A New Bayesian Bootstrap for Quantitative Trade and Spatial Models

Bas Sanders, *Harvard University**

May 2025

Abstract

Economists use quantitative trade and spatial models to make counterfactual predictions. Because such predictions often inform policy decisions, it is important to communicate the uncertainty surrounding them. Three key challenges arise in this setting: the data are dyadic and exhibit complex dependence; the number of interacting units is typically small; and counterfactual predictions depend on the data in two distinct ways—through the estimation of structural parameters and through their role as inputs into the model's counterfactual equilibrium. I address these challenges by proposing a new Bayesian bootstrap procedure tailored to this context. The method is simple to implement and provides both finite-sample Bayesian and asymptotic frequentist guarantees. Revisiting the results in Waugh (2010), Caliendo and Parro (2015), and Artuç, Chaudhuri, and McLaren (2010) illustrates the practical advantages of the approach.

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

*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, Marc Melitz, Neil Shephard, Rahul Singh, Elie Tamer, Davide Viviano and Chris Walker for helpful discussions. I am also grateful for comments from participants of the Harvard Graduate Student Workshops in Econometrics and Trade. All errors are mine.

predictions often inform policy decisions, it is important to communicate the uncertainty surrounding them. However, in practice, counterfactuals are often reported without any measure of uncertainty. For instance, in a survey 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—together with the observed data—is used 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, and the second step combines the estimated elasticity 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 (Sanders, 2023).

To address these challenges and support more informed policy decisions, I propose a Bayesian approach. 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 method amounts to reweighting the data using products of draws from an exponential distribution. It readily extends to settings where only a 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). Moreover, because the approach is Bayesian, uncertainty quantification for the counterfactual prediction follows automatically from the posterior distribution of the structural parameter.

The key theoretical contribution of this paper is to introduce and justify a new Bayesian bootstrap procedure tailored to polyadic data structures. The method extends the classical

[1] The survey includes all papers published between 2015 and 2024 in five general-interest economics journals (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.

Bayesian bootstrap (Rubin, 1981; Chamberlain and Imbens, 2003) to settings where each observation involves more than a single unit. A central result is that this procedure admits a finite-sample Bayesian interpretation. Specifically, under a particular choice of model and prior, the posterior distribution converges to the Bayesian bootstrap distribution as a prior informativeness parameter tends to zero. The model assumes that the polyadic data are generated as functions of unit-specific latent variables, which are drawn independently from a common distribution. A Dirichlet process prior is placed on this distribution, and the Bayesian bootstrap distribution emerges as the limiting posterior when the prior becomes uninformative. In the main text, I provide several additional motivations for the model and prior underlying my results.

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 small samples, posterior distributions are often non-Gaussian, and asymmetries in their shape can have important policy implications: for instance, right-skewness could suggest greater potential for large welfare gains, while left-skewness indicates a chance of substantial welfare losses. Capturing this asymmetry is critical for informed decision-making.

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—which are reviewed later in the introduction—that rely exclusively on asymptotic approximations. For frequentist uncertainty quantification of the counterfactual prediction, I provide a delta method-type result that accounts for the fact that counterfactual predictions 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. 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 the economic conclusions are robust.

To further illustrate the flexibility of the method, 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 regions of the parameter space where model assumptions are violated. 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. Skewness in the posterior distributions of the elasticities induces asymmetry in the posterior distribution of welfare changes, shifting probability mass away from zero. The credible intervals around welfare predictions reflect substantial uncertainty and considerable heterogeneity across countries, although the ranking of welfare effects 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 with a posterior probability of 25%—a finding not visible from point estimates alone.

The method I propose substantially improves how uncertainty is quantified for both the structural parameter estimator 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, clustering either on dyads or only on the origin or destination unit. A more flexible alternative, used in several papers, is two-way clustering on both the origin and destination units. While this 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. 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 method to two alternatives from the existing literature. The closest 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. 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. Another difference is that theoretical guarantees for the pigeonhole bootstrap rely on asymptotic approximations that assume a large number of interacting units. However, the applications I study involve small numbers of units. In these settings, I show that the pigeonhole bootstrap can be numerically unstable and tends to produce wider confidence intervals than the Bayesian bootstrap procedure. By contrast, in settings with a large number of interacting units, the approaches are equivalent.

A second alternative for uncertainty quantification is to derive frequentist standard errors. Graham (2020a,b) build 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, which may perform poorly when the number of interacting units is small. Again, when the number of interacting units is large, using analytic standard errors is equivalent to using my approach.

Both Davezies, D'haultfœuille, and Guyonvarch (2021) and Graham (2020a,b) only focus on uncertainty quantification for the estimator of the structural parameter.[3] Since the 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, the Bayesian bootstrap procedure provides valid finite-sample

[2] To date, among all papers citing this literature, 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.

uncertainty quantification for counterfactual predictions, supporting better-informed policy decisions in small-sample settings.

This paper contributes to several literatures. First, it adds to a 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, it advances research on bootstrap methods designed for settings in which standard resampling approaches fail (Janssen, 1994; Davezies, D'haultfœuille, and Guyonvarch, 2021; Menzel, 2021). Third, it engages with the emerging literature on uncertainty quantification in polyadic settings (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. The next section introduces the setting and the proposed Bayesian bootstrap procedure. Sections 3 and 4 present the main theoretical contributions, focusing on finite-sample Bayesian results and asymptotic validity, respectively. Section 5 discusses several extensions of the core framework. The method is then applied in Section 6 to the empirical settings studied in Caliendo and Parro (2015) and Artuç, Chaudhuri, and McLaren (2010). Section 7 compares the proposed approach to alternative methods for uncertainty quantification. Section 8 concludes.

2 Setting and Proposed Method

In this section I introduce the setting and goal of the paper. I lay out my proposed method 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 3 and 4.

2.1 Setting and Goal

2.1.1 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 $k \neq \ell$.[4] 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 in country k. $\triangle$

2.1.2 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 assigns mass $\frac{1}{n(n-1)}$ to each observation $X_{k\ell}$ and because there are $n(n-1)$ observations this is a valid distribution.

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[5]:

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:

$$T_{\text{average}}\left(\mathbb{P}_{n,X_{ij}}\right) = \mathbb{E}_{\mathbb{P}_{n,X_{ij}}}[X_{ij}] = \sum_{k \neq \ell} \frac{1}{n(n-1)} \cdot X_{k\ell}$$

$$T_{\text{regression}}\left(\mathbb{P}_{n,(F_{ij},R_{ij})}\right) = \arg\min_{\vartheta \in \Theta} \mathbb{E}_{\mathbb{P}_{n,\left(F_{ij},R_{ij}\right)}}\left[(F_{ij} - \vartheta R_{ij})^2\right] = \frac{\sum_{k \neq \ell} F_{k\ell} R_{k\ell}}{\sum_{k \neq \ell} R_{k\ell}^2}$$

$$T_{\text{GMM}}\left(\mathbb{P}_{n,X_{ij}}\right) = \arg\min_{\vartheta \in \Theta} \mathbb{E}_{\mathbb{P}_{n,X_{ij}}}[\psi(X_{ij}; \vartheta)]' \, \Omega \mathbb{E}_{\mathbb{P}_{n,X_{ij}}}[\psi(X_{ij}; \vartheta)].$$

[4] 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 2.2.1.

[5] For ease of exposition I assume $\hat{\theta}$ is a scalar, but the same arguments apply to vector-valued $\hat{\theta}$.

Estimators that are not covered by Assumption 1 are estimators that involve multiple flows, such as the average flow over triads of units in a graph.

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

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

The estimator $\hat{\theta}$ indeed 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

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

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

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

for any permutation $\sigma : \{1,...,n\} \to \{1,...,n\}$.[6] Joint exchangeability of the data implies that the elements of $\left\{X_{k\ell}\right\}_{k\neq\ell}$ have a common *marginal* probability distribution, which I will

[6]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.

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 this 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

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

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 there is country-level uncertainty due to country-level productivity shocks. The regression equation will only recover the true structural parameter θ_M under the assumption that the productivity shocks are exogenous and follow a Fréchet distribution. Nevertheless, Waugh (2010) conducts estimation using (3), so going forward, I focus on the estimand θ rather than θ_M. $\triangle$

2.1.3 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 $\hat{\theta}$:

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 a function only of the distribution of the data, rather than the actual observations. The dependence of γ on both the realized bilateral data and their distribution creates a non-classical setting, as was noted in Sanders (2023).

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

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

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 levels of specific countries or summary statistics across the wage vector.

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 ($\tau_{ij}^{\mathrm{cf}} = \infty$ for all $i \neq j$), symmetry ($\tau_{ij}^{\mathrm{cf}} = \min\{\tau_{ij}, \tau_{ji}\}$ for all $i \neq j$) and free trade ($\tau_{ij}^{\mathrm{cf}} = 1$ for 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 2.2, I present a Bayesian bootstrap procedure to quantify uncertainty for $\hat{\theta}$. Then, in Section 2.3, I use Assumption 2 to quantify uncertainty for $\hat{\gamma}$.

2.2 Bayesian Uncertainty Quantification for $\hat{\theta}$

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 just requires drawing from an exponential distribution and reweighting the data appropriately.

In Sections 3 and 4 I will provide various theoretical motivations for the Bayesian bootstrap procedure. The key takeaway from Section 3 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

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 $\left\{\hat{\theta}^{*,(1)}, ..., \hat{\theta}^{*,(B)}\right\}$.

prior. In addition to this finite-sample Bayesian motivation, the key takeaway from Section 4 is that the bootstrap procedure is also asymptotically valid in a frequentist sense.

