A New Bayesian Bootstrap for Quantitative Trade and Spatial Models

Bas Sanders, *Harvard University**

August 2026

Abstract

Economists use quantitative trade and spatial models to make counterfactual predictions that inform policy. Quantifying uncertainty in these predictions is challenging because the data are dyadic, the number of interacting units is typically small, and counterfactuals depend on both estimated structural parameters and the observed status quo. I propose a Bayesian bootstrap tailored to this setting. The procedure is simple to implement and provides both finite-sample Bayesian and asymptotic frequentist guarantees. I illustrate the practical advantages of this approach by revisiting the applications in Waugh (2010), Caliendo and Parro (2015), and Artuç, Chaudhuri, and McLaren (2010).

1 Introduction

Economists use quantitative trade and spatial models to answer counterfactual questions. For example, what is the effect on welfare levels and inequality when trade costs or tariffs between a set of countries change? What happens to employment shares and wages across sectors after a sudden liberalization of the manufacturing sector? Since such counterfactual predictions aim to inform policy decisions, it is important to communicate the uncertainty

*E-mail: bas_sanders@g.harvard.edu. I thank my advisors, Isaiah Andrews, Pol Antràs, Anna Mikusheva and Jesse Shapiro, for their guidance and generous support. I also thank Dmitry Arkhangelsky, Kirill Borusyak, Dave Donaldson, Tilman Graff, Xavier Jaravel, Marc Melitz, Neil Shephard, Rahul Singh, Elie Tamer, Davide Viviano and Chris Walker for helpful discussions. I am also grateful for comments from seminar participants at University of Amsterdam, Erasmus University Rotterdam, Tilburg University, Free University Amsterdam, the Conference of the International Association for Applied Econometrics, Radboud University Nijmegen, the CPB Netherlands Bureau for Economic Policy Analysis and the Econometrics of Equilibrium Effects Conference.

surrounding them. However, in practice, counterfactuals are often reported without any measure of uncertainty. For instance, in a survey of recently published articles only 2 out of 36 papers report any uncertainty quantification for their counterfactual predictions.[1]

Counterfactual predictions are typically constructed in two steps. First, the data are used to estimate a finite-dimensional structural parameter, for example using the generalized method of moments (GMM). Second, the estimator is combined with the observed data—which reflect the current state of the world—to compute the counterfactual prediction. For instance, in the canonical Armington model (Armington, 1969), the first step involves estimating a trade elasticity using observed bilateral trade flows. In the second step, the estimated elasticity is combined with the trade flows to predict welfare changes under a hypothetical shift in trade costs.

Quantifying uncertainty for such a counterfactual raises three main challenges. First, the data are often dyadic, meaning that each observation reflects an interaction between two units. This induces a strong dependence structure across observations. Second, the number of interacting units—such as countries or sectors—is typically small, making it important to use methods that retain a clear interpretation in small samples. Third, the data enter both the estimation of the structural parameter and the computation of the counterfactual prediction, so that the prediction depends on the data in two distinct ways. This creates a non-classical setting for uncertainty quantification, as was discussed in Sanders (2023).

The practical contribution of this paper is to address these challenges by providing a unified method for inference and counterfactual analysis, which supports more informed policy decisions. Specifically, to quantify uncertainty for the estimator of the structural parameter, I introduce a new Bayesian bootstrap procedure that is intuitive, easy to implement, and theoretically grounded. The procedure amounts to reweighting the data using products of draws from an exponential distribution. It readily extends to settings where only subset of all possible flows is observed (e.g. because observations that equal zero are dropped), or where the data are polyadic (i.e., each observation involves more than two units). Because the approach is Bayesian, it also implies a posterior distribution for the counterfactual prediction with a finite-sample interpretation. Provided one is satisfied with the Bayesian interpretation, the shape of the posterior conveys valuable information for decision-making. For instance, right-skewness could suggest greater potential for large welfare gains, while left-skewness indicates a chance of substantial welfare losses.

[1]The survey includes all papers published between 2015 and 2024 in *American Economic Review*, *Econometrica*, *Journal of Political Economy*, *Quarterly Journal of Economics*, and *Review of Economic Studies*) that contain the phrase "bilateral trade flows" or "bilateral flows" and conduct a counterfactual exercise.

The key theoretical contribution of this paper is to introduce and justify a new Bayesian bootstrap procedure tailored to polyadic data structures. The procedure extends the classical Bayesian bootstrap (Rubin, 1981; Chamberlain and Imbens, 2003) to settings where each observation involves more than a single unit. One main result of this paper is to show that this procedure admits a finite-sample Bayesian interpretation. Specifically, I show that the Bayesian bootstrap distribution arises as the limit of a sequence of Bayesian posteriors, under a particular model and class of priors. The underlying model assumes that the polyadic data are generated as functions of unit-specific latent variables, which are drawn independently from a common distribution. The distribution of latent variables acts as an unknown parameter and is endowed with a Dirichlet process prior.

The fact that the Bayesian bootstrap procedure admits a finite-sample Bayesian interpretation is particularly relevant in applications with a small number of units. In addition to its finite-sample Bayesian validity, the procedure is also asymptotically valid in a frequentist sense under mild regularity conditions. These conditions are generally satisfied, for example, by the class of GMM estimators, including the Pseudo Poisson Maximum Likelihood (PPML) estimator of Silva and Tenreyro (2006). This dual validity makes the procedure competitive with existing methods in the literature—reviewed below—whose sole justification is asymptotic. For frequentist uncertainty quantification on the counterfactual prediction, I provide a delta method-type result that accounts for the fact that a counterfactual prediction can depend on the data in two distinct ways.

Throughout the paper, I use the application in Waugh (2010) as a running example. In this setting, the structural parameter is a productivity parameter common across countries; the interacting units are 43 countries; and the estimation method is simple OLS on dyadic trade flows—a special case of GMM. The posterior variance implied by my procedure is considerably larger than the heteroskedasticity-robust variance reported in Waugh (2010), which does not account for dependence across dyads—such as trade flows that share a country in common. The counterfactual objects of interest are various inequality statistics under alternative trade cost schedules, for which I construct credible intervals; these intervals are narrow, and there is not much economically meaningful uncertainty in the counterfactuals. Thus, despite yielding much more uncertainty about model parameters, my approach delivers precise inference on counterfactuals.

To further illustrate the flexibility of the procedure, I also revisit results in Caliendo and Parro (2015) and Artuç, Chaudhuri, and McLaren (2010). In Caliendo and Parro (2015), the structural parameters are sector-specific trade elasticities; the interacting units are countries;

and the estimation method is simple OLS on *triadic flows*. The number of countries per sector ranges from 11 to 15. The credible intervals for the elasticities are substantially wider than the heteroskedasticity-robust confidence intervals reported in the original paper, and they often include zero. For some sectors, the posterior distribution of the elasticity is approximately normal; for others, it is skewed or heavy-tailed. The counterfactual objects of interest are changes in welfare due to NAFTA, originally reported in Caliendo and Parro (2015) without uncertainty quantification. The credible intervals around welfare predictions reflect substantial uncertainty and considerable heterogeneity across countries, although the ranking of welfare effects across countries remains unchanged.

In the setting of Artuç, Chaudhuri, and McLaren (2010), the structural parameters are the mean and variance of workers' switching costs between sectors; the interacting units are *six* sectors, and the estimation method is *over-identified GMM* with three instruments. The posterior distributions for both parameters are non-normal and exhibit heavy right tails, indicating substantial uncertainty—particularly regarding the possibility of large switching costs. The counterfactual objects of interest are changes in various economic outcomes following a liberalization of the manufacturing sector, and the resulting credible intervals again reveal substantial uncertainty. Notably, accounting for this uncertainty reveals that equilibrium wages may plausibly increase as a result of liberalization—a finding not visible from point estimates alone.

The procedure I propose substantially improves how uncertainty is quantified for both the structural parameter and the counterfactual prediction, relative to current practice in quantitative trade and spatial economics. In the survey mentioned above, 24 out of 36 papers report a standard error for the estimator of the structural parameter. However, the most common approach is to compute heteroskedasticity-robust standard errors by clustering either on dyads or only on the origin or destination unit, implicitly ignoring important dependence across flows. A more flexible alternative, used in several papers, is two-way clustering on both the origin and destination units. While two-way clustering allows for richer dependence, it still fails to capture key dyadic correlations—for example, between the trade flow from Germany to the United States and the trade flow from France to Germany, since these share neither an exporter or importer. Ideally, one would allow for dependence between flows that have at least one unit in common. A small literature proposes methods to account for such dyadic dependence (Fafchamps and Gubert, 2007; Cameron and Miller, 2014; Aronow, Samii, and Assenova, 2015; Graham, 2020a,b; Davezies, D'haultfœuille, and Guyonvarch, 2021). However, it remains rare for published papers in quantitative trade and

spatial economics to adopt these tools.[2]

I compare my procedure to two alternatives from the recent econometrics literature using my applications and a Monte Carlo study. The closest alternative is the pigeonhole bootstrap introduced by Davezies, D'haultfœuille, and Guyonvarch (2021), which extends the standard resampling bootstrap to polyadic settings. Both my approach and the pigeonhole bootstrap reweight polyadic observations using specific weights. Practically, one key difference is that the Bayesian bootstrap assigns continuous and strictly positive weights to all observations, whereas the pigeonhole bootstrap draws discrete weights and may assign zero weight to some. A second difference is theoretical. Existing guarantees for the pigeonhole bootstrap rely on asymptotic approximations that require a large number of interacting units. By contrast, the finite-sample Bayesian interpretation developed here applies directly to settings with a small number of interacting units, which are common in quantitative trade and spatial applications. In line with these theoretical considerations, the Monte Carlo study finds that the pigeonhole bootstrap is more conservative than the proposed procedure when the number of interacting units is small. In the empirical applications below, it also tends to produce wider confidence intervals and can be numerically unstable. As the number of interacting units grows, the two procedures become asymptotically equivalent under standard regularity conditions.

A second alternative for uncertainty quantification is to derive frequentist standard errors. Graham (2020a,b) builds on earlier work (Fafchamps and Gubert, 2007; Cameron and Miller, 2014; Aronow, Samii, and Assenova, 2015) to develop consistent variance estimators for maximum likelihood estimators. I extend these results to Z-estimators—that is, estimators defined as the solution to a system of estimating equations. As with the pigeonhole bootstrap, the validity of these frequentist standard errors relies on asymptotic approximations that require a large number of interacting units. When the number of interacting units is small, the procedure may become numerically unstable, as illustrated in the Monte Carlo study. Again, when the number of interacting units is large, using analytic standard errors is approximately equivalent to using my approach under standard regularity conditions.

Both Davezies, D'haultfœuille, and Guyonvarch (2021) and Graham (2020a,b) exclusively focus on uncertainty quantification for the estimator of the structural parameter.[3] Since the

[2]To date, among all papers citing the literature on accounting for dyadic dependence, only two papers both contain the phrase "bilateral trade flows" or "bilateral flows" and explicitly account for dyadic dependence: Rosendorf (2023) and Wigton-Jones (2024).

[3]As mentioned above, only 2 out of 36 papers in the survey report uncertainty quantification for their counterfactual prediction. Adao, Costinot, and Donaldson (2017) samples from the asymptotic distribution of the estimator, while Allen, Arkolakis, and Takahashi (2020) samples uniformly over its confidence interval.

counterfactual prediction depends on this estimator as an input, it inherits the challenges associated with dyadic data and a small number of interacting units. As a result, valid uncertainty quantification for counterfactual predictions using these methods also relies on asymptotic approximations. In contrast, my Bayesian bootstrap procedure provides valid finite-sample uncertainty quantification for counterfactual predictions.

This paper contributes to several literatures. First, it contributes a new method for uncertainty quantification to the growing body of work aimed at improving counterfactual analysis in quantitative trade and spatial economics (Kehoe, Pujolas, and Rossbach, 2017; Adao, Costinot, and Donaldson, 2017; Dingel and Tintelnot, 2020; Adão, Costinot, and Donaldson, 2023; Sanders, 2023; Ansari, Donaldson, and Wiles, 2024). Second, by introducing a Bayesian procedure with a finite-sample interpretation in a non-standard setting, it advances research on bootstrap methods designed for situations where standard resampling approaches fail (Janssen, 1994; Davezies, D'haultfœuille, and Guyonvarch, 2021; Menzel, 2021; Zylkin, 2024). Third, it contributes to the emerging literature on uncertainty quantification in polyadic settings by offering a method that is both easy to implement and valid in finite samples (Snijders, Borgatti et al., 1999; Graham, 2020a,b; Menzel, 2021; Davezies, D'haultfœuille, and Guyonvarch, 2021; Graham, 2024). While the Bayesian bootstrap is briefly mentioned in Graham (2020b) in the context of dyadic regression, it has not been further developed or applied in polyadic settings.

The rest of the paper is organized as follows. Section 2 illustrates the method using a canonical gravity model. Section 3 introduces the setting and the proposed Bayesian bootstrap procedure. Sections 4 and 5 present the main theoretical contributions, focusing on finite-sample Bayesian results and asymptotic validity, respectively. Section 6 discusses several extensions of the core framework. Section 7 applies the proposed procedure to the empirical setting studied in Caliendo and Parro (2015). Section 8 compares the proposed procedure to alternative methods for uncertainty quantification. Section 9 concludes.

2 Overview Through Estimation in a Canonical Gravity Model

To fix ideas, before introducing the formal framework, consider the canonical gravity model from Silva and Tenreyro (2006) with n countries. Let C_i collect the *country-level primitives*, such as productivity, endowments, market size, geographic location, linguistic characteristics, historical characteristics, policy variables, and other observed or latent fundamentals. Next, let X_{ij} collect the *observed bilateral variables entering the estimating equation*. In Silva and

Tenreyro (2006), these are the bilateral trade flow F_{ij}, log GDP levels GDP_i and GDP_j, and log distance dist_{ij}. Throughout the paper I assume that the observed bilateral variables entering the estimation equation are measurable functions of the underlying country primitives, so that

$$X_{ij} = \left(F_{ij}, \text{GDP}_i, \text{GDP}_j, \text{dist}_{ij}\right)' = h\left(C_i, C_j\right),$$

for some function h.[4] This assumption implies that the empirical distribution of bilateral observations is determined by the empirical distribution of countries, rather than being treated as primitive.

Silva and Tenreyro (2006) uses the observed bilateral data to run a Poisson Pseudo-Maximum Likelihood (PPML) regression of bilateral trade flows on a constant, the exporter's log GDP, the importer's log GDP and the log distance. The corresponding estimator $\hat{\theta}$ solves the sample moment condition

$$\sum_{k\neq\ell} \frac{1}{n\left(n-1\right)} \left(F_{k\ell} - \exp\left\{\begin{pmatrix} 1 & \text{GDP}_k & \text{GDP}_\ell & \text{dist}_{k\ell} \end{pmatrix} \hat{\theta}\right\}\right) \begin{pmatrix} 1 \\ \text{GDP}_k \\ \text{GDP}_\ell \\ \text{dist}_{k\ell} \end{pmatrix} = 0.$$

In this paper I quantify uncertainty by placing a *prior on the distribution from which the country-level primitives C_i are drawn*. This implies a bootstrap procedure that assigns exponentially distributed random weights $V_1, ..., V_n$ to countries. Since the bilateral observation X_{ij} is generated jointly by the pair (C_i, C_j), the induced weight on dyad (i, j) is proportional to the product $V_i V_j$. This implies that the b-th bootstrap draw $\hat{\theta}^{*,(b)}$ solves

$$\sum_{k\neq\ell} \frac{V_k^{(b)} \cdot V_\ell^{(b)}}{\sum_{s\neq t} V_s^{(b)} \cdot V_t^{(b)}} \left(F_{k\ell} - \exp\left\{\begin{pmatrix} 1 & \text{GDP}_k & \text{GDP}_\ell & \text{dist}_{k\ell} \end{pmatrix} \hat{\theta}^{*,(b)}\right\}\right) \begin{pmatrix} 1 \\ \text{GDP}_k \\ \text{GDP}_\ell \\ \text{dist}_{k\ell} \end{pmatrix} = 0,$$

where $\left(V_1^{(b)}, ..., V_n^{(b)}\right) \overset{\text{iid}}{\sim} \text{Exp}\left(1\right)$. Thus, the reweighting operates at the level of country primitives, not at the level of individual bilateral observations. If country i receives a larger weight, all dyads involving i receive larger weight as well. This reflects the fact that bilateral observations inherit their dependence structure from the shared country primitives from

[4] Allowing irreducible dyad-specific heterogeneity, so that $X_{ij} = \tilde{h}\left(C_i, C_j, D_{ij}\right)$ requires additional modeling. I will come back to this in Section 4.1.

which they are generated.

3 Setting and Proposed Procedure

In this section I introduce the setting and goal of the paper. I lay out my proposed procedure and illustrate it using my running example. I consider misspecification-robust uncertainty quantification for over-identified GMM as a special case. Theoretical justifications are deferred to Sections 4 and 5.

3.1 Setting and Goal

3.1.1 Data Environment

We observe a sample of bilateral data $\{X_{k\ell}\}_{k\neq\ell} \in \mathcal{X}^{n(n-1)}$ with $\mathcal{X} \subseteq \mathbb{R}^{d_X}$, with $n \in \mathbb{N}$ the *number of interacting units*. Since we consider bilateral data, the effective *sample size* is $n\,(n-1)$.

Example (Waugh, 2010)**.** In Waugh (2010), the interacting units are 43 countries so $n = 43$. The data are

$$X_{k\ell} = (\lambda_{k\ell}, \lambda_{kk}, \tau_{k\ell}, p_k, p_\ell) \in \mathcal{X} = [0,1]^2 \times (1,\infty] \times \mathbb{R}_+^2,$$

for $k \neq \ell$.[5] Here, $\lambda_{k\ell}$ denotes country ℓ's expenditure share on goods from country k, $\tau_{k\ell}$ denotes estimated iceberg trade costs from country k to country ℓ, and p_k denotes the aggregate price of goods in country k. $\triangle$

3.1.2 Structural Estimator and Estimand

Denote the empirical distribution of the data by

$$\mathbb{P}_{n,X_{ij}} = \sum_{k\neq\ell} \frac{1}{n\,(n-1)} \cdot \delta_{X_{k\ell}}, \tag{1}$$

for δ_x the Dirac measure at x. The Dirac measure at a single observation $X_{k\ell}$ corresponds to a degenerate probability distribution which puts a mass of 1 at that observation. The empirical distribution hence assigns mass $\frac{1}{n(n-1)}$ to each observation $X_{k\ell}$.

[5] The sample size in Waugh (2010) is not actually $43 \cdot 42 = 1806$ but 1373, because observations with $\lambda_{k\ell} = 0$ are dropped. I will come back to this in Section 3.2.1.

The researcher aims to estimate a structural parameter using the observed data $\{X_{k\ell}\}_{k\neq\ell}$. I assume the estimator $\hat{\theta}$ is a function of the empirical distribution. For ease of exposition I assume $\hat{\theta}$ is a scalar, but the same arguments apply to vector-valued $\hat{\theta}$.

Assumption 1 (Structural estimator)**.** *We have*

$$\hat{\theta} = T\left(\mathbb{P}_{n,X_{ij}}\right), \tag{2}$$

for a known function $T : \Delta(\mathcal{X}) \to \Theta \subseteq \mathbb{R}$.

Here, $\Delta(\mathcal{X})$ denotes the set of all probability distributions over $\mathcal{X}$. Assumption 1 covers many common estimators such as averages, regression estimators and generalized method of moments (GMM) estimators:

$$\begin{aligned}
T_{\text{AVG}}\left(\mathbb{P}_{n,X_{ij}}\right) &= \mathbb{E}_{\mathbb{P}_{n,X_{ij}}}\left[X_{ij}\right] = \sum_{k\neq\ell} \frac{1}{n(n-1)} \cdot X_{k\ell} \\
T_{\text{OLS}}\left(\mathbb{P}_{n,(F_{ij},R_{ij})}\right) &= \arg\min_{\vartheta\in\Theta} \mathbb{E}_{\mathbb{P}_{n,\left(F_{ij},R_{ij}\right)}}\left[\left(F_{ij} - \vartheta R_{ij}\right)^2\right] = \frac{\sum_{k\neq\ell} F_{k\ell}R_{k\ell}}{\sum_{k\neq\ell} R_{k\ell}^2} \\
T_{\text{GMM}}\left(\mathbb{P}_{n,X_{ij}}\right) &= \arg\min_{\vartheta\in\Theta} \mathbb{E}_{\mathbb{P}_{n,X_{ij}}}\left[\psi\left(X_{ij};\vartheta\right)\right]' \Omega \mathbb{E}_{\mathbb{P}_{n,X_{ij}}}\left[\psi\left(X_{ij};\vartheta\right)\right].
\end{aligned}$$

Estimators that are not covered by Assumption 1 are estimators that involve multiple flows, such as the network moments discussed in Graham (2020b).[6]

Example (Waugh, 2010)**.** The relevant empirical distribution in Waugh (2010) is

$$\mathbb{P}_{n,X_{ij}} = \sum_{k\neq\ell} \frac{1}{n(n-1)} \cdot \delta_{(\lambda_{k\ell},\lambda_{kk},\tau_{k\ell},p_k,p_\ell)}.$$

The author aims to estimate a productivity parameter which governs the dispersion of efficiency levels across countries. An arbitrage condition motivates the simple linear regression

[6] Assumption 1 can be extended to accommodate weighted empirical distributions $\mathbb{P}^{\omega}_{n,X_{ij}} = \sum_{k\neq\ell} \omega_{k\ell} \cdot \delta_{X_{k\ell}}$. This may be desirable, for example, when assigning lower weight to trade flows between small and distant economies. In such cases, the weight can be absorbed into the definition of the observation. Specifically, make the restriction that $\hat{\theta} = \varphi\left(\mathbb{E}_{\mathbb{P}^{\omega}_{n,X_{ij}}}\left[f\left(X_{ij}\right)\right]\right)$, for some function f. Then we can write $\hat{\theta} = \varphi\left(\mathbb{E}_{\mathbb{P}_{n,Y_{ij}}}\left[Y_{ij}\right]\right)$ for $Y_{ij} = n(n-1)\,\omega_{ij} f\left(X_{ij}\right)$.

using $\left\{\log\left(\frac{\lambda_{k\ell}}{\lambda_{kk}}\right)\right\}_{k\neq\ell}$ and $\left\{\log\left(\tau_{k\ell}\frac{p_k}{p_\ell}\right)\right\}_{k\neq\ell}$:

$$\begin{aligned}\hat{\theta} &= -T_{\text{Waugh}}\left(\mathbb{P}_{n,X_{ij}}\right) = -\arg\min_{\vartheta\in\Theta}\ \mathbb{E}_{\mathbb{P}_{n,X_{ij}}}\left[\left(\log\left(\frac{\lambda_{ij}}{\lambda_{ii}}\right) - \vartheta\log\left(\tau_{ij}\frac{p_i}{p_j}\right)\right)^2\right] \\ &= -\frac{\sum_{k\neq\ell}\log\left(\frac{\lambda_{k\ell}}{\lambda_{kk}}\right)\log\left(\tau_{k\ell}\frac{p_k}{p_\ell}\right)}{\sum_{k\neq\ell}\left(\log\left(\tau_{k\ell}\frac{p_k}{p_\ell}\right)\right)^2}. \end{aligned} \tag{3}$$

The estimator $\hat{\theta}$ satisfies Assumption 1. $\triangle$

Note that estimators that can be written as in Equation (2) are *permutation invariant* with respect to the observed data $\{X_{k\ell}\}_{k\neq\ell}$, because for any permutation $\sigma : \{1,...,n\} \to \{1,...,n\}$, we have

$$\hat{\theta} = T\left(\sum_{k\neq\ell}\frac{1}{n\left(n-1\right)}\cdot\delta_{X_{k\ell}}\right) = T\left(\sum_{k\neq\ell}\frac{1}{n\left(n-1\right)}\cdot\delta_{X_{\sigma(k)\sigma(\ell)}}\right).$$

For the purposes of structural estimation, it is then without loss to assume that the observed data are *jointly exchangeable*, which means that the joint distribution does not change when we relabel the indices, so that

$$\{X_{k\ell}\}_{k\neq\ell} \overset{d}{=} \left\{X_{\sigma(k)\sigma(\ell)}\right\}_{k\neq\ell},$$

for any permutation $\sigma : \{1,...,n\} \to \{1,...,n\}$.[7] Joint exchangeability of the data implies that the elements of $\{X_{k\ell}\}_{k\neq\ell}$ have a common *marginal* probability distribution, which I will denote by $\mathbb{P}_{X_{ij}}$. The *structural estimand of interest* then is

$$\theta \equiv T\left(\mathbb{P}_{X_{ij}}\right). \tag{4}$$

Note that while the functional T is not n-specific, the estimand θ can depend on n.[8] More-

[7] In other papers concerning dyadic dependence (Graham, 2020a,b; Davezies, D'haultfœuille, and Guyonvarch, 2021), joint exchangeability is used as a primitive assumption. I instead motivate it by focusing on the class of estimators that satisfy Assumption 1. Relatedly, one could relax Assumption 1 by instead assuming that the estimator is symmetric under relabeling of the indices. However, this complicates subsequent steps considerably, and estimators that satisfy such a symmetry property but are not functions of the empirical distribution come up rarely in quantitative trade and spatial models.

[8] To see this, suppose we know the joint distribution of $\{X_{k\ell}\}_{k\neq\ell}$, denoted by $\mathbb{P}^n$, with marginals $\left(\mathbb{P}^n_{X_{12}},...,\mathbb{P}^n_{X_{n(n-1)}}\right)$ and no restriction on the copula. This joint distribution may vary with n; for example, the distribution of observations from an economy with three countries may differ from that of an

over, the estimand might differ from the structural parameter of interest if the model is misspecified, as illustrated in the next example. In such cases, my results deliver valid inference for the estimand θ.

Example (Waugh, 2010)**.** Given that Waugh (2010) considers a simple regression, for the purposes of estimation, it is without loss to assume that the observed data $\{X_{k\ell}\}_{k\neq\ell}$ are jointly exchangeable. Here, joint exchangeability means that the joint distribution of bilateral data remains unchanged if we relabel the countries. Joint exchangeability implies that there exists some marginal distribution $\mathbb{P}_{X_{ij}}$ from which all the observations are drawn. The structural estimand of interest then equals the negative of the function T_{Waugh} applied to $\mathbb{P}_{X_{ij}}$, so that

$$\theta \equiv -T_{\text{Waugh}}\left(\mathbb{P}_{X_{ij}}\right) = -\arg\min_{\vartheta\in\Theta} \mathbb{E}_{\mathbb{P}_{X_{ij}}}\left[\left(\log\left(\frac{\lambda_{ij}}{\lambda_{ii}}\right) - \vartheta \log\left(\tau_{ij}\frac{p_i}{p_j}\right)\right)^2\right].$$

This estimand corresponds to the coefficient in the regression

$$\log\left(\frac{\lambda_{ij}}{\lambda_{ii}}\right) = -\theta \log\left(\tau_{ij}\frac{p_i}{p_j}\right) + \varepsilon_{ij}, \tag{5}$$

for an orthogonal error term ε_{ij}. The regression coefficient in Equation (3) was motivated by the model equation

$$\log\left(\frac{\lambda_{ij}}{\lambda_{ii}}\right) = -\theta_M \log\left(\tau_{ij}\frac{p_i}{p_j}\right),$$

which does not hold exactly in-sample because of country-level productivity shocks. The regression equation recovers the true structural parameter θ_M under the assumption that the productivity shocks are exogenous and follow a Fréchet distribution. Nevertheless, since Waugh (2010) conducts estimation based on Equation (3), I focus going forward on the resulting estimand θ rather than the structural parameter θ_M.$\triangle$

3.1.3 Counterfactual Predictions

In quantitative trade and spatial models, researchers are interested in forming counterfactual predictions. Since these predictions are relative to some observed factual situation, they are functions of the realized bilateral data $\{X_{k\ell}\}_{k\neq\ell}$ and the structural estimator $\hat{\theta}$:

economy with four. In such cases, $\mathbb{P}_{X_{ij}}$ denotes the marginal of a randomly selected observation, and thus both $\mathbb{P}_{X_{ij}}$ and θ generally depend on n.

Assumption 2 (Counterfactual prediction)**.** *The reported counterfactual prediction of interest can be written as*

$$\hat{\gamma} = g\left(\{X_{k\ell}\}_{k\neq\ell}, \hat{\theta}\right), \tag{6}$$

for a known function $g : \mathcal{X}^{n(n-1)} \times \Theta \to \mathbb{R}$*.*

The corresponding estimand is

$$\gamma \equiv g\left(\{X_{k\ell}\}_{k\neq\ell}, \theta\right) \equiv g\left(\{X_{k\ell}\}_{k\neq\ell}, T\left(\mathbb{P}_{X_{ij}}\right)\right).$$

In conventional economic models the estimand of interest is typically defined as a function of the population distribution of the data, not the specific realized observations. In contrast, γ depends both on the realized bilateral data and on a population distribution, which creates a non-classical setting (Sanders, 2023). Note that even when the estimand θ does not depend on n, the estimand γ will generally vary with n, as it depends on the realized data.[9]

Example (Waugh, 2010)**.** After obtaining the estimator $\hat{\theta}$, we can use the model in Waugh (2010) to find the equilibrium wage vector for a given counterfactual trade cost schedule. The relevant counterfactual mapping is

$$\{X_{k\ell}\}_{k\neq\ell}, \hat{\theta}, \left\{\tau_{k\ell}^{\text{cf}}\right\} \mapsto \left\{\hat{w}_k^{\text{cf}}\right\}. \tag{7}$$

It maps the realized data, the structural estimator and a counterfactual trade cost schedule to a vector which contains the counterfactual wage for all 43 countries. Appendix A.1 outlines the details of this mapping. From the counterfactual wage vector we can compute various scalar objects of interest, such as the wage level of a specific country or a summary statistic across countries. Each such object corresponds to a particular function g, as described in Assumption 2.

