Bias in Poisson Pseudo-Maximum Likelihood Estimation of Structural Gravity Models: How Much of a Problem for Applied Research?

Ben Shepherd. Principal, Developing Trade Consultants. Ben@Developing-Trade.Com.[1]

June 23rd, 2023.

Abstract: Theory-consistent ("structural") gravity models are commonly used to model bilateral trade. Extensive simulations suggest that the Poisson Pseudo-Maximum Likelihood (PPML) estimator performs well in that setting. However, an influential review by Head and Mayer (2014) ("HM") includes a simulation where PPML exhibits significant bias, which leads them to recommend a toolbox approach, including PPML and other estimators. I show that PPML's bias is related to the noisiness of the data.. A more realistic error variance assumption reveals that even for the same distribution type as HM, PPML's bias is 4% rather than the 27% they report.

Keywords: International trade; Gravity model; Poisson Pseudo-Maximum Likelihood estimation.

JEL Codes: F14; F15; C13.

[1] I am grateful to Thierry Mayer for graciously providing replication data and code, and to Tom Zylkin for helpful discussions.

1 INTRODUCTION

Santos Silva and Tenreyro (2006) show that OLS produces inconsistent estimates of log-linearized models when the error term is heteroscedastic. Given that trade data are typically heteroskedastic, they recommend the Poisson Pseudo-Maximum Likelihood (PPML) estimator as the workhorse for theory-consistent ("structural") gravity models, where bilateral flows are a function of market size, trade costs, and unobserved relative prices (Anderson and Van Wincoop, 2003). PPML provides consistent estimates in a wide variety of settings. Arvis and Shepherd (2013) and Fally (2015) demonstrate that PPML is unique among Generalized Linear Models (GLMs) in producing estimates that accord exactly with theory, as in Anderson and Van Wincoop (2003). Weidner and Zylkin (2021) prove that PPML estimates with two- and three-dimensional fixed effects are not inconsistent due to the incidental parameters problem, although bias corrections may be necessary.

Simulation evidence supports this analysis. Santos Silva and Tenreyro (2006) consider strictly positive trade, while Santos Silva and Tenreyro (2011) show that PPML performs well even with a high proportion of zeros in the dependent variable. Santos Silva and Tenreyro (2021) review 15 years of literature, and conclude that the recommendation to use PPML as the workhorse for structural gravity has held up well to scrutiny. Of course, they recognize that no estimator is optimal in every set of empirical conditions. Their recommendation is that PPML represents a sensible first candidate in the absence of further knowledge about the data generating process.

Head and Mayer (2014) ("HM") is a widely cited review of gravity modeling, focusing on theoretical underpinnings and estimation. It presents simulation evidence using a data generating process consistent with structural gravity under a firm heterogeneity framework with fixed market entry costs, in which some trade flows are zero. As part of this simulation, they find that PPML performs poorly when the error term is a homoscedastic log-normal variable (HM Table 7). As a result, they recommended a toolbox approach to estimation, involving comparison of results from various estimators including PPML and the use of specification tests to ensure robustness.

This paper shows that the HM simulation results for PPML are dependent on the magnitude of the error variance, which does not reflect the observed data upon which their simulation is based. A more realistic specification of the error term, even as a homoscedastic log-normal variable, produces only a minor degree of bias. The objective of this paper is therefore to rebalance applied researchers' perceptions of the different estimators in HM's toolkit, by showing where PPML's bias comes from, and how it plays out in a simulation context that closely mirrors their observed source data.

PPML estimates are in general biased even though they are consistent, and the bias depends on sample size and the error variance; however, this result has not been addressed in the applied international trade literature due to the fact that trade data typically have large N (Section 2). Section 3 contextualizes the HM simulation by considering a variety of values for the error variance, thereby showing that their results are not driven by homoscedasticity or zeros. In Section 4, I conduct a fresh simulation where sample moments accord more closely with HM's observed trade flows, and show that all estimators in the HM toolkit in fact perform similarly: PPML's bias is around 5% rather than the 27% HM report. The final section discusses directions for further research.