Example (Waugh, 2010)**.** The productivity parameter in Waugh (2010) is estimated for the full sample and for the two subsets of OECD countries and non-OECD countries. Using Algorithm 1, we can obtain draws from the posterior distribution of the structural parameter given the bilateral data. Specifically, for each bootstrap iteration I compute

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

It turns out that in the case of simple OLS, for each bootstrap draw we can just re-weight both the dependent and independent variables by $\sqrt{V_k^{(b)} \cdot V_\ell^{(b)}}$ for all $k \neq \ell$, and compute the corresponding OLS coefficient. Because the bootstrap distribution corresponds to a limiting posterior distribution, we can interpret $\left\{\hat{\theta}^{*,(1)}, ..., \hat{\theta}^{*,(B)}\right\}$ as posterior draws. From these posterior draws, we can obtain a $100(1-\alpha)\%$ *credible interval* by computing the $\alpha/2$ and $1-\alpha/2$ quantiles. The 95% credible intervals are reported in Table 1 and the posterior distributions are 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 $\hat{\Sigma}_\theta$.[7] 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	As in paper	Bayesian bootstrap
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 1: Uncertainty quantification for productivity parameters as in Waugh (2010).

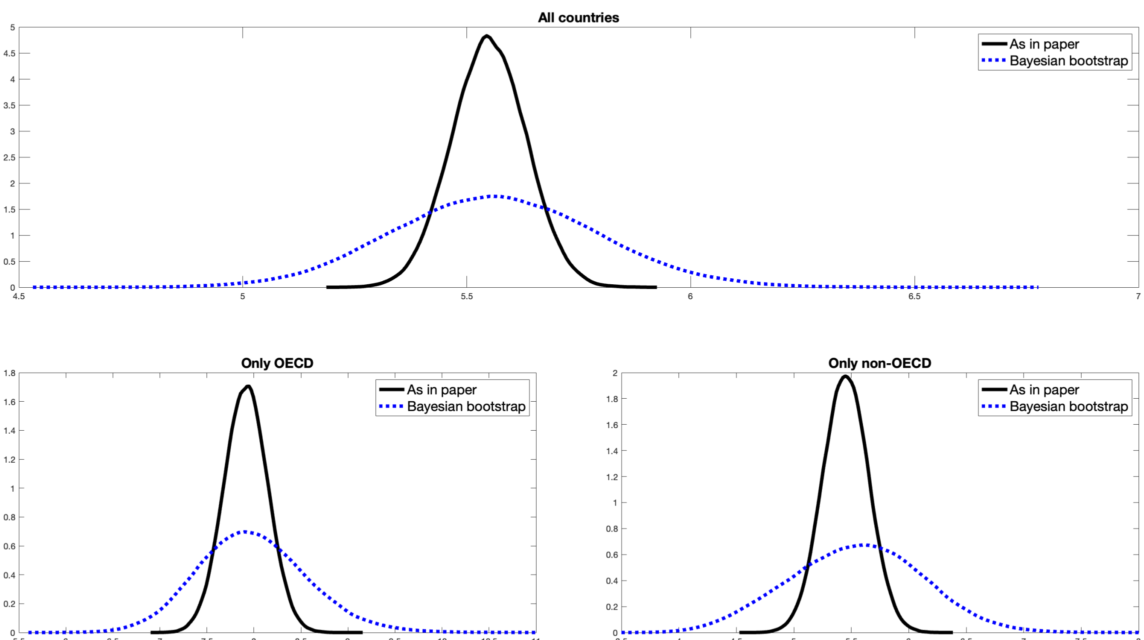

Figure 1: Distributions for productivity parameters as in Waugh (2010).

Table 1 shows that 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 7.1. This setup enables a direct evaluation of the coverage performance of various uncertainty quantification

[7]Formally, this normality could follow from assumptions on the underlying data generating process such that a Bernstein-von Mises type result holds (Van der Vaart, 2000). In that case the influence of the prior distribution on θ becomes negligible and the posterior distribution approximately equals a normal distribution centered at the maximum likelihood estimator.

methods. Table 2 shows that using heteroskedasticity-robust standard errors leads to under-coverage. $\triangle$

	As in paper	Bayesian bootstrap
All countries, $n = 43$	0.498	0.979
Only OECD, $n = 19$	0.533	0.954
Only non-OECD, $n = 24$	0.416	0.913

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

2.2.1 Special Case: Misspecification-Robust Uncertainty Quantification for GMM

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

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

where $\psi : \mathcal{X} \to \mathbb{R}^L$ with $L \geq K$, $X_{ij} \in \mathcal{X}$, $\theta \in \Theta \subseteq \mathbb{R}^K$ 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, $\vartheta \in \mathbb{R}^{d_\theta}$ is a vector of parameters and $R_{ij} \in \mathbb{R}^{d_\theta}$ is a vector of regressors.

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:

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

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

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:[8]

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

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

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

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

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 4.1.3. 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

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

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

[8] I follow Lee (2014) and use the centered weight matrix, rather than the uncentered version $\hat{\Omega}^{\mathrm{uncentered}}(\vartheta) = \left(\mathbb{E}_{\mathbb{P}_{n,X_{ij}}}\left[\psi(X_{k\ell};\vartheta)\,\psi(X_{k\ell};\vartheta)'\right]\right)^{-1}$. 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

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

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

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

In Appendix A.2 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. △

2.3 Bayesian Uncertainty Quantification for $\hat{\gamma}$

Taking a Bayesian perspective on uncertainty quantification, in Section 3 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 $\{\hat{\gamma}^{*,(1)}, ..., \hat{\gamma}^{*,(B)}\}$.

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. △

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

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

Table 3: Bayesian uncertainty quantification for counterfactual predictions as in Waugh (2010).

3 Theory— Finite-Sample Bayesian Results

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.

3.1 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. I assume the following model:

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

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

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

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

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

3.1.1 Theoretical Motivation for Model

To motivate Assumption 3, first note that it implies joint exchangeability of the data $\{X_{k\ell}\}_{k\neq\ell}$, as discussed in Section 2.1.2. Conversely, starting from joint exchangeability and viewing the realized data as being sampled from a superpopulation, we can use the Aldous-Hoover representation (Aldous, 1981; Hoover, 1979) to motivate Assumption 3.

Specifically, suppose the data $\{X_{k\ell}\}_{k\neq\ell}$ are sampled from the infinite random array $\{X_{ij}\}_{i,j\in\mathbb{N}, i\neq j}$. That is, we sample $\{1, ..., n\}$ from the natural numbers $\mathbb{N}$ and only keep the corresponding rows and columns.

Example (Waugh, 2010)**.** The corresponding thought experiment in the application of Waugh (2010) is that we observe a draw of 43 countries from an infinite superpopulation of countries. Since there are only 195 countries in the world, the existence of an infinite superpopulation might feel unnatural. This is however the underlying thought experiment that is necessary for existing asymptotic justifications (Graham, 2020a,b; Davezies, D'haultfœuille, and Guyonvarch, 2021; Menzel, 2021). △

Since the superpopulation is an infinite random array, we have the following result from Aldous (1981):

Lemma 1 (Theorem 1.4 in Aldous, 1981)**.** *If for every permutation* $\sigma : \mathbb{N} \to \mathbb{N}$ *we have*

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

then there exists another array $\left\{X^*_{ij}\right\}_{i,j\in\mathbb{N},i\neq j}$ *generated according to*

$$X^*_{ij} = \tilde{h}^{AH}\left(U, C_i, C_j, D_{ij}\right), \tag{20}$$

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

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

Here, U is a common "mixture variable" that is unidentifiable (Bickel and Chen, 2009; Graham, 2020a). Conditioning on this random variable yields $h^{AH}\left(C_i, C_j, D_{ij}\right) \equiv \tilde{h}^{AH}\left(U, C_i, C_j, D_{ij}\right)$. The only difference between the models $h\left(C_i, C_j\right)$ and $h^{AH}\left(C_i, C_j, D_{ij}\right)$ is then the idiosyncratic component D_{ij}. Although the Aldous-Hoover representation is more general and can hence generate more distributions for the bilateral data, given observed data $\{X_{k\ell}\}_{k\neq\ell}$ with finite sample size and arbitrarily flexible h, the models are observationally equivalent. That is, we cannot reject $h\left(C_i, C_j\right)$ relative to $h^{AH}\left(C_i, C_j, D_{ij}\right)$ observing only $\{X_{k\ell}\}_{k\neq\ell}$.