Waugh (2010) considers a series of counterfactuals that calculate inequality statistics of the equilibrium wage vector for different trade cost schedules. The inequality statistics are the variance of log wages, the ratio of the 90th and 10th percentile of wages, and the mean percentage change in wages. The different counterfactual trade cost schedules are autarky ($\tau_{ij}^{\text{cf}} = \infty$ for all $i \neq j$), symmetry ($\tau_{ij}^{\text{cf}} = \min\{\tau_{ij}, \tau_{ji}\}$ for all $i \neq j$) and free trade ($\tau_{ij}^{\text{cf}} = 1$ for

[9]Furthermore, Assumption 2 encompasses "invertible" quantitative trade and spatial models (Redding and Rossi-Hansberg, 2017), in which a subset of structural parameters—often referred to as "model fundamentals"—are first backed out from observed data. These fundamentals are then held fixed in any counterfactual exercise that follows, so they do not influence Bayesian uncertainty quantification.

all $i \neq j$). Using the equilibrium mapping in Equation (7), it follows that each counterfactual prediction can be written as in Equation (6). The resulting point estimates are reported in Table 4 of Waugh (2010) without any uncertainty quantification. $\triangle$

The discussion in the previous two sections highlights two distinct statistical objects of interest: the structural estimator $\hat{\theta}$ and the counterfactual prediction $\hat{\gamma}$. To quantify uncertainty for each, I proceed in two steps. First, in Section 3.2, I present a Bayesian bootstrap procedure to quantify uncertainty for $\hat{\theta}$. Then, in Section 3.3, I use Assumption 2 to quantify uncertainty for $\hat{\gamma}$.

3.2 Bayesian Uncertainty Quantification for the Structural Parameter

To quantify uncertainty for $\hat{\theta}$, I consider a bootstrap procedure. Specifically, in each bootstrap iteration $b = 1, ..., B$, $\hat{\theta}^{*,(b)}$ is computed by replacing the empirical distribution in Equation (1) by a weighted version of this empirical distribution,

$$\mathbb{P}^{*,(b)}_{n,X_{ij}} = \sum_{k \neq \ell} \omega^{(b)}_{k\ell} \cdot \delta_{X_{k\ell}}.$$

The weights $\left\{\omega^{(b)}_{k\ell}\right\}_{k\neq\ell}$ are computed using draws from a Dirichlet distribution,

$$\omega^{(b)}_{k\ell} = \frac{W^{(b)}_k \cdot W^{(b)}_\ell}{\sum_{s\neq t} W^{(b)}_s \cdot W^{(b)}_t}, \quad \left(W^{(b)}_1, ..., W^{(b)}_n\right) \sim \text{Dir}\left(n; 1, ..., 1\right). \tag{8}$$

In practice, it is convenient that the Dirichlet distribution $\text{Dir}\left(n; 1, ..., 1\right)$ can be constructed from i.i.d. draws from an exponential distribution:

$$\begin{aligned}
\left(V^{(b)}_1, ..., V^{(b)}_n\right) &\overset{\text{iid}}{\sim} \text{Exp}\left(1\right) \\
W^{(b)}_k &= \frac{V^{(b)}_k}{\sum_{s=1}^n V^{(b)}_s}, \quad k = 1, ..., n \\
\omega^{(b)}_{k\ell} &= \frac{V^{(b)}_k \cdot V^{(b)}_\ell}{\sum_{s\neq t} V^{(b)}_s \cdot V^{(b)}_t}, \quad k, \ell = 1, ..., n.
\end{aligned}$$

The procedure to quantify uncertainty for the estimator $\hat{\theta}$ is summarized in Algorithm 1.

This procedure is a natural generalization of the univariate Bayesian bootstrap (Rubin, 1981; Chamberlain and Imbens, 2003). It is intuitive and easy to implement, as it requires

Algorithm 1 Bayesian bootstrap procedure

1. Input: Bilateral data $\{X_{k\ell}\}_{k\neq\ell}$ and estimator function $T : \Delta(\mathcal{X}) \to \Theta$.
2. For each bootstrap draw $b = 1, ..., B$:
   (a) Sample $\left(V_1^{(b)}, ..., V_n^{(b)}\right) \overset{\text{iid}}{\sim} \text{Exp}(1)$.
   (b) Compute
   $$\hat{\theta}^{*,(b)} = T\left(\sum_{k\neq\ell} \frac{V_k^{(b)} \cdot V_\ell^{(b)}}{\sum_{s\neq t} V_s^{(b)} \cdot V_t^{(b)}} \cdot \delta_{X_{k\ell}}\right).$$
3. Report the quantiles of interest of $\left\{\hat{\theta}^{*,(1)}, ..., \hat{\theta}^{*,(B)}\right\}$.

only reweighting the data by products of standard exponential draws.

In Sections 4 and 5 I will provide various theoretical motivations for the Bayesian bootstrap procedure. The key takeaway from Section 4 is that Algorithm 1 produces draws from a limiting posterior for θ given the bilateral data $\{X_{k\ell}\}_{k\neq\ell}$ for a well-motivated model and prior. In addition to this finite-sample Bayesian motivation, the key takeaway from Section 5 is that the bootstrap procedure is also asymptotically valid in a frequentist sense.

Example (Waugh, 2010)**.** Using Algorithm 1, we can obtain draws from the posterior distribution of the the productivity parameter in Waugh (2010) given the bilateral data. Specifically, for each bootstrap iteration b, I compute a weighted regression coefficient:

$$\begin{aligned}\hat{\theta}^{*,(b)} &= -T_{\text{Waugh}}\left(\sum_{k\neq\ell} \frac{V_k^{(b)} \cdot V_\ell^{(b)}}{\sum_{s\neq t} V_s^{(b)} \cdot V_t^{(b)}} \cdot \delta_{X_{k\ell}}\right) \\ &= -\arg\min_{\vartheta\in\Theta} \sum_{k\neq\ell} V_k^{(b)} \cdot V_\ell^{(b)} \cdot \left(\log\left(\frac{\lambda_{k\ell}}{\lambda_{kk}}\right) - \vartheta \log\left(\tau_{k\ell}\frac{p_k}{p_\ell}\right)\right)^2.\end{aligned}$$

Because the bootstrap distribution corresponds to a limiting posterior distribution, we can interpret $\left\{\hat{\theta}^{*,(1)}, ..., \hat{\theta}^{*,(B)}\right\}$ as posterior draws. A $100(1-\alpha)\%$ *credible interval* can then be constructed by taking the empirical $\alpha/2$ and $1-\alpha/2$ quantiles of these draws. The 95% credible interval is reported in Table 1 and the posterior distribution is plotted in Figure 1.

In Waugh (2010), no uncertainty quantification is discussed for $\hat{\theta}$. In the accompanying code, the author computes the dyad-level heteroskedastic-robust standard error $\sqrt{\hat{\Sigma}_\theta}$. In Table 1 I add the corresponding 95% confidence intervals, computed using the familiar $\left[\hat{\theta} \pm 1.96 \cdot \sqrt{\hat{\Sigma}_\theta}\right]$. In Figure 1, I also plot a normal distribution with mean $\hat{\theta}$ and standard

error $\sqrt{\hat{\Sigma}_\theta}$, since the standard confidence intervals rely on $\hat{\theta}$ to be approximately normal centered at θ with variance $\hat{\Sigma}_\theta$. We observe that the posterior is approximately normal but has larger variance than reported in the accompanying code of the paper, which suggests that considering dyadic dependence is important.

	Point estimate	95% confidence interval based on Waugh (2010)	95% Bayesian bootstrap credible interval
Productivity parameter, $n = 43$	5.55	[5.39, 5.71]	[5.12, 6.02]

Table 1: Uncertainty quantification for productivity parameter in Waugh (2010).

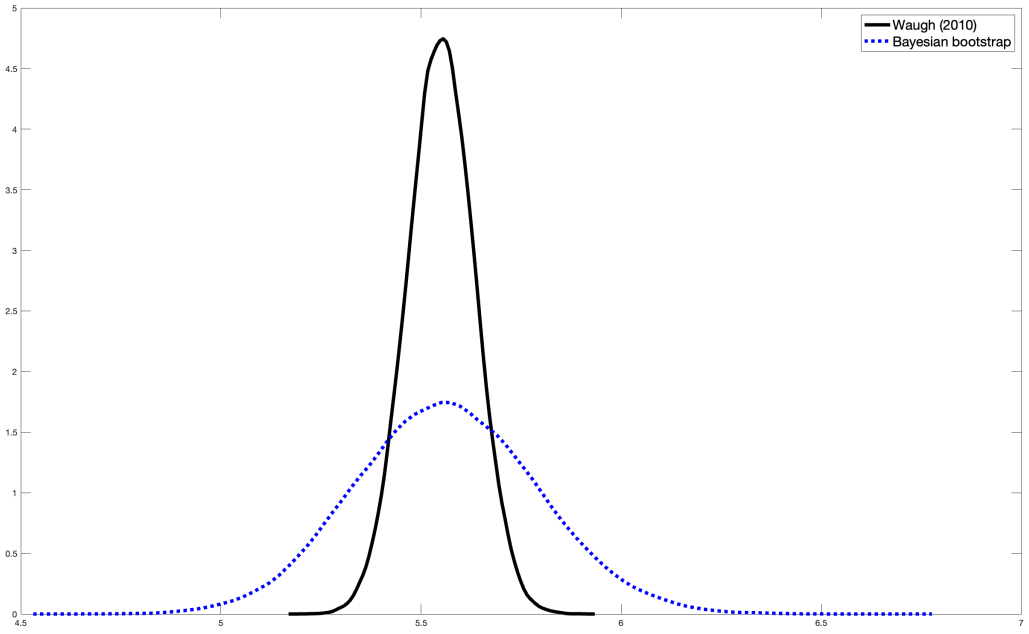

Figure 1: Distributions for productivity parameter in Waugh (2010). "Waugh (2010)" corresponds to the normal approximation as implied by the standard error reported in Waugh (2010), and "Bayesian bootstrap" corresponds to the smoothed Bayesian bootstrap distribution.

Table 1 shows thats the confidence intervals and credible intervals differ substantially. To better understand this discrepancy, Appendix B presents a data-calibrated simulation exercise based on the pigeonhole bootstrap—a method introduced in Section 8.1. This setup enables a direct evaluation of the coverage performance of various uncertainty quantification

methods. Table 2 shows that confidence intervals based on heteroskedastic-robust standard errors have below-nominal coverage. $\triangle$

	Based on Waugh (2010)	Bayesian bootstrap
Productivity parameter, $n = 43$	0.498	0.979

Table 2: Coverage for the approach used in Waugh (2010) and the Bayesian bootstrap using the pigeonhole bootstrap DGP as described in Appendix B.

3.2.1 Special Case: Misspecification-Robust Uncertainty Quantification for GMM

Often, researchers are interested in over-identified GMM estimators of the form

$$\hat{\theta} = \underset{\vartheta \in \Theta}{\arg\min}\ \mathbb{E}_{\mathbb{P}_{n,X_{ij}}}\left[\psi\left(X_{ij};\vartheta\right)\right]' \hat{\Omega} \mathbb{E}_{\mathbb{P}_{n,X_{ij}}}\left[\psi\left(X_{ij};\vartheta\right)\right],$$

where $\psi : \mathcal{X} \to \mathbb{R}^L$ with $L \geq 1$, $X_{ij} \in \mathcal{X}$, $\theta \in \Theta \subseteq \mathbb{R}$ and $\hat{\Omega}$ is an estimated weight matrix. For example, the PPML estimator in Silva and Tenreyro (2006) corresponds to the moment function

$$\psi\left(F_{ij}, R_{ij}; \vartheta\right) = \left(F_{ij} - \exp\left\{R_{ij}\vartheta\right\}\right) R_{ij}, \tag{9}$$

where $F_{ij} \in \mathbb{R}_+$ is the dependent variable and $R_{ij} \in \mathbb{R}$ a regressor.

We know that the optimal weight matrix is the inverse of the variance-covariance matrix of the moments at θ (Hansen, 1982; Chamberlain, 1987). In practice, since we require an estimate of this optimal weight matrix, researchers often use a two-step procedure. In the first step the identity matrix is used as a weight matrix:

$$\psi_n\left(\vartheta\right) = \mathbb{E}_{\mathbb{P}_{n,X_{ij}}}\left[\psi\left(X_{ij};\vartheta\right)\right] \tag{10}$$

$$\hat{\theta}^{1-\mathrm{GMM}} = \underset{\vartheta \in \Theta}{\arg\min}\ \psi_n\left(\vartheta\right)' \psi_n\left(\vartheta\right). \tag{11}$$

The resulting estimator is plugged in to find an estimator of the optimal weight matrix,

which is then used to find the two-step GMM estimator:[10]

$$\hat{\Omega}(\vartheta) = \left(\mathbb{E}_{\mathbb{P}_{n,X_{ij}}}\left[\{\psi(X_{ij};\vartheta) - \psi_n(\vartheta)\}\{\psi(X_{ij};\vartheta) - \psi_n(\vartheta)\}'\right]\right)^{-1} \tag{12}$$

$$\hat{\theta}^{2-\text{GMM}} = \underset{\vartheta\in\Theta}{\arg\min}\ \psi_n(\vartheta)'\,\hat{\Omega}\left(\hat{\theta}^{1-\text{GMM}}\right)\psi_n(\vartheta). \tag{13}$$

The two-step estimator satisfies Assumption 1, which implies that for the purposes of estimation it is without loss to assume that the elements of $\{X_{k\ell}\}_{k\neq\ell}$ have a common marginal distribution $\mathbb{P}_{X_{ij}}$. The relevant moment conditions then are

$$\mathbb{E}_{\mathbb{P}_{X_{ij}}}[\psi(X_{ij};\theta)] = 0. \tag{14}$$

These moments might be misspecified, meaning that there exists no $\theta \in \Theta$ such that the moment equations in Equation (14) hold. In this case, we might still be interested in doing uncertainty quantification for the probability limit of the two-step GMM estimator in Equation (13)—the pseudo-true parameter. However, valid uncertainty quantification using the conventional GMM standard errors hinges on the moments being well-specified (Hall and Inoue, 2003; Lee, 2014).

The Bayesian bootstrap procedure from Algorithm 1 is robust to misspecification of the two-step GMM estimator. This means that it yields valid uncertainty quantification in both the finite-sample Bayesian and asymptotic frequentist senses. I will make the claim of valid asymptotic frequentist uncertainty quantification precise in Section 5.1.3. The resulting Bayesian bootstrap procedure is summarized in Algorithm 2. Effectively, each empirical distribution $\mathbb{P}_{n,X_{ij}}$ in Equations (10)-(13) is replaced by its weighted analog.

Example (Waugh, 2010)**.** The estimator in Equation (3) has corresponding moment function

$$\psi_{\text{Waugh}}(X_{ij};\vartheta) = \left(\log\left(\frac{\lambda_{ij}}{\lambda_{ii}}\right) - (-\vartheta)\log\left(\tau_{ij}\frac{p_i}{p_j}\right)\right)\log\left(\tau_{ij}\frac{p_i}{p_j}\right). \tag{15}$$

In Waugh (2010), whenever $\lambda_{k\ell} = 0$ for countries k and ℓ, the corresponding observation $X_{k\ell}$ is omitted. This results in removing 433 out of the possible $43 \cdot 42 = 1806$ bilateral observations. To avoid removing these observations, one could adapt the simple OLS estimator to

[10] I follow Lee (2014) and use the centered weight matrix, rather than the uncentered version $\hat{\Omega}^{\text{uncentered}}(\vartheta) = \left(\mathbb{E}_{\mathbb{P}_{n,X_{ij}}}\left[\psi(X_{k\ell};\vartheta)\,\psi(X_{k\ell};\vartheta)'\right]\right)^{-1}$. The choice of weight matrix affects the resulting pseudo-true value. As outlined in Hall (2000), the uncentered version includes bias terms of the moment function, which makes the centered version better behaved under misspecification.

Algorithm 2 Bayesian bootstrap procedure for GMM

1. Input: Bilateral data $\{X_{k\ell}\}_{k\neq\ell}$, moment equations $\psi : \mathcal{X} \to \Theta$.
2. For each bootstrap draw $b = 1, ..., B$:
   (a) Sample $\left(V_1^{(b)}, ..., V_n^{(b)}\right) \overset{\text{iid}}{\sim} \text{Exp}(1)$.
   (b) Construct $\omega_{k\ell}^{(b)} = V_k^{(b)} \cdot V_\ell^{(b)} / \left(\sum_{s\neq t} V_s^{(b)} \cdot V_t^{(b)}\right)$, for $k, \ell = 1, ..., n$.
   (c) Solve for $\hat{\theta}^{*,2-\text{GMM},(b)}$ from

$$\begin{aligned}
\psi_n^{(b)}(\vartheta) &= \sum_{k\neq\ell} \omega_{k\ell}^{(b)} \cdot \psi(X_{k\ell}; \vartheta) \\
\hat{\theta}^{*,1-\text{GMM},(b)} &= \arg\min_{\vartheta\in\Theta} \psi_n^{(b)}(\vartheta)' \psi_n^{(b)}(\vartheta) \\
\hat{\Omega}^{(b)}(\vartheta) &= \left(\sum_{k\neq\ell} \omega_{k\ell}^{(b)} \cdot \left\{\psi(X_{k\ell}; \vartheta) - \psi_n^{(b)}(\vartheta)\right\} \left\{\psi(X_{k\ell}; \vartheta) - \psi_n^{(b)}(\vartheta)\right\}'\right)^{-1} \\
\hat{\theta}^{*,2-\text{GMM},(b)} &= \arg\min_{\vartheta\in\Theta} \psi_n^{(b)}(\vartheta)' \hat{\Omega}^{(b)}\left(\hat{\theta}^{*,1-\text{GMM},(b)}\right) \psi_n^{(b)}(\vartheta).
\end{aligned}$$

3. Report the quantiles of interest of $\left\{\hat{\theta}^{*,2-\text{GMM},(1)}, ..., \hat{\theta}^{*,2-\text{GMM},(B)}\right\}$.

a PPML estimator as in Equation (9), with corresponding sample moment condition

$$\psi_{\text{Waugh,PPML}}(X_{ij}; \vartheta) = \left(\frac{\lambda_{ij}}{\lambda_{ii}} - \exp\left\{-\vartheta \log\left(\tau_{ij}\frac{p_i}{p_j}\right)\right\}\right) \log\left(\tau_{ij}\frac{p_i}{p_j}\right). \tag{16}$$

In Appendix A.3 I compute the point estimates and posterior distributions while not omitting zeros and using PPML. The point estimates drop considerably and there is more uncertainty. △

3.3 Bayesian Uncertainty Quantification for the Counterfactual

Taking a Bayesian perspective on uncertainty quantification, in Section 4 I show that for a specific choice of model and prior, the posterior for θ given the bilateral data $\{X_{k\ell}\}_{k\neq\ell}$ converges to the Bayesian bootstrap distribution as a certain informativeness parameter is taken to zero.

Since we are also interested in uncertainty quantification for the counterfactual prediction, we aim to find the corresponding limiting posterior for γ given the realized bilateral data

$\{X_{k\ell}\}_{k\neq\ell}$. Towards this end, note that, conditional on the realized data $\{X_{k\ell}\}_{k\neq\ell}$, the only randomness is coming from the posterior for the structural parameter. So having obtained draws $\left\{\hat{\theta}^{*,(1)},...,\hat{\theta}^{*,(B)}\right\}$ from the limiting posterior distribution for θ given the bilateral data $\{X_{k\ell}\}_{k\neq\ell}$ using the Bayesian bootstrap procedure, we can use Assumption 2 to obtain draws from the limiting posterior distribution for γ given the bilateral data $\{X_{k\ell}\}_{k\neq\ell}$,

$$\hat{\gamma}^{*,(b)} = g\left(\{X_{k\ell}\}_{k\neq\ell}, \hat{\theta}^{*,(b)}\right), \quad b = 1, ..., B. \tag{17}$$

To construct Bayesian credible intervals, we can then report the relevant quantiles of the draws $\left\{\hat{\gamma}^{*,(1)},...,\hat{\gamma}^{*,(B)}\right\}$.

Example (Waugh, 2010)**.** In Table 3, I reproduce Table 4 of Waugh (2010), but include 95% Bayesian credible intervals. The resulting intervals are small, implying there is not much economically meaningful uncertainty in the counterfactuals. $\triangle$

Scenario	Baseline	Autarky	Symmetry	Free trade
τ_{ij}^{cf}	τ_{ij}	$\infty \cdot \mathbb{I}\{i \neq j\}$	$\min\{\tau_{ij}, \tau_{ji}\}$	1
Variance of log wages	1.30 [1.28, 1.32]	1.35 [1.31, 1.38]	1.05 [1.05, 1.05]	0.76 [0.75, 0.78]
90th/10th percentile of wages	25.7 [25.1, 26.2]	23.5 [22.6, 24.2]	17.3 [17.2, 17.4]	11.4 [11.0, 11.9]
Mean % change in wages	-	-10.5 [-11.4, -9.6]	24.2 [22.4, 25.8]	128.0 [114.4, 140.7]

Table 3: Bayesian uncertainty quantification for counterfactual predictions as in Table 4 of Waugh (2010). The numbers in brackets correspond to 95% Bayesian bootstrap credible intervals.

Remark. Accompanying the paper, I provide an easy-to-use toolkit.[11] It implements Algorithm 2, and takes as inputs a dataset and a GMM moment function, and outputs bootstrap draws $\left\{\hat{\theta}^{*,2-\text{GMM},(b)}\right\}$. It also includes a vignette, which illustrates the procedure by performing uncertainty quantification for the gains from trade estimates for the multi-sector model with perfect competition in Costinot and Rodríguez-Clare (2014). These counterfactual predictions take as inputs the sector-level trade elasticities from Caliendo and Parro (2015) that I will discuss in Section 7.1.

[11]The toolkit is written in MATLAB and can be found on my website, https://sandersbas.github.io/.

4 Theory for Finite-Sample Bayesian Interpretation

In this section I formally introduce and motivate the model and prior. I then present the key result of the paper: the bootstrap procedure in Algorithm 1 admits a finite-sample Bayesian interpretation.

4.1 Model

We observe a sample of bilateral data $\{X_{k\ell}\}_{k\neq\ell} \in \mathcal{X}^{n(n-1)}$, with $\mathcal{X} \subseteq \mathbb{R}^{d_X}$. I adopt a Bayesian approach, which requires specifying both a model and a prior.

The model is motivated by the view that dyadic data are generated as functions of latent unit-specific characteristics. The units are sampled independently from a common population, and bilateral observations are determined by the characteristics of the two sampled units.

Assumption 3 (Model)**.** *The data $\{X_{k\ell}\}_{k\neq\ell}$ are generated according to*

$$C_1, ..., C_n | h, \mathbb{P}_C \overset{\text{iid}}{\sim} \mathbb{P}_C \tag{18}$$

$$X_{ij} = h\left(C_i, C_j\right), \quad \text{for } C_i \neq C_j, \tag{19}$$

where the latent variables $\{C_k\}$ are continuous and take values in $\mathcal{C} \subseteq \mathbb{R}^{d_C}$ for finite d_C and $h : \mathcal{C}^2 \to \mathcal{X}$ is some measurable function.

The parameters of the model for which I will later specify prior distributions are the function h, which maps latent characteristics into observables, and the distribution $\mathbb{P}_C$, from which those latent characteristics are drawn.[12]

Example (Waugh, 2010)**.** Recall that in Waugh (2010) we had $X_{k\ell} = (\lambda_{k\ell}, \lambda_{kk}, \tau_{k\ell}, p_k, p_\ell)$. In this case, C_i is a vector of country-specific primitives (also called "exogenous variables"

[12]Section 6 discusses how the framework can be adapted to accommodate richer dependence structures. The guiding principle is to modify the model in Assumption 3 so that it reflects the exchangeability assumptions appropriate for the application. As an illustration, Boehm, Levchenko, and Pandalai-Nayar (2023) estimates trade elasticities using a panel of bilateral trade flows indexed by exporter i, importer j, product p, year t, and horizon s. One possible model is

$$X_{ijpt,s} = h_{pt,s}\left(C_i, C_j\right), \quad C_1, ..., C_n | \left\{h_{pt,s}\right\}, \mathbb{P}_C \overset{\text{iid}}{\sim} \mathbb{P}_C,$$

so that, conditional on product, year, and horizon, the dyadic observations are jointly exchangeable under relabelings of countries. More generally, if products or time periods are also viewed as sampled from larger populations, additional latent variables can be introduced for those dimensions, yielding exchangeability along multiple indices.

or "fundamentals") from which we can generate the observed data. As further outlined in Appendix A.1, these primitives are trade costs, labor endowments and productivity parameters. Labor endowments and productivity parameters are all country-specific, so it remains to find country-specific primitives that, when paired across countries, can generate trade costs. Trade costs are affected by, for example, distance, shared language, and shared colonial history (Eaton and Kortum, 2002), all of which can be represented as functions of country-specific primitives (location, language and colonial history). It follows that C_i could contain, among other things, a country's rental rate, labor endowment, productivity parameter, latitude, longitude, and dummy vectors of spoken languages and previous colonizers. $\triangle$

The observation X_{ij} may also depend on general equilibrium effects, captured by a common random variable U. This results in the more general model:

$$\begin{aligned} U|\tilde{h}, \mathbb{P}_C &\sim U[0,1] \\ C_1, ..., C_n|\tilde{h}, \mathbb{P}_C, U &\overset{\text{iid}}{\sim} \mathbb{P}_C \\ X_{ij} &= \tilde{h}(U, C_i, C_j), \quad \text{for } C_i \neq C_j. \end{aligned}$$

The variable U can be thought of as capturing general equilibrium effects or system-wide interdependencies. Following Graham (2020a), I condition on U and suppress it going forward. Specifically, I define

$$h(C_i, C_j) \equiv \tilde{h}(U, C_i, C_j),$$

so that the model reduces to the one in Assumption 3.

4.1.1 Motivation for Model

Assumption 3 implies joint exchangeability of the data $\{X_{k\ell}\}_{k\neq\ell}$, as discussed in Section 3.1.2. Thus, the model is consistent with the exchangeability assumption commonly adopted for dyadic data (Graham, 2020a,b; Davezies, D'haultfœuille, and Guyonvarch, 2021).

The assumption is motivated by the economic structure of the applications considered in this paper. In quantitative trade and spatial equilibrium models, the primitive objects are countries, each characterized by latent fundamentals such as productivity, factor endowments, geography, language, or historical characteristics. Bilateral outcomes are then generated as equilibrium objects from the characteristics of the two countries. It is therefore natural to view the sampling experiment as first drawing countries from a common

population and then constructing all bilateral observations from the sampled countries.[13]

From a probabilistic perspective, Assumption 3 is a restriction of the general jointly exchangeable model characterized by the Aldous-Hoover representation theorem (Aldous, 1981; Hoover, 1979). Specifically, every jointly exchangeable array admits a representation of the form

$$X_{ij} = \tilde{h}^{AH}\left(U, C_i, C_j, D_{ij}\right),$$

for $U, \{C_i\}, \{D_{ij}\} \overset{\text{iid}}{\sim} U[0,1]$. Conditioning on U yields the representation

$$X_{ij} = h^{AH}\left(C_i, C_j, D_{ij}\right).$$

Assumption 3 can therefore be viewed as the restriction of the general Aldous–Hoover representation obtained by excluding the dyad-specific latent variables D_{ij}. This restriction does not preclude rich heterogeneity across dyads: since h is unrestricted, bilateral outcomes may vary arbitrarily across pairs of units. Rather, the maintained assumption is that the dependence and heterogeneity across bilateral outcomes are induced by latent unit-specific characteristics, as in the quantitative trade and spatial models motivating this paper, rather than by an additional irreducible pair-specific latent variable.

4.2 Dirichlet Process Prior and Bayesian Interpretation

Having specified the model in Assumption 3, I assume the following prior on h and $\mathbb{P}_C$:

Assumption 4 (Prior)**.** *We have that h and $\mathbb{P}_C$ are independently drawn according to*

$$(h, \mathbb{P}_C) \sim \pi(h) \cdot \pi(\mathbb{P}_C) = \pi(h) \cdot DP(Q, \alpha). \tag{20}$$

Note that $\pi(h)$ is a distribution over functions, while $\pi(\mathbb{P}_C)$ is a distribution over distributions. Here, $DP(Q, \alpha)$ denotes a Dirichlet process, where Q is a probability measure on $\mathcal{C}$, referred to as the *center measure,* and $\alpha > 0$ is a scalar known as the *prior precision.* The

[13] As an illustration, consider the gravity model $F_{ij} = \exp\left\{\delta_i^{\text{orig}} + \delta_j^{\text{dest}} - \theta \log \tau_{ij} + \delta_{ij}\right\}$ as outlined in Costinot and Rodríguez-Clare (2014). In this specification, δ_i^{orig} and δ_j^{dest} are origin and destination fixed effects, respectively; θ is the trade elasticity; τ_{ij} denotes bilateral trade costs; and δ_{ij} is an idiosyncratic preference shock orthogonal to the other components. Since in Equation (19), h may be non-symmetric and C_i and C_j may be vector-valued, Assumption 3 accommodates gravity specifications in which bilateral trade costs and preference heterogeneity are induced by country-specific latent characteristics.

Dirichlet process prior implies that for any partition $\{A_1, ..., A_R\}$ of $\mathcal{C}$, we have

$$(\mathbb{P}_C(A_1), ..., \mathbb{P}_C(A_R)) \sim \text{Dir}(R; \alpha \cdot Q(A_1), ..., \alpha \cdot Q(A_R)).$$

In Appendix C, I further motivate this class of priors by showing it is uninformative in a specific sense and investigating how much information is drawn from the prior in applications.