2 BIAS IN PPML ESTIMATION

It is well-known in the econometrics and statistics literatures that PPML estimates are biased in small samples, as is the case for nonlinear maximum likelihood estimates in general(Cox and Hinkley, 1974).

This section presents the intuition behind that result, which has not been addressed in the applied international trade literature because trade data typically have large sample sizes.

The score function U for PPML with a single explanatory variable is:

$$(1)\ U(b) = \sum_{i=1}^{N} (y_i - \exp(x_i b)\, x_i$$

Let $\hat{b}$ be the parameter estimate obtained by setting (1) equal to zero and solving, with b as the true parameter. A second order Taylor series expansion around the true parameter gives:

$$(2)\ U(\hat{b}) = 0 \approx U(b) + (\hat{b} - b)U'(b) + \frac{1}{2}(\hat{b} - b)^2 U''(b)$$

Taking expectations shows that the second and third terms vanish as N grows without bound, but that in finite samples, the third term is non-zero, which means that the estimate $\hat{b}$ is biased. A second implication is that the third term depends on the variance of $\hat{b}$, which means that it also shrinks as the error variance declines.

3 Unpacking HM's Simulation

HM's simulation only considers one set of parameter values. But the above analysis suggests that sample size and data noisiness should influence PPML's empirical bias. This section therefore contextualizes their simulations by considering a range of parameter values and sample sizes. I use their data as the basis for each simulation, construct trade flows consistent with structural gravity in the same way, and use the same assumptions for error terms. As in their simulations, observable trade costs are log distance (coefficient = -1.000) and RTA membership (coefficient = 0.500). I consider square datasets with 50, 100, and 150 countries.

System (1a)-(1d) is a standard structural gravity model (Anderson and Van Wincoop, 2003) with exporter and importer fixed effects S and M, where Y is output and E is expenditure; t is trade costs, and the cs are its observable components; theta is the trade elasticity; i and j index exporters and importers; and e is an error term with unitary mean and finite variance, which captures unobservable trade costs:

$$(1a)\ X_{ij} = \exp\left(S_i + M_j + t_{ij}^{-\theta}\right) e_{ij}$$

$$(1b)\ \exp(S_i) = \frac{Y_i}{\sum_l \frac{E_l}{t_{il} e_{il}}}$$

$$(1c)\ \exp(M_j) = \frac{E_j}{\sum_l S_l\, t_{lj} e_{lj}}$$

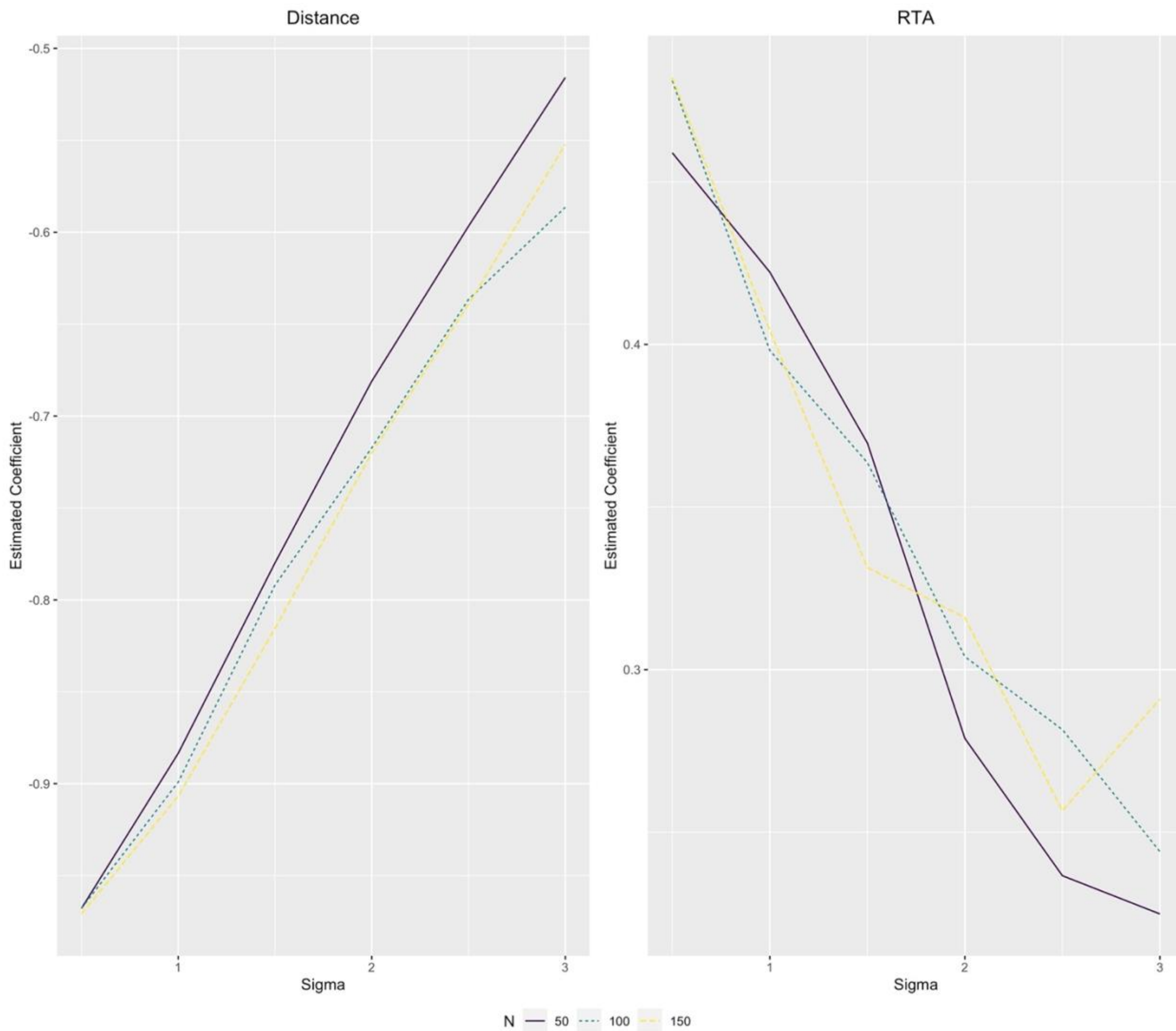
$$(1d)\ \log t_{ij} = \sum\nolimits_k b_k c_{ij}^k$$