Example (Waugh, 2010)**.** Applying Lemma 1 to the setting in Waugh (2010) yields the representation

$$X_{k\ell} = \tilde{h}^{AH}\left(U, C_k, C_\ell, D_{k\ell}\right),$$

for each $k \neq \ell$. Here, U is a latent variable common to all countries, which can be interpreted as capturing economy-wide spillovers. C_k is a latent variable specific to country k, and $D_{k\ell}$ is a latent variable specific to the country pair (k, ℓ). Conditional on U, and given the observed data $\{X_{k\ell}\}_{k\neq\ell}$, we cannot distinguish a model that includes pair-specific latent variables from one that flexibly combines the country-specific latent variables. $\triangle$

3.2 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*

$$(h, \mathbb{P}_C) \sim \pi(h) \cdot \pi(\mathbb{P}_C) = \pi(h) \cdot DP(Q, \alpha). \tag{21}$$

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

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

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 $\alpha \downarrow 0$, the posterior on θ given the realized data $\{X_{k\ell}\}_{k\neq\ell}$ converges in distribution to the distribution of the draws produced by the Bayesian bootstrap procedure in Algorithm 1.*

All proofs can be found in Appendix C. The proof of Theorem 1 proceeds in five steps. First, I find the posterior of $\mathbb{P}_C$ 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 (21) and uses the conjugacy of the Dirichlet process. Second, I find the posterior which corresponds to the case where $\alpha \downarrow 0$, and argue that it is proper. Importantly, it does not depend on the center measure Q. Denoting with π_0 the probability under the limiting posterior as $\alpha \downarrow 0$, we have

$$\pi_0(\mathbb{P}_C|h, \{C_k\}) = DP\left(\sum_{k=1}^{n} \frac{1}{n} \cdot \delta_{C_k}, n\right). \tag{22}$$

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 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 $\pi_0 \left(\mathbb{P}_{X_{ij}} | \{X_{k\ell}\}_{k\neq\ell}\right)$, which I show corresponds to

$$\begin{aligned}
&\mathbb{P}^*_{n,X_{ij}} \sim \pi_0 \left(\mathbb{P}_{X_{ij}} | \{X_{k\ell}\}_{k\neq\ell}\right) \\
&\Rightarrow \mathbb{P}^*_{n,X_{ij}} = \sum_{k\neq\ell} \frac{W_k \cdot W_\ell}{\sum_{s\neq t} W_s \cdot W_t} \cdot \delta_{X_{k\ell}}, \quad (W_1, ..., W_n) \sim \text{Dir}\,(n; 1, ..., 1)\,,
\end{aligned} \tag{23}$$

which is exactly the distribution we saw in Algorithm 1. Lastly, since $\theta = T\left(\mathbb{P}_{X_{ij}}\right)$, this also implies a limiting posterior on the structural parameter, $\pi_0 \left(\theta | \{X_{k\ell}\}_{k\neq\ell}\right)$, and the conclusion 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 $\pi_0 \left(\gamma | \{X_{k\ell}\}_{k\neq\ell}\right)$ converges to the Bayesian bootstrap distribution characterized by Algorithm 1 and Equation (17).

3.2.1 Theoretical Motivation for Dirichlet Process Prior

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 1 (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

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

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 2 (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 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.*[9]

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

[9] 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 $a, b \in \mathcal{C}$ an arbitrary random variable Z with values in $[0, 1]$.

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.

3.2.2 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 $\pi_0\left(\theta \middle| \{X_{k\ell}\}_{k\neq\ell}\right)$. By comparing these two distributions, we can see how much information is drawn from the prior.

However, 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 (21) to

$$\mathbb{P}_C \sim DP\left(\sum_{k=1}^{n} \frac{1}{n} \cdot \delta_{C_k}, \alpha\right). \tag{24}$$

This specific choice of the center measure Q implies that mass is supported only on the latent variables $\{C_k\}$.[10] This is also the case for the posterior in Equation (22), which can be recovered by setting $\alpha = n$. We then have the following result:

Theorem 3 (Limiting marginal prior)**.** *Under Assumptions 3 and using the Dirichlet process prior as in Equation* (24)*, if $\hat{\theta}$ is of the form*

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

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

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

[10] 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 $n \to \infty$.

Theorem 3 shows how to characterize the limiting object 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 6.1, continuity of χ is satisfied. For estimators that can not be written in this way, Appendix D presents an algorithm for plotting proper priors along the limit sequence.

Example (Waugh, 2010)**.** In Figure 2, I plot the bootstrap posterior and the limiting marginal prior using Theorem 3, where we have

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

and continuity of χ is satisfied. We observe that the limiting marginal prior is much flatter than the bootstrap posterior. Its diffuse shape reflects weak prior information, allowing for a wide range of plausible values for the productivity parameter. △

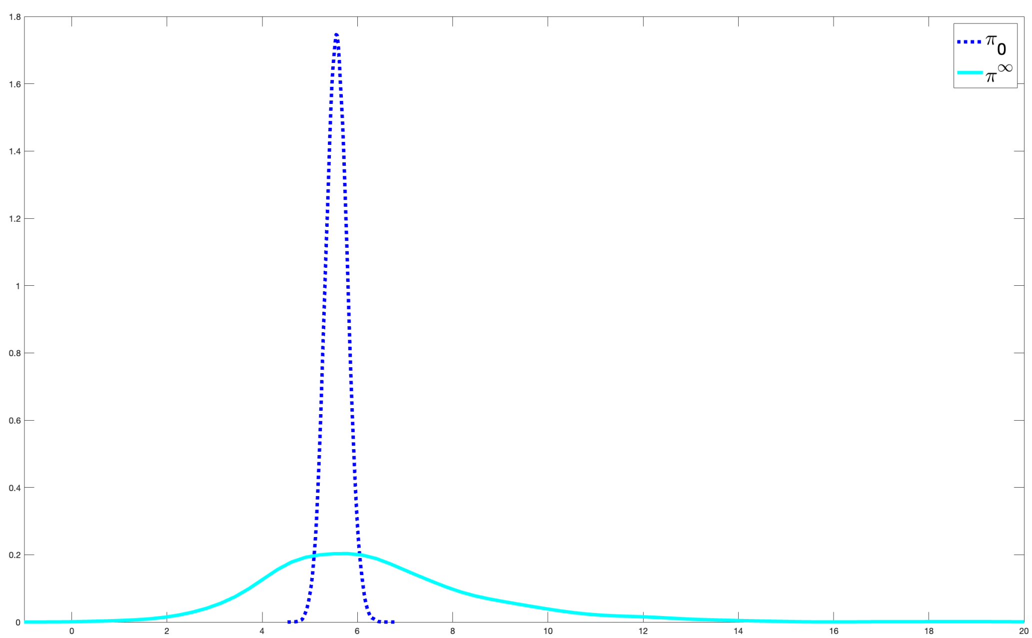

Figure 2: Limiting marginal prior for productivity parameter using the full sample as in Waugh (2010).

4 Theory— Asymptotic Results

In this section I provide conditions on $\hat{\theta}$ that guarantee asymptotic validity of the proposed Bayesian bootstrap procedure. I again consider misspecification-robust uncertainty quantification for over-identified GMM as a special case.

4.1 Frequentist Uncertainty Quantification for $\hat{\theta}$

4.1.1 Sampling Thought Experiment

In Section 3.1.1 I introduced the thought experiment that the data $\{X_{k\ell}\}_{k\neq\ell}$ are sampled from a superpopulation. In this section I will state this as an assumption:

Assumption 5 (Sampling thought experiment)**.** *The infinite random array $\{X_{ij}\}_{i,j\in\mathbb{N},i\neq j}$ is jointly exchangeable, so that for every permutation $\sigma:\mathbb{N}\to\mathbb{N}$ we have*

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

The data $\{X_{k\ell}\}_{k\neq\ell}$ are generated by sampling $\{1,...,n\}$ from the natural numbers $\mathbb{N}$ and only keeping the corresponding rows and columns.

Assumption 5 implies that as we sample more observations from this superpopulation, the resulting data $\{X_{k\ell}\}_{k\neq\ell}$ always will be jointly exchangeable, and that all observations will have the same marginal distribution, denoted by $\mathbb{P}_{X_{ij}}$.

4.1.2 Asymptotic Bootstrap Validity

The goal of this section is to prove asymptotic validity of the bootstrap procedure in Algorithm 1 for a given estimator

$$\hat{\theta} = T\left(\mathbb{P}_{n,X_{ij}}\right) = T\left(\sum_{k\neq\ell}\frac{1}{n\,(n-1)}\cdot\delta_{X_{k\ell}}\right). \tag{25}$$

Going forward, let $\mathbb{P}^*_{n,X_{ij}}$ be a given drawn distribution from $\pi_0\left(\mathbb{P}_{X_{ij}}\middle|\,\{X_{k\ell}\}_{k\neq\ell}\right)$, so that

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

Definition 2 (Asymptotic bootstrap validity). *The bootstrap procedure is* asymptotically valid *for the estimator $\hat{\theta}$ as defined in Equation* (25) *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 mean zero normal random variable.*

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.

To show asymptotic validity of the bootstrap for a structural estimator, I will take a two-step approach. First I show convergence of the empirical process, and then use the functional delta method to argue validity of the bootstrap for certain classes of estimators.

The relevant empirical processes, defined on a class of real-valued functions $\mathcal{F}$, are

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

for $f \in \mathcal{F}$. Here, $\mathbb{P}_{X_{ij}} f$ denotes $\mathbb{E}_{\mathbb{P}_{X_{ij}}}\left[f\left(X_{ij}\right)\right]$, and $\mathbb{P}_{n,X_{ij}} f$ and $\mathbb{P}^*_{n,X_{ij}} f$ are defined analogously. We want to show weak convergence over $\ell^\infty\left(\mathcal{F}\right)$ of both $\mathbb{G}_n$ and $\mathbb{G}^*_n$ to the same 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, for $\ell^\infty\left(\mathcal{F}\right)$ the set of bounded functions on $\mathcal{F}$. A formal definition of weak convergence is given in Definition 1.3.3 in Van Der Vaart and Wellner (1996). To ensure this convergence, we require some regularity conditions on the function class $\mathcal{F}$.

Assumption 6 (Regularity conditions on $\mathcal{F}$). *Let $\mathcal{F} \subseteq \mathcal{X}^{\mathbb{R}}$ be a measurable class of functions such that:*

(i) $\mathcal{F}$ is permissible (see page 196 in Pollard, 1984) and admits a positive envelope F with $\mathbb{P}_{X_{ij}} F^2 < \infty$.

(ii) We have non-degeneracy, meaning that the covariance kernel is positive for all elements of $\mathcal{F}$:

$$K\left(f_1, f_2\right) = \mathrm{Cov}\left(f_1\left(X_{12}\right) + f_1\left(X_{21}\right), f_2\left(X_{12'}\right) + f_2\left(X_{2'1}\right)\right) > 0 \ \forall f_1, f_2 \in \mathcal{F}.$$

(iii) There exist $0 < c, v < \infty$ such that for every $\epsilon > 0$ and probability measure Q with

$QF^2 < \infty$, we have

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

Condition (i) captures regularity condition 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 F for a class $\mathcal{F}$ means that $|f(x)| \leq F(x)$ for all $f \in \mathcal{F}$ and all $x \in \mathcal{X}$.