Given the model in Assumption 3 and the prior in Assumption 4, we are interested in finding the posterior of θ given the observed data $\{X_{k\ell}\}_{k\neq\ell}$. The key result of this paper is that we can interpret the draws of the Bayesian bootstrap procedure in Algorithm 1 as draws from this posterior in the uninformative limit where $\alpha \downarrow 0$:

Theorem 1 (Finite-sample Bayesian interpretation)**.** *Under Assumptions 3 and 4, in the uninformative limit $\alpha \downarrow 0$, if T is continuous with respect to the topology of weak convergence, then the posterior on θ given the realized data $\{X_{k\ell}\}_{k\neq\ell}$ converges weakly to the distribution induced by the Bayesian bootstrap procedure in Algorithm 1.*

All proofs can be found in Appendix E. The proof of Theorem 1 proceeds in five steps. First, I find the posterior on $\mathbb{P}_C$ given the function h and draws $\{C_k\}$ for a given center measure Q and precision parameter α, and denote it by $\pi_\alpha(\mathbb{P}_C|h, \{C_k\})$. This step combines the model in Equation (18) and the prior in Equation (20) and uses the conjugacy of the Dirichlet process. Second, denoting with π_0 the probability under the limiting posterior as $\alpha \downarrow 0$, I find the limiting posterior on $\mathbb{P}_C$ given the function h and draws $\{C_k\}$:

$$\pi_0(\mathbb{P}_C|h, \{C_k\}) = DP\left(\sum_{k=1}^{n} \frac{1}{n} \cdot \delta_{C_k}, n\right). \tag{21}$$

Note that this limiting posterior is proper and does not depend on the center measure Q. Third, I use the model and properties of Dirichlet processes to find an expression for $\pi_0\left(\mathbb{P}_{X_{ij}}|h, \{C_k\}\right)$, the limiting posterior on the marginal distribution of the observed data given the latent variables $\{C_k\}$ and the function h.

The first three steps all consider the thought experiment where we observe the latent variables $\{C_k\}$ and know the function h. However, in practice we do not observe the latent variables $\{C_k\}$ and do not know the function h; we only observe $\{X_{k\ell}\}_{k\neq\ell}$. In the fourth step I therefore find an expression for $\pi_0\left(\mathbb{P}_{X_{ij}}|\{X_{k\ell}\}_{k\neq\ell}\right)$, the limiting posterior on the distribution of the marginal distribution of the data given the observed data, which I show

corresponds to

$$\begin{aligned}
&\mathbb{P}^*_{n,X_{ij}} \sim \pi_0\left(\mathbb{P}_{X_{ij}} | \left\{X_{k\ell}\right\}_{k\neq\ell}\right) \\
&\Rightarrow \mathbb{P}^*_{n,X_{ij}} = \sum_{k\neq\ell} \frac{W_k \cdot W_\ell}{\sum_{s\neq t} W_s \cdot W_t} \cdot \delta_{X_{k\ell}}, \quad (W_1, ..., W_n) \sim \text{Dir}\left(n; 1, ..., 1\right), \qquad (22)
\end{aligned}$$

which is exactly the distribution we saw in Algorithm 1. Lastly, since $\theta = T\left(\mathbb{P}_{X_{ij}}\right)$, a limiting posterior on the structural parameter, $\pi_0\left(\theta | \left\{X_{k\ell}\right\}_{k\neq\ell}\right)$, is also induced and the conclusion of Theorem 1 follows.

Concerning the Bayesian interpretation of the counterfactual prediction, note that—conditional on the realized data $\{X_{k\ell}\}_{k\neq\ell}$—the only remaining source of randomness arises from the posterior distribution for the structural parameter. Then, combining Theorem 1 and Assumption 2, it follows that $\pi_0\left(\gamma | \left\{X_{k\ell}\right\}_{k\neq\ell}\right)$ converges in distribution to the distribution induced by the Bayesian bootstrap procedure in Algorithm 1 and Equation (17).[14]

5 Theory for Asymptotic Interpretation

In this section I discuss asymptotic validity of the proposed Bayesian bootstrap procedure. I again consider misspecification-robust uncertainty quantification for over-identified GMM as a special case.

5.1 Frequentist Uncertainty Quantification for the Structural Parameter

5.1.1 Sampling Thought Experiment

Assumption 3 specifies the data-generating process for a fixed number of units. To study its large-sample properties, I embed this model in an infinite sequence of latent units and let the number of observed units increase. Throughout this section, the measurable function h and the probability distribution $\mathbb{P}_C$ are treated as fixed but unknown.

Assumption 5 (Sampling thought experiment)**.** *Let*

$$C_1, C_2, ... \overset{\text{iid}}{\sim} \mathbb{P}_C,$$

[14]While the latent variables $\{C_i\}$ may differ across counterfactual scenarios, this does not affect inference because the structural estimand θ is assumed to be fixed, and the estimand for the counterfactual prediction γ is expressed as a function of the observed data $\{X_{k\ell}\}_{k\neq\ell}$ and θ. As a result, we do not need to model how $\{C_i\}$ would change under the counterfactual when quantifying uncertainty for θ or γ.

and define

$$X_{ij} = h\left(C_i, C_j\right), \ i, j \in \mathbb{N}, \ i \neq j.$$

For each sample size n, the observed data consists of the subarray

$$\left\{X_{k\ell} : 1 \leq k, \ell \leq n, \ k \neq \ell\right\}.$$

Assumption 5 formalizes the thought experiment that increasing the sample size corresponds to drawing additional units from the same population and observing all bilateral outcomes among the sampled units. Since the latent characteristics are independently and identically distributed, the resulting array is jointly exchangeable for every n, and each observation X_{ij} has the same marginal distribution, denoted by $\mathbb{P}_{X_{ij}}$.

5.1.2 Asymptotic Bootstrap Validity

The purpose of this section is to establish the frequentist large-sample properties of the Bayesian bootstrap introduced in Algorithm 1. Throughout this section, let

$$\hat{\theta} = T\left(\mathbb{P}_{n,X_{ij}}\right) = T\left(\sum_{k \neq \ell} \frac{1}{n\left(n-1\right)} \cdot \delta_{X_{k\ell}}\right). \tag{23}$$

and let

$$\mathbb{P}^*_{n,X_{ij}} \sim \pi_0\left(\mathbb{P}_{X_{ij}} \middle| \left\{X_{k\ell}\right\}_{k \neq \ell}\right)$$

denote a draw from the Bayesian bootstrap posterior.

Definition 1 (Asymptotic bootstrap validity)**.** *The bootstrap procedure is* asymptotically valid *for the estimator $\hat{\theta}$ if, conditional on the data $\{X_{k\ell}\}_{k\neq\ell}$ and almost surely, $\sqrt{n}\left(T\left(\mathbb{P}^*_{n,X_{ij}}\right) - T\left(\mathbb{P}_{n,X_{ij}}\right)\right)$ and $\sqrt{n}\left(T\left(\mathbb{P}_{n,X_{ij}}\right) - T\left(\mathbb{P}_{X_{ij}}\right)\right)$ converge in distribution to the same limit.*

The main appeal of bootstrap validity for $\hat{\theta}$ is that it implies asymptotic validity of confidence intervals based on the bootstrap, because if n grows large, we can approximate the normal distribution to which $\sqrt{n}\left(\hat{\theta} - \theta\right)$ converges in distribution sufficiently well.[15] Theorem 2 in Appendix D establishes the asymptotic validity of the bootstrap procedure in Algorithm 1 for $\hat{\theta}$ under mild regularity conditions. Corollary 1 provides sufficient conditions for these regularity conditions to hold for Z-estimators, the key requirement being Hadamard differentiability of the zero-extraction map.

[15] The definition of asymptotic bootstrap validity is based on Theorem 23.9 in Van der Vaart (2000).

5.1.3 Special Case: Misspecification-Robust Uncertainty Quantification for GMM

Following Imbens (1997), the two estimation steps of the two-step GMM estimator from Section 3.2.1 can be combined into a single just-identified system,

$$\begin{aligned}
&\phi\left(X_{ij};\theta^{1-\mathrm{GMM}},\theta^{2-\mathrm{GMM}},m,\mathrm{vec}\left\{\Omega\right\},\mathrm{vec}\left\{G_1\right\},\mathrm{vec}\left\{G_2\right\}\right)\\
&=\begin{pmatrix}
\mathrm{vec}\left\{G_1-\frac{\partial}{\partial\theta}\psi\left(X_{ij};\theta^{1-\mathrm{GMM}}\right)\right\}\\
G_1'\psi\left(X_{ij};\theta^{1-\mathrm{GMM}}\right)\\
\psi\left(X_{ij};\theta^{1-\mathrm{GMM}}\right)-m\\
\mathrm{vec}\left\{\Omega-\left[\psi\left(X_{ij};\theta^{1-\mathrm{GMM}}\right)-m\right]\left[\psi\left(X_{ij};\theta^{1-\mathrm{GMM}}\right)-m\right]'\right\}\\
\mathrm{vec}\left\{G_2-\frac{\partial}{\partial\theta}\psi\left(X_{ij};\theta^{2-\mathrm{GMM}}\right)\right\}\\
G_2'\Omega\psi\left(X_{ij};\theta^{2-\mathrm{GMM}}\right)
\end{pmatrix},
\end{aligned}$$

where

$$\mathbb{E}_{\mathbb{P}_{X_{ij}}}\left[\phi\left(X_{ij};\theta^{1-\mathrm{GMM}},\theta^{2-\mathrm{GMM}},m,\mathrm{vec}\left\{\Omega\right\},\mathrm{vec}\left\{G_1\right\},\mathrm{vec}\left\{G_2\right\}\right)\right]=0. \tag{24}$$

Note that the moment equations in Equation (24) hold regardless of whether the moments equations in Equation (14) hold for some $\theta\in\Theta$.[16] Importantly, running this just-identified GMM procedure is numerically equivalent to running the two-step GMM procedure. Since the resulting estimator is a Z-estimator, Corollary 1 in Appendix D implies that Algorithm 2 is asymptotically valid under standard regularity conditions on the stacked moment functions and Hadamard differentiability of the associated zero-extraction map.

Example (Waugh, 2010)**.** Given the moment condition in Equation (15), the estimating function

$$\psi\left(X_{ij};\vartheta\right)=\left(\log\left(\frac{\lambda_{ij}}{\lambda_{ii}}\right)+\vartheta\log\left(\tau_{ij}\frac{p_i}{p_j}\right)\right)\log\left(\tau_{ij}\frac{p_i}{p_j}\right)$$

is linear in ϑ. Standard results for Z-estimators therefore imply the required Hadamard differentiability of the associated zero-extraction map under the usual identification and smoothness conditions. Since there is only a single linear moment function, the remaining regularity conditions in Corollary 1 are straightforward to verify. $\triangle$

[16] Imbens (1997) also shows that iterated GMM estimator (Hansen and Lee, 2021) can be written as a just-identified GMM estimator.

5.2 Frequentist Uncertainty Quantification for the Counterfactual

Recall the estimand $\gamma = g\left(\{X_{k\ell}\}_{k\neq\ell}, \theta\right)$, which is random because it depends on the data $\{X_{k\ell}\}_{k\neq\ell}$. Given that $\hat{\theta}$ is approximately asymptotically normally distributed, we can use a uniform delta method-type result (see for example Chapter 3.4 in Van der Vaart, 2000) to find a valid confidence interval:

Theorem 2 (Delta method for random object)**.** *Suppose we have* $\sqrt{n}\left(\hat{\theta} - \theta\right) \overset{d}{\approx} \mathcal{N}(0, \Sigma)$*, and we can consistently estimate its asymptotic variance by* $\hat{\Sigma}$*. Then, for* $G(\cdot) = \nabla_\theta g\left(\{X_{k\ell}\}_{k\neq\ell}, \cdot\right)$*, if we have*

$$\forall c > 0, \sup_{\tilde{\theta}: \left\|\tilde{\theta}-\theta\right\| \leq \frac{c}{\sqrt{n}}} \left|G\left(\tilde{\theta}\right) - G(\theta)\right| \xrightarrow{p} 0, \tag{25}$$

a valid confidence interval for γ *is given by*

$$\left[\hat{\gamma} \pm \Phi^{-1}(1-\alpha/2) \cdot \sqrt{\frac{1}{n} G\left(\hat{\theta}\right)^2 \hat{\Sigma}}\right].$$

This implies that reporting the quantiles of the bootstrap draws in Equation (17) is an asymptotically valid approach to uncertainty quantification for the counterfactual prediction in a frequentist sense.

6 Extensions

The Bayesian bootstrap procedure in Algorithm 1 can easily be adapted to accommodate various extensions. In this section I consider two such extensions and provide the corresponding changes to the bootstrap procedure, the model and the priors. In Appendix F I additionally discuss multiway clustering and conditional exchangeability.

6.1 Polyadic data

The data do not necessarily have to be dyadic. For example in Section 7.1 we see that the estimation in Caliendo and Parro (2015) corresponds to a triadic regression.

For the general case with polyadic data of order P, denote by $\mathbb{K}_P$ the set of all P-tuples of $\{1, ..., n\}$ without repetition. In this case, we would sample $\left(V_1^{(b)}, ..., V_n^{(b)}\right) \overset{\text{iid}}{\sim} \text{Exp}(1)$, and

compute bootstrap draws according to

$$\hat{\theta}^{*,(b)} = T\left(\sum_{k\in\mathbb{K}_P} \frac{V_{k_1}^{(b)} \cdot ... \cdot V_{k_P}^{(b)}}{\sum_{s\in\mathbb{K}_P} V_{s_1}^{(b)} \cdot ... \cdot V_{s_P}^{(b)}} \cdot \delta_{X_k}\right).$$

The priors from Assumption 4 do not change, and in the model from Assumption 3 only the link function changes, so that we have

$$\begin{aligned} &C_1, ..., C_n | h, \mathbb{P}_C \overset{\text{iid}}{\sim} \mathbb{P}_C \\ &\qquad X_i = h\left(C_{i_1}, ..., C_{i_P}\right), \quad \text{for } C_{i_1} \neq ... \neq C_{i_P}. \end{aligned}$$

6.2 Missing data

When we observe the full matrix of bilateral observations, we observe dyads indexed by the elements of some index set $\mathcal{I}_{\text{non-diag}} = \left\{(i,j) \in \{1,...,n\}^2 : i \neq j\right\}$. However, sometimes non-diagonal observations are missing. In quantitative trade and spatial models, the most common reason for these missing observations is that zero flows are omitted, as is the case for the running example based on Waugh (2010) and in the application based on Caliendo and Parro (2015) in Section 7.1.

To illustrate how to adapt the procedure of Algorithm 1, suppose that we only observe dyads in the set $\mathcal{I} \subset \mathcal{I}_{\text{non-diag}}$. We would then sample $\left(V_1^{(b)}, ..., V_n^{(b)}\right) \overset{\text{iid}}{\sim} \text{Exp}(1)$, and compute bootstrap draws according to

$$\hat{\theta}^{*,(b)} = T\left(\sum_{(k,\ell)\in\mathcal{I}} \frac{V_k^{(b)} \cdot V_\ell^{(b)}}{\sum_{(s,t)\in\mathcal{I}} V_s^{(b)} \cdot V_t^{(b)}} \cdot \delta_{X_{k\ell}}\right).$$

The model in Assumption 3 can be adapted by assuming that the function h maps to an empty set if (C_i, C_j) corresponds to a tuple of indices (i,j) that was not observed. We then have

$$\begin{aligned} &C_1, ..., C_n | h, \mathbb{P}_C \overset{\text{iid}}{\sim} \mathbb{P}_C \\ &\qquad X_{ij} = h\left(C_i, C_j\right) \in \mathcal{X} \cup \emptyset, \quad \text{for } C_i \neq C_i \text{ and } h\left(C_i, C_j\right) \neq \emptyset. \end{aligned}$$

The priors from Assumption 4 do not change.[17]

[17] If we observe a random sample of dyads, we could view the index set $\mathcal{I}$ as random and consider priors on h and $\mathbb{P}_C$ conditional on this index set, so that $\mathcal{I} \sim \pi(\mathcal{I})$ and $(h, \mathbb{P}_C)|\mathcal{I} \sim \pi(h|\mathcal{I}) \cdot DP(Q_{\mathcal{I}}, \alpha)$. The

7 Application: Caliendo and Parro (2015)

In this section I discuss the application in Caliendo and Parro (2015), for which the number of interacting units n ranges between 11 and 15. In Appendix G I discuss the application in Artuç, Chaudhuri, and McLaren (2010), where the number of interacting units is 6.

7.1 Application 1: Caliendo and Parro (2015)

7.1.1 Parameter Estimation

Caliendo and Parro (2015) introduces a new method to estimate trade elasticities. Denoting with F_{ij}^s and t_{ij}^s the trade flow and tariff rate between country i and j in sector s, respectively, the method amounts to running the triadic regressions

$$\log\left(\frac{F_{ij}^s F_{jr}^s F_{ri}^s}{F_{ji}^s F_{rj}^s F_{ir}^s}\right) = -\theta^s \log\left(\frac{t_{ij}^s t_{jr}^s t_{ri}^s}{t_{ji}^s t_{rj}^s t_{ir}^s}\right) + \varepsilon_{ijr}^s,$$

with the identification restriction that the random disturbance term ε_{ijr}^s is orthogonal to the regressor. The number of interacting units n ranges between 11 and 15 across different sector-specific regressions. Using insights from Section 6, the bootstrap procedure can easily be adapted to this triadic setting, where now for each bootstrap draw we compute

$$\hat{\theta}^{s,*,(b)} = T^s \left(\sum_{(k,\ell,m)\in\mathcal{I}^s} \frac{V_k^{(b)} \cdot V_\ell^{(b)} \cdot V_m^{(b)}}{\sum_{(t,u,v)\in\mathcal{I}^s} V_t^{(b)} \cdot V_u^{(b)} \cdot V_v^{(b)}} \cdot \delta_{X_{k\ell m}^s} \right).$$

Note that $\mathcal{I}^s$ is a strict subset of $\left\{(i,j,r) \in \{1,...,n\}^3 : i \neq j \neq r\right\}$, because $\mathcal{I}^s$ only contains observations with $\frac{F_{k\ell}^s F_{\ell m}^s F_{mk}^s}{F_{\ell k}^s F_{m\ell}^s F_{km}^s} > 0$. Table 4 gives the corresponding 95% Bayesian credible intervals and 95% confidence intervals constructed using the point estimates and heteroskedastic-robust standard errors as reported in the paper. Figure 2 plots the corresponding posterior distributions and implied normal distributions. It is noteworthy that many of the credible intervals include zero, which violates the model assumption that $\theta^s > 0$ for all sectors s, since θ^s represents a Fréchet shape parameter and must be strictly positive. Appendix B presents a data-calibrated simulation exercise, which highlights that using heteroskedastic-robust standard errors for uncertainty quantification results in under-

model equations then change to $C_1,...,C_n|h,\mathbb{P}_C,\mathcal{I} \overset{\text{iid}}{\sim} \mathbb{P}_C$ and $X_{ij} = h(C_i, C_j)$, for $(i,j) \in \mathcal{I}$. However, the corresponding bootstrap distribution will not change, so using such a different underlying Bayesian model has no practical implications.

coverage.

Figure 2 highlights that, using the Bayesian bootstrap procedure, we do not have to ex ante think about which cases will result in Gaussian posteriors. For example the posterior for the elasticity for paper looks approximately normal, but the posterior for the elasticity for mining is skewed with a heavy right tail—indicating greater uncertainty about large elasticity values than about small ones.

	Point estimate	95% confidence interval based on Caliendo and Parro (2015)	95% Bayesian bootstrap credible interval
Agriculture, $n = 15$	9.11	[5.17, 13.05]	[-4.05, 25.63]
Mining, $n = 13$	13.53	[6.34, 20.73]	[0.69, 42.35]
Food, $n = 15$	2.62	[1.43, 3.81]	[-1.26, 6.83]
Textile, $n = 14$	8.10	[5.58, 10.61]	[0.52, 16.76]
Wood, $n = 12$	11.50	[5.87, 17.12]	[-11.30, 22.88]
Paper, $n = 14$	16.52	[11.33, 21.71]	[1.70, 31.32]
Petroleum, $n = 12$	64.44	[33.84, 95.04]	[-6.41, 128.87]
Chemicals, $n = 14$	3.13	[-0.37, 6.62]	[-8.49, 13.72]
Plastic, $n = 13$	1.67	[-2.69, 6.03]	[-12.65, 14.01]
Minerals, $n = 14$	2.41	[-0.72, 5.55]	[-3.17, 9.47]
Basic Metals, $n = 14$	3.28	[-1.64, 8.19]	[-11.32, 15.91]
Metal products, $n = 14$	6.99	[2.82, 11.15]	[-5.75, 19.46]
Machinery, $n = 14$	1.45	[-4.04, 6.93]	[-12.75, 17.24]
Office, $n = 14$	12.95	[4.07, 21.83]	[-7.71, 36.25]
Electrical, $n = 14$	12.91	[9.70, 16.12]	[0.20, 21.37]
Communication, $n = 11$	3.95	[0.48, 7.43]	[-5.25, 10.98]
Medical, $n = 14$	8.71	[5.65, 11.78]	[-0.66, 26.37]
Auto, $n = 12$	1.84	[0.04, 3.64]	[-3.80, 5.48]
Other Transport, $n = 14$	0.39	[-1.73, 2.51]	[-5.84, 5.67]
Other, $n = 13$	3.98	[1.86, 6.11]	[-2.11, 9.68]

Table 4: Uncertainty quantification for the benchmark estimates (which remove the countries with the lowest 1% share of trade for each sector) in Table 1 of Caliendo and Parro (2015).

7.1.2 Counterfactual Prediction

The main counterfactual question in Caliendo and Parro (2015) concerns the effects of the NAFTA trade agreement on welfare in Mexico, Canada and the United States. These welfare predictions, which depend on both the data and the estimated trade elasticities, are reported in the abstract and in Table 2 of Caliendo and Parro (2015) without any uncertainty quan-

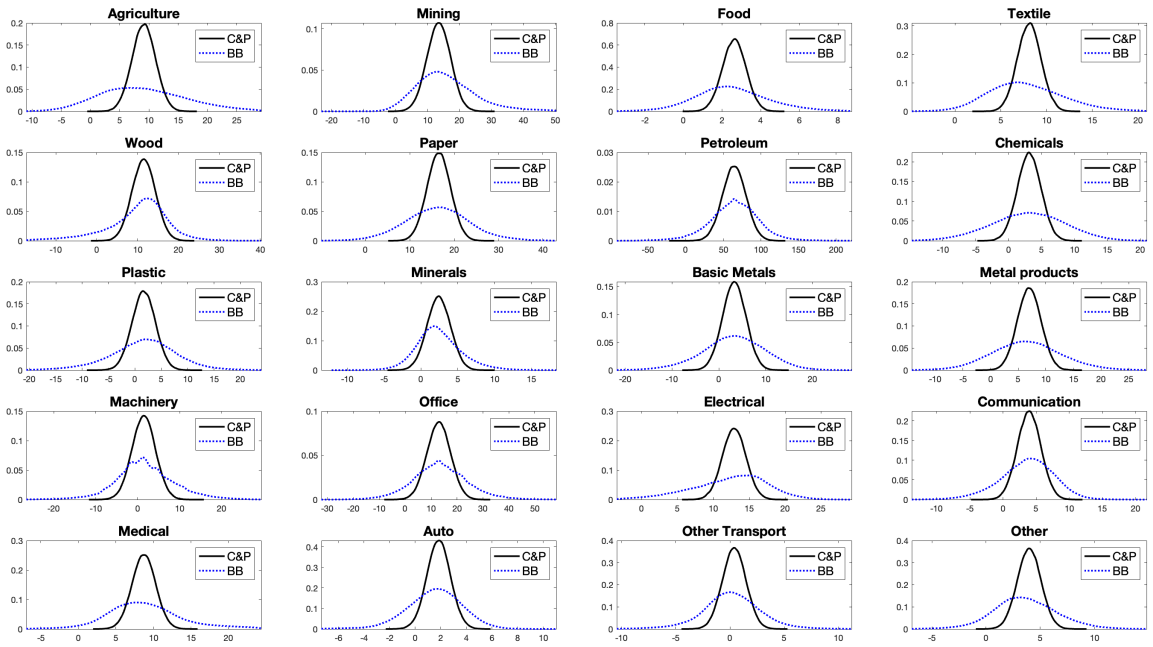

Figure 2: Smoothed densities of the benchmark estimates (which remove the countries with the lowest 1% share of trade for each sector) in Table 1 of Caliendo and Parro (2015). "C&P" corresponds to the normal approximation as implied by the standard errors reported in Caliendo and Parro (2015), and "BB" corresponds to the smoothed Bayesian bootstrap posterior.

tification. In Table 5, I reproduce these results and include 95% Bayesian credible intervals. Figure 3 displays the corresponding posterior distributions. Implementation details and additional results are provided in Appendix I.1.

	Welfare effect
Mexico	1.31% [0.65%, 2.51%]
Canada	-0.06% [-0.10%, -0.02%]
U.S.	0.08% [0.07%, 0.11%]

Table 5: Bayesian uncertainty quantification for welfare effects as in Table 2 of Caliendo and Parro (2015). The numbers in brackets correspond to 95% Bayesian bootstrap credible intervals.

The credible intervals and posterior distributions show asymmetry in the distribution of welfare changes, shifting probability mass away from zero relative to Gaussian posteriors. Furthermore, we observe there is much more uncertainty around the welfare effect for Mexico than around the welfare effects for Canada and the United States. However, since none of

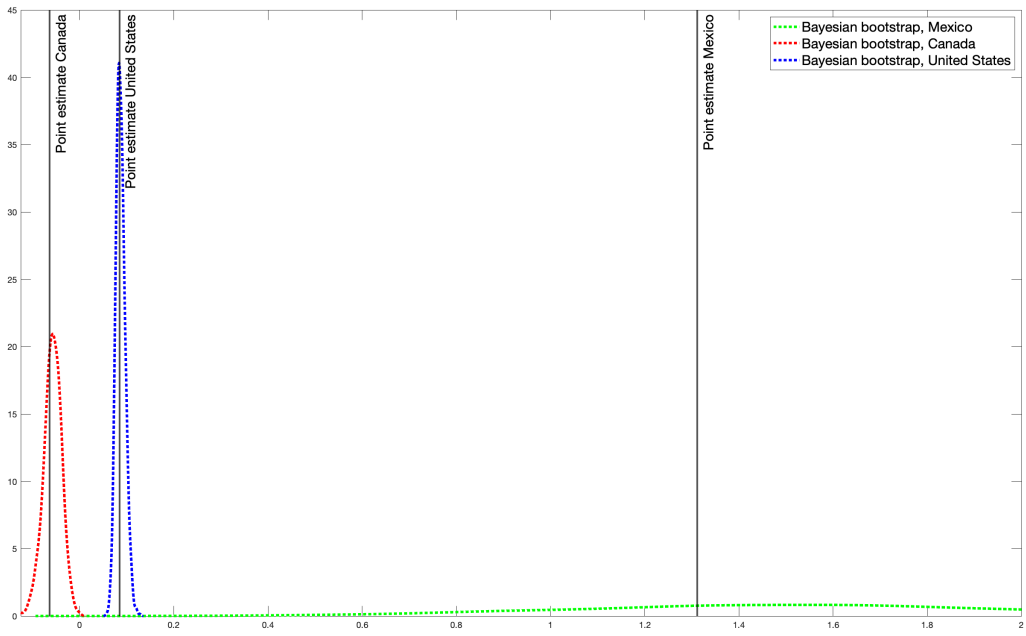

Figure 3: Smoothed Bayesian bootstrap posterior distributions for welfare effects as in Table 2 of Caliendo and Parro (2015).

the credible intervals include zero, the direction of the effect is clearly determined. This is also true for the ranking of welfare effects among the three countries, since for none of the bootstrap draws the ranking is different from the ranking corresponding to the point estimates.

8 Comparison with Alternative Methods

As discussed in the introduction, there exist various alternatives for uncertainty quantification. Here, I discuss an alternative bootstrap from Davezies, D'haultfœuille, and Guyonvarch (2021) based on resampling, and analytic standard errors based on Graham (2020a,b). I compare the various methods using a Monte Carlo study.

8.1 Pigeonhole Bootstrap

The closest method for uncertainty quantification for $\hat{\theta}$ that is theoretically grounded is the pigeonhole bootstrap from Davezies, D'haultfœuille, and Guyonvarch (2021). The method

is summarized in Algorithm 3. For quantitative trade and spatial models, the most important disadvantage of the pigeonhole bootstrap is that its existing theoretical guarantees rely on approximations that envision large number of units. However, as illustrated by the applications in Section 7, relevant applications often include a small number of units.

Algorithm 3 Pigeonhole bootstrap procedure

1. Input: Bilateral data $\{X_{k\ell}\}_{k\neq\ell}$ and estimator function $T : \Delta(\mathcal{X}) \to \Theta$.
2. For each bootstrap draw $b = 1, ..., B$:
   (a) Sample n units *independently with replacement* from $\{1, ..., n\}$ with equal probability. Let $W_k^{\text{pb},(b)}$ denote the number of times that k is sampled.
   (b) Compute
   $$\hat{\theta}^{*,\text{pb},(b)} = T\left(\sum_{k\neq\ell} \frac{W_k^{\text{pb},(b)} \cdot W_\ell^{\text{pb},(b)}}{n(n-1)} \cdot \delta_{X_{k\ell}}\right).$$
3. Report the quantiles of interest of $\left\{\hat{\theta}^{*,\text{pb},(1)}, ..., \hat{\theta}^{*,\text{pb},(B)}\right\}$.

Example (Waugh, 2010)**.** For the application in Waugh (2010), the pigeonhole bootstrap resamples countries with replacement, so a given bootstrap draw may include repeated countries and omit others entirely. For instance, with 43 countries, one draw might include multiple copies of Australia and no copy of Belgium. In contrast, the Bayesian bootstrap procedure in Algorithm 1 assigns continuous and strictly positive weights to all 43 countries in every draw. $\triangle$

Towards uncertainty quantification for the counterfactual prediction $\hat{\gamma}$, the pigeonhole bootstrap procedure again only delivers asymptotic frequentist guarantees. If one is confident in the asymptotic approximation and the validity of the resulting coverage interval for θ, then uncertainty can be propagated using a delta method or bootstrap approximation. Specifically, one could compute bootstrap draws as

$$\hat{\gamma}^{*,\text{pb},(b)} = g\left(\{X_{k\ell}\}_{k\neq\ell}, \hat{\theta}^{*,\text{pb},(b)}\right), \tag{26}$$

for $b = 1, ..., B$, and construct a coverage interval for γ using these draws. The validity of this approach follows from Theorem 2. In Appendix B I perform a simulation exercise that for all my applications compares coverage across methods, assuming the data are generated according to the pigeonhole bootstrap.