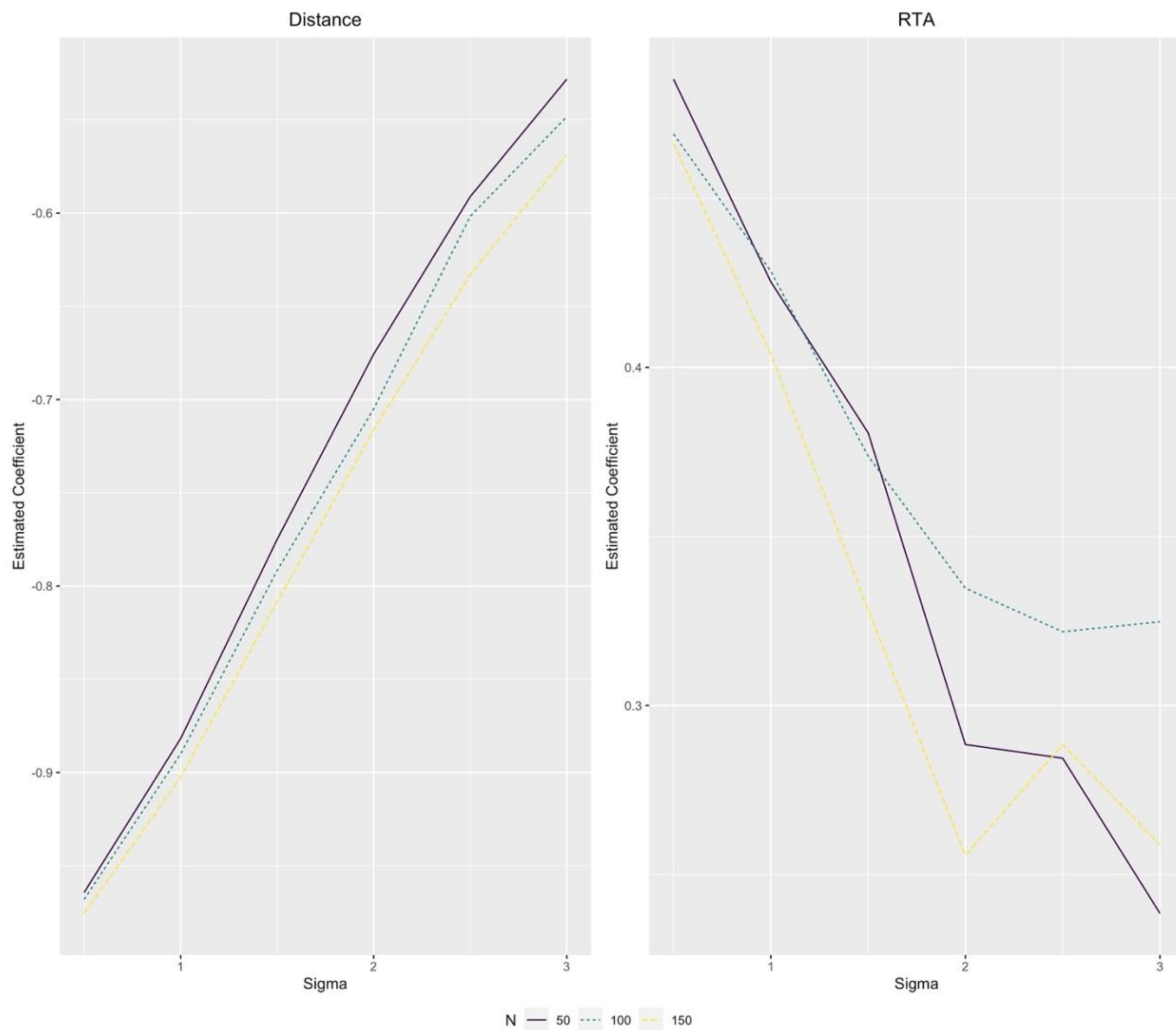
To conduct simulations that are comparable to HM, I focus on e. The upper panel of Figure 1 shows results for a homoscedastic log-normal error term[2] with zero trade flows determined consistently with

[2] $e_{ij} = \exp\,(n_{ij})$, where $n_{ij} \sim N(0, \sigma)$.

a heterogeneous firms model with fixed market entry costs, as used for HM Table 7. The standard deviation of the logarithm of the error term ranges from 0.500 to 3.000 in increments of 0.500; HM's value is 2.000. Comparing the two panels shows that structural zeros are not determinative: the bias is similar with and without structural zeros. The figure evidences bias in PPML estimates for HM's error variance, but shows that its extent is increasing in the dependent variable's noisiness, and decreasing in the number of countries in the dataset, as predicted.

Figure 1: Simulation results for log-normal homoscedastic errors with (top) and without (bottom) structural zeros.

