–196.
- Helpman, E., Melitz, M., Rubinstein, Y., 2008. Estimating trade flows: trading partners and trading volumes. *Q. J. Econ.* 123 (2), 441–487.
- Hinz, J., Stammann, A., Wanner, J., 2020. State dependence and unobserved heterogeneity in the extensive margin of trade arXiv preprint arXiv:2004.12655.
- Imbens, G.W., Kolesar, M., 2016. Robust standard errors in small samples: some practical advice. *Rev. Econ. Stat.* 98 (4), 701–712.
- Jochmans, K., 2017. Two-way models for gravity. *Rev. Econ. Stat.* 99 (3), 478–485.
- Kato, K., Galvao Jr., A.F., Montes-Rojas, G.V., 2012. Asymptotics for panel Quantile regression models with individual effects. *J. Econ.* 170 (1), 76–91.
- Kauermann, G., Carroll, R.J., 2001. A note on the efficiency of sandwich covariance matrix estimation. *J. Am. Stat. Assoc.* 96 (456), 1387–1396.
- Lancaster, T., 2002. Orthogonal parameters and panel data. *Rev. Econ. Stud.* 69 (3), 647–666.
- Larch, M., Wanner, J., Yotov, Y.V., Zylkin, T., 2019. Currency unions and trade: a PPML re-assessment with high-dimensional fixed effects. *Oxf. Bull. Econ. Stat.* 81 (3), 487–510.
- Mackinnon, J.G., White, H., 1985. Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *J. Econ.* 29 (3), 305–325.
- Moon, H.R., Weidner, M., 2017. Dynamic linear panel regression models with interactive fixed effects. *Econom. Theory* 33 (1), 158–195.
- Neyman, J., Scott, E.L., 1948. Consistent estimates based on partially consistent observations. *Econometrica* 16 (1), 1–32.
- Palmgren, J., 1981. The fisher information matrix for log linear models arguing conditionally on observed explanatory variables. *Biometrika* 68 (2), 563–566.
- Pesaran, M.H., 2006. Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica* 74 (4), 967–1012.

- Pfaffermayr, M., 2019. Gravity models, PPML estimation and the bias of the robust standard errors. *Appl. Econ. Lett.* 1–5.
- Pfaffermayr, M., 2021. Confidence intervals for the trade cost parameters of cross-section gravity models. *Econ. Lett.* 201, 109787.
- Phillips, P.C.B., Moon, H.R., 1999. Linear regression limit theory for nonstationary panel data. *Econometrica* 67 (5), 1057–1111.
- Rose, A.K., 2019. Soft power and exports. *Rev. Int. Econ.* 27 (5), 1573–1590.
- Santos Silva, J.M.C., Tenreyro, S., 2006. The log of gravity. *Rev. Econ. Stat.* 88 (4), 641–658.
- Stammann, A., 2018. Fast and feasible estimation of generalized linear models with high-dimensional k-way fixed effects. arXiv preprint. *arXiv:1707.01815*.
- Verdier, V., 2020. Estimation and inference for linear models with two-way fixed effects and sparsely matched data. *Rev. Econ. Stat.* 102 (1), 1–16 (forthcoming).
- Wooldridge, J.M., 1999. Distribution-free estimation of some nonlinear panel data models. *J. Econ.* 90 (1), 77–97.
- Woutersen, T., 2002. Robustness against Incidental Parameters. University of Western Ontario, Department of Economics, Discussion paper.
- Yotov, Y.V., Piermartini, R., Monteiro, J.-A., Larch, M., 2016. An Advanced Guide to Trade Policy Analysis: The Structural Gravity Model. World Trade Organization, Geneva.