To illustrate the differences between the Bayesian bootstrap and the pigeonhole bootstrap, consider the application in Artuç, Chaudhuri, and McLaren (2010) discussed in Appendix G, which uses over-identified GMM to estimate the mean and standard deviation of workers' switching cost and the number of interacting units is $n = 6$. Table 6 shows that the credible intervals obtained from the Bayesian bootstrap are narrower than the coverage intervals obtained from the pigeonhole bootstrap. Figure 10 displays the corresponding bootstrap distributions, omitting draws outside of the considered ranges. There is a non-negligible probability that the pigeonhole bootstrap distribution only has bilateral flows between two industries (around 2% for $n = 6$). The pigeonhole bootstrap also produces more extreme outliers. Specifically, approximately 0.8% of the bootstrap draws for the mean and 0.7% for the variance fall outside the plotted ranges. For the Bayesian bootstrap, the corresponding rates are 0.01% and 0.005%, respectively.

	Point estimate	95% Bayesian bootstrap credible interval	95% pigeonhole bootstrap confidence interval
Mean	5.98	[4.31, 10.13]	[4.06, 11.39]
Standard deviation	1.93	[1.35, 3.04]	[1.18, 3.45]

Table 6: Uncertainty quantification for Panel IV in Table 3 in Artuç, Chaudhuri, and McLaren (2010) for $\zeta = 0.97$.

8.2 Analytic Standard Errors

A second alternative approach for uncertainty quantification for $\hat{\theta}$ is to find frequentist standard errors. I adapt the likelihood setting in Graham (2020b) to obtain a new result for Z-estimators:

Proposition 1 (Analytic standard error for Z-estimators)**.** *Suppose* $\hat{\theta}$ *solves* $\mathbb{E}_{\mathbb{P}_{n,X_{ij}}}\left[\phi\left(X_{ij};\hat{\theta}\right)\right] = 0$ *and* θ *solves* $\mathbb{E}_{\mathbb{P}_{X_{ij}}}\left[\phi\left(X_{ij};\theta\right)\right] = 0$*. Then a consistent variance estimator for* $\hat{\theta}$ *is given by*

$$\widehat{\mathrm{Var}}_{\mathrm{Graham}}\left(\hat{\theta}\right) = \frac{1}{n}\hat{\Sigma}_1^{-1}\left(4\hat{\Sigma}_2 + \frac{2}{n-1}\left(\hat{\Sigma}_3 - 2\hat{\Sigma}_2\right)\right)\left(\hat{\Sigma}_1^{-1}\right)'$$

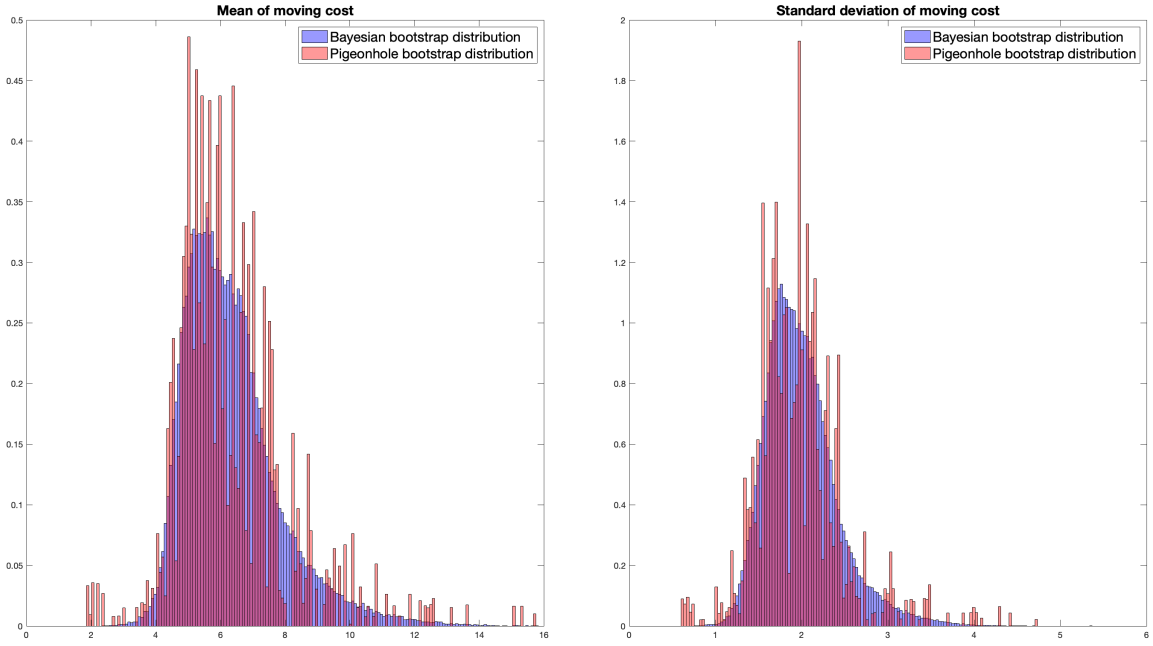

Figure 4: Bootstrap distributions of estimators for Panel IV in Table 3 in Artuç, Chaudhuri, and McLaren (2010) for $\zeta = 0.97$. Draws outside the considered ranges are omitted.

where

$$\hat{\Sigma}_1 = \frac{1}{n\left(n-1\right)} \sum_{k \neq \ell} \frac{\partial \phi\left(X_{k\ell}; \theta\right)}{\partial \theta}|_{\theta=\hat{\theta}}$$

$$\hat{\Sigma}_2 = \binom{n}{3}^{-1} \sum_{k=1}^{n-2} \sum_{\ell=k+1}^{n-1} \sum_{s=\ell+1}^{n} \frac{1}{3} \left\{ \left(\frac{\hat{\phi}_{k\ell} + \hat{\phi}_{\ell k}}{2} \right) \left(\frac{\hat{\phi}_{ks} + \hat{\phi}_{sk}}{2} \right)' \right.$$
$$\left. \left(\frac{\hat{\phi}_{k\ell} + \hat{\phi}_{\ell k}}{2} \right) \left(\frac{\hat{\phi}_{\ell s} + \hat{\phi}_{s\ell}}{2} \right)' + \left(\frac{\hat{\phi}_{ks} + \hat{\phi}_{sk}}{2} \right) \left(\frac{\hat{\phi}_{\ell s} + \hat{\phi}_{s\ell}}{2} \right)' \right\}$$

$$\hat{\Sigma}_3 = \binom{n}{2}^{-1} \sum_{k=1}^{n-1} \sum_{\ell=k+1}^{n} \left(\frac{\hat{\phi}_{k\ell} + \hat{\phi}_{\ell k}}{2} \right) \left(\frac{\hat{\phi}_{k\ell} + \hat{\phi}_{\ell k}}{2} \right)',$$

with $\hat{\phi}_{ij} = \phi\left(X_{ij}; \hat{\theta}\right)$.

For $\hat{\theta}$ a Z-estimator, one can then report the confidence interval $\left[\hat{\theta} \pm 1.96 \cdot \sqrt{\widehat{\text{Var}}_{\text{Graham}}\left(\hat{\theta}\right)}\right]$. As argued in Theorem 2, under some regularity conditions we can use the delta method to find a valid confidence interval for γ.

There are various reasons why one might prefer using the Bayesian bootstrap procedure instead of these analytic standard errors. Firstly, similar to the pigeonhole bootstrap, the validity of the standard errors relies on asymptotic approximations that envision a large number of units. Secondly, it is non-trivial how to adjust the analytic approach to various extensions as discussed in Section 6 (Graham, 2024). Lastly, the approach can be difficult to implement. For example, for over-identified GMM, this approach requires computing many numerical derivatives.

8.3 Comparison when Sample Size is Large

That being said, when the sample size is large, the data are dyadic and there are no missing values, both the pigeonhole bootstrap and the analytic standard errors result in uncertainty quantification that is similar to the Bayesian bootstrap procedure for a Z-estimator. This follows from Corollary 1 and Proposition 1 in the current paper and Theorem 2.4 in Davezies, D'haultfœuille, and Guyonvarch (2021). To illustrate this, in Appendix H I revisit the application that was considered in both Graham (2020a) and Davezies, D'haultfœuille, and Guyonvarch (2021), namely a PPML regression based on data from Silva and Tenreyro (2006). In that setting, despite the Bayesian bootstrap being the only procedure with a finite-sample interpretation, all three methods yield similar uncertainty quantification.

8.4 Monte Carlo Study

8.4.1 Design

I assess finite-sample performance using a gravity-style dyadic Monte Carlo design based on Silva and Tenreyro (2006) that satisfies Assumption 3 exactly. The exact data generating process is described in Appendix J.

For each simulated dataset, I estimate the same dyadic PPML specification as in Appendix H, namely a regression of trade flows on a constant, exporter size, importer size, and log distance. I use exactly the same PPML moment condition and inference code as used for the results in Appendix H. The Monte Carlo design varies the number of countries $n \in \{5, 10, 25, 50, 100\}$ and a dependence parameter $\rho \in \{0, 0.25, 0.5\}$. The parameter ρ controls the extent to which residual variation is shared across dyads with a common exporter or importer. Larger values of ρ therefore correspond to stronger dependence across bilateral observations while holding the overall residual variance fixed.

8.4.2 Results

For each design cell, I use $M = 200$ Monte Carlo replications and $B = 200$ bootstrap draws. I report coverage and mean interval length for four inference procedures: (i) naive analytic standard errors that cluster on dyads, (ii) the Bayesian bootstrap from Algorithm 1, (iii) the pigeonhole bootstrap from Algorithm 3, and (iv) the analytic variance estimator from Proposition 1.

Figure 5 reports Monte Carlo coverage and mean interval length for the log-distance coefficient in the gravity PPML design. Several patterns stand out. First, naive analytic standard errors that cluster on dyads undercover systematically. Averaging across $\rho \in \{0, 0.25, 0.5\}$, their coverage is 0.66 at $n = 5$, 0.76 at $n = 10$, 0.80 at $n = 25$, 0.88 at $n = 50$, and 0.90 at $n = 100$. Second, the Bayesian bootstrap substantially improves finite-sample coverage without becoming excessively conservative. Its average coverage is 0.83 at $n = 5$, 0.94 at $n = 10$, 0.95 at $n = 25$, 0.97 at $n = 50$, and 0.98 at $n = 100$. Third, the pigeonhole bootstrap is highly conservative, especially when the number of countries is small. For example, when $\rho = 0.5$ and $n = 5$, its mean interval length is 19.49, compared with 0.61 for the Bayesian bootstrap. Finally, the analytic variance estimator from Proposition 1 improves with n but is numerically fragile in small samples: its success rate is 0.31 at $n = 5$, 0.58 at $n = 10$, 0.88 at $n = 25$, 0.98 at $n = 50$, and 0.99 at $n = 100$.

Averaging across the positive-dependence cells $\rho \in \{0.25, 0.5\}$, the Bayesian bootstrap attains coverage of 0.94 with mean interval length 0.30, compared with 0.80 and 0.22 for naive dyad-clustered inference and 0.99 and 5.37 for the pigeonhole bootstrap. Overall, the Monte Carlo suggests that the Bayesian bootstrap delivers the best coverage-length tradeoff in this design. The full results can be found in Appendix J.

9 Conclusion

This paper considers uncertainty quantification for counterfactual predictions in polyadic settings. I propose a Bayesian bootstrap procedure to quantify uncertainty around estimators for structural parameters. This also implies valid uncertainty quantification for the point estimates of counterfactual predictions. The procedure is especially appealing in applications with a small number of interacting units, as it admits a finite-sample Bayesian interpretation. At the same time, it provides frequentist asymptotic guarantees under mild conditions. By revisiting the applications in Waugh (2010), Caliendo and Parro (2015) and

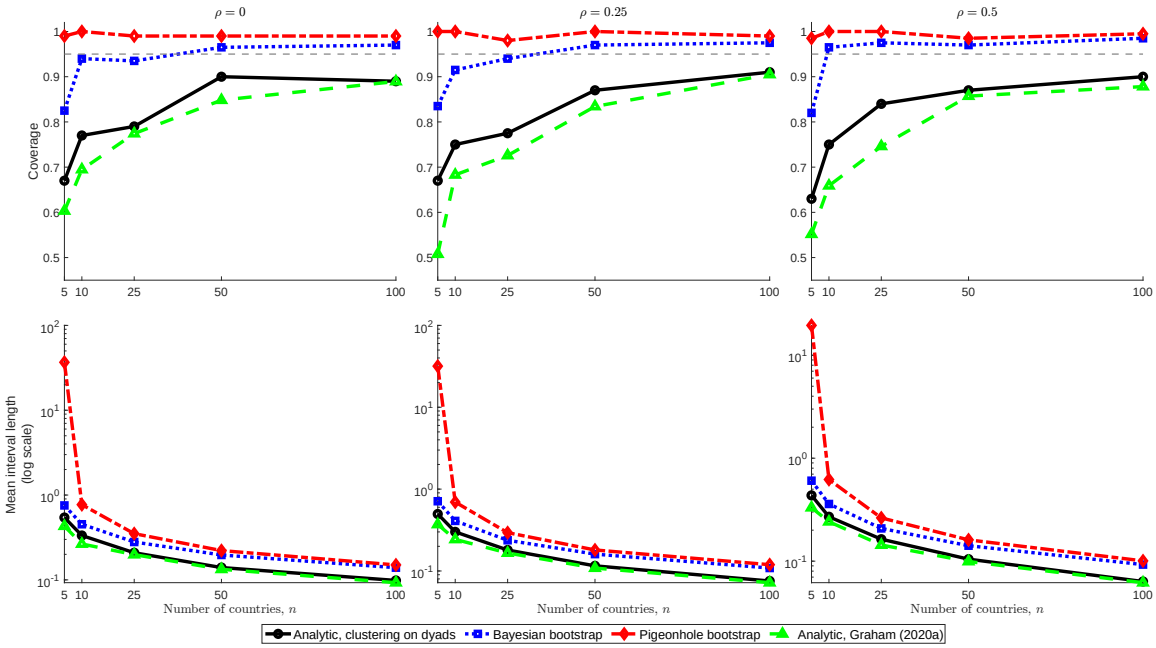

Figure 5: Monte Carlo performance for the log-distance coefficient in the Assumption 3 gravity PPML design. The top row plots empirical coverage and the bottom row plots mean interval length on a log scale. Each column corresponds to a different dependence level $\rho \in \{0, 0.25, 0.5\}$, where ρ is the conditional correlation between log trade flows that share an exporter or importer. The horizontal line in the top row marks nominal 95% coverage. The Bayesian bootstrap tracks nominal coverage more closely than naive dyad-clustered inference, while the pigeonhole bootstrap is markedly more conservative and produces substantially wider intervals, especially when n is small.

Artuç, Chaudhuri, and McLaren (2010), I illustrate the practical advantages of the proposed approach.

References

Adao, R., A. Costinot, and D. Donaldson (2017): "Nonparametric counterfactual predictions in neoclassical models of international trade," *American Economic Review*, 107, 633–689.

Adão, R., A. Costinot, and D. Donaldson (2023): "Putting Quantitative Models to the Test: An Application to Trump's Trade War,"Technical report, National Bureau of Economic Research.

Aldous, D. J. (1981): "Representations for partially exchangeable arrays of random variables," *Journal of Multivariate Analysis*, 11, 581–598.

Alexander, K. S. (1987): "The central limit theorem for empirical processes on Vapnik–Červonenkis classes," *The Annals of Probability*, 178–203.

Allen, T., C. Arkolakis, and Y. Takahashi (2020): "Universal gravity," *Journal of Political Economy*, 128, 393–433.

Andrews, I., and J. M. Shapiro (2024): "Bootstrap Diagnostics for Irregular Estimators,"Technical report, National Bureau of Economic Research.

Ansari, H., D. Donaldson, and E. Wiles (2024): "Quantifying the Sensitivity of Quantitative Trade Models."

Arcones, M. A., and E. Giné (1993): "Limit theorems for U-processes," *The Annals of Probability*, 1494–1542.

Armington, P. S. (1969): "A Theory of Demand for Products Distinguished by Place of Production," *Staff Papers-International Monetary Fund*, 159–178.

Aronow, P. M., C. Samii, and V. A. Assenova (2015): "Cluster–robust variance estimation for dyadic data," *Political analysis*, 23, 564–577.

Artuç, E., S. Chaudhuri, and J. McLaren (2010): "Trade shocks and labor adjustment: A structural empirical approach," *American economic review*, 100, 1008–1045.

Boehm, C. E., A. A. Levchenko, and N. Pandalai-Nayar (2023): "The long and short (run) of trade elasticities," *American Economic Review*, 113, 861–905.

Caliendo, L., and F. Parro (2015): "Estimates of the Trade and Welfare Effects of NAFTA," *The Review of Economic Studies*, 82, 1–44.

Cameron, A. C., and D. L. Miller (2014): "Robust inference for dyadic data," *Unpublished manuscript, University of California-Davis*, 15.

Chamberlain, G. (1987): "Asymptotic efficiency in estimation with conditional moment restrictions," *Journal of econometrics*, 34, 305–334.

Chamberlain, G., and G. W. Imbens (2003): "Nonparametric applications of Bayesian inference," *Journal of Business & Economic Statistics*, 21, 12–18.

Costinot, A., and A. Rodríguez-Clare (2014): "Trade theory with numbers: Quantifying the consequences of globalization," in *Handbook of international economics* Volume 4: Elsevier, 197–261.

Crane, H., and H. Towsner (2018): "Relatively exchangeable structures," *The Journal of Symbolic Logic*, 83, 416–442.

Davezies, L., X. D'haultfœuille, and Y. Guyonvarch (2021): "Empirical process results for exchangeable arrays."

Dingel, J. I., and F. Tintelnot (2020): "Spatial economics for granular settings,"Technical report, National Bureau of Economic Research.

Eaton, J., and S. Kortum (2002): "Technology, geography, and trade," *Econometrica*, 70, 1741–1779.

Fafchamps, M., and F. Gubert (2007): "The formation of risk sharing networks," *Journal of development Economics*, 83, 326–350.

Ghosal, S., and A. W. van der Vaart (2017): *Fundamentals of nonparametric Bayesian inference* Volume 44: Cambridge University Press.

Graham, B. S. (2020a): "Dyadic regression," *The econometric analysis of network data*, 23–40.

——— (2020b): "Network data," in *Handbook of econometrics* Volume 7: Elsevier, 111–218.

——— (2024): "Sparse network asymptotics for logistic regression under possible misspecification," *Econometrica*, 92, 1837–1868.

Hall, A. R. (2000): "Covariance matrix estimation and the power of the overidentifying restrictions test," *Econometrica*, 68, 1517–1527.

Hall, A. R., and A. Inoue (2003): "The large sample behaviour of the generalized method of moments estimator in misspecified models," *Journal of Econometrics*, 114, 361–394.

Hansen, B. E., and S. Lee (2021): "Inference for iterated GMM under misspecification," *Econometrica*, 89, 1419–1447.

Hansen, L. P. (1982): "Large sample properties of generalized method of moments estimators," *Econometrica: Journal of the econometric society*, 1029–1054.

Hoover, D. N. (1979): "Relations on Probability Spaces and Arrays of," *t, Institute for Advanced Study*.

Imbens, G. W. (1997): "One-step estimators for over-identified generalized method of moments models," *The Review of Economic Studies*, 64, 359–383.

Janssen, P. (1994): "Weighted bootstrapping of U-statistics," *Journal of statistical planning and inference*, 38, 31–41.

Kehoe, T. J., P. S. Pujolas, and J. Rossbach (2017): "Quantitative trade models: Developments and challenges," *Annual Review of Economics*, 9, 295–325.

Kosorok, M. R. (2008): *Introduction to empirical processes and semiparametric inference* Volume 61: Springer.

Lee, S. (2014): "Asymptotic refinements of a misspecification-robust bootstrap for generalized method of moments estimators," *Journal of Econometrics*, 178, 398–413.

Menzel, K. (2021): "Bootstrap with cluster-dependence in two or more dimensions," *Econometrica*, 89, 2143–2188.

Pollard, D. (1984): *Convergence of stochastic processes*: Springer Science & Business Media.

Redding, S. J., and E. Rossi-Hansberg (2017): "Quantitative spatial economics," *Annual Review of Economics*, 9, 21–58.

Rosendorf, O. (2023): "Alliance Complements or Substitutes? Explaining Bilateral Intergovernmental Strategic Partnership Ties," *Czech Journal of International Relations*, 58, 7–41.

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

Sanders, B. (2023): "Counterfactual Sensitivity in Equilibrium Models," *arXiv preprint arXiv:2311.14032*.

Silva, J. S., and S. Tenreyro (2006): "The log of gravity," *The Review of Economics and statistics*, 641–658.

Snijders, T. A., S. P. Borgatti et al. (1999): "Non-parametric standard errors and tests for network statistics," *Connections*, 22, 161–170.

Van der Vaart, A. W. (2000): *Asymptotic statistics* Volume 3: Cambridge university press.

Van Der Vaart, A. W., and J. A. Wellner (1996): *Weak convergence*: Springer.

Waugh, M. E. (2010): "International trade and income differences," *American Economic Review*, 100, 2093–2124.

Wigton-Jones, E. (2024): "Lexical distance and the diffusion of technology," *The World Economy*, 47, 3034–3075.

Zhang, D. (2001): "Bayesian bootstraps for U-processes, hypothesis tests and convergence of Dirichlet U-processes," *Statistica Sinica*, 463–478.

Zylkin, T. (2024): "Bootstrap for Gravity Models," *Unpublished Manuscript, University of Richmond*.

Supplemental Appendix

A Extra Results for Running Example: Waugh (2010)

A.1 Model Details

In Section 3.1.3 I introduced the counterfactual mapping

$$\{X_{k\ell}\}_{k\neq\ell}, \hat{\theta}, \{\tau^{\text{cf}}_{k\ell}\} \mapsto \{\hat{w}^{\text{cf}}_k\}.$$

Here,

$$\begin{aligned}\{X_{k\ell}\}_{k\neq\ell} &= \{(\lambda_{k\ell}, \lambda_{kk}, \tau_{k\ell}, p_k, p_\ell)\}_{k\neq\ell} \\ &= (\{\lambda_{k\ell}\}, \{\tau_{k\ell}\}, \{p_k\}),\end{aligned}$$

since $\tau_{kk} = 1$ for all k. Recall that $\lambda_{k\ell}$ denotes country ℓ's expenditure share on goods from country k, $\tau_{k\ell}$ denotes estimated iceberg trade costs from country k to country ℓ, and p_k denotes the aggregate price in country k.

The equilibrium conditions in Waugh (2010) can be viewed as mapping rental rates, trade costs, labor endowments, production parameters and the productivity parameter to aggregated prices, expenditure shares and wages:

$$\{r_k\}, \{\tau_{k\ell}\}, \{L_k\}, \{Q_k\}, \alpha, \beta, \theta_M \mapsto \{p_k\}, \{\lambda_{k\ell}\}, \{w_k\}. \tag{27}$$

Currently, $X_{k\ell}$ only contains the variables that are relevant for constructing the estimator $\hat{\theta}$. The other variables that are inputs to the counterfactual analysis are implicit in the counterfactual mapping. These are labor endowments $\{L_k\}$, aggregate capital-labor ratios $\{K_k\}$ and the production parameters (α, β). So the "data" that we have in hand are

$$\left(\{X_{k\ell}\}_{k\neq\ell}, \{L_k\}, \{K_k\}, \alpha, \beta\right).$$

It follows that we require a "calibration-mapping" that maps observed variables and parameters to rental rates and production parameters:

$$\{X_{k\ell}\}_{k\neq\ell}, \{L_k\}, \{K_k\}, \alpha, \beta, \hat{\theta} \mapsto \{\hat{r}_k\}, \left\{\hat{Q}_k\right\}.$$

Such a mapping exists and we can use the equilibrium mapping in Equation (27) to arrive

at:

$$\left\{\hat{r}_k\right\}, \left\{\tau_{k\ell}^{\text{cf}}\right\}, \left\{L_k\right\}, \left\{\hat{Q}_k\right\}, \alpha, \beta, \hat{\theta} \mapsto \left\{\hat{p}_k^{\text{cf}}\right\}, \left\{\hat{\lambda}_{k\ell}^{\text{cf}}\right\}, \left\{\hat{w}_k^{\text{cf}}\right\}.$$

Once we have obtained the counterfactual wage vector $\left\{\hat{w}_k^{\text{cf}}\right\}$, we can calculate the various inequality statistics.

A.2 Different Subsets of the Data

The productivity parameter in Waugh (2010) is estimated for the full sample and for the two subsets of OECD countries and non-OECD countries. Table 7 and Figure 6 add these subsets to Table 1 and Figure 1, respectively.

A.3 Using PPML instead of OLS

Equation (16) gives the moment function for the application in Waugh (2010) when not omitting zeros and using PPML. Table 8 and Figure 7 add the resulting credible intervals and posterior distributions to Table 7 and Figure 6, respectively. The point estimates drop considerably and there is more uncertainty.

A.4 Implementation details

To compute the limiting marginal prior according to Theorem 4, since there are missing data I use $\delta_{\frac{\chi(\varrho(X_{k\ell}))+\chi(\varrho(X_{\ell k}))}{2}}$ when both (k,ℓ) and (ℓ,k) are observed, $\delta_{\chi(\varrho(X_{k\ell}))}$ when (k,ℓ) but not (ℓ,k) is observed, and $\delta_{\chi(\varrho(X_{\ell k}))}$ when (ℓ,k) but not (k,ℓ) is observed.

A.5 Alternative Methods

Table 9, Figure 8 and Table 10 reproduce Table 7, Figure 6 and Table 3, respectively, but add the results corresponding to the pigeonhole bootstrap from Section 8.1. There are some small differences, especially for the non-OECD sample, but overall the economic conclusions do not change. The approach using analytic standard errors from Section 8.2 cannot be applied here because a substantial share of the observations are missing.

B Comparing Methods using Pigeonhole Bootstrap DGP

Recall the model in Assumption 3:

$$C_1, ..., C_n | h, \mathbb{P}_C \overset{\text{iid}}{\sim} \mathbb{P}_C$$
$$X_{ij} = h(C_i, C_j), \quad \text{for } C_i \neq C_j.$$

To test the performance of the various methods discussed in Section 8, I will use a simulation DGP. Specifically, consider the thought experiment where we observe the latent variables $\{C_k\}$; resample them *with replacement*; and then construct the corresponding data. As summarized in Algorithm 4, this corresponds exactly to the pigeonhole bootstrap. After constructing such a dataset, we can use the various available approaches and check whether the resulting confidence or credible interval covers the structural estimator $\hat{\theta}$. By repeating this procedure many times, we can compute the coverage for each method. Table 11 reports

Algorithm 4 Pigeonhole bootstrap DGP

1. Input: Bilateral data $\{X_{k\ell}\}_{k \neq \ell}$.
2. Sample n units *independently with replacement* from $\{1, ..., n\}$ with equal probability. Let $W_k^{\text{pb},(b)}$ denote the number of times that k is sampled.
3. Construct a new dataset by replicating the observation $X_{k\ell}$ a specific number of times, namely $W_k^{\text{pb},(b)} \cdot W_\ell^{\text{pb},(b)}$, for all $k \neq \ell$.

coverage results for the structural estimators considered in Waugh (2010) and Caliendo and Parro (2015), extending the results previously shown in Table 2.

C Motivation for Dirichlet Process Prior

C.1 Smoothing Across Events

Theorem 1 shows that the choice of the Dirichlet process prior implies a finite-sample Bayesian interpretation for Algorithm 1. Moreover, by considering the limit as the prior precision tends to zero, the procedure becomes agnostic to the choice of center measure Q and prior on h. I further motivate this class of priors by showing it is uninformative in a specific sense:

Definition 2 (Smoothing across events)**.** *Say the posterior* $\pi\left(\mathbb{P}_{X_{ij}} | \{X_{k\ell}\}_{k \neq \ell}\right)$ *does not* smooth

across events *if for every measurable partition* $\{B_1, ..., B_R\}$ *of the support* $\mathcal{X}$ *and*

$$\mathbb{P}^*_{n,X_{ij}} \sim \pi\left(\mathbb{P}_{X_{ij}} | \{X_{k\ell}\}_{k\neq\ell}\right),$$

the distribution of

$$\left(\mathbb{P}^*_{n,X_{ij}}(B_1), ..., \mathbb{P}^*_{n,X_{ij}}(B_R)\right),$$

only depends on the indicators $1^r_{k\ell} = \mathbb{I}\{X_{k\ell} \in B_r\}$ *for* $r = 1, ..., R$.

If a posterior does not smooth across events, to calculate the posterior probability for a given event B, we can replace the data $\{X_{k\ell}\}_{k\neq\ell}$ with its binarized version $\{1_{k\ell}\}_{k\neq\ell}$ for $1_{k\ell} = \mathbb{I}\{X_{k\ell} \in B\}$.