Non-degeneracy in condition (ii) ensures that the limiting processes of $\mathbb{G}_n$ and $\mathbb{G}_n^*$ are Gaussian with non-zero variance. Degeneracy may arise, for instance, if the data $\{X_{k\ell}\}_{k\neq\ell}$ are in fact i.i.d, in which case the limiting process of $\mathbb{G}_n$ is a Gaussian chaos process. In such settings, $\mathbb{G}_n^*$ will not converge to the correct (non-Gaussian) limit under standard bootstrap procedures. Alternative bootstrap methods have been developed to handle degeneracy, including those proposed by Hušková and Janssen (1993), Menzel (2021) and Han (2022).

Condition (iii) bounds the complexity of $\mathcal{F}$. Here, the covering number $N\left(\epsilon, \mathcal{F}, \left\|\cdot\right\|_{L_2(Q)}\right)$ is the minimal number of $L_2(Q)$-balls of radius ε needed to cover $\mathcal{F}$. This condition is for example satisfied for VC classes of functions by Lemma 4.4 in Alexander (1987).[11]

Example (Smooth functionals of empirical cdf)**.** Consider the class of estimators that are smooth functionals of the empirical cdf $F_{n,X_{ij}}$ and suppose for exposition that X_{ij} is a scalar. For some function φ, we have $\hat{\theta} = \varphi\left(F_{n,X_{ij}}\right)$, $\theta = \varphi\left(F_{X_{ij}}\right)$ and the relevant function class is

$$\mathcal{F}_{\text{cdf}} \equiv \left\{u \mapsto \mathbb{I}\left\{u \leq x\right\} : x \in \mathbb{R}\right\}.$$

As an envelope function we can take the constant function $F_{\text{cdf}} \equiv 1$. The covariance kernel is

$$K_{\text{cdf}}(x, y) = \text{Cov}\left(\mathbb{I}\left\{X_{12} \leq x\right\} + \mathbb{I}\left\{X_{21} \leq x\right\}, \mathbb{I}\left\{X_{12'} \leq y\right\} + \mathbb{I}\left\{X_{2'1} \leq y\right\}\right), \tag{26}$$

which we require to be non-zero for all $x, y \in \mathbb{R}$. Lastly, we know $\mathcal{F}_{\text{cdf}}$ satisfies condition (iii) in Assumption 6 from Example 19.16 in Van der Vaart (2000).

We have the following result for the empirical processes:

[11] Alternatively, one could assume that $\mathcal{F}$ 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).

Theorem 4 (Weak convergence of empirical processes)**.** *If $\mathcal{F}$ 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 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 having a sample size of $n(n-1)$, as is also the case for non-degenerate U-statistics. The proof of Theorem 4 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 we have established convergence of the empirical process, we can appeal to the functional delta method for the bootstrap to argue asymptotic validity of the bootstrap for a given estimator. We require the estimator to be sufficiently smooth:

Assumption 7 (Smoothness)**.** *Suppose $\hat{\theta}$ is of the form $T\left(\mathbb{P}_{n,X_{ij}}\right) = \varphi\left(\mathbb{P}_{n,X_{ij}} f\right)$ for $f \in \mathcal{F}$, where $\varphi : \ell^{\infty}(\mathcal{F}) \mapsto \Theta$ with derivative φ'. The function φ is Hadamard differentiable at $\mathbb{P}_{X_{ij}} f$ tangentially to a subspace $\ell_0^{\infty}(\mathcal{F}) \subset \ell^{\infty}(\mathcal{F})$.*

The precise definition of Hadamard differentiability is given in Section 20.2 of Van der Vaart (2000). Section 20.3 of Van der Vaart (2000) give examples of Hadamard differentiable functions.

Example (Smooth functionals of empirical cdf)**.** Consider again the class of estimators that are smooth functionals of the empirical cdf, so that $\hat{\theta} = \varphi\left(F_{n,X_{ij}}\right)$. For Assumption 7 to hold we require φ to be Hadamard differentiable tangentially to a subspace $\ell_0^{\infty}(\mathcal{F}_{\text{cdf}})$. For example, from Lemma 21.3 in Van der Vaart (2000) we know this is the case for the empirical quantiles under mild differentiability conditions on $F_{X_{ij}}$.

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

Theorem 5 (Bootstrap validity)**.** *Under Assumption 5, if $\mathcal{F}$ and $\hat{\theta}$ satisfy Assumptions 6 and 7, then the bootstrap procedure in Algorithm 1 is asymptotically valid for $\hat{\theta}$.*

From the examples throughout this section, we then have the following corollary:

Corollary 1 (Asymptotic bootstrap validity for smooth functionals of empirical cdf)**.** *The bootstrap procedure in Algorithm 1 is asymptotically valid for estimators of the form $\hat{\theta} = \varphi\left(F_{n,X_{ij}}\right)$ if $K_{\text{cdf}}(x,y)$ in Equation (26) is positive for all $x, y \in \mathbb{R}$ and φ is Hadamard differentiable tangentially to a subspace $\ell_0^{\infty}(\mathcal{F}_{\text{cdf}})$.*

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

Corollary 2 (Asymptotic bootstrap validity for Z-estimators)**.** *Suppose $\hat{\theta}$ and θ solve*

$$0 = \Psi_n(\vartheta) \equiv \sup_{\eta \in \mathcal{H}} \left|\Psi_n(\vartheta)(\eta)\right| = \sup_{\eta \in \mathcal{H}} \left|\mathbb{P}_{n,X_{ij}} \nu_{\vartheta,\eta}\right|$$
$$0 = \Psi(\vartheta) \equiv \sup_{\eta \in \mathcal{H}} \left|\Psi(\vartheta)(\eta)\right| = \sup_{\eta \in \mathcal{H}} \left|\mathbb{P}_{X_{ij}} \nu_{\vartheta,\eta}\right|,$$

and uppose the following conditions hold:

(i) $\Psi : \Theta \mapsto \mathbb{R}^L$ *is uniformly norm-bounded over* Θ*, and satisfies* $\Psi(\theta) = 0$*.*

(ii) Ψ *is Fréchet differentiable at* θ *with continuously invertible derivative* $\dot{\Psi}_\theta$*.*

(iii) $\sup_{\eta \in \mathcal{H}} \left|\Psi(\theta_w)\right| \to 0$ *implies* $\|\theta_w - \theta\| \to 0$ *for every sequence* $\{\theta_w\}$ *in* Θ*.*

(iv) Ψ_n *has at least one zero for all n large enough, outer almost surely (see Section 18.2 in Van der Vaart (2000) for a formal definition).*

(v) The limit of $\vartheta \mapsto \sqrt{n}\left(\Psi_n(\vartheta) - \Psi(\vartheta)\right)$ *is almost surely continuous at* θ*.*

(vi) The function class $\mathcal{F}_Z \equiv \{\nu_{\vartheta,\eta} : (\vartheta, \eta) \in \Theta \times \mathcal{H}\}$ *satisfies Assumption 6.*

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

4.1.3 Special Case: Misspecification-Robust Uncertainty Quantification for GMM

Following Imbens (1997), the two estimation steps of the two-step GMM estimator from Section 2.2.1 can be combined into a single just-identified system,

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

where

$$\mathbb{E}_{\mathbb{P}_{X_{ij}}}\left[\phi\left(X_{ij};\theta^{1-\mathrm{GMM}},\theta^{2-\mathrm{GMM}},m,\mathrm{vec}\left\{\Omega\right\},\mathrm{vec}\left\{G_1\right\},\mathrm{vec}\left\{G_2\right\}\right)\right]=0. \tag{27}$$

Note that the moment equations in Equation (27) hold regardless of whether the moments equations in Equation (14) hold for some $\theta \in \Theta$.[12] Importantly, running this just-identified GMM procedure is numerically equivalent to running the two-step GMM procedure. Since the just-identified GMM estimator is a Z-estimator, we can apply Corollary 2 with

$$\mathcal{H}=\left\{1,...,2LK+2K+L^2\right\}$$
$$\nu_{\vartheta,\eta}\left(X_{ij}\right)=\phi_\eta\left(X_{ij};\vartheta\right),$$

and asymptotic validity of the bootstrap in Algorithm 2 amounts to checking relevant conditions on the moment functions.

Example (Waugh, 2010)**.** Given the moment condition in Equation (15), we should check whether

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

is Fréchet differentiable in ϑ. This is trivially the case because $\psi\left(X_{ij};\cdot\right)$ is linear. The complexity condition (iii) in Assumption 6 is also satisfied for this just-identified case with a single linear moment function. △

4.2 Frequentist Uncertainty Quantification for $\hat{\gamma}$

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

Theorem 6 (Delta method for random object)**.** *Suppose we have* $\sqrt{n}\left(\hat{\theta}-\theta\right)\overset{d}{\approx}\mathcal{N}\left(0,\Sigma\right)$, *and we can consistently estimate its asymptotic variance by* $\hat{\Sigma}$*. Then, for* $G\left(\cdot\right)=\nabla_\theta g\left(\left\{X_{k\ell}\right\}_{k\neq\ell},\cdot\right)$, *if we have*

$$\forall c>0,\ \sup_{\tilde{\theta}:\left\|\tilde{\theta}-\theta\right\|\leq\frac{c}{\sqrt{n}}}\left|G\left(\tilde{\theta}\right)-G\left(\theta\right)\right|\overset{p}{\to}0, \tag{28}$$

[12] Imbens (1997) also shows that iterated GMM estimator (Hansen and Lee, 2021) can be written as a just-identified GMM estimator, but the continuously updated GMM estimator cannot.

a valid confidence interval for γ *is given by*

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

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.