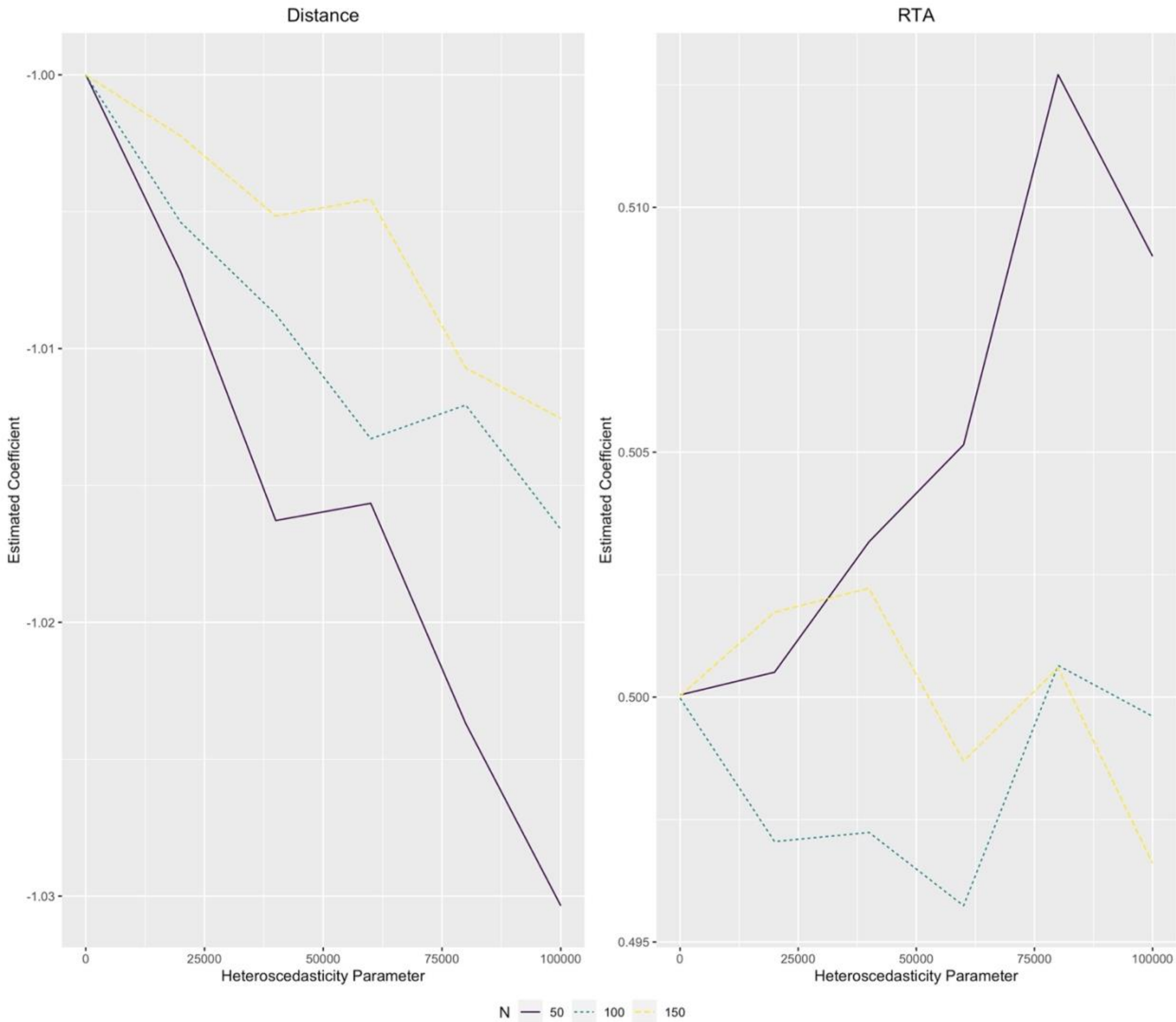




HM report bias in PPML estimates with a log-normal homoscedastic error, but not with a heteroscedastic error. Figure 2 contextualizes this result by varying the heteroscedasticity parameter H (100 in their simulation, ranging from 1 to 100,001 in mine, in increments of 20,000), where a larger value increases the error variance.[3] The conclusion is the same: there are parameter values for which PPML's bias is minimal, but there are also values for which it is significant. For a given sample size, the determinative factor is the noisiness of the dependent variable, which depends on the heteroscedasticity parameter.

[3] $e_{ij} = \exp(n_{ij})$ where $n_{ij} \sim N\left(-\frac{\log\left(1+{}^{H}/_{\hat{X}_{ij}}\right)}{2}, \sqrt{\log\left(1+{}^{H}/_{\hat{X}_{ij}}\right)}\right)$..

Figure 2: Simulation results for log-normal heteroscedastic errors without structural zeros.



4 Practical Implications

The previous sections showed that PPML estimates are generally biased but consistent, and that the degree of bias depends on sample size and noisiness. From an applied perspective, the key question is: how serious is PPML's bias in a real world setting?