Example (Waugh, 2010)**.** In Waugh (2010), if the posterior does not smooth across events, to compute the posterior probability that a bilateral observation drawn from $\mathbb{P}_{X_{ij}}$ lies in a certain subset $B \subset \mathcal{X}$, we can binarize the observations $\{X_{k\ell}\}_{k\neq\ell}$ into those that lie within B and those that do not. For example, if we would want to predict the probability that a new observation will have an own-country trade share λ_{kk} less than 0.5, we can binarize the observations according to

$$\begin{aligned} 1_{k\ell} &= \mathbb{I}\{X_{k\ell} \in B\} = \mathbb{I}\{\lambda_{kk} < 0.5\} \\ &= \mathbb{I}\{k \in \{\text{Belgium, Benin, Ireland, Mali, Sierra Leone}\}\}. \end{aligned}$$

In particular, for computation of this posterior probability, all countries with own-country trade shares above 0.5 are treated identically. For example, there is no distinction between Denmark ($\lambda_{kk} = 0.523$) and the United States ($\lambda_{kk} = 0.897$). $\triangle$

We have the following theorem:

Theorem 3 (Smoothing across events and Dirichlet process priors)**.** *Under Assumption 3 and the generic priors*

$$(h, \mathbb{P}_C) \sim \pi(h) \cdot \pi(\mathbb{P}_C),$$

we have:

1. If $\pi(\mathbb{P}_C)$ *is a Dirichlet process prior and the prior precision* α *is taken to zero, then the resulting limiting posterior* $\pi_0\left(\mathbb{P}_{X_{ij}} | \{X_{k\ell}\}_{k\neq\ell}\right)$ *does not smooth across events for all* $\pi(h)$.

2. *There exists a prior* $\pi(h)$ *such that the corresponding posterior* $\pi\left(\mathbb{P}_{X_{ij}} | \{X_{k\ell}\}_{k\neq\ell}\right)$ *does not smooth across events* if and only if $\pi(\mathbb{P}_C)$ *is a Dirichlet process prior or a trivial*

process.[18]

So if we want a prior for $\mathbb{P}_C$ that ensures the posterior probability assigned to a set depends only on the data observed within that set, then this mechanically leads us to use a Dirichlet process prior. Such a prior reflects a situation where we have no prior reason to smooth across regions of $\mathcal{X}$; posterior beliefs about a region of the sample space are updated solely based on whether observed data fall inside that region.

C.2 Limiting Marginal Prior for θ

I am taking a Bayesian approach by specifying a prior in Assumption 4. One might wonder how informative Dirichlet process priors are. Specifically, it is of interest to plot the implied limiting marginal prior $\pi(\theta)$ and compare it to the limiting posterior $\pi_0\left(\theta|\left\{X_{k\ell}\right\}_{k\neq\ell}\right)$. By comparing these two distributions, we can see how much information is drawn from the prior.

Theorem 1 shows that the relevant posterior corresponds to an uninformative limit of posteriors for any choice of Q, which implies that there is not a unique well-defined implied limiting marginal prior for θ. In this subsection, I consider a specific choice for the center measure Q and a class of estimators for which we can plot the limiting distribution of $\pi(\theta)$.

Concretely, I constrain the Dirichlet process prior in Equation (20) to

$$\mathbb{P}_C \sim DP\left(\sum_{k=1}^{n}\frac{1}{n}\cdot\delta_{C_k},\alpha\right). \tag{28}$$

This specific choice of the center measure Q implies that mass is supported only on the latent variables $\{C_k\}$.[19] This is also the case for the posterior in Equation (21), which can be recovered by setting $\alpha = n$. We then have the following result:

Theorem 4 (Limiting marginal prior)**.** *Under Assumption 3 and Assumption 8 in Appendix E.3, and using the Dirichlet process prior as in Equation* (28)*, if $\hat{\theta}$ is of the form*

$$\hat{\theta} = T\left(\mathbb{P}_{n,X_{ij}}\right) = \chi\left(\mathbb{E}_{\mathbb{P}_{n,X_{ij}}}\left[\varrho\left(X_{ij}\right)\right]\right),$$

[18]The three trivial processes, as discussed in Section 4.4 of Ghosal and van der Vaart (2017), are: (1) $\pi(\mathbb{P}_C) = \rho$ a.s., for a deterministic probability measure ρ, (2) $\pi(\mathbb{P}_C) = \delta_Y$, for a random variable $Y \sim \rho$, (3) $\pi(\mathbb{P}_C) = Z\delta_a + (1-Z)\delta_b$, for deterministic $a, b \in \mathcal{C}$ and an arbitrary random variable Z with values in $[0,1]$.

[19]One can generalize this to a center measure of $\sum_{k=1}^{n}\omega_k\delta_{C_k}$ for weights $\{\omega_k\}$ that sum up to 1. Then the assumption that mass is supported only on $\{C_k\}$ becomes less restrictive as the number of units n grows large. In particular, Andrews and Shapiro (2024) shows that if $\mathcal{C}$ is a Polish space and $\mathbb{P}_C$ has full support, then for every $\mathbb{P} \in \Delta(\mathcal{C})$ and almost every sequence of draws $\{C_1, C_2, ...\}$ from $\mathbb{P}_C$ there exists a sequence of weights $\{\omega_k^n\}$ such that $\sum_{k=1}^{n}\omega_k^n\delta_{C_k}$ converges weakly to $\mathbb{P}$ as $n \to \infty$.

for known functions $\varrho : \mathcal{X} \to \mathcal{R}$ *and* $\chi : \mathcal{R} \to \Theta$*, and* $\chi(\cdot)$ *is continuous at* $\varrho(X_{k\ell})$ *for all* $k \neq \ell$*, then, as* $\alpha \downarrow 0$*, the implied marginal prior* $\pi(\theta)$ *converges weakly to*

$$\pi^{\infty}(\theta) = \frac{2}{n(n-1)} \sum_{k>\ell} \delta_{\frac{\chi(\varrho(X_{k\ell}))+\chi(\varrho(X_{\ell k}))}{2}}.$$

Theorem 4 shows how to characterize the limiting marginal prior for the class of estimators that can be written as functions of means. For example for the case of simple OLS without an intercept as in the running example and the application in Section 7.1, continuity of χ is satisfied. For estimators that cannot be written in this way, Section C.3 presents an algorithm for plotting proper priors along the limit sequence.

Example (Waugh, 2010)**.** In Figure 9, I plot the bootstrap posterior and the limiting marginal prior using Theorem 4, where we have

$$\varrho(X_{ij}) = \begin{pmatrix} \log\left(\tau_{ij}\frac{p_j}{p_i}\right)^2 \\ -\log\left(\tau_{ij}\frac{p_j}{p_i}\right) \cdot \log\left(\frac{\lambda_{ij}}{\lambda_{ii}}\right) \end{pmatrix}, \ \chi\left(\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}\right) = \frac{a_2}{a_1},$$

and continuity of χ is satisfied. We observe that the limiting marginal prior π^{∞} is much flatter than the bootstrap posterior π_0. Its diffuse shape reflects weak prior information, allowing for a wide range of plausible values for the productivity parameter. $\triangle$

C.3 Plotting Proper Priors along the Limit Sequence

Since the Dirichlet process prior in Equation (28) is only supported on $\{C_k\}$, we have

$$\begin{aligned}(\mathbb{P}_C(C_1), ..., \mathbb{P}_C(C_n)) &\sim \text{Dir}\left(n; \alpha \cdot \sum_{k=1}^{n} \frac{1}{n} \cdot \delta_{C_k}(C_1), ..., \alpha \cdot \sum_{k=1}^{n} \frac{1}{n} \cdot \delta_{C_k}(C_n)\right) \\ &\sim \text{Dir}\left(n; \frac{\alpha}{n}, ..., \frac{\alpha}{n}\right),\end{aligned}$$

which collapses to the Bayesian bootstrap *posterior* in Equation (22) by setting $\alpha = n$. From an analogous argument as in the proof of Theorem 1, it now follows that for a given choice of α we can sample from the corresponding marginal prior for $\mathbb{P}_{X_{ij}}$:

$$\begin{aligned}&\mathbb{P}_{X_{ij}} \sim \pi\left(\mathbb{P}_{X_{ij}}\right) \\ &\Rightarrow \mathbb{P}_{X_{ij}} = \sum_{k\neq\ell} \frac{W_k \cdot W_\ell}{\sum_{s\neq t} W_s \cdot W_t} \cdot \delta_{X_{k\ell}}, \quad (W_1, ..., W_n) \sim \text{Dir}\left(n; \frac{\alpha}{n}, ..., \frac{\alpha}{n}\right).\end{aligned}$$

Since $\theta = T\left(\mathbb{P}_{X_{ij}}\right)$, a marginal prior for θ is also implied for a given choice of α, as is summarized in Algorithm 5.

Algorithm 5 A marginal prior of θ along the uninformative limit sequence

1. Input: Bilateral data $\{X_{k\ell}\}_{k\neq\ell}$, estimator function $T : \Delta(\mathcal{X}) \to \Theta$ and prior precision α.
2. For each draw $b = 1, ..., B$:
   (a) Sample $\left(V_1^{(b)}, ..., V_n^{(b)}\right) \overset{\text{iid}}{\sim} \text{Ga}\left(\frac{\alpha}{n}, 1\right)$.
   (b) Construct $\omega_{k\ell}^{(b)} = V_k^{(b)} \cdot V_\ell^{(b)} / \left(\sum_{s\neq t} V_s^{(b)} \cdot V_t^{(b)}\right)$, for $k, \ell = 1, ..., n$.
   (c) Compute
   $$\hat{\theta}^{*,(b)} = T\left(\sum_{k\neq\ell} \omega_{k\ell}^{(b)} \cdot \delta_{X_{k\ell}}\right).$$
3. Plot the histogram $\left\{\hat{\theta}^{*,(1)}, ..., \hat{\theta}^{*,(B)}\right\}$.

D Asymptotic Bootstrap Validity

To establish bootstrap validity, I first study the associated empirical processes. Let $\mathbb{P}_{X_{ij}} f$ denote $\mathbb{E}_{\mathbb{P}_{X_{ij}}}[f(X_{ij})]$, and define $\mathbb{P}_{n,X_{ij}} f$ and $\mathbb{P}^*_{n,X_{ij}} f$ analogously. For a class of measurable functions $\mathcal{F}$, define

$$\mathbb{G}_n f = \sqrt{n}\left\{\mathbb{P}_{n,X_{ij}} f - \mathbb{P}_{X_{ij}} f\right\}$$
$$\mathbb{G}^*_n f = \sqrt{n}\left\{\mathbb{P}^*_{n,X_{ij}} f - \mathbb{P}_{n,X_{ij}} f\right\}.$$

The goal is to establish weak convergence of both processes in $\ell^\infty(\mathcal{F})$, the space of bounded real-valued functions on $\mathcal{F}$, where the convergence of $\mathbb{G}^*_n$ holds conditional on the data $\{X_{k\ell}\}_{k\neq\ell}$ and outer almost surely. A formal definition of weak convergence is given in Definition 1.3.3 in Van Der Vaart and Wellner (1996).

Assumption 6 (Regularity conditions on $\mathcal{F}$)**.** *Let $\mathcal{F} \subseteq \mathbb{R}^{\mathcal{X}}$ be a measurable class of functions. For each $f \in \mathcal{F}$, define the symmetrized kernel*

$$g_f(c_1, c_2) = \frac{1}{2}\left[f(h(c_1, c_2)) + f(h(c_2, c_1))\right],$$

and let $\mathcal{G} = \{g_f : f \in \mathcal{F}\}$. Suppose

(i) $\mathcal{G}$ *is permissible (see page 196 in Pollard, 1984) and admits a positive envelope* G *with* $\mathbb{P}_C^2 G^2 < \infty$*, where* $\mathbb{P}_C^2 \equiv \mathbb{P}_C \otimes \mathbb{P}_C$*.*

(ii) There exist $0 < c, v < \infty$ *such that for every* $\epsilon > 0$ *and probability measure* Q *on* $\mathcal{C}^2$ *with* $QG^2 < \infty$*, we have*

$$N\left(\epsilon \left\|G\right\|_{L_2(Q)}, \mathcal{G}, \left\|\cdot\right\|_{L_2(Q)}\right) \leq c\epsilon^{-v}.$$

Condition (i) captures regularity conditions on the function class. Permissibility is a mild measure-theoretic regularity condition that ensures function classes meet minimal requirements for measurability and integration, making them suitable for empirical process analysis. The existence of an envelope function G for a class $\mathcal{G}$ means that $|g(c_1, c_2)| \leq G(c_1, c_2)$ for all $g \in \mathcal{G}$ and all $(c_1, c_2) \in \mathcal{C}^2$.

Condition (ii) bounds the complexity of $\mathcal{G}$ through its covering numbers. Here, the covering number $N\left(\epsilon, \mathcal{G}, \left\|\cdot\right\|_{L_2(Q)}\right)$ is the minimal number of $L_2(Q)$-balls of radius ε needed to cover $\mathcal{G}$. This condition is for example satisfied for VC classes of functions by Lemma 4.4 in Alexander (1987).[20]

Theorem 5 (Weak convergence of empirical processes)**.** *Suppose Assumption 5 holds and* $\mathcal{F}$ *satisfies Assumption 6. Then we have weak convergence over* $\ell^\infty(\mathcal{F})$ *of both* $\mathbb{G}_n$ *and* $\mathbb{G}_n^*$ *to the same tight centered Gaussian process* $\mathbb{G}$*, where the convergence of* $\mathbb{G}_n^*$ *holds conditional on the data* $\{X_{k\ell}\}_{k \neq \ell}$ *and outer almost surely.*

Note that the convergence rate is $\sqrt{n}$ despite the sample containing $n(n-1)$ dyads, reflecting the dependence induced by sampling n latent units. The proof of Theorem 5 builds on results from Arcones and Giné (1993) and Zhang (2001), which present a uniform CLT for U-processes and a bootstrap uniform CLT for U-processes, respectively.

Once weak convergence of the empirical processes has been established, bootstrap validity for smooth estimators follows from the functional delta method.

Assumption 7 (Smoothness)**.** *View* $\mathbb{P}_{n,X_{ij}}$ *and* $\mathbb{P}_{X_{ij}}$ *as elements of* $\ell^\infty(\mathcal{F})$*, defined by* $f \mapsto \mathbb{P}_{n,X_{ij}} f$ *and* $f \mapsto \mathbb{P}_{X_{ij}} f$*, respectively. Suppose the estimator is of the form*

$$\hat{\theta} = T\left(\mathbb{P}_{n,X_{ij}}\right) = \varphi\left(\mathbb{P}_{n,X_{ij}}\right),$$

[20] Alternatively, one could assume that $\mathcal{G}$ has polynomial discrimination, defined on page 17 of Pollard (1984). By Lemma II.25 in Pollard (1984), this is a sufficient condition for condition (iii). Also, finite VC-dimension implies polynomial discrimination due to the Sauer-Shelah lemma, see page 275 in Van der Vaart (2000).

where $\varphi : \mathbb{D}_\varphi \subseteq \ell^\infty(\mathcal{F}) \to \Theta$. *The map* φ *is Hadamard differentiable at* $\mathbb{P}_{X_{ij}}$, *tangentially to a subspace* $\ell_0^\infty(\mathcal{F}) \subset \ell^\infty(\mathcal{F})$ *that contains the sample paths of the limiting process* $\mathbb{G}$. *Denote its derivative by* φ'.

The precise definition of Hadamard differentiability is given in Section 20.2 of Van der Vaart (2000), while Section 20.3 provides several examples.

Application of the functional delta method for the bootstrap (Theorem 23.9 in Van der Vaart, 2000) then yields the following theorem:

Theorem 6 (Bootstrap validity)**.** *Suppose Assumption 5 holds. If* $\mathcal{F}$ *satisfies Assumption 6 and the estimator* $\hat{\theta}$ *satisfies Assumption 7, then the bootstrap procedure in Algorithm 1 is asymptotically valid for* $\hat{\theta}$.

It will be useful to gather sufficient conditions for Assumptions 6 and 7 for Z-estimators.

Corollary 1 (Asymptotic bootstrap validity for Z-estimators)**.** *Let* $\Theta \subseteq \mathbb{R}^p$ *and* $\nu_\theta : \mathcal{X} \to \mathbb{R}^p$. *Define*

$$\Psi(\vartheta) = \mathbb{P}_{X_{ij}}\nu_\vartheta, \quad \Psi_n(\vartheta) = \mathbb{P}_{n,X_{ij}}\nu_\vartheta, \quad \Psi_n^*(\vartheta) = \mathbb{P}_{n,X_{ij}}^*\nu_\vartheta.$$

Suppose that θ, $\hat{\theta}$ *and* $\hat{\theta}^*$ *satisfy*

$$\Psi(\theta) = 0, \quad \Psi_n\left(\hat{\theta}\right) = 0, \quad \Psi_n^*\left(\hat{\theta}^*\right) = 0.$$

Assume further that

(i) The scalar function class

$$\mathcal{F}_Z = \{\nu_{\theta,r} : \vartheta \in \Theta, r \in \{1, ..., p\}\}$$

satisfies Assumption 6.

(ii) There exists a neighborhood $\mathcal{N}$ *of* Ψ *and a zero-extraction map* $\varphi : \mathcal{N} \to \Theta$ *such that*

$$\theta = \varphi(\Psi), \quad \hat{\theta} = \varphi(\Psi_n), \quad \hat{\theta}^* = \varphi(\Psi_n^*)$$

with probability approaching one, and φ *is Hadamard differentiable at* Ψ, *tangentially to a subspace containing the sample paths of the limiting process.*[21]

Then the bootstrap procedure in Algorithm 1 is asymptotically valid for $\hat{\theta}$.

[21] Standard sufficient conditions for condition (ii) include local identification of θ, Fréchet differentiability of Ψ at θ, nonsingularity of its derivative, and consistency of the sample and bootstrap roots. See, for example, Theorem 13.5 and Corollary 13.6 in Kosorok (2008).

E Proofs

E.1 Proof of Theorem 1

The proof proceeds in five steps. The first three steps consider the thought experiment where we observe the latent variables $\{C_k\}$ and know the function h. The fourth and fifth step incorporate that in practice we only observe $\{X_{k\ell}\}_{k\neq\ell}$.

Finding $\pi_\alpha(\mathbb{P}_C|h,\{C_k\})$. Combining Equations (18) and (20), we know from Theorem 4.6 in Ghosal and van der Vaart (2017) that the posterior for $\mathbb{P}_C$ is

$$\pi_\alpha(\mathbb{P}_C|h,\{C_k\}) = DP\left(\frac{\alpha}{\alpha+n}Q + \frac{n}{\alpha+n}\frac{1}{n}\sum_{k=1}^{n}\delta_{C_k}, \alpha+n\right). \tag{29}$$

Finding $\pi_0(\mathbb{P}_C|h,\{C_k\})$. Applying Corollary 4.17 in Ghosal and van der Vaart (2017), we can find the weak limit as the precision parameter α is taken to zero:

$$\pi_\alpha(\mathbb{P}_C|h,\{C_k\}) \underset{\alpha\downarrow 0}{\rightsquigarrow} DP\left(\sum_{k=1}^{n}\frac{1}{n}\cdot\delta_{C_k}, n\right),$$

almost surely. Going forward, denote with π_0 the probability under the limiting posterior as $\alpha\downarrow 0$, so that

$$\pi_0(\mathbb{P}_C|h,\{C_k\}) = DP\left(\sum_{k=1}^{n}\frac{1}{n}\cdot\delta_{C_k}, n\right).$$

Note that this limiting posterior distribution on $\mathbb{P}_C$ given the latent variables $\{C_k\}$ and the function h is proper. Furthermore, a random probability distribution $\mathbb{P}^*_{n,C}$ drawn from $\pi_0(\mathbb{P}_C|h,\{C_k\})$ is necessarily supported on the observation points $\{C_k\} = \{C_1,...,C_n\}$. Hence, by definition of a Dirichlet process (e.g. Definition 4.1 in Ghosal and van der Vaart, 2017), we have

$$\begin{aligned}\left(\mathbb{P}^*_{n,C}(C_1),....,\mathbb{P}^*_{n,C}(C_n)\right) &\sim \text{Dir}\left(n; n\cdot\left(\sum_{k=1}^{n}\frac{1}{n}\cdot\delta_{C_k}\right)(C_1),...,n\cdot\left(\sum_{k=1}^{n}\frac{1}{n}\cdot\delta_{C_k}\right)(C_n)\right)\\ &\sim \text{Dir}(n;1,...,1).\end{aligned}$$

It follows that

$$\begin{aligned}
&\mathbb{P}^*_{n,C} \sim \pi_0\left(\mathbb{P}_C | h, \{C_k\}\right) \\
&\Rightarrow \mathbb{P}^*_{n,C} = \sum_{k=1}^{n} W_k \cdot \delta_{C_k}, \quad (W_1, ..., W_n) \sim \mathrm{Dir}\left(n; 1, ..., 1\right),
\end{aligned} \tag{30}$$

and we have

$$Pr_{\pi_0}\left\{C_i \in B | h, \{C_k\}\right\} = \sum_{k=1}^{n} W_k \cdot \mathbb{I}\left\{C_k \in B\right\}. \tag{31}$$

Finding $\pi_0\left(\mathbb{P}_{X_{ij}} | h, \{C_k\}\right)$. Next, combining the model from Assumption 3 and the limiting posterior in Equation (30), we find

$$\begin{aligned}
Pr_{\pi_0}\left\{X_{ij} \in A | h, \{C_k\}\right\} &= Pr_{\pi_0}\left\{h\left(C_i, C_j\right) \in A | C_i \neq C_j, h, \{C_k\}\right\} \\
&= \frac{\mathbb{E}_{\pi_0}\left[\mathbb{I}\left\{h\left(C_i, C_j\right) \in A\right\} \cdot \mathbb{I}\left\{C_i \neq C_j\right\} | h, \{C_k\}\right]}{Pr_{\pi_0}\left\{C_i \neq C_j | h, \{C_k\}\right\}} \\
&= \frac{\sum_{k \neq \ell} W_k \cdot W_\ell \cdot \mathbb{I}\left\{h\left(C_k, C_\ell\right) \in A\right\}}{1 - \sum_s W_s^2} \\
&= \frac{\sum_{k \neq \ell} W_k \cdot W_\ell \cdot \mathbb{I}\left\{X_{k\ell} \in A\right\}}{\sum_{s \neq t} W_s \cdot W_t}.
\end{aligned}$$

This implies that

$$\begin{aligned}
&\mathbb{P}^*_{n,X_{ij}} \sim \pi_0\left(\mathbb{P}_{X_{ij}} | h, \{C_k\}\right) \\
&\Rightarrow \mathbb{P}^*_{n,X_{ij}} = \sum_{k \neq \ell} \frac{W_k \cdot W_\ell}{\sum_{s \neq t} W_s \cdot W_t} \cdot \delta_{X_{k\ell}}, \quad (W_1, ..., W_n) \sim \mathrm{Dir}\left(n; 1, ..., 1\right).
\end{aligned} \tag{32}$$

So we have found an expression for the limiting posterior on the marginal distribution of the observed data given the latent variables $\{C_k\}$ and the function h.

Finding $\pi_0\left(\mathbb{P}_{X_{ij}} | \{X_{k\ell}\}_{k \neq \ell}\right)$. However, in practice we do not observe $\{C_k\}$ and the function h is unknown. We only observe $\{X_{k\ell}\}_{k \neq \ell}$, so we are interested in the limiting posterior on the marginal distribution of the observed data given the observed data, $\pi_0\left(\mathbb{P}_{X_{ij}} | \{X_{k\ell}\}_{k \neq \ell}\right)$. But using the fact that a random probability distribution drawn from $\pi_0\left(\mathbb{P}_{X_{ij}} | h, \{C_k\}\right)$ can be expressed using solely the observed data $\{X_{k\ell}\}_{k \neq \ell}$, we can find an expression for this

limiting posterior:

$$\begin{aligned}\pi_0\left(\mathbb{P}_{X_{ij}}|\left\{X_{k\ell}\right\}_{k\neq\ell}\right) &= \int \pi_0\left(\mathbb{P}_{X_{ij}}|h,\left\{C_k\right\},\left\{X_{k\ell}\right\}_{k\neq\ell}\right) d\pi_0\left(h,\left\{C_k\right\}|\left\{X_{k\ell}\right\}_{k\neq\ell}\right) \\ &= \int \pi_0\left(\mathbb{P}_{X_{ij}}|h,\left\{C_k\right\}\right) d\pi_0\left(h,\left\{C_k\right\}|\left\{X_{k\ell}\right\}_{k\neq\ell}\right) \\ &= \pi_0\left(\mathbb{P}_{X_{ij}}|h,\left\{C_k\right\}\right)\int d\pi_0\left(h,\left\{C_k\right\}|\left\{X_{k\ell}\right\}_{k\neq\ell}\right) \\ &= \pi_0\left(\mathbb{P}_{X_{ij}}|h,\left\{C_k\right\}\right).\end{aligned}$$

The second equality follows from that knowing $(h,\{C_k\})$ implies knowing $\{X_{k\ell}\}_{k\neq\ell}$. The third equality follows from noting that in Equation (32), the limiting posterior $\pi_0\left(\mathbb{P}_{X_{ij}}|h,\{C_k\}\right)$ does not depend on $\{C_k\}$ or h. In conclusion, we have

$$\begin{aligned}&\mathbb{P}^*_{n,X_{ij}} \sim \pi_0\left(\mathbb{P}_{X_{ij}}|\left\{X_{k\ell}\right\}_{k\neq\ell}\right) \\ &\Rightarrow \mathbb{P}^*_{n,X_{ij}} = \sum_{k\neq\ell}\frac{W_k\cdot W_\ell}{\sum_{s\neq t}W_s\cdot W_t}\cdot\delta_{X_{k\ell}}, \quad (W_1,...,W_n)\sim \mathrm{Dir}\left(n;1,...,1\right).\end{aligned}$$

Finding $\pi_0\left(\theta|\left\{X_{k\ell}\right\}_{k\neq\ell}\right)$. Lastly, since $\theta = T\left(\mathbb{P}_{X_{ij}}\right)$ and T is continuous with respect to the topology of weak convergence, the limiting posterior $\pi_0\left(\mathbb{P}_{X_{ij}}|\left\{X_{k\ell}\right\}_{k\neq\ell}\right)$ also implies a limiting posterior on structural estimand, $\pi_0\left(\theta|\left\{X_{k\ell}\right\}_{k\neq\ell}\right)$. So indeed, the procedure from Algorithm 1 has a Bayesian interpretation.

E.2 Proof of Theorem 3

Statement 1. The first part of Theorem 3 follows from the derivations in the proof of Theorem 1, where we noted that the limiting posterior $\pi_0\left(\mathbb{P}_{X_{ij}}|h,\{C_k\}\right)$ did not depend on h or Q, which implies the influences of $\pi(h)$ and the center measure on the limiting posterior drop out when we take the prior precision parameter to zero. Furthermore we can write

$$Pr_{\pi_0}\left\{X_{ij}\in A|\left\{X_{k\ell}\right\}_{k\neq\ell}\right\} = \sum_{k\neq\ell}\frac{W_k\cdot W_\ell}{\sum_{s\neq t}W_s\cdot W_t}\cdot\mathbb{I}\left\{X_{k\ell}\in A\right\},$$

for any A, from which it follows that $\pi_0\left(\mathbb{P}_{X_{ij}}|\left\{X_{k\ell}\right\}_{k\neq\ell}\right)$ does not smooth across events.

Statement 2. The second part of Theorem 3 builds on Corollary 4.29 in Ghosal and van der Vaart (2017), applied to the prior $\pi(\mathbb{P}_C)$. This corollary states that if for every n and every

measurable partition $\{A_1, ..., A_{R_C}\}$ of $\mathcal{C}$, the vector $\left(\mathbb{P}^*_{n,C}(A_1), ..., \mathbb{P}^*_{n,C}(A_{R_C})\right)$, where

$$\mathbb{P}^*_{n,C} \sim \pi\left(\mathbb{P}_C | h, \{C_k\}\right),$$

depends only on the counts $\left(N_1^C, ..., N_{R_C}^C\right)$, for $N_r^C = \sum_{k=1}^n \mathbb{I}\{C_k \in A_r\}$, if and only if the prior $\pi(\mathbb{P}_C)$ is a Dirichlet process or a trivial process.

Given this result, consider the prior $\pi(h)$ that only puts probability mass on the function $h : \mathcal{C}^2 \to \mathcal{X}$ defined by $h(C_i, C_j) = C_i$. In that case $\mathcal{X} = \mathcal{C}$ and for a given partition $\{B_1, ..., B_{R_X}\}$, we have

$$\begin{aligned}
&\left(Pr_\pi\left\{X_{ij} \in B_1 | \{X_{k\ell}\}_{k\neq\ell}\right\}, ..., Pr_\pi\left\{X_{ij} \in B_{R_X} | \{X_{k\ell}\}_{k\neq\ell}\right\}\right) \\
&= \left(Pr_\pi\left\{C_i \in B_1 | \underbrace{C_1, ..., C_1}_{n-1 \text{ times}}, C_2, ..., C_2, ..., C_n, ..., C_n\right\}, ..., \right. \\
&\qquad \left. Pr_\pi\{C_i \in B_{R_X} | C_1, ..., C_1, C_2, ..., C_2, ..., C_n, ..., C_n\}\right).
\end{aligned}$$

For this choice of prior, we then know from Corollary 4.29 in Ghosal and van der Vaart (2017), that the vector

$$\begin{aligned}
&\left(Pr_\pi\left\{C_i \in B_1 | \underbrace{C_1, ..., C_1}_{n-1 \text{ times}}, C_2, ..., C_2, ..., C_n, ..., C_n\right\}, ..., \right. \\
&\qquad \left. Pr_\pi\{C_i \in B_{R_X} | C_1, ..., C_1, C_2, ..., C_2, ..., C_n, ..., C_n\}\right)
\end{aligned}$$

depends only on the counts $\left(N_1^C, ..., N_{R_X}^C\right)$ that use

$$\{C_1, ..., C_1, C_2, ..., C_2, ..., C_n, ..., C_n\},$$

if and only if $\pi(\mathbb{P}_C)$ is a Dirichlet process or a trivial process.