5 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 E I additionally discuss multiway clustering and conditional exchangeability.

5.1 Polyadic data

The data do not necessarily have to be dyadic. For example in Section 6.1 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

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

5.2 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 6.1.

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

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

The priors from Assumption 4 do not change.[13]

6 Applications

In this section I discuss the applications in Caliendo and Parro (2015) and Artuç, Chaudhuri, and McLaren (2010). For both, the number of interacting units is small, which makes the Bayesian bootstrap procedure an appealing approach for uncertainty quantification.

[13] There are cases where we would want to model this differently, for example if we observe a random sample of dyads. In that case 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 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.

6.1 Application 1: Caliendo and Parro (2015)

6.1.1 Parameter Estimation

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

$$\log\left(\frac{F^s_{ij}F^s_{jr}F^s_{ri}}{F^s_{ji}F^s_{rj}F^s_{ir}}\right) = -\theta^s \log\left(\frac{t^s_{ij}t^s_{jr}t^s_{ri}}{t^s_{ji}t^s_{rj}t^s_{ir}}\right) + \varepsilon^s_{ijr},$$

with the identification restriction that the random disturbance term ε^s_{ijr} 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 5, the bootstrap procedure can easily be adapted to this triadic setting, where now for each bootstrap draw we compute

$$\hat{\theta}^{s,*,(b)} = T^s\left(\sum_{(k,\ell,m)\in\mathcal{I}^s} \frac{V^{(b)}_k \cdot V^{(b)}_\ell \cdot V^{(b)}_m}{\sum_{(t,u,v)\in\mathcal{I}^s} V^{(b)}_t \cdot V^{(b)}_u \cdot V^{(b)}_v} \cdot \delta_{X^s_{k\ell m}}\right).$$

Note that I sum over the subset $\mathcal{I}^s \subset \left\{(i,j,r) \in \{1,...,n\}^3 : i \neq j \neq r\right\}$, because in this application observations with $\frac{F^s_{k\ell}F^s_{\ell m}F^s_{mk}}{F^s_{\ell k}F^s_{m\ell}F^s_{km}} = 0$ are dropped. 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 3 plots the corresponding posterior distributions and implied normal distributions. It is alarming that many credible intervals include -1, which violates the model assumption that $\theta^s > -1$ for all sectors s.[14] Appendix B presents a data-calibrated simulation exercise, which highlights that using heteroskedastic-robust standard errors for uncertainty quantification results in under-coverage.

Figure 3 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.

[14] Specifically, the sector-specific Fréchet shape parameter θ^s are assumed to be at least one greater than the within-sector elasticities of substitution, which is assumed to be strictly positive.

	Point estimate	As in paper	Bayesian bootstrap
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).

6.1.2 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 quantification. In Table 5, I reproduce these results and include 95% Bayesian credible intervals. Figure 4 displays the corresponding posterior distributions. Implementation details and additional results are provided in Appendix F.1.

The credible intervals and posterior distributions show asymmetry in the distribution of welfare changes, shifting probability mass away from zero. 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 the credible intervals include zero, the signs of the effects are robust to uncertainty. This is also true for the ranking of

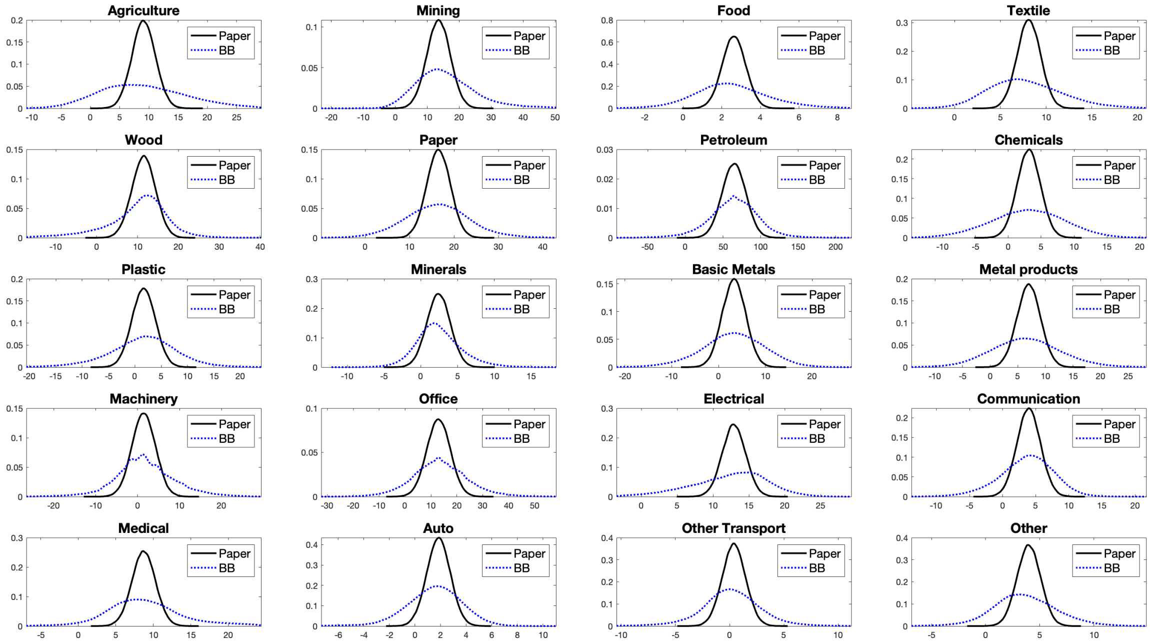

Figure 3: Distributions of the benchmark estimates (which remove the countries with the lowest 1% share of trade for each sector) in Table 1 of Caliendo and Parro (2015). “Paper” corresponds to the normal approximation as implied by the standard errors reported in the paper, and “BB” corresponds to the Bayesian bootstrap posterior.

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.

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

6.2.1 Parameter Estimation

Artuç, Chaudhuri, and McLaren (2010) uses over-identified GMM to estimate the mean and variance of workers' switching cost, denoted with μ and σ^2, respectively. 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

$$\psi^{\text{industry}-\text{year}}\left(X_{ij,t};\theta\right) = \left(Y_{ij,t} - \begin{pmatrix} \frac{\zeta-1}{\sigma^2}\mu & \frac{\zeta}{\sigma^2} & \zeta \end{pmatrix} R_{ij,t}\right) Z_{ij,t}, \tag{29}$$

	Point estimate	Bayesian bootstrap
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).

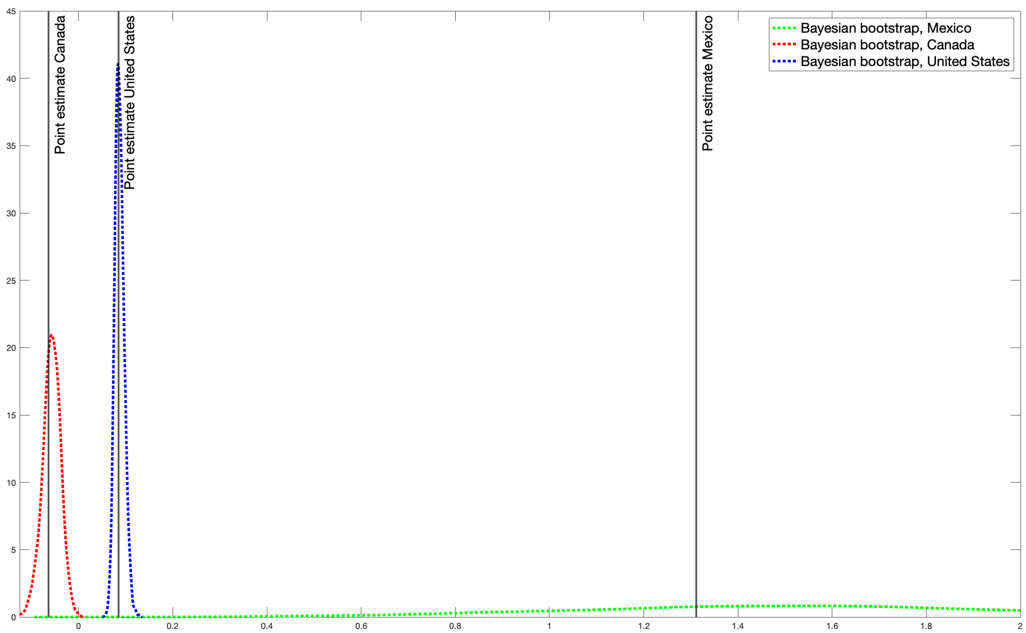

Figure 4: Posterior distributions for welfare effects as in Table 2 of Caliendo and Parro (2015)

with

$$Y_{ij,t} = \log m_{ij,t} - \log m_{ii,t}$$

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

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

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 reproduce the estimates for μ and σ^2 in Panel IV of Table 3 in Artuç, Chaudhuri, and McLaren (2010), which fixes $\zeta = 0.97$ and corresponds to the authors' preferred specification.

Instead of exchangeability across all observations, one might argue a more plausible assumption is exchangeability across industries. The corresponding moment function is

$$\psi^{\text{industry}}\left(X_{ij};\theta\right)=\frac{1}{T}\sum_{t=1}^{T}\left(Y_{ij,t}-\left(\begin{array}{ccc}\frac{\zeta-1}{\sigma^2}\mu & \frac{\zeta}{\sigma^2} & \zeta\end{array}\right)R_{ij,t}\right)Z_{ij,t}. \tag{30}$$

Given the moment functions in Equations (29) and (30), I consider three different approaches to uncertainty quantification. The first approach follows Artuç, Chaudhuri, and McLaren (2010) and uses Equation (30) to compute the analytic GMM standard error assuming all observations are exchangeable. The second approach is an intermediary case; it still computes an analytic GMM standard error but uses Equation (30) and hence assumes exchangeability only across industries. The third approach is my preferred approach, and it uses the Bayesian bootstrap procedure from Algorithm 2 and Equation (30).[15] The resulting 95% confidence intervals and credible intervals are given in Table 6.[16] The corresponding implied normal distributions and posterior distributions are plotted in Figure 5. 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. Implementation details and extra results can be found in Appendix F.2. Furthermore, a data-calibrated simulation exercise in Appendix B shows that standard GMM standard errors lead to under-coverage.