The HM dataset includes observed trade data in addition to simulated values. Comparing the two shows that the simulateds are much noisier than the observations (variance to mean ratio of 1,555,074 versus 53,327). For a given sample size, PPML's simulated bias is therefore likely to be substantially worse than its bias when using observed data.

I conduct a simulation that accords more closely with the moments of observed trade in the HM dataset. I fill in missing observations in CIF imports by mirroring, then keep only non-missing observations.. I perform 1,000 simulations using the same trade cost coefficients as HM and observed trade costs to create simulated bilateral flows for non-missing observations. I use total exports and imports as proxies for production and expenditure. I set the standard deviation of the logarithm of the error to 0.6, compared with 2.0 in HM, based on a grid search. The simulated and observed variances accord closely (ratio = 1.051) and the means are identical.

Table 1 reports results for the estimators in HM's toolkit approach, including PPML. First, the realistic error variance assumption results in much smaller bias for PPML and MPML (numerically equivalent to PPML in expenditure shares rather than levels). For distance, HM report a PPML bias of 27%, whereas in the new simulation, it is only 4%. For the RTA dummy, the HM bias is 42%, compared with 5% in the new simulation. MPML estimates show even smaller bias. While there is no single estimator that is optimal in every context, this exercise shows that PPML's bias is much less severe than previously thought.

Table 1: Simulation results for the coefficients of log distance (upper panel) and RTA dummy (lower panel).

	$\sigma = 0.6$		$\sigma = 2.0$	
	Estimate	Std. Dev.	Estimate	Std. Dev.
LSDV	-0.968	(0.008)	-0.81	(0.02)
ETT	-0.978	(0.009)	-0.94	(0.02)
EKT	-0.996	(0.008)	-0.99	(0.02)
PPML	-0.960	(0.052)	-0.73	(0.14)
GPML	-1.192	(0.012)	-1.05	(0.04)
MPML	-0.991	(0.021)	-0.79	(0.06)

	$\sigma = 0.6$		$\sigma = 2.0$	
	Estimate	Std. Dev.	Estimate	Std. Dev.
LSDV	0.527	(0.017)	0.63	(0.06)
ETT	0.489	(0.018)	0.53	(0.06)
EKT	0.503	(0.017)	0.50	(0.06)
PPML	0.476	(0.116)	0.29	(0.43)
GPML	0.280	(0.026)	0.41	(0.11)
MPML	0.480	(0.048)	0.36	(0.15)

Note: Abbreviations follow HM. Results for σ=2.0 come from HM Table 7 (reported precision). GPML results should interpreted cautiously due to convergence issues.

5 Conclusion

While PPML estimates are consistent under weak assumptions, they are generally biased. When simulated data are closely matched to observed noisiness, however, that bias is minimal. While the econometrics literature discloses attempts to remedy it (Firth, 1993), they do not prove effective with the type of bias considered here. Developing further corrections could therefore be a fruitful avenue for research.

References

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

Arvis, J.-F., and B. Shepherd. 2013. "The Poisson Quasi-Maximum Likelihood Estimator: A Solution to the 'Adding Up' Problem in Gravity Models." *Applied Economics Letters*, 20(6): 515-519.

Cox, D., and D. Hinkley. 1974. *Theoretical Statistics*, New York: Wiley.

Fally, T. 2015. "Structural Gravity and Fixed Effects." *Journal of International Economics*, 97(1): 76-85.

Firth, D. 1993. "Bias Reduction of Maximum Likelihood Estimates." *Biometrika*, 80(1): 27-38.

Head, K., and T. Mayer. 2014. "Gravity Equations: Workhorse, Toolkit, and Cookbook." In G. Gopinath, E. Helpman, and K. Rogoff (eds.) *Handbook of International Economics*, 4: 131-195.

Santos Silva, J., and S. Tenreyro. 2006. "The Log of Gravity." *Review of Economics and Statistics*, 88(4): 641-658.

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

Santos Silva, J., and S. Tenreyro. 2021. "The Log of Gravity at 15." *Portuguese Economic Journal*, 21: 423-437.

Weidner, M., and T. Zylkin. 2021. "Bias and Consistency in Three-Way Gravity Models." *Journal of International Economics*, 132: 103513.