We can relate these counts back to $\{X_{k\ell}\}_{k\neq\ell}$:

$$\begin{aligned}
&\left(N_1^C, ..., N_{R_X}^C\right) \\
&= \left((n-1) \cdot \sum_{k=1}^n \mathbb{I}\{C_k \in B_1\}, ..., (n-1) \cdot \sum_{k=1}^n \mathbb{I}\{C_k \in B_{R_X}\}\right) \\
&= \left(\sum_{k\neq\ell} \mathbb{I}\{X_{k\ell} \in B_1\}, ..., \sum_{k\neq\ell} \mathbb{I}\{X_{k\ell} \in B_{R_X}\}\right).
\end{aligned}$$

We then have that the vector

$$\left(Pr_\pi\left\{X_{ij} \in B_1 | \left\{X_{k\ell}\right\}_{k\neq\ell}\right\}, ..., Pr_\pi\left\{X_{ij} \in B_{R_X} | \left\{X_{k\ell}\right\}_{k\neq\ell}\right\}\right)$$

only depends on the counts

$$\left(\sum_{k\neq\ell}\mathbb{I}\left\{X_{k\ell} \in B_1\right\}, ..., \sum_{k\neq\ell}\mathbb{I}\left\{X_{k\ell} \in B_{R_X}\right\}\right)$$

if and only if $\pi\left(\mathbb{P}_C\right)$ is a Dirichlet process or a trivial process. The conclusion follows.

E.3 Proof of Theorem 4

We first have to check whether the limit exists. We hence require the function $T : \Delta\left(\mathcal{X}\right) \to \Theta$ to be well-behaved in some sense. To formalize this, denote the function that maps a given empirical distribution supported on $\{C_k\}$ to an element of the parameter space by $T_C : \Delta\left(\{C_k\}\right) \mapsto \Theta$. I require this function to be well-behaved when evaluated on weighted empirical distributions where the weights approach a degenerate limit.

Assumption 8 (Condition for existence of limiting marginal prior)**.** *For any permutation $\sigma : \{1, ..., n\} \to \{1, ..., n\}$ and any sequence $\{\omega_{k,w}\}_w \in \Delta\left(\{1, .., n\}\right)$ with $\omega_{k,w} > 0$ and $\lim_{w\to\infty}\omega_{k+1,w}/\omega_{k,w} = 0$ for all k, we have that*

$$\lim_{w\to\infty} T_C\left(\sum_{k=1}^{n}\omega_{\sigma(k),w}\delta_{C_{\sigma(k)}}\right) = \bar{T}_C\left(C_{\sigma(1)}, ..., C_{\sigma(n)}\right),$$

for some limit $\bar{T}_C\left(\cdot\right)$.

Under Assumption 8, the limiting marginal prior as $\alpha \downarrow 0$ takes a simple form:

Lemma 1 (Existence)**.** *Under Assumptions 3 and 8, and using the Dirichlet process prior*

$$\mathbb{P}_C \sim DP\left(\sum_{k=1}^{n}\frac{1}{n}\cdot\delta_{C_k}, \alpha\right),$$

as $\alpha \downarrow 0$ the implied marginal prior $\pi\left(\theta\right)$ converges weakly to $\pi^\infty \in \Delta\left(\Theta\right)$, where

$$\pi^\infty\left(\theta\right) = \sum_{\sigma\in S}\frac{1}{n!}\mathbb{I}\left\{\bar{T}_C\left(C_{\sigma(1)}, C_{\sigma(2)}, ..., C_{\sigma(n)}\right) = \theta\right\},$$

for S the set of permutations $\sigma : \{1, ..., n\} \to \{1, ..., n\}$.

Lemma 1 shows existence of the limiting marginal prior. For the class of estimators that can be written as functions of means, we can actually characterize the limiting object. For this class, we have

$$T_C\left(\mathbb{P}_C\right) = \chi\left(\mathbb{E}_{C_i,C_j\sim\mathbb{P}_C|C_i\neq C_j}\left[\varrho\left(h\left(C_i,C_j\right)\right)\right]\right),$$

and since

$$T_C\left(\sum_{k=1}^{n}\omega_{\sigma(k),w}\delta_{C_{\sigma(k)}}\right) = \chi\left(\sum_{k\neq\ell}\frac{\omega_{\sigma(k),w}\cdot\omega_{\sigma(\ell),w}}{\sum_{s\neq t}\omega_{\sigma(s),w}\cdot\omega_{\sigma(t),w}}\cdot\varrho\left(h\left(C_{\sigma(k)},C_{\sigma(\ell)}\right)\right)\right),$$

Assumption 8 is satisfied for

$$\bar{T}\left(c_1,...,c_n\right) = \frac{\chi\left(\varrho\left(h\left(c_1,c_2\right)\right)\right)+\chi\left(\varrho\left(h\left(c_2,c_1\right)\right)\right)}{2}.$$

This is the case because we have that $\sum_{k\neq\ell}\frac{\omega_{k,w}\cdot\omega_{\ell,w}}{\sum_{s\neq t}\omega_{s,w}\cdot\omega_{t,w}} = 1$ and for $k>\ell$ we have

$$\frac{\omega_{k,w}\cdot\omega_{\ell,w}}{\sum_{s\neq t}\omega_{s,w}\cdot\omega_{t,w}} = \frac{\omega_{1,w}\cdot\omega_{2,w}}{\sum_{s\neq t}\omega_{s,w}\cdot\omega_{t,w}}\underbrace{\frac{\omega_{k,w}}{\omega_{k-1,w}}\cdots\frac{\omega_{3,w}}{\omega_{2,w}}}_{\to 0}\cdot\underbrace{\frac{\omega_{\ell,w}}{\omega_{\ell-1,w}}\cdots\frac{\omega_{2,w}}{\omega_{1,w}}}_{\to 0}.$$

Applying Lemma 1 implies that the marginal prior limit $\pi^{\infty}\left(\theta\right)$ equals

$$\frac{2}{n\left(n-1\right)}\sum_{k>\ell}\delta_{\frac{\chi(\varrho(X_{k\ell}))+\chi(\varrho(X_{\ell k}))}{2}}.$$

E.4 Proof of Lemma 1

The proof follows that of Theorem 3 in Andrews and Shapiro (2024). The stick-breaking representation of Dirichlet processes (see e.g. Theorem 4.12 of Ghosal and van der Vaart, 2017) implies that we can write draws from the prior $\pi\left(\mathbb{P}_C\right)$ as

$$\mathbb{P}_C = \sum_{m=1}^{\infty}V_m\left(\alpha\right)\delta_{\tilde{C}_m},$$

where the random variables $\tilde{C}_m$ are drawn i.i.d. from $\sum_{k=1}^{n}\frac{1}{n}\cdot\delta_{C_k}$, and

$$V_m\left(\alpha\right) = \left(1-U_m^{\frac{1}{\alpha}}\right)\prod_{r=1}^{m-1}U_r^{\frac{1}{\alpha}}$$

where the random variables U_r are i.i.d. standard uniform. Note that $Pr\left\{U_r\in\left(0,1\right)\text{ for all }r\right\} = 1$, and that conditional on this event $V_m\left(\alpha\right)\in\left(0,1\right)$ for all m and all $\alpha>0$, while

$V_{m+1}(\alpha)/V_m(\alpha) \to 0$ as $\alpha \downarrow 0$.

Let $\tau(1) \in \{1, ..., n\}$ be the index for the observation in the original (latent) data with $C_{\tau(1)} = \tilde{C}_1$. For $r \in \{2, ..., n\}$, let $s(r)$ be the smallest s such that $\tilde{C}_s$ is distinct from $\left\{C_{\tau(1)}, ..., C_{\tau(r-1)}\right\}$, and let $C_{\tau(r)} = \tilde{C}_{s(r)}$. We can then equivalently write

$$\mathbb{P}_C = \sum_{k=1}^{n} \omega_k(\tau, \alpha)\, \delta_{C_{\tau(k)}}, \quad \omega_k(\tau, \alpha) = \sum_{m=1}^{\infty} V_m(\alpha)\, \mathbb{I}\left\{\tilde{C}_m = C_{\tau(k)}\right\}.$$

By construction $\mathbb{P}_C \in \Delta(\{C_k\})$, and $\omega_k(\tau, \alpha) \in (0, 1)$ with probability one for all $\alpha > 0$. Moreover, as $\alpha \downarrow 0$ we have $\omega_{k+1}(\tau, \alpha)/\omega_k(\tau, \alpha) \to 0$ for all k, so

$$\lim_{\alpha \downarrow 0} T_C(\mathbb{P}_C) = \lim_{\alpha \downarrow 0} T_C\left(\sum_{k=1}^{n} \omega_k(\tau, \alpha)\, \delta_{C_{\tau(k)}}\right) = \bar{T}_C\left(C_{\tau(1)}, C_{\tau(2)}, ..., C_{\tau(n)}\right)$$

by Assumption 8. The fact that we have to multiply by $1/n!$ then follows from the definition of τ.

E.5 Proof of Theorem 5

Recall the definitions of $\mathbb{G}_n$ and $\mathbb{G}_n^*$, now using Dirichlet draws $(W_1, ..., W_n) \sim \text{Dir}(n; 1, ..., 1)$ instead of exponential draws:

$$\begin{aligned}
\mathbb{G}_n f &= \sqrt{n}\left\{\sum_{k \neq \ell} \frac{1}{n(n-1)} \cdot f(X_{k\ell}) - \mathbb{E}_{\mathbb{P}_{X_{ij}}}[f(X_{ij})]\right\} \\
\mathbb{G}_n^* f &= \sqrt{n}\left\{\sum_{k \neq \ell} \frac{W_k \cdot W_\ell}{\sum_{s \neq t} W_s \cdot W_t} \cdot f(X_{k\ell}) - \sum_{k \neq \ell} \frac{1}{n(n-1)} \cdot f(X_{k\ell})\right\}. \qquad (33)
\end{aligned}$$

I first state a result for U-processes based on Arcones and Giné (1993) and Zhang (2001).

Lemma 2 (Weak convergence of U-processes)**.** *Let $\tilde{\mathcal{F}} \subseteq \mathbb{R}^{\mathcal{C}^2}$ be a measurable class of* symmetric *measurable functions, and let*

$$C_1, ..., C_n \overset{\text{iid}}{\sim} \mathbb{P}_C.$$

The U-process based on $\mathbb{P}_C$ and indexed by $\tilde{\mathcal{F}}$ is

$$U_n\left(\tilde{f}\right) = \sum_{k \neq \ell} \frac{1}{n(n-1)} \cdot \tilde{f}(C_k, C_\ell).$$

Since each $\tilde{f}$ is symmetric, this is equivalently

$$U_n\left(\tilde{f}\right) = \binom{n}{2}^{-1} \sum_{k<\ell} \tilde{f}\left(C_k, C_\ell\right).$$

Suppose that:

(i) $\tilde{\mathcal{F}}$ is permissible (see page 196 in Pollard, 1984) and admits a positive envelope $\tilde{F}$ with $\mathbb{P}_C^2 \tilde{F}^2 < \infty$, where $\mathbb{P}_C^2 \equiv \mathbb{P}_C \otimes \mathbb{P}_C$.

(ii) There exist $0 < c, v < \infty$ such that for every $\epsilon > 0$ and probability measure $\tilde{Q}$ on $\mathcal{C}^2$ with $\tilde{Q}\tilde{F}^2 < \infty$, we have

$$N\left(\epsilon \left\|\tilde{F}\right\|_{L_2(\tilde{Q})}, \tilde{\mathcal{F}}, \|\cdot\|_{L_2(\tilde{Q})}\right) \leq c\epsilon^{-v}.$$

Define

$$\tilde{\mathbb{G}}_n \tilde{f} = \sqrt{n}\left\{U_n\left(\tilde{f}\right) - \mathbb{P}_C^2 \tilde{f}\right\}$$
$$\tilde{\mathbb{G}}_n^* \tilde{f} = \sqrt{n}\left\{\sum_{k\neq\ell} W_k \cdot W_\ell \cdot \tilde{f}\left(C_k, C_\ell\right) - U_n\left(\tilde{f}\right)\right\}.$$

Then $\tilde{\mathbb{G}}_n$ and $\tilde{\mathbb{G}}_n^*$ converge weakly in $\ell^\infty\left(\tilde{\mathcal{F}}\right)$ to the same tight centered Gaussian process $\tilde{\mathbb{G}}$, where the convergence of $\tilde{\mathbb{G}}_n^*$ holds conditional on the data $\{C_k\}$ and outer almost surely.

To apply the lemma, for each $f \in \mathcal{F}$ define

$$\widetilde{f}\left(c_1, c_2\right) = \frac{f\left(h\left(c_1, c_2\right)\right) + f\left(h\left(c_2, c_1\right)\right)}{2},$$

and let $\widetilde{\mathcal{F}} = \left\{\widetilde{f} : f \in \mathcal{F}\right\}$. Every $\widetilde{f}$ is symmetric. This symmetrization does not require h itself to be symmetric.

The empirical average can be written as

$$\begin{aligned}
\sum_{k\neq\ell} \frac{1}{n\left(n-1\right)} f\left(X_{k\ell}\right) &= \sum_{k\neq\ell} \frac{1}{n\left(n-1\right)} \frac{f\left(X_{k\ell}\right) + f\left(X_{\ell k}\right)}{2} \\
&= \sum_{k\neq\ell} \frac{1}{n\left(n-1\right)} \widetilde{f}\left(C_k, C_\ell\right) \\
&= U_n\left(\widetilde{f}\right)
\end{aligned}$$

The first equality follows because exchanging k and ℓ leaves the ordered-pair sum unchanged. Similarly, since (C_1, C_2) and (C_2, C_1) have the same distribution,

$$\begin{aligned}\mathbb{P}_{X_{ij}} f &= \mathbb{E}\left[f\left(h\left(C_1, C_2\right)\right)\right] \\ &= \mathbb{E}\left[\frac{f\left(h\left(C_1, C_2\right)\right) + f\left(h\left(C_2, C_1\right)\right)}{2}\right] \\ &= \mathbb{P}_C^2 \tilde{f}.\end{aligned}$$

It follows that $\mathbb{G}_n f = \widetilde{\mathbb{G}}_n \widetilde{f}$.

By Assumption 6, the induced class $\widetilde{\mathcal{F}}$ satisfies the assumptions of Lemma 2. The lemma therefore implies that $\mathbb{G}_n \rightsquigarrow \mathbb{G}$ in $\ell^\infty(\mathcal{F})$, where $\mathbb{G}$ is a tight centered Gaussian process.

For an order-two U-process, the covariance kernel of the limiting process is

$$4\operatorname{Cov}\left(\mathbb{E}\left[\tilde{f}_1\left(C_1, C_2\right)|C_1\right], \mathbb{E}\left[\tilde{f}_2\left(C_1, C_2\right)|C_1\right]\right).$$

By the definition of the symmetrized kernels,

$$\mathbb{E}\left[\tilde{f}_r\left(C_1, C_2\right)|C_1\right] = \frac{1}{2}\mathbb{E}\left[f_r\left(X_{12}\right) + f_r\left(X_{21}\right)|C_1\right], \quad r \in \{1, 2\}.$$

Hence the covariance kernel can be written as

$$K\left(f_1, f_2\right) = \operatorname{Cov}\left(\mathbb{E}\left[f_1\left(X_{12}\right) + f_1\left(X_{21}\right)|C_1\right], \mathbb{E}\left[f_2\left(X_{12}\right) + f_2\left(X_{21}\right)|C_1\right]\right).$$

Equivalently, let $C_{2'}$ be an independent copy of C_2, and define

$$X_{12'} = h\left(C_1, C_{2'}\right), \quad X_{2'1} = h\left(C_{2'}, C_1\right).$$

Conditional independence of C_2 and $C_{2'}$ given C_1 implies

$$K\left(f_1, f_2\right) = \operatorname{Cov}\left(f_1\left(X_{12}\right) + f_1\left(X_{21}\right), f_2\left(X_{12'}\right) + f_2\left(X_{2'1}\right)\right).$$

We next consider the bootstrap process. Define the unnormalized bootstrap measure

$$\overline{\mathbb{P}}^*_{n, X_{ij}} = \sum_{k \neq \ell} W_k W_\ell \delta_{X_{k\ell}},$$

and the corresponding process

$$\overline{\mathbb{G}}^*_n f = \sqrt{n}\left\{\overline{\mathbb{P}}^*_{n, X_{ij}} f - \mathbb{P}_{n, X_{ij}} f\right\}.$$

Because $W_k W_\ell = W_\ell W_k$,

$$\begin{aligned}\overline{\mathbb{P}}^*_{n,X_{ij}} f &= \sum_{k \neq \ell} W_k W_\ell \frac{f\left(X_{k\ell}\right) + f\left(X_{\ell k}\right)}{2} \\ &= \sum_{k \neq \ell} W_k W_\ell \tilde{f}\left(C_k, C_\ell\right).\end{aligned}$$

Therefore, $\overline{\mathbb{G}}^*_n f = \widetilde{\mathbb{G}}^*_n \widetilde{f}$ in the notation of Lemma 2. The lemma consequently implies that $\overline{\mathbb{G}}^*_n \rightsquigarrow \mathbb{G}$ conditionally on $\{C_k\}$ and outer almost surely. Conditional on the latent sequence, the law of $\overline{\mathbb{G}}^*_n$ is determined by the realized array $\{X_{k\ell}\}_{k \neq \ell}$, since the bootstrap weights are independent of the latent sequence and the process is constructed from the array and those weights. Consequently, the conditional convergence also holds given the observed array $\{X_{k\ell}\}_{k \neq \ell}$.

Finally, it remains to account for the normalization in $\mathbb{G}^*_n$. Let

$$A_n = \sum_{s \neq t} W_s W_t = 1 - \sum_{s=1}^n W_s^2.$$

For $W_1, ..., W_n \sim \text{Dir}\left(n; 1, ..., 1\right)$, we have

$$\mathbb{E}^*\left[\sum_{s=1}^n W_s^2\right] = \frac{2}{n+1}.$$

Hence, by Markov's inequality,

$$\sqrt{n}(1 - A_n) = \sqrt{n} \sum_{s=1}^n W_s^2 = o_{P^*}(1).$$

It follows that

$$A_n^{-1} = 1 + o_{P^*}(1), \qquad \sqrt{n}(A_n^{-1} - 1) = o_{P^*}(1).$$

The normalized bootstrap measure satisfies

$$\mathbb{P}^*_{n,X_{ij}} = A_n^{-1} \overline{\mathbb{P}}^*_{n,X_{ij}}.$$

It follows that

$$\mathbb{G}^*_n = A_n^{-1} \overline{\mathbb{G}}^*_n + \sqrt{n}(A_n^{-1} - 1)\mathbb{P}_{n,X_{ij}}.$$

Consequently,

$$\left\|\mathbb{G}_n^* - \overline{\mathbb{G}}_n^*\right\|_{\mathcal{F}} \leq \ |A_n^{-1} - 1| \left\|\overline{\mathbb{G}}_n^*\right\|_{\mathcal{F}} + \sqrt{n}|A_n^{-1} - 1| \sup_{f \in \mathcal{F}} |\mathbb{P}_{n,X_{ij}} f|.$$

Conditional weak convergence of $\overline{\mathbb{G}}_n^*$ implies $\left\|\overline{\mathbb{G}}_n^*\right\|_{\mathcal{F}} = O_{P^*}(1)$ outer almost surely. Moreover, using G, the envelop introduced in Assumption 6,

$$\sup_{f \in \mathcal{F}} |\mathbb{P}_{n,X_{ij}} f| = \sup_{g \in \mathcal{G}} |U_n(g)| \leq U_n(G).$$

By the strong law of large numbers for U-statistics,

$$U_n(G) \longrightarrow \mathbb{P}_C^2 G < \infty \qquad \text{almost surely,}$$

where finiteness follows from

$$\mathbb{P}_C^2 G \leq \left(\mathbb{P}_C^2 G^2\right)^{1/2} < \infty.$$

Thus,

$$\sup_{f \in \mathcal{F}} |\mathbb{P}_{n,X_{ij}} f| = O(1) \qquad \text{almost surely.}$$

Combining these results yields

$$\left\|\mathbb{G}_n^* - \overline{\mathbb{G}}_n^*\right\|_{\mathcal{F}} = o_{P^*}(1)$$

outer almost surely. Conditional Slutsky's theorem therefore gives $\mathbb{G}_n^* \rightsquigarrow \mathbb{G}$ in $\ell^\infty(\mathcal{F})$, conditional on the observed data and outer almost surely.

E.6 Proof of Lemma 2

The result for the ordinary U-process follows from Theorem 4.9 in Arcones and Giné (1993). The corresponding conditional convergence result for the Dirichlet U-process follows from Corollary 1 in Zhang (2001). Together, these results imply that $\widetilde{\mathbb{G}}_n$ and $\widetilde{\mathbb{G}}_n^*$ converge weakly to the same tight centered Gaussian process, with the convergence of $\widetilde{\mathbb{G}}_n^*$ holding conditional on the latent data and outer almost surely.

Condition (ii) in Lemma 2 differs from Condition 2 in Zhang (2001), as the author assumes $\tilde{\mathcal{F}}$ has polynomial discrimination (defined on page 17 of Pollard, 1984) rather than the condition that the covering numbers are bounded by a polynomial in $1/\varepsilon$. However, in the proofs of Theorem 2.1 and Corollary 1 of Zhang (2001), polynomial discrimination is only used to bound covering numbers using Lemma II.25 and II.36 in Pollard (1984). So

assuming the more familiar bound on the covering numbers directly is without loss.

E.7 Proof of Theorem 6

From Theorem 5, under Assumptions 5 and 6, the empirical process $\mathbb{G}_n = \sqrt{n}\left\{\mathbb{P}_{n,X_{ij}} - \mathbb{P}_{X_{ij}}\right\}$ converges weakly in $\ell^\infty(\mathcal{F})$ to a tight centered Gaussian process $\mathbb{G}$. Moreover, the bootstrap empirical process $\mathbb{G}_n^* = \sqrt{n}\left\{\mathbb{P}^*_{n,X_{ij}} - \mathbb{P}_{n,X_{ij}}\right\}$ converges conditionally on $\{X_{k\ell}\}_{k\neq\ell}$ and outer almost surely, to the same limit. Equivalently,

$$\sup_{\kappa\in \mathrm{BL}_1}\left|\mathbb{E}\left[\kappa\left(\mathbb{G}_n^*\right)\mid\{X_{k\ell}\}_{k\neq\ell}\right] - \mathbb{E}\left[\kappa\left(\mathbb{G}\right)\right]\right| \overset{\text{as}*}{\to} 0,$$

where BL_1 denotes the class of bounded Lipschitz functions from $\ell^\infty(\mathcal{F})$ to $[0,1]$ (page 332 in Van der Vaart, 2000).

By Assumption 7, the estimator satisfies $\hat{\theta} = \varphi\left(\mathbb{P}_{n,X_{ij}}\right)$, where φ is tangentially to a subspace containing the sample paths of the limiting process $\mathbb{G}$. The conclusion then follows from the functional delta method for the bootstrap, Theorem 23.9 in Van der Vaart (2000).

E.8 Proof of Corollary 1

By condition (i) and Theorem 5, $\sqrt{n}\left(\Psi_n - \Psi\right)$ and $\sqrt{n}\left(\Psi_n^* - \Psi_n\right)$ converge, unconditionally and conditionally respectively, to the same Gaussian limit in the relevant function space.

By condition (ii), $\hat{\theta}$ and $\hat{\theta}^*$ are obtained by applying the same Hadamard differentiable zero-extraction map φ to Ψ_n and Ψ_n^*. The result therefore follows directly from Theorem 2.

E.9 Proof of Theorem 2

We can Taylor expand $\hat{\gamma} = \left(\{X_{k\ell}\}_{k\neq\ell}, \hat{\theta}\right)$ around θ to find

$$\sqrt{n}\left(\hat{\gamma} - \gamma\right) = G\left(\bar{\theta}\right)\sqrt{n}\left(\hat{\theta} - \theta\right),$$

for $G(\cdot) = \nabla_\theta g\left(\{X_{k\ell}\}_{k\neq\ell}, \cdot\right)$ and $\bar{\theta}$ an intermediate value between $\hat{\theta}$ and θ. The gradient term is random because it depends on the data $\{X_{k\ell}\}_{k\neq\ell}$. However, under the condition in Equation (25), we have that $G\left(\hat{\theta}\right) = G\left(\bar{\theta}\right) + o_p(1)$. This leads to the approximation

$$\frac{\sqrt{n}\left(\hat{\gamma} - \gamma\right)}{\sqrt{G\left(\hat{\theta}\right)^2\hat{\Sigma}}} \overset{d}{\approx} \mathcal{N}(0,1),$$

and the result follows.

F Other Extensions

F.1 Multiway Clustering

One might want to incorporate another dimension of clustering. For example, in addition to country-heterogeneity one might want to add sector-heterogeneity or time-heterogeneity. We can accommodate this by using an additional, separate exchangeability assumption.

To illustrate, if the observed data $\{X_{k\ell,s}\}_{k\neq\ell,s}$ has a time component and we separately want to allow for clustering across time periods, we would sample

$$\left(W_1^{(b)},...,W_n^{(b)}\right) \sim \text{Dir}\left(n;1,...,1\right)$$
$$\left(\check{W}_1^{(b)},...,\check{W}_T^{(b)}\right) \sim \text{Dir}\left(T;1,...,1\right),$$

and compute bootstrap draws according to

$$\hat{\theta}^{*,(b)} = T\left(\sum_{k\neq\ell,s} \frac{W_k^{(b)}\cdot W_\ell^{(b)}}{\sum_{u\neq v} W_u^{(b)}\cdot W_v^{(b)}} \cdot \check{W}_s^{(b)} \cdot \delta_{X_{k\ell,s}}\right).$$

In the model, adding another dimension of heterogeneity corresponds to independently sampling another set of latent variables. For the example where we also have time-heterogeneity, we have

$$C_1,...,C_n|h,\mathbb{P}_C,\mathbb{P}_{\check{C}} \overset{\text{iid}}{\sim} \mathbb{P}_C$$
$$\check{C}_1,...,\check{C}_T|h,\mathbb{P}_C,\mathbb{P}_{\check{C}} \overset{\text{iid}}{\sim} \mathbb{P}_{\check{C}}$$
$$X_{ij,t} = h\left(C_i,C_j,\check{C}_t\right), \quad \text{for } C_i\neq C_j \text{ for each } t,$$

with corresponding priors

$$(h,\mathbb{P}_C,\mathbb{P}_{\check{C}}) \sim \pi\left(h\right)\cdot DP\left(Q_C,\alpha_C\right)\cdot DP\left(Q_{\check{C}},\alpha_{\check{C}}\right).$$

F.2 Conditional Exchangeability

The key underlying model assumption, as discussed in Section 4.1, is that latent characteristics of the units are drawn i.i.d. from some distribution, or that "units are exchangeable". One might believe this exchangeability assumption only conditional on a set of covariates.

For example one might argue latent characteristics of countries are only i.i.d. within continent or within trade agreement. In this case, there exist different "types" within which agents are exchangeable.

To illustrate, with two types we would independently sample

$$\begin{aligned}\left(W_1^{(b)}, ..., W_{n_1}^{(b)}\right) &\sim \text{Dir}\left(n_1; 1, ..., 1\right) \\ \left(W_{n_1+1}^{(b)}, ..., W_{n_1+n_2}^{(b)}\right) &\sim \text{Dir}\left(n_2; 1, ..., 1\right),\end{aligned}$$

and compute bootstrap draws according to

$$\hat{\theta}^{*,(b)} = T\left(\sum_{k=1}^{n_1+n_2}\sum_{\ell=1,\ell\neq k}^{n_1+n_2}\frac{W_k^{(b)}\cdot W_\ell^{(b)}}{\sum_{s=1}^{n_1+n_2}\sum_{t=1,t\neq s}^{n_1+n_2} W_s^{(b)}\cdot W_t^{(b)}}\cdot\delta_{X_{k\ell}}\right).$$

The corresponding model is

$$\begin{aligned}C_1, ..., C_{n_1}|h, \mathbb{P}_{C_1}, \mathbb{P}_{C_2} &\overset{\text{iid}}{\sim} \mathbb{P}_{C_1} \\ C_{n_1+1}, ..., C_{n_1+n_2}|h, \mathbb{P}_{C_1}, \mathbb{P}_{C_2} &\overset{\text{iid}}{\sim} \mathbb{P}_{C_2} \\ X_{ij} &= h\left(C_i, C_j\right), \quad \text{for } C_i \neq C_j,\end{aligned}$$

and the priors change to

$$\left(h, \mathbb{P}_{C_1}, \mathbb{P}_{C_2}\right) \sim \pi\left(h\right)\cdot DP\left(Q_1, \alpha_1\right)\cdot DP\left(Q_2, \alpha_2\right).$$

As in Section 4.1.1, we can again motivate the model using an Aldous-Hoover representation. Now, $\{X_{ij}\}_{i,j\in\mathbb{N},i\neq j}$ is assumed to be *relatively exchangeable* with respect to "types" R, which means that there exist subpopulations within which agents are exchangeable. For the bootstrap procedure, for each of these types we would obtain a separate vector of Dirichlet draws. In this case,

$$\{X_{ij}\}_{i,j\in\mathbb{N},i\neq j} \overset{d}{=} \left\{X_{\sigma_R(i)\sigma_R(j)}\right\}_{i,j\in\mathbb{N},i\neq j},$$

for any within-type relabeling operation $\sigma_R : \mathbb{N} \to \mathbb{N}$. Graham (2020a) uses results from Crane and Towsner (2018) to show that in this case there exists another array $\left\{X_{ij}^*\right\}_{i,j\in\mathbb{N},i\neq j}$ generated according to

$$X_{ij}^* = \tilde{h}^{AH}\left(U, R_i, R_j, C_i, C_j, D_{ij}\right), \tag{34}$$

for $U, \{C_i\}, \{D_{ij}\} \overset{\text{iid}}{\sim} U\left[0,1\right]$, such that

$$\{X_{ij}\}_{i,j\in\mathbb{N},i\neq j} \overset{d}{=} \left\{X_{ij}^*\right\}_{i,j\in\mathbb{N},i\neq j}.$$

G Application 2: Artuç, Chaudhuri, and McLaren (2010)

G.1 Parameter Estimation

Artuç, Chaudhuri, and McLaren (2010) uses over-identified GMM to estimate the mean and standard deviation of workers' switching cost, denoted with μ and σ, respectively.[22] The data consists of a panel of dyadic data across industries. There are $n = 6$ industries and $T = 23$ years. Towards uncertainty quantification, Artuç, Chaudhuri, and McLaren (2010) ignores the dependence across years and industries and uses the standard GMM asymptotic variance formula. Implicitly, this imposes the assumption that all 690 $(= n \cdot (n-1) \cdot T)$ observations are exchangeable. The corresponding moment function is

$$\psi^{\text{ACM}}\left(X_{ij,t};\theta\right) = \left(Y_{ij,t} - \left(\begin{array}{ccc} \frac{\zeta-1}{\sigma}\mu & \frac{\zeta}{\sigma} & \zeta \end{array}\right) R_{ij,t}\right) Z_{ij,t}, \tag{35}$$

with

$$\begin{aligned} Y_{ij,t} &= \log m_{ij,t} - \log m_{ii,t} \\ R_{ij,t} &= \left(\begin{array}{ccc} 1 & w_{j,t+1} - w_{i,t+1} & \log m_{ij,t+1} - \log m_{jj,t+1} \end{array}\right)' \\ Z_{ij,t} &= \left(\begin{array}{ccc} 1 & w_{j,t-1} - w_{i,t-1} & \log m_{ij,t-1} - \log m_{jj,t-1} \end{array}\right)'. \end{aligned}$$

Here, $m_{ij,t}$ denotes the fraction of the labor force in industry i at time t that chooses to move to industry j and $w_{i,t}$ denotes the wage in industry i at time t. The parameter ζ denotes the discount factor, which is fixed ex ante. I focus on the estimates for μ and σ in Panel IV of Table 3 in Artuç, Chaudhuri, and McLaren (2010), which fixes $\zeta = 0.97$ and corresponds to the authors' preferred specification.