	Mean	Variance
Point estimate	5.33	1.48
As in paper: analytic errors, exchangeability across all observations	[2.98, 7.67]	[0.87, 2.09]
Intermediary case: analytic errors, exchangeability across industries	[4.14, 6.51]	[1.08, 1.88]
Preferred approach: Bayesian bootstrap, exchangeability across industries	[3.64, 9.39]	[1.09, 2.57]

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

[15] Equivalently to using the moment function in Equation (30) with weights $\omega_{k\ell}^{(b)} = V_k^{(b)} \cdot V_\ell^{(b)} / \left(\sum_{u\neq v} V_u^{(b)} \cdot V_v^{(b)}\right)$ for $k,\ell = 1,...,n$, one could use the moment function in Equation (29) with weights $\omega_{k\ell,s}^{(b)} = V_k^{(b)} \cdot V_\ell^{(b)} / \left(T \cdot \sum_{u\neq v} V_u^{(b)} \cdot V_v^{(b)}\right)$ for $k,\ell = 1,...,n$ and $s = 1,...,T$.

[16] The point estimates differ slightly from those in Artuç, Chaudhuri, and McLaren (2010). This is because there the authors use iterated GMM rather than two-step GMM and they use a different weight matrix. In Appendix F.2.2 I consider their exact setup and the conclusions do not change.

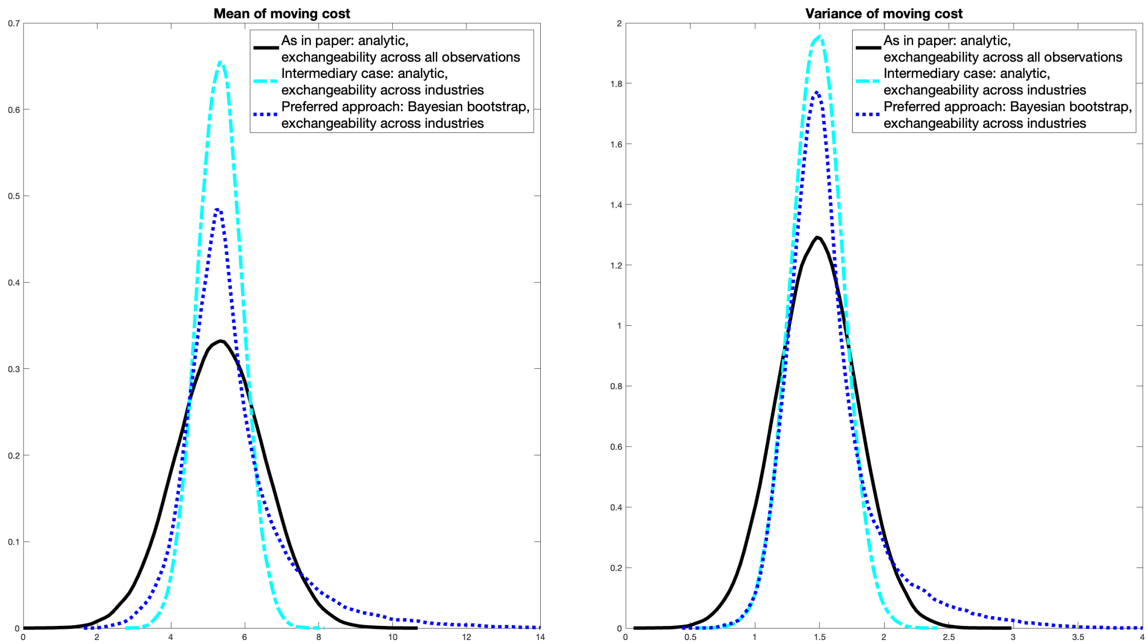

Figure 5: Distribution of estimators for Panel IV in Table 3 in Artuç, Chaudhuri, and McLaren (2010) for $\zeta = 0.97$.

6.2.2 Counterfactual Prediction

The estimated mean and variance 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 pre- and post-employment share of the manufacturing sector, the pre- and post-wage of the manufacturing sector, and the expected discounted lifetime utilities before and after the announcement of liberalization. These counterfactual predictions are reported in Figures 3, 4 and 5 in Artuç, Chaudhuri, and McLaren (2010) without any uncertainty quantification. The 95% Bayesian credible intervals (using my preferred approach, assuming only exchangeability across industries) for these quantities are given in Table 7.

	Employment share	Wage	Utility
Before liberalization	25.3% [23.3%, 26.0%]	1.043 [1.028, 1.085]	39.4 [36.7, 48.2]
After liberalization	15.7% [15.5%, 16.1%]	1.036 [1.020, 1.043]	40.8 [38.2, 49.3]

Table 7: Uncertainty quantification for relevant economic quantities from Figures 3, 4 and 5 in Artuç, Chaudhuri, and McLaren (2010) for $\zeta = 0.97$.

The credible intervals are again asymmetric around the point estimates. In all of the bootstrap draws, the employment share goes down and the lifetime utility goes up. Notably, in around 25% of bootstrap draws, the equilibrium wage after liberalization is higher than the equilibrium wage before liberalization. To investigate this further, Figure 6 plots the posterior distribution of the *difference* between the post- and pre-wage of the manufacturing sector, which has a heavy left tail but 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.

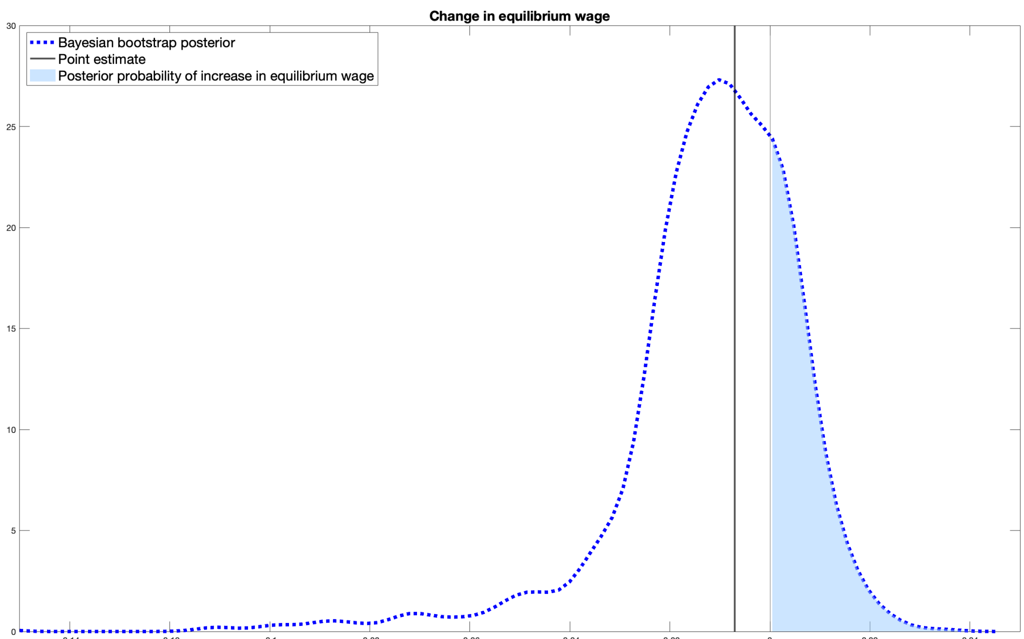

Figure 6: Posterior distribution for the change in wages based on Figure 4 in Artuç, Chaudhuri, and McLaren (2010) for $\zeta = 0.97$.

7 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).

7.1 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, 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), using the pigeonhole bootstrap, if we run sufficiently many iterations, we will eventually draw a world with three copies of Australia and no Belgium. In contrast, every bootstrap draw in the Bayesian procedure in Algorithm 1 will have all 43 countries, but they are reweighted using continuous and strictly positive weights. $\triangle$

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

$$\hat{\gamma}^{*,\text{pb},(b)} = g\left(\{X_{k\ell}\}_{k\neq\ell}, \hat{\theta}^{*,\text{pb},(b)}\right), \tag{31}$$

for $b = 1, ..., B$, and construct a coverage interval for γ using these draws. The validity of this approach follows from Theorem 6. 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 Section 6.2, where, for my preferred specification, the number of interacting units is $n = 6$. Table 8 shows that the credible intervals obtained from the Bayesian bootstrap are narrower than the coverage intervals obtained from the pigeonhole bootstrap. Figure 5 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 $n = 6$), in which case the optimal weight-matrix is reported to be near-singular. The pigeonhole bootstrap also produces more extreme outliers. Specifically, approximately 3% of the bootstrap draws for the mean and 1% for the variance fall more than 10 standard deviations (as measured by the GMM standard error) away from the point estimate. For the Bayesian bootstrap, the corresponding rates are 0.5% and 0.1%, respectively. In addition, the Bayesian bootstrap draws for the variance estimator are always nonnegative, whereas about 0.5% of the pigeonhole bootstrap draws yield negative values.

	Mean	Variance
Point estimate	5.33	1.48
Bayesian bootstrap	[3.64, 9.39]	[1.09, 2.57]
Pigeonhole bootstrap	[2.76, 11.17]	[0.59, 3.01]

Table 8: Uncertainty quantification for Panel IV in Table 3 in Artuç, Chaudhuri, and McLaren (2010) for $\zeta = 0.97$.