The authors use iterated GMM rather than two-step GMM, and rely on a different weight matrix than the centered weight matrix discussed in Section 3.2.1. Specifically, their weight matrix relies on the assumption that the residual $e_{ij,t}\left(\theta\right) \equiv Y_{ij,t} - \left(\begin{array}{ccc} \frac{\zeta-1}{\sigma}\mu & \frac{\zeta}{\sigma} & \zeta \end{array}\right) R_{ij,t}$ is independent of the instrument $Z_{ij,t}$ *for each dyad-year-observation* (ij,t). Their full procedure is summarized in Algorithm 6.

Exchangeability across all observations is unlikely to hold because of dependence across time and industries. Additionally, one might want to use a weight-matrix that does not rely on the assumption that $e_{ij,t}\left(\theta\right)$ and $Z_{ij,t}$ are independent for each (ij,t). Hence, my preferred approach assumes exchangeability across industries, allowing arbitrary dependence across

[22] To be precise, idiosyncratic moving shocks are assumed to follow an extreme-value distribution indexed by parameter σ, which implies a variance of workers' switching cost of $\pi^2\sigma^2/6$.

Algorithm 6 Iterated GMM procedure used by Artuç, Chaudhuri, and McLaren (2010)

1. Set $\hat{\Omega}_{(0)} = I_3$.
2. Until convergence, compute

$$\hat{\theta}_{(w+1)} = \underset{\vartheta \in \Theta}{\arg\min} \left[\frac{1}{n(n-1)T} \sum_{k \neq \ell, s} e_{k\ell,s}(\vartheta) Z_{k\ell,s} \right]' \hat{\Omega}_{(w)} \left[\frac{1}{n(n-1)T} \sum_{k \neq \ell, s} e_{k\ell,s}(\vartheta) Z_{k\ell,s} \right]$$

$$\hat{\Omega}_{(w+1)} = \Omega^{\text{ACM}}\left(\hat{\theta}_{(w+1)}\right) \equiv \left(\left\{ \frac{1}{n(n-1)T} \sum_{k \neq \ell, s} e_{k\ell,s}\left(\hat{\theta}_{(w+1)}\right)^2 \right\} \left\{ \frac{1}{n(n-1)T} \sum_{k \neq \ell, s} Z_{k\ell,s} Z_{k\ell,s}' \right\} \right)^{-1}.$$

3. Denote with $\hat{\theta}$ and $\hat{\Omega}$ the converged versions.
4. For inference, report standard errors obtained from the variance covariance matrix $\hat{\Sigma} = \frac{1}{n(n-1)T} \left(\hat{G}' \hat{\Omega} \hat{G} \right)^{-1}$, with $\hat{G} = \frac{1}{n(n-1)T} \sum_{k \neq \ell, s} \frac{\partial e_{k\ell,s}\left(\hat{\theta}\right) Z_{k\ell,s}}{\partial \theta}$.

years, and uses the centered weight matrix from Section 3.2.1. This preferred procedure is summarized in Algorithm 7.

The use of a different weight matrix implies a different pseudo-true parameter, and different exchangeability assumptions will yield different uncertainty quantification. However, since my procedure does not rely on correction specification, I can provide valid estimation and uncertainty quantification for the pseudo-true parameter corresponding to any given weight matrix and exchangeability assumption. Given this discussion and using iterated GMM throughout, I consider four approaches to estimation and uncertainty quantification.

The first approach replicates the results from Artuç, Chaudhuri, and McLaren (2010) using Algorithm 6. The second and third approaches are intermediate cases between this first approach and my preferred approach. The second approach still assumes exchangeability across all observations and uses the weight matrix $\Omega^{\text{ACM}}(\cdot)$ from Algorithm 6, but uses a Bayesian bootstrap procedure instead of GMM asymptotics. The third approach uses $\Omega^{\text{ACM}}(\cdot)$, assumes exchangeability across industries, and uses a Bayesian bootstrap procedure. Lastly, the fourth approach implements my preferred procedure as outline in Algorithm 7, which uses a Bayesian bootstrap procedure with a centered weight matrix, and assumes exchangeability across industries.

The resulting 95% confidence intervals and credible intervals are given in Table 12. The corresponding implied normal distributions and posterior distributions are plotted in Fig-

Algorithm 7 Bayesian bootstrap procedure for iterated GMM.

1. For each bootstrap draw $b = 1, ..., B$:

   (a) Sample $\left(V_1^{(b)}, ..., V_6^{(b)}\right) \overset{\text{iid}}{\sim} \text{Exp}(1)$.

   (b) Compute $\omega_{k\ell,s}^{(b)} = V_k^{(b)} \cdot V_\ell^{(b)} / \left(\sum_{u \neq v} V_u^{(b)} \cdot V_v^{(b)}\right)$ for $k, \ell = 1, ..., n$.

   (c) Denote

$$\mathbb{P}_{n,X_{ij}}^{*,(b)} = \sum_{k \neq \ell} \omega_{k\ell}^{(b)} \cdot \delta_{X_{k\ell}} = \sum_{k \neq \ell} \omega_{k\ell}^{(b)} \cdot \delta_{\left\{X_{k\ell,s}\right\}_{s=1}^{T}}.$$

$$\psi\left(X_{ij}; \vartheta\right) = \frac{1}{T} \sum_{t=1}^{T} e_{ij,t}\left(\vartheta\right) Z_{ij,t}$$

$$\psi_n^{(b)}\left(\vartheta\right) = \mathbb{E}_{\mathbb{P}_{n,X_{ij}}^{*,(b)}}\left[\psi\left(X_{ij}; \vartheta\right)\right] = \sum_{k \neq \ell} \omega_{k\ell}^{(b)} \cdot \frac{1}{T} \sum_{t=1}^{T} e_{k\ell,t}\left(\vartheta\right) Z_{k\ell,t}.$$

   (d) Set $\hat{\Omega}_{(0)}^{*,(b)} = I_3$.

   (e) Until convergence, compute

$$\hat{\theta}_{(w+1)}^{*,(b)} = \underset{\vartheta \in \Theta}{\arg\min}\ \psi_n^{(b)}\left(\vartheta\right)' \hat{\Omega}_{(w)}^{*,(b)} \psi_n^{(b)}\left(\vartheta\right)$$

$$\hat{\Omega}_{(w+1)}^{*,(b)} = \mathbb{E}_{\mathbb{P}_{n,X_{ij}}^{*,(b)}}\left[\left\{\psi\left(X_{ij}; \hat{\theta}_{(w+1)}^{*,(b)}\right) - \psi_n^{(b)}\left(\hat{\theta}_{(w+1)}^{*,(b)}\right)\right\}\left\{\psi\left(X_{ij}; \hat{\theta}_{(w+1)}^{*,(b)}\right) - \psi_n^{(b)}\left(\hat{\theta}_{(w+1)}^{*,(b)}\right)\right\}'\right]^{-1}.$$

   (f) Denote with $\hat{\theta}^{*,(b)}$ and $\hat{\Omega}^{*,(b)}$ the converged versions.

2. Report the quantiles of interest of $\left\{\hat{\theta}^{*,(1)}, ..., \hat{\theta}^{*,(B)}\right\}$.

ure 10. The posterior distributions for both the mean and standard deviation of workers' switching costs are non-normal and exhibit heavy right tails, indicating substantial uncertainty—particularly regarding the possibility of large switching costs. Implementation details and extra results can be found in Appendix I.2.

G.2 Counterfactual Prediction

The estimated mean and standard deviation of the moving cost are then used for a simulation exercise. The counterfactual scenario of interest is a sudden liberalization of the manufacturing sector. The main economic quantities of interest are the changes in the employment share of the manufacturing sector, the wage of the manufacturing sector, and the expected

discounted lifetime utility. These counterfactual predictions are reported in Figures 3, 4 and 5 in Artuç, Chaudhuri, and McLaren (2010) without any uncertainty quantification. The 95% confidence intervals and credible intervals for these quantities, under the four sets of assumptions discussed above, are given in Table 13. For uncertainty quantification under the assumptions made in Artuç, Chaudhuri, and McLaren (2010), I use both a standard delta method and an asymptotic bootstrap. For the latter, I draw parameter vectors from the normal distribution implied by Algorithm 6, compute the corresponding counterfactual prediction, and report the quantiles of the resulting bootstrap distribution. The asymptotic validity of both methods follows from Theorem 2.

All intervals are asymmetric around the point estimates. For all approaches, for each of the bootstrap draws, the employment share goes down and the lifetime utility goes up. Notably, this is not the case for the change in the equilibrium wage. To investigate this further, Figure 11 plots normal approximation as implied by the standard errors and the smoothed bootstrap distributions corresponding to the rows of Table 13. The posterior distributions corresponding to the second, third and fourth approaches have heavy left tails, indicating a small probability of a large decrease in the equilibrium wage. Furthermore, all distributions have non-negligible mass above zero. In footnote 26 of Artuç, Chaudhuri, and McLaren (2010) it is mentioned that in principle it could happen that the equilibrium wage rises but "that does not happen in this case". However, when we account for uncertainty this turns out to be an economically important scenario that should be taken into consideration—a finding not visible from point estimates alone.

H Application 3: Silva and Tenreyro (2006)

I follow the empirical illustrations in Graham (2020a) and Davezies, D'haultfœuille, and Guyonvarch (2021), which both consider the dyadic PPML regression from Silva and Tenreyro (2006). Specifically, I consider the fitted regression function of bilateral trade flows $F_{k\ell}$ on a constant, the exporter's log GDP, the importer's log GDP and the log distance. By taking the first order condition of the log likelihood, we can obtain the sample moment condition

$$\mathbb{E}_{\mathbb{P}_{n,X_{ij}}}\left[\left(F_{ij}-\exp\left\{\begin{pmatrix}1 & \text{GDP}_i & \text{GDP}_j & \text{dist}_{ij}\end{pmatrix}\theta\right\}\right)\begin{pmatrix}1\\ \text{GDP}_i\\ \text{GDP}_j\\ \text{dist}_{ij}\end{pmatrix}\right]=0.$$

The basic specification has $n = 136$ countries so $n \cdot (n-1) = 18,360$ bilateral trade flows. For each of the four regression coefficients, I compute a coverage or credible interval in Table 16 and plot the resulting posterior or implied distributions in Figure 14 using (i) naive analytic standard errors that cluster on dyads; (ii) the Bayesian bootstrap procedure in Algorithm 1; (iii) the pigeonhole-type bootstrap in Algorithm 3; and (iv) analytic standard errors from Proposition 1. Reassuringly, all methods with a principled basis in large-sample theory yield comparable results for uncertainty quantification.

I Extra Results for Applications

I.1 Extra Results for Caliendo and Parro (2015)

I.1.1 Details for Implementation

For the Bayesian bootstrap procedure, for each draw I impose a lower bound of $\min\left\{\hat{\theta}^s\right\} = 1.67$ to ensure the code runs. From a Bayesian perspective, one can interpret this as a dogmatic requirement that $\theta^s > 1.67$ for all s. I also follow the authors and replace the elasticity for sectors "Auto" and "Other Transport" by the average elasticity of the other sectors.

I.1.2 Marginal Priors on $\{\theta^s\}$

We can use Theorem 4 with

$$\varrho\left(X_{k\ell m}\right) = \begin{pmatrix} \log\left(\frac{F^s_{k\ell}F^s_{\ell m}F^s_{mk}}{F^s_{\ell k}F^s_{m\ell}F^s_{km}}\right)^2 \\ -\log\left(\frac{F^s_{k\ell}F^s_{\ell m}F^s_{mk}}{F^s_{\ell k}F^s_{m\ell}F^s_{km}}\right)\cdot\log\left(\frac{t^s_{k\ell}t^s_{\ell m}t^s_{mk}}{t^s_{\ell k}t^s_{m\ell}t^s_{km}}\right) \end{pmatrix},\ \chi\left(\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}\right) = \frac{a_2}{a_1},$$

and continuity of χ is satisfied. Figure 12 plots the bootstrap posterior and the limiting marginal prior using Theorem 4. For almost all cases, the marginal priors have some outliers in the right tail, so I only consider (normalized) prior mass within ten standard deviations. We observe that the prior is much flatter than the bootstrap posterior.

I.1.3 Alternative Methods

Table 14, Figure 13 and Table 15 reproduce Table 4, Figure 2 and Table 5, respectively, but add the results corresponding to the pigeonhole bootstrap from Section 8.1. The confidence intervals for the sectoral elasticities constructed using the pigeonhole bootstrap are consistently larger and all include zero. The confidence intervals for the welfare predictions are

somewhat larger, especially the upper bound on the welfare effect of Mexico. The approach using analytic standard errors from Section 8.2 cannot be applied here because data are triadic and a small share of observations are missing.

I.2 Extra Results for Artuç, Chaudhuri, and McLaren (2010)

I.2.1 Details for Implementation

In a small number of counterfactual draws, the welfare effects are complex numbers. I omit these draws, and one can again interpret this as a dogmatic requirement that welfare effects must be real numbers.

I.2.2 Alternative Methods

In principle one could apply the approach using analytic standard errors from Section 8.2 by interpreting the two-step GMM estimator as a just-identified GMM estimator as per the discussion in Section 5.1.3. However, since this amounts to taking many numerical derivatives, the procedure is numerically unstable.

J Details for Monte Carlo

J.1 Data-Generating Process

For each country $i = 1, ..., n$, let

$$C_i = \left(G_i, S_i, U_i^o, U_i^d, M_i^o, M_i^d\right), \qquad C_1, \ldots, C_n \overset{\text{iid}}{\sim} \mathbb{P}_C,$$

where G_i is a country-size variable, $S_i \in [0,1]^2$ is a geographic location, and U_i^o, U_i^d, M_i^o, M_i^d are latent country characteristics. In the implementation, G_i, U_i^o, U_i^d, M_i^o and M_i^d are independent standard normal random variables and S_i is drawn uniformly on the unit square.

For each ordered dyad (i, j) with $i \neq j$, define bilateral log trade costs as

$$\ell_{ij} = \log\left(1 + 100 \left\|S_i - S_j\right\|\right).$$

I then construct three multiplicative latent components. First, exporter-specific and importer-specific effects are given by

$$A_i^o\left(\rho, \sigma\right) = \exp\left(\sigma\sqrt{\rho}U_i^o - \sigma^2\rho/2\right), \qquad A_i^d\left(\rho, \sigma\right) = \exp\left(\sigma\sqrt{\rho}U_i^d - \sigma^2\rho/2\right).$$

Second, I introduce a bilateral term built entirely from country-level primitives,

$$D_{ij}(\rho,\sigma)=\frac{\exp\left(\sigma\sqrt{1-2\rho}M_i^oM_j^d\right)}{\mathbb{E}\left[\exp\left(\sigma\sqrt{1-2\rho}M_i^oM_j^d\right)\right]},\qquad 0\le\rho\le\frac{1}{2}.$$

Trade flows are then generated according to

$$F_{ij}=\exp\left(\beta_0+\beta_1G_i+\beta_2G_j+\beta_3\ell_{ij}\right)A_i^o(\rho,\sigma)A_j^d(\rho,\sigma)D_{ij}(\rho,\sigma),$$

with baseline parameter vector $\beta=(1,0.5,0.5,-1)'$. In the baseline implementation, I fix $\sigma=0.5$ throughout and vary ρ across design cells.

The observed dyadic record is

$$X_{ij}=(F_{ij},G_i,G_j,\ell_{ij}),$$

so that $X_{ij}=h(C_i,C_j)$ for a measurable function h. Hence the design fits Assumption 3 exactly: all dyadic dependence is induced by shared country-level latent variables. Importantly, $D_{ij}(\rho,\sigma)$ is not an unrestricted dyad-specific shock; it is itself a deterministic function of country-level primitives.

Because each multiplicative latent component is normalized to have mean one,

$$\mathrm{E}[A_i^o(\rho,\sigma)]=\mathrm{E}\left[A_i^d(\rho,\sigma)\right]=\mathrm{E}[D_{ij}(\rho,\sigma)]=1,$$

the conditional mean remains correctly specified:

$$\mathbb{E}[F_{ij}\mid G_i,G_j,S_i,S_j]=\exp\left(\beta_0+\beta_1G_i+\beta_2G_j+\beta_3\ell_{ij}\right).$$

Although the simulated trade flows are continuous and strictly positive, PPML is therefore correctly interpreted as a quasi-likelihood estimator.

This parameterization separates overall residual dispersion from dependence. The scale parameter σ governs the overall variance of the latent component, while ρ governs pairwise dependence across observations that share a country. In particular, conditional on the observed regressors,

$$\mathrm{Corr}(\log F_{ij},\log F_{ik}\mid G,S)=\rho,\qquad \mathrm{Corr}(\log F_{ji},\log F_{ki}\mid G,S)=\rho.$$

Thus ρ has a direct interpretation as the correlation between two flows that share an exporter or importer.[23]

[23]This design can also be interpreted as a structured special case of Graham's multiplicative gravity specification,

J.2 Results

Table 17 shows the empirical coverage and mean interval length for the Monte Carlo exercise.

K Appendix Tables

	Point estimate	95% confidence interval based on Waugh (2010)	95% Bayesian bootstrap credible interval
All countries, $n = 43$	5.55	[5.39, 5.71]	[5.12, 6.02]
Only OECD, $n = 19$	7.91	[7.46, 8.37]	[6.91, 9.21]
Only non-OECD, $n = 24$	5.45	[5.06, 5.84]	[4.42, 6.65]

Table 7: Uncertainty quantification for productivity parameters in Waugh (2010).

$$Y_{ij} = \exp\left(R_{ij}'\beta\right) A_i B_j V_{ij},$$

obtained by setting

$$R_{ij} = (1, G_i, G_j, \ell_{ij}), \qquad A_i = A_i^o(\rho, \sigma), \qquad B_j = A_j^d(\rho, \sigma), \qquad V_{ij} = D_{ij}(\rho, \sigma).$$

Unlike an unrestricted dyad shock, however, V_{ij} is itself constructed from country-level latent variables, so that the Assumption 3 representation $X_{ij} = h(C_i, C_j)$ continues to hold. In the notation of footnote 13, the mapping is

$$\tau_{ij} = 1 + 100\,\|S_i - S_j\|, \qquad \delta_i^{\text{orig}} = \beta_1 G_i + \sigma\sqrt{\rho}\,U_i^o - \sigma^2\rho/2, \qquad \delta_j^{\text{dest}} = \beta_2 G_j + \sigma\sqrt{\rho}\,U_j^d - \sigma^2\rho/2,$$

and

$$\delta_{ij} = \log D_{ij}(\rho, \sigma),$$

up to the common intercept β_0. Thus the bilateral term in the gravity equation is present, but it is generated by country-level primitives rather than by an unrestricted dyad-specific residual.

	Point estimate	95% confidence interval based on Waugh (2010)	95% Bayesian bootstrap credible interval
All countries, OLS, $n = 43$	5.55	[5.39, 5.71]	[5.12, 6.02]
Only OECD, OLS, $n = 19$	7.91	[7.46, 8.37]	[6.91, 9.21]
Only non-OECD, OLS, $n = 24$	5.45	[5.06, 5.84]	[4.42, 6.65]
All countries, PPML, $n = 43$	4.19	-	[3.42, 5.12]
Only OECD, PPML, $n = 19$	5.81	-	[4.43, 7.41]
Only non-OECD, PPML, $n = 24$	4.49	-	[3.17, 6.68]

Table 8: Uncertainty quantification for productivity parameters in Waugh (2010) using OLS and PPML.

	Point estimate	95% confidence interval based on Waugh (2010)	95% Bayesian bootstrap credible interval	95% pigeonhole bootstrap confidence interval
All countries, $n = 43$	5.55	[5.39, 5.71]	[5.12, 6.02]	[5.12, 6.05]
Only OECD, $n = 19$	7.91	[7.46, 8.37]	[6.91, 9.21]	[6.86, 9.41]
Only non-OECD, $n = 24$	5.45	[5.06, 5.84]	[4.42, 6.65]	[4.43, 6.91]

Table 9: Uncertainty quantification for productivity parameters in Waugh (2010).

Scenario	Baseline		Autarky	
τ_{ij}^{cf}	τ_{ij}		$\infty \cdot \mathbb{I}\{i \neq j\}$	
Method	BB	PB	BB	PB
Variance of log wages	1.30 [1.28, 1.32]	1.30 [1.28, 1.32]	1.35 [1.31, 1.38]	1.35 [1.31, 1.38]
90th/10th percentile of wages	25.7 [25.1, 26.2]	25.7 [25.1, 26.2]	23.5 [22.6, 24.2]	23.5 [22.6, 24.2]
Mean % change in wages	-	-	-10.5 [-11.4, -9.6]	-10.5 [-11.4, -9.5]

Scenario	Symmetry		Free trade	
τ_{ij}^{cf}	$\min\{\tau_{ij}, \tau_{ji}\}$		1	
Method	BB	PB	BB	PB
Variance of log wages	1.05 [1.05, 1.05]	1.05 [1.05, 1.05]	0.76 [0.75, 0.78]	0.76 [0.75, 0.78]
90th/10th percentile of wages	17.3 [17.2, 17.4]	17.3 [17.2, 17.4]	11.4 [11.0, 11.9]	11.4 [11.0, 11.9]
Mean % change in wages	24.2 [22.4, 25.8]	24.2 [22.3, 25.8]	128.0 [114.4, 140.7]	128.0 [113.5, 140.7]

Table 10: Bayesian uncertainty quantification for counterfactual predictions as in Table 4 of Waugh (2010). The numbers in brackets in the columns "BB" correspond to 95% Bayesian bootstrap credible intervals, and the numbers in brackets in the columns "PB" correspond to 95% pigeonhole bootstrap confidence intervals.

	Based on method in paper	Bayesian bootstrap	Pigeonhole bootstrap
Waugh (2010), all countries, $n = 43$	0.498	0.979	0.986
Waugh (2010), only OECD, $n = 19$	0.533	0.954	0.984
Waugh (2010), only non-OECD, $n = 24$	0.416	0.913	0.941
Caliendo and Parro (2015), θ^1, $n = 15$	0.295	0.911	0.991
Caliendo and Parro (2015), θ^2, $n = 13$	0.467	0.933	0.996
Caliendo and Parro (2015), θ^3, $n = 15$	0.360	0.950	0.999
Caliendo and Parro (2015), θ^4, $n = 14$	0.377	0.932	1.000

Table 11: Coverage for approach based on the method in paper, Bayesian bootstrap and pigeonhole bootstrap using the pigeonhole bootstrap DGP from Algorithm 4, using 1000 simulated datasets and $B = 1000$. For both Waugh (2010) and Caliendo and Parro (2015), I use heteroskedastic-robust standard errors.

	Mean	Standard deviation
Artuç et al (2010): weight matrix $\Omega^{\text{ACM}}(\cdot)$, exchangeability across (ij,t), GMM asymptotics	6.56 [3.07, 10.06]	1.88 [1.05, 2.72]
Intermediate case 1: weight matrix $\Omega^{\text{ACM}}(\cdot)$, exchangeability across (ij,t), Bayesian bootstrap	6.56 [4.02, 13.79]	1.88 [1.25, 4.13]
Intermediate case 2: weight matrix $\Omega^{\text{ACM}}(\cdot)$, exchangeability across industries, Bayesian bootstrap	6.56 [4.49, 10.09]	1.88 [1.35, 2.84]
Preferred approach: centered weight matrix, exchangeability across industries, Bayesian bootstrap	5.98 [4.31, 10.13]	1.93 [1.35, 3.04]

Table 12: Uncertainty quantification for Panel IV in Table 3 in Artuç, Chaudhuri, and McLaren (2010) for $\zeta = 0.97$. For the first row the numbers in brackets correspond to 95% confidence intervals. For the other rows, the numbers in brackets correspond to 95% Bayesian bootstrap credible intervals.

	Change in employment share	Change in wage	Change in utility
Delta method: weight matrix $\Omega^{\text{ACM}}(\cdot)$, exchangeability across (ij,t), GMM asymptotics	-0.088 [-0.113, -0.063]	-0.026 [-0.085, 0.032]	1.27 [1.06, 1.49]
Asymptotic bootstrap: weight matrix $\Omega^{\text{ACM}}(\cdot)$, exchangeability across (ij,t), GMM asymptotics	-0.088 [-0.109, -0.062]	-0.026 [-0.091, 0.024]	1.27 [1.07, 1.51]
Intermediate case 1: weight matrix $\Omega^{\text{ACM}}(\cdot)$, exchangeability across (ij,t), Bayesian bootstrap	-0.088 [-0.108, -0.050]	-0.026 [-0.122, 0.019]	1.27 [0.90, 1.46]
Intermediate case 2: weight matrix $\Omega^{\text{ACM}}(\cdot)$, exchangeability across industries, Bayesian bootstrap	-0.088 [-0.098, -0.064]	-0.026 [-0.085, -0.003]	1.27 [1.06, 1.42]
Preferred approach: centered weight matrix, exchangeability across industries, Bayesian bootstrap	-0.078 [-0.094, -0.059]	-0.049 [-0.099, -0.013]	1.25 [1.03, 1.40]

Table 13: Uncertainty quantification for relevant economic quantities from Figures 3, 4 and 5 in Artuç, Chaudhuri, and McLaren (2010) for $\zeta = 0.97$. For the first row the numbers in brackets correspond to 95% confidence intervals. For the other rows, the numbers in brackets correspond to 95% Bayesian bootstrap credible intervals.

	Point estimate	95% confidence interval based on Caliendo and Parro (2015)	95% Bayesian bootstrap credible interval	95% pigeonhole bootstrap confidence interval
Agriculture, $n = 15$	9.11	[5.17, 13.05]	[-4.05, 25.63]	[-7.98, 30.19]
Mining, $n = 13$	13.53	[6.34, 20.73]	[0.69, 42.35]	[-0.89, 71.13]
Food, $n = 15$	2.62	[1.43, 3.81]	[-1.26, 6.83]	[-3.88, 8.17]
Textile, $n = 14$	8.10	[5.58, 10.61]	[0.52, 16.76]	[-2.43, 18.95]
Wood, $n = 12$	11.50	[5.87, 17.12]	[-11.30, 22.88]	[-30.90, 31.29]
Paper, $n = 14$	16.52	[11.33, 21.71]	[1.70, 31.32]	[-7.18, 39.60]
Petroleum, $n = 12$	64.44	[33.84, 95.04]	[-6.41, 128.87]	[-30.93, 128.88]
Chemicals, $n = 14$	3.13	[-0.37, 6.62]	[-8.49, 13.72]	[-11.59, 17.50]
Plastic, $n = 13$	1.67	[-2.69, 6.03]	[-12.65, 14.01]	[-20.33, 18.74]
Minerals, $n = 14$	2.41	[-0.72, 5.55]	[-3.17, 9.47]	[-5.99, 13.42]
Basic Metals, $n = 14$	3.28	[-1.64, 8.19]	[-11.32, 15.91]	[-15.76, 20.16]
Metal products, $n = 14$	6.99	[2.82, 11.15]	[-5.75, 19.46]	[-11.41, 25.91]
Machinery, $n = 14$	1.45	[-4.04, 6.93]	[-12.75, 17.24]	[-22.56, 26.82]
Office, $n = 14$	12.95	[4.07, 21.83]	[-7.71, 36.25]	[-14.35, 45.52]
Electrical, $n = 14$	12.91	[9.70, 16.12]	[0.20, 21.37]	[-5.35, 25.43]
Communication, $n = 11$	3.95	[0.48, 7.43]	[-5.25, 10.98]	[-10.96, 14.68]
Medical, $n = 14$	8.71	[5.65, 11.78]	[-0.66, 26.37]	[-10.96, 14.68]
Auto, $n = 12$	1.84	[0.04, 3.64]	[-3.80, 5.48]	[-46.82, 10.08]
Other Transport, $n = 14$	0.39	[-1.73, 2.51]	[-5.84, 5.67]	[-13.30, 10.14]
Other, $n = 13$	3.98	[1.86, 6.11]	[-2.11, 9.68]	[-6.80, 11.83]

Table 14: Uncertainty quantification for the benchmark estimates (which remove the countries with the lowest 1% share of trade for each sector) in Table 1 of Caliendo and Parro (2015).