7.2 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*

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

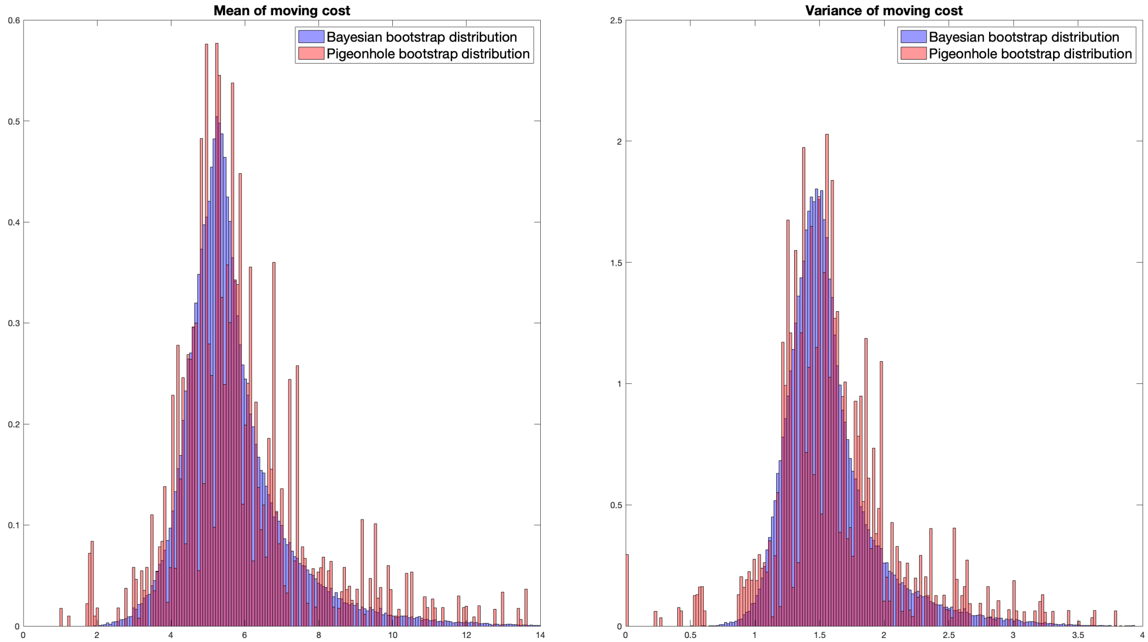

Figure 7: Bootstrap distributions of estimators for Panel IV in Table 3 in Artuç, Chaudhuri, and McLaren (2010) for $\zeta = 0.97$.

where

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

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

$$\left. \left(\frac{\hat{\phi}_{k\ell} + \hat{\phi}_{\ell k}}{2} \right) \left(\frac{\hat{\phi}_{\ell s} + \hat{\phi}_{s\ell}}{2} \right)' + \left(\frac{\hat{\phi}_{ks} + \hat{\phi}_{sk}}{2} \right) \left(\frac{\hat{\phi}_{\ell s} + \hat{\phi}_{s\ell}}{2} \right)' \right\}$$

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

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

For $\hat{\theta}$ a Z-estimator, one can then report the confidence interval $\left[\hat{\theta} \pm 1.96 \cdot \sqrt{\widehat{\mathrm{Var}}_{\mathrm{Graham}}\left(\hat{\theta}\right)}\right]$. As argued in Theorem 6, 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 5 (Graham, 2024). Lastly, the approach can be difficult to implement. For example, for over-identified GMM, this approach requires computing many numerical derivatives.

7.3 Application 3: Silva and Tenreyro (2006)

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 2 and Proposition 1 in the current paper and Theorem 2.4 in Davezies, D'haultfœuille, and Guyonvarch (2021). To illustrate this, in Appendix G 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 method with a finite-sample guarantee, all three methods yield similar uncertainty quantification.

8 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 method 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 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.

Bickel, P. J., and A. Chen (2009): "A nonparametric view of network models and Newman–Girvan and other modularities," *Proceedings of the National Academy of Sciences*, 106, 21068–21073.

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.

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.

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.

Han, Q. (2022): "Multiplier u-processes: Sharp bounds and applications," *Bernoulli*, 28, 87–124.

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.*

Hušková, M., and P. Janssen (1993): "Consistency of the generalized bootstrap for degenerate U-statistics," *The Annals of Statistics*, 1811–1823.

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.

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.

Appendix

A Extra Results for Running Example: Waugh (2010)

A.1 Model Details

In Section 2.1.3 I introduced the counterfactual mapping

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

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 prices in country k.

The equilibrium conditions map rental rates, trade costs, labor endowments, production parameters and the productivity parameter to aggregated prices, expenditure shares and wages:

$$\{r_k\}, \{\tau_{k\ell}\}, \{L_k\}, \{Q_k\}, \alpha, \beta, \theta_M \mapsto \{p_k\}, \{\lambda_{k\ell}\}, \{w_k\}. \tag{32}$$

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 subsumed into g. 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

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

It follows that we require a "calibration-mapping"

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

Such a mapping exists and we can use the equilibrium mapping in Equation (32) to arrive at:

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

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

A.2 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 9 and Figure 8 add the resulting credible intervals and posterior distributions to Table 1 and Figure 1, respectively. The point estimates drop considerably and there is more uncertainty.

	Point estimate	As in paper	Bayesian bootstrap
All countries, OLS	5.55	[5.39, 5.71]	[5.12, 6.02]
Only OECD, OLS	7.91	[7.46, 8.37]	[6.91, 9.21]
Only non-OECD, OLS	5.45	[5.06, 5.84]	[4.42, 6.65]
All countries, PPML	4.19	-	[3.42, 5.12]
Only OECD, PPML	5.81	-	[4.43, 7.41]
Only non-OECD, PPML	4.49	-	[3.17, 6.68]

Table 9: Uncertainty quantification for productivity parameters as in Waugh (2010) using OLS and PPML.

A.3 Implementation details

To compute the limiting marginal prior according to Theorem 3, 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.4 Alternative Methods

Table 10, Figure 9 and Table 11 reproduce Table 1, Figure 1 and Table 3, respectively, but add the results corresponding to the pigeonhole bootstrap from Section 7.1. 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 7.2 cannot be applied here because a substantial share of the observations are missing.

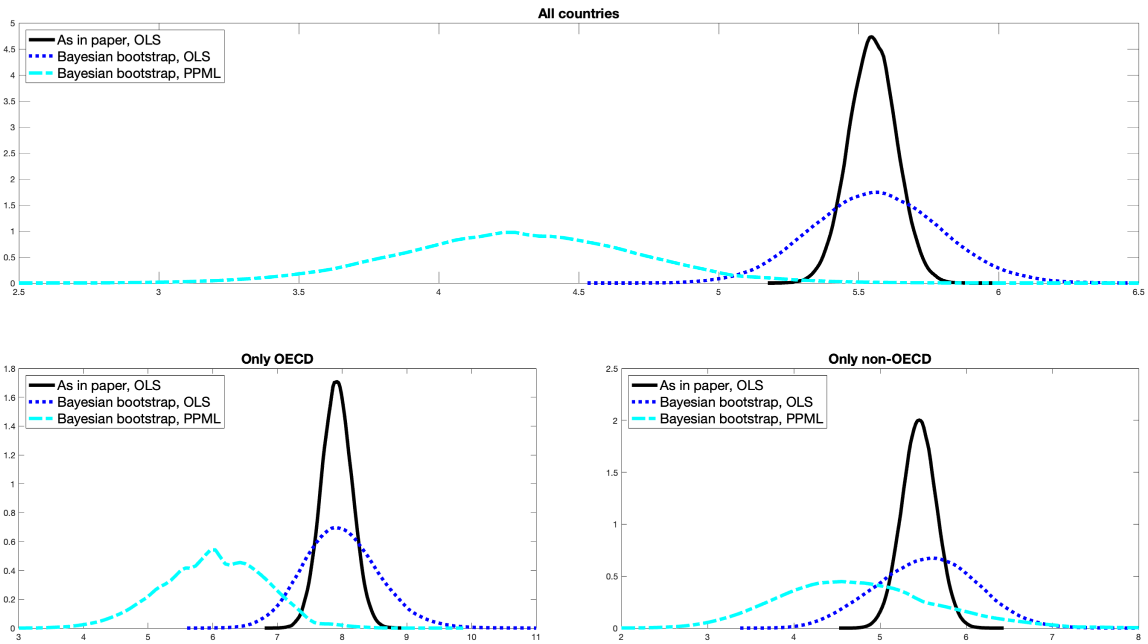

Figure 8: Distributions for productivity parameters as in Waugh (2010) using OLS and PPML.

	Point estimate	As in paper	Bayesian bootstrap	Pigeonhole bootstrap
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 10: Uncertainty quantification for productivity parameters as in Waugh (2010).