	Point estimate	95% Bayesian bootstrap credible interval	95% pigeonhole bootstrap confidence interval
Mexico	1.31%	[0.65%, 2.51%]	[0.68%, 3.38%]
Canada	-0.06%	[-0.10%, -0.02%]	[-0.10%, -0.01%]
U.S.	0.08%	[0.07%, 0.11%]	[0.07%, 0.13%]

Table 15: Uncertainty quantification for welfare effects as in Table 2 of Caliendo and Parro (2015).

	Point estimate	Analytic, clustering on dyads	Bayesian bootstrap	Pigeonhole bootstrap	Analytic, Graham (2020a)
Constant	1.22	[-2.58, 5.02]	[-5.12, 8.34]	[-5.77, 9.69]	[-5.99, 8.43]
Exporter GDP	0.90	[0.76, 1.05]	[0.63, 1.16]	[0.59, 1.19]	[0.65, 1.16]
Importer GDP	0.89	[0.76, 1.02]	[0.63, 1.14]	[0.58, 1.18]	[0.63, 1.16]
Distance	-0.57	[-0.76, -0.38]	[-0.97, -0.21]	[-1.08, -0.17]	[-1.00, -0.14]

Table 16: Uncertainty quantification when regressing bilateral trade flows on a constant, log exporter and importer GDP, and log distance using PPML. "Analytic, clustering on dyads" corresponds to a 95% analytic confidence interval clustering on dyads, "Bayesian bootstrap" corresponds to the 95% Bayesian bootstrap credible interval, "Pigeonhole bootstrap" corresponds to 95% pigeonhole bootstrap confidence interval, and "Analytic,Graham (2020a)" correspond to the 95% analytic confidence interval based on Graham (2020a).

	Analytic, clustering on dyads	Bayesian bootstrap	Pigeonhole bootstrap	Analytic, Graham (2020a)	Success rate
$\rho = 0.00$, $n = 5$	0.67 [0.54]	0.82 [0.75]	0.99 [36.73]	0.60 [0.43]	0.32
$\rho = 0.00$, $n = 10$	0.77 [0.33]	0.94 [0.45]	1.00 [0.77]	0.69 [0.27]	0.47
$\rho = 0.00$, $n = 25$	0.79 [0.21]	0.94 [0.28]	0.99 [0.35]	0.77 [0.20]	0.78
$\rho = 0.00$, $n = 50$	0.90 [0.14]	0.96 [0.20]	0.99 [0.22]	0.85 [0.13]	0.95
$\rho = 0.00$, $n = 100$	0.89 [0.10]	0.97 [0.14]	0.99 [0.15]	0.89 [0.09]	0.99
$\rho = 0.25$, $n = 5$	0.67 [0.50]	0.83 [0.71]	1.00 [31.78]	0.51 [0.37]	0.33
$\rho = 0.25$, $n = 10$	0.75 [0.30]	0.92 [0.41]	1.00 [0.69]	0.68 [0.24]	0.61
$\rho = 0.25$, $n = 25$	0.78 [0.18]	0.94 [0.24]	0.98 [0.29]	0.73 [0.17]	0.93
$\rho = 0.25$, $n = 50$	0.87 [0.12]	0.97 [0.16]	1.00 [0.18]	0.83 [0.11]	0.99
$\rho = 0.25$, $n = 100$	0.91 [0.08]	0.97 [0.11]	0.99 [0.12]	0.91 [0.07]	1.00
$\rho = 0.50$, $n = 5$	0.63 [0.44]	0.82 [0.60]	0.98 [19.49]	0.55 [0.33]	0.29
$\rho = 0.50$, $n = 10$	0.75 [0.27]	0.96 [0.36]	1.00 [0.62]	0.66 [0.24]	0.65
$\rho = 0.50$, $n = 25$	0.84 [0.16]	0.97 [0.21]	1.00 [0.26]	0.75 [0.14]	0.94
$\rho = 0.50$, $n = 50$	0.87 [0.10]	0.97 [0.14]	0.98 [0.16]	0.86 [0.10]	0.98
$\rho = 0.50$, $n = 100$	0.90 [0.06]	0.98 [0.09]	0.99 [0.10]	0.88 [0.06]	0.98

Table 17: Monte Carlo performance for the log-distance coefficient in the gravity PPML design based on Assumption 3. Each entry reports empirical coverage with mean interval length in brackets. The columns correspond to "Analytic, clustering on dyads", "Bayesian bootstrap", "Pigeonhole bootstrap", and "Analytic, Graham (2020a)". The design fixes $\sigma = 0.50$ and varies ρ over $\{0, 0.25, 0.5\}$, where ρ is the conditional correlation between log trade flows that share an exporter or importer. The final column reports the share of Monte Carlo replications for which "Analytic, Graham (2020a)" returned a finite interval. Results are based on $M = 200$ Monte Carlo replications and $B = 200$ bootstrap draws.

L Appendix Figures

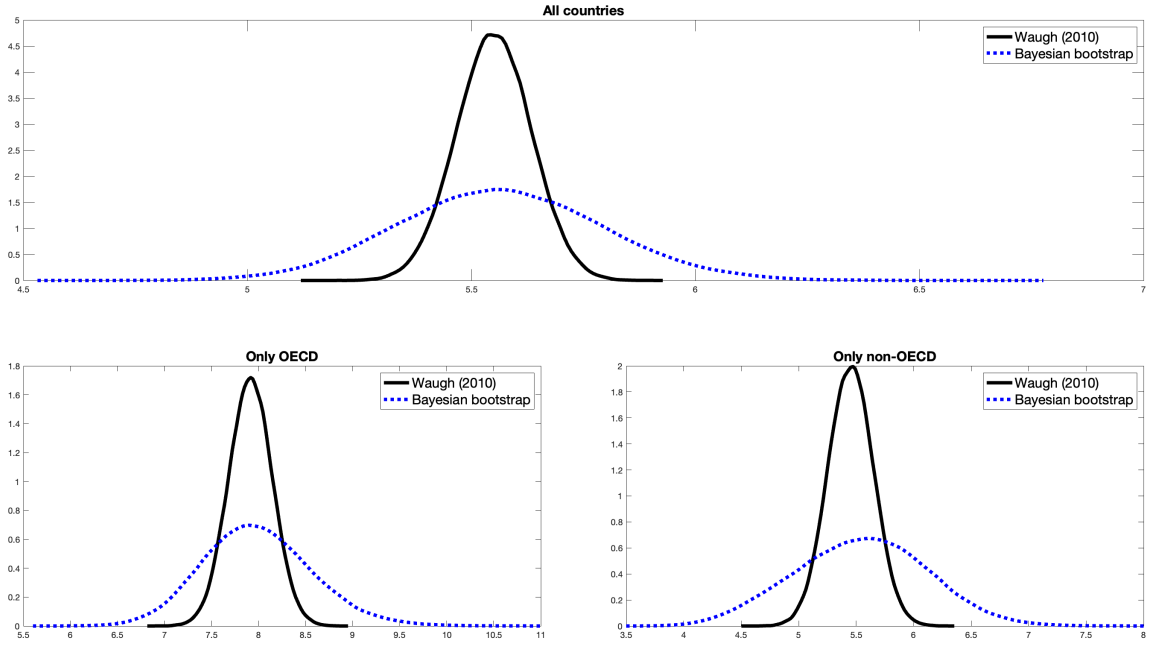

Figure 6: Distributions for productivity parameters in Waugh (2010). "Waugh (2010)" corresponds to the normal approximation as implied by the standard error reported in Waugh (2010), and "Bayesian bootstrap" corresponds to the smoothed Bayesian bootstrap distribution.

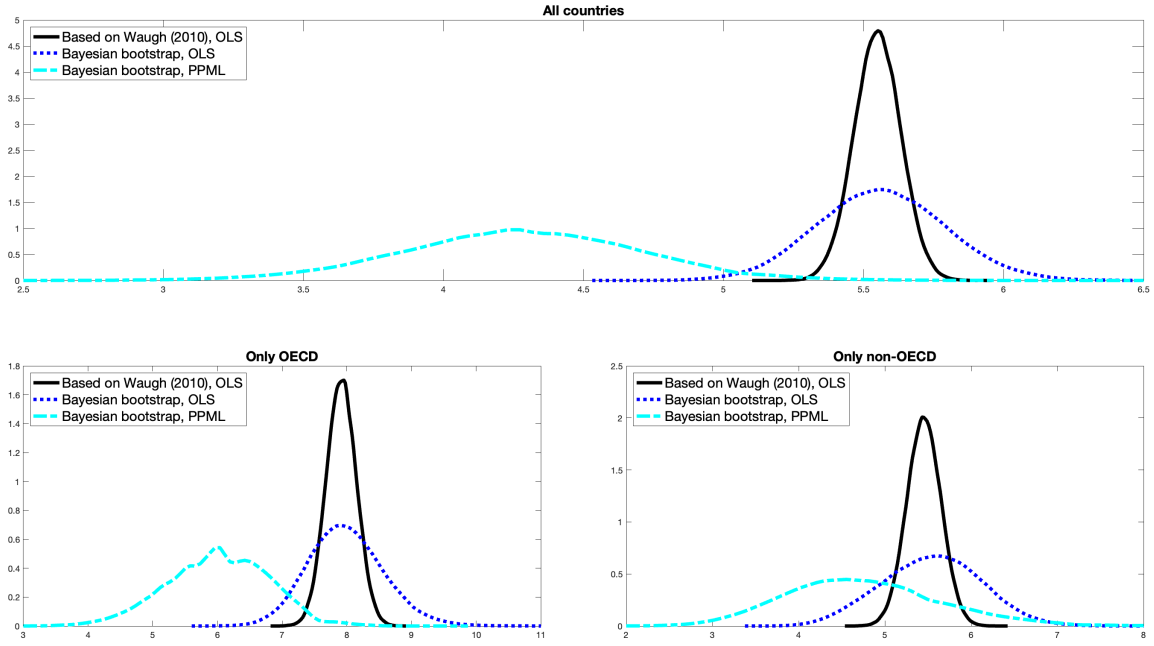

Figure 7: Distributions for productivity parameters in Waugh (2010). "Waugh (2010), OLS" corresponds to the normal approximation as implied by the standard error reported in Waugh (2010) using OLS, "Bayesian bootstrap, OLS" corresponds to the smoothed Bayesian bootstrap distribution using OLS, and "Bayesian bootstrap, PPML" corresponds to the smoothed Bayesian bootstrap distribution using PPML.

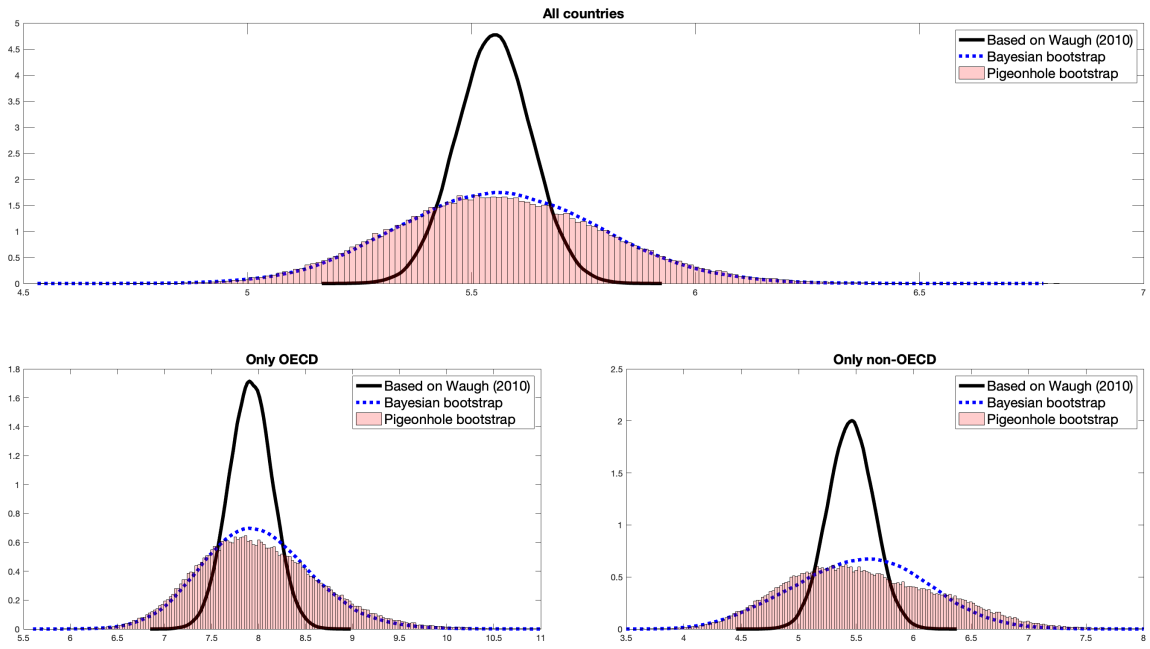

Figure 8: Distributions for productivity parameters in Waugh (2010). "Waugh (2010)" corresponds to the normal approximation as implied by the standard error reported in Waugh (2010), "Bayesian bootstrap" corresponds to the smoothed Bayesian bootstrap distribution, and "Pigeonhole bootstrap" corresponds to the pigeonhole bootstrap distribution.

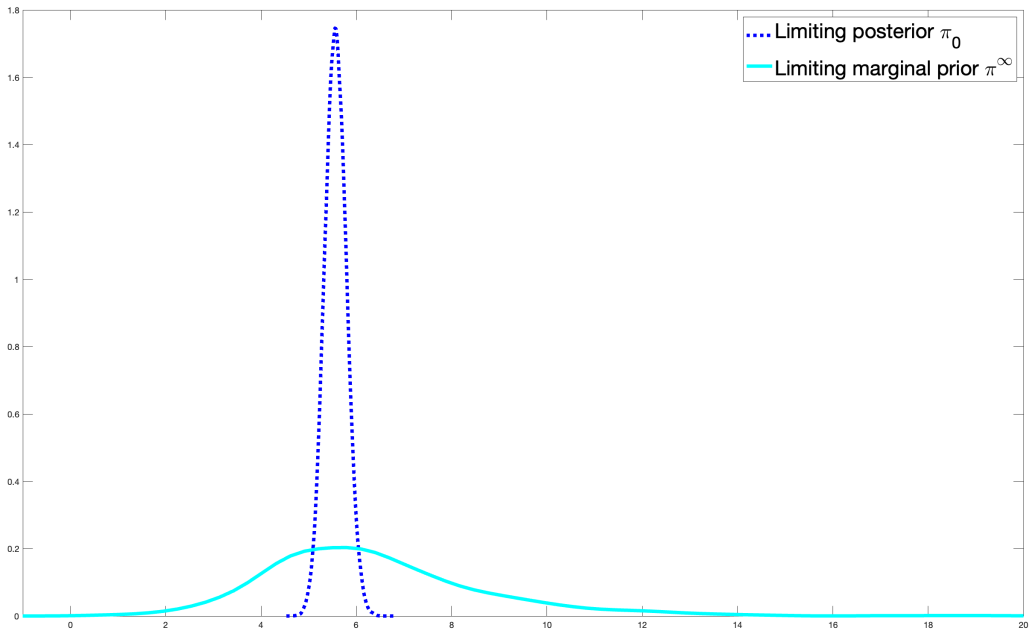

Figure 9: Smoothed limiting posterior and marginal prior for productivity parameter in Waugh (2010).

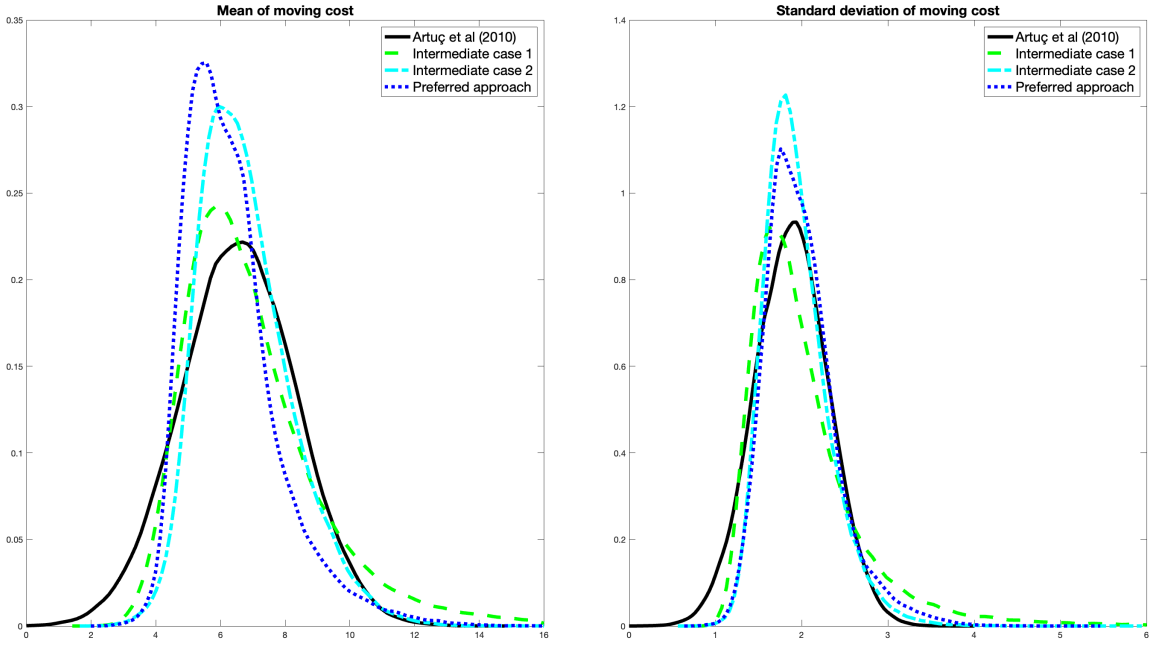

Figure 10: Distributions for estimators in Panel IV in Table 3 in Artuç, Chaudhuri, and McLaren (2010) for $\zeta = 0.97$. "Artuç et al (2010)" corresponds to the normal approximation based on the analytic approach using the weight matrix $\Omega^{\text{ACM}}(\cdot)$ from Algorithm 6 and assuming exchangeability across all observations, "Intermediate case 1" corresponds to the smoothed Bayesian bootstrap posterior using the weight matrix $\Omega^{\text{ACM}}(\cdot)$ from Algorithm 6 and assuming exchangeability across all observations, "Intermediate case 2" corresponds to the smoothed Bayesian bootstrap posterior using the weight matrix $\Omega^{\text{ACM}}(\cdot)$ from Algorithm 6 and assuming exchangeability across industries, and "Preferred approach" corresponds to the smoothed Bayesian bootstrap posterior using a centered weight matrix and assuming exchangeability across industries.

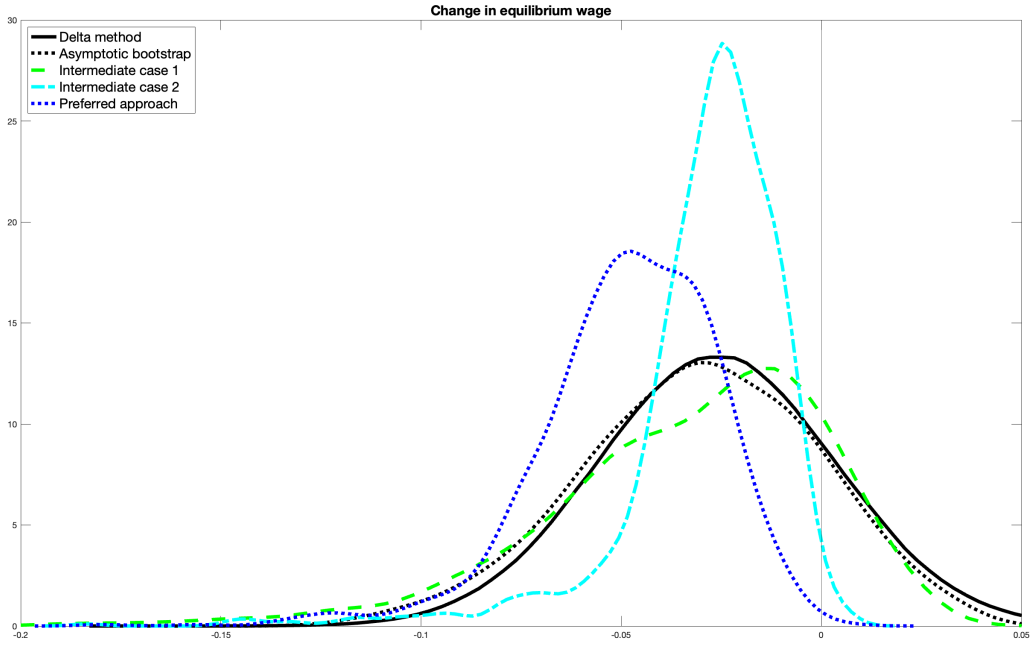

Figure 11: Smoothed Bayesian bootstrap posterior distribution for the change in wages based on Figure 4 in Artuç, Chaudhuri, and McLaren (2010) for $\zeta = 0.97$. "Delta method" and "Asymptotic bootstrap" use the normal approximation based on the analytic approach using the weight matrix $\Omega^{\text{ACM}}(\cdot)$ from Algorithm 6 and assuming exchangeability across all observations, "Intermediate case 1" corresponds to the smoothed Bayesian bootstrap posterior using the weight matrix $\Omega^{\text{ACM}}(\cdot)$ from Algorithm 6 and assuming exchangeability across all observations, "Intermediate case 2" corresponds to the smoothed Bayesian bootstrap posterior using the weight matrix $\Omega^{\text{ACM}}(\cdot)$ from Algorithm 6 and assuming exchangeability across industries, and "Preferred approach" corresponds to the smoothed Bayesian bootstrap posterior using a centered weight matrix and assuming exchangeability across industries.

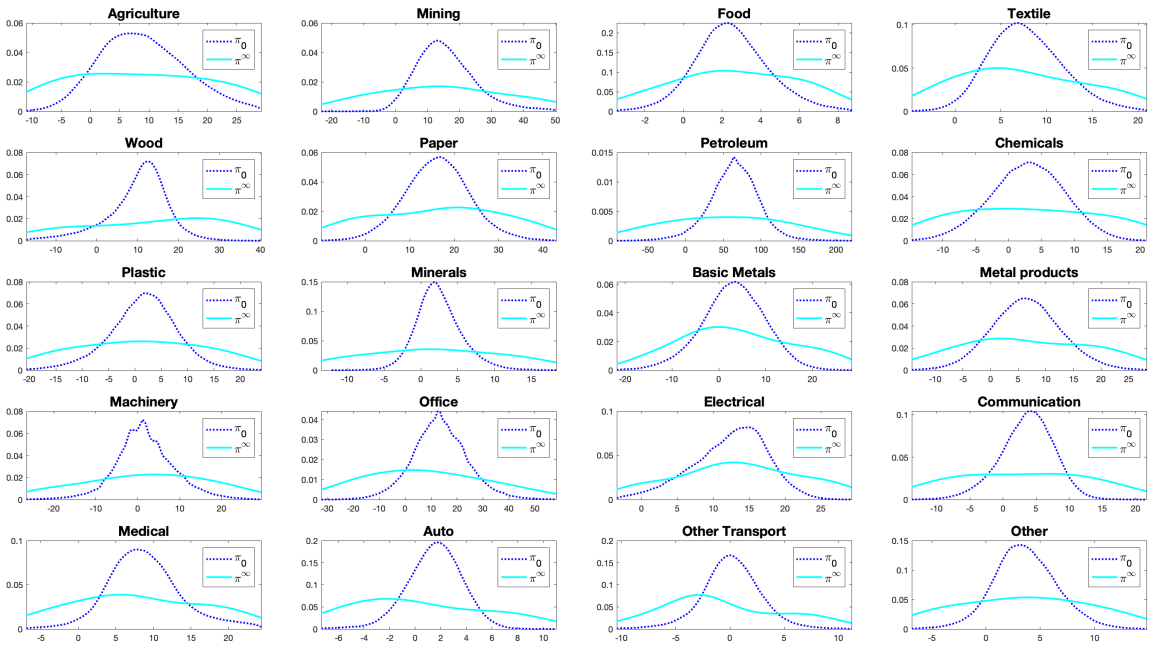

Figure 12: Smoothed limiting posterior (π_0) and marginal prior (π^∞) for trade elasticities as in Caliendo and Parro (2015).

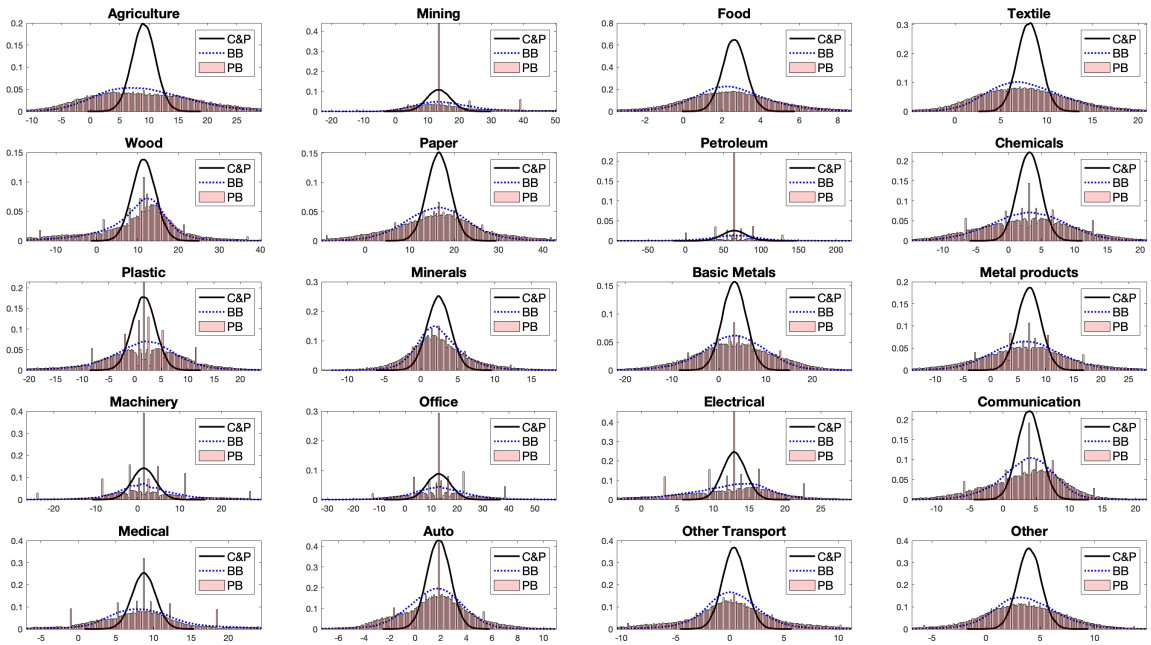

Figure 13: Distributions for the benchmark estimates (which remove the countries with the lowest 1% share of trade for each sector) in Table 1 of Caliendo and Parro (2015). "C&P" corresponds to the normal approximation as implied by the standard errors reported in Caliendo and Parro (2015), "BB" corresponds to the smoothed Bayesian bootstrap posterior, and "PB" corresponds to the pigeonhole bootstrap distribution.

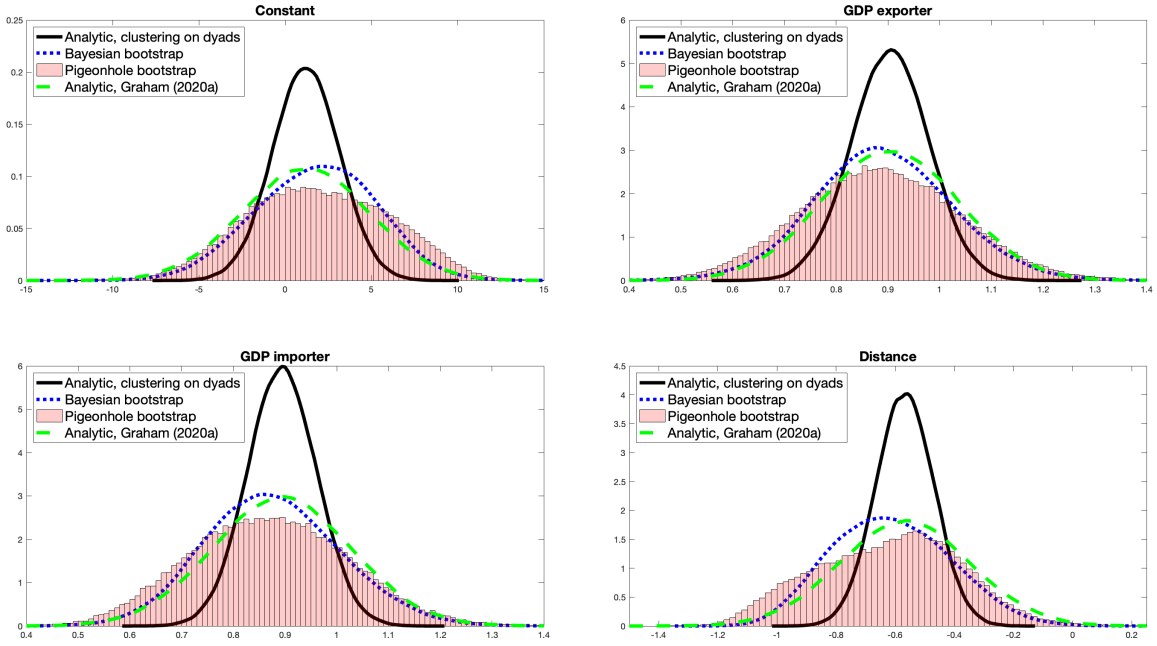

Figure 14: Distributions when regressing bilateral trade flows on a constant, log exporter and importer GDP, and log distance using PPML. "Analytic, clustering on dyads" corresponds to the normal approximation based on the standard errors using clustering on dyads, "Bayesian bootstrap" corresponds to the smoothed Bayesian bootstrap posterior distribution, "Pigeonhole bootstrap" corresponds to the pigeonhole bootstrap distribution, and "Analytic, Graham (2020a)" correspond to the normal approximation based on the standard errors using Graham (2020a).