B Comparing Methods using Pigeonhole Bootstrap DGP

Recall the model in Assumption 3:

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

To test the performance of the various methods discussed in Section 7, 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

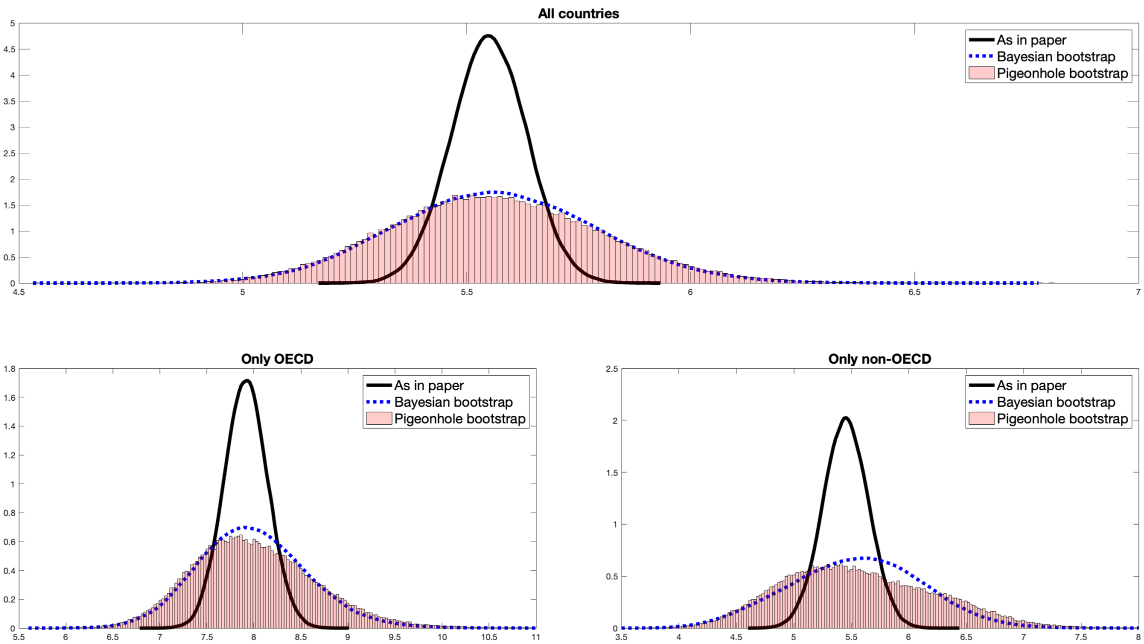

Figure 9: Distributions for productivity parameters as in Waugh (2010).

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. The coverage

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 $k \neq \ell$.

results for the structural estimators considered in the body of the paper can be found in Table 12.

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

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

Table 11: Bayesian uncertainty quantification for counterfactual predictions as in Waugh (2010). Here, "p.e." denotes point estimate, "b.b." denotes Bayesian bootstrap, and "p.b." denotes pigeonhole bootstrap.

C Proofs

C.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 (21), we know from Theorem 4.6 in Ghosal and van der Vaart (2017) that the posterior for $\mathbb{P}_C$ is

$$\pi_\alpha(\mathbb{P}_C|h,\{C_k\}) = DP\left(\frac{\alpha}{\alpha+n}Q + \frac{n}{\alpha+n}\frac{1}{n}\sum_{k=1}^{n}\delta_{C_k}, \alpha+n\right). \tag{33}$$

Finding $\pi_0(\mathbb{P}_C|h,\{C_k\})$. Applying Theorem 4.16 in Ghosal and van der Vaart (2017) yields the posterior distribution that sends the precision parameter α to zero:

$$\pi_0(\mathbb{P}_C|h,\{C_k\}) = \lim_{\alpha\downarrow 0}\pi_\alpha(\mathbb{P}_C|h,\{C_k\}) = DP\left(\sum_{k=1}^{n}\frac{1}{n}\cdot\delta_{C_k}, n\right).$$

	As 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
Artuç, Chaudhuri, and McLaren (2010), μ, $n = 6$	0.733	0.897	0.884
Artuç, Chaudhuri, and McLaren (2010), σ^2, $n = 6$	0.825	0.925	0.902

Table 12: Coverage for approach used in paper, Bayesian bootstrap and pigeonhole bootstrap using the pigeonhole bootstrap DGP from Algorithm 4, using 1000 simulated datasets and $B = 1000$. For Waugh (2010) and Caliendo and Parro (2015), I use heteroskedastic-robust standard errors. For Artuç, Chaudhuri, and McLaren (2010) I used the standard GMM variance formula, assuming only exchangeability across industries.

Note that this posterior distribution is proper. Furthermore, a random probability distribution $\mathbb{P}^*_{n,C}$ drawn from $\pi_0 \left(\mathbb{P}_C | h, \{C_k\}\right)$ is necessarily supported on the observation points $\{C_k\} = \{C_1, ..., C_n\}$. Hence, by definition of a Dirichlet process, we have

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

It follows that

$$\begin{aligned}&\mathbb{P}^*_{n,C} \sim \pi_0\left(\mathbb{P}_C | h, \{C_k\}\right) \\ &\Rightarrow \mathbb{P}^*_{n,C} = \sum_{k=1}^{n} W_k \cdot \delta_{C_k}, \quad \left(W_1, ..., W_n\right) \sim \text{Dir}\left(n; 1, ..., 1\right),\end{aligned} \tag{34}$$

and we have

$$Pr_{\pi_0}\left\{C_i \in B | h, \{C_k\}\right\} = \sum_{k=1}^{n} W_k \cdot \mathbb{I}\left\{C_k \in B\right\}. \tag{35}$$

Finding $\pi_0\left(\mathbb{P}_{X_{ij}}|h,\{C_k\}\right)$**.** Next, combining the model from Assumption 3 and the Bayesian bootstrap posterior in Equation (34), we find

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

This implies that

$$\begin{aligned}
&\mathbb{P}^*_{n,X_{ij}}\sim\pi_0\left(\mathbb{P}_{X_{ij}}|h,\{C_k\}\right)\\
&\Rightarrow\mathbb{P}^*_{n,X_{ij}}=\sum_{k\neq\ell}\frac{W_k\cdot W_\ell}{\sum_{s\neq t}W_s\cdot W_t}\cdot\delta_{X_{k\ell}},\quad (W_1,...,W_n)\sim\mathrm{Dir}(n;1,...,1)\,. \qquad (36)
\end{aligned}$$

So we have found an expression for the Bayesian bootstrap posterior for $\mathbb{P}_{X_{ij}}$ conditional on observing $\{C_k\}$ and knowing 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 the posterior distribution of interest is $\pi_0\left(\mathbb{P}_{X_{ij}}|\{X_{k\ell}\}_{k\neq\ell}\right)$. But using the fact that draws from $\pi_0\left(\mathbb{P}_{X_{ij}}|h,\{C_k\}\right)$ only depend on $\{X_{k\ell}\}_{k\neq\ell}$, we can find an expression for this posterior:

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

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 (36), the posterior $\pi_0\left(\mathbb{P}_{X_{ij}}|h,\{C_k\}\right)$ does not depend on $\{C_k\}$ or h. In conclusion, we have

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

Finding $\pi_0\left(\theta \middle| \{X_{k\ell}\}_{k\neq\ell}\right)$. Lastly, since θ is a function of $\mathbb{P}_{X_{ij}}$, the limiting posterior $\pi_0\left(\mathbb{P}_{X_{ij}} \middle| \{X_{k\ell}\}_{k\neq\ell}\right)$ also implies a limiting posterior on structural estimand, $\pi_0\left(\theta \middle| \{X_{k\ell}\}_{k\neq\ell}\right)$. So indeed, the procedure from Algorithm 1 has a Bayesian interpretation.

C.2 Proof of Theorem 2

Statement 1. The first part of Theorem 2 follows from the derivations in the proof of Theorem 1, where we noted that the 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 posterior drop out when we take the prior precision parameter to zero. Furthermore we can write

$$Pr_{\pi_0}\{X_{ij} \in B|h, \{C_k\}\} = \sum_{k\neq\ell} \frac{W_k \cdot W_\ell}{\sum_{s\neq t} W_s \cdot W_t} \cdot \mathbb{I}\{X_{k\ell} \in B\}\,,$$

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

Statement 2. The second part of Theorem 2 builds on Corollary 4.29 in Ghosal and van der Vaart (2017), applied to the prior $\pi(\mathbb{P}_C)$. This result 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. Consider a 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

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

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

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

which we can relate back to $\{X_{k\ell}\}_{k\neq\ell}$:

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

C.3 Proof of Theorem 3

We first have to check whether the limit exists. We hence require the function $T : \Delta(\mathcal{X}) \to \Theta$ 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 $T_C : \Delta(\{C_k\}) \mapsto \Theta$. 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 $\{\omega_{k,w}\}_w \in \Delta(\{1, .., n\})$ 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 $\bar{T}_C(\cdot)$.

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

Lemma 2 (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 $\pi(\theta)$ converges weakly to $\pi^\infty \in \Delta(\Theta)$, where

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

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

Lemma 2 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

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

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

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

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 2 implies that the marginal prior limit $\pi^\infty(\theta)$ equals

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

C.4 Proof of Lemma 2

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

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

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

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

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

Let $\tau(1) \in\{1, \ldots, n\}$ be the index for the observation in the original (latent) data with $C_{\tau(1)}=\tilde{C}_1$. For $r \in\{2, \ldots, n\}$, let $s(r)$ be the smallest s such that $\tilde{C}_s$ is distinct from $\left\{C_{\tau(1)}, \ldots, C_{\tau(r-1)}\right\}$, 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 $\mathbb{P}_C \in \Delta\left(\left\{C_k\right\}\right)$, and $\omega_k(\tau, \alpha) \in(0,1)$ with probability one for all $\alpha>0$. Moreover, as $\alpha \downarrow 0$ we have $\omega_{k+1}(\tau, \alpha) / \omega_k(\tau, \alpha) \rightarrow 0$ for all k, so

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

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

C.5 Proof of Theorem 4

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

$$\begin{aligned}
\mathbb{G}_n f & =\sqrt{n}\left\{\sum_{k \neq \ell} \frac{1}{n(n-1)} \cdot f\left(X_{k \ell}\right)-\mathbb{E}_{\mathbb{P}_{X_{ij}}}\left[f\left(X_{ij}\right)\right]\right\} \\
\mathbb{G}_n^* f & =\sqrt{n}\left\{\sum_{k \neq \ell} \frac{W_k \cdot W_\ell}{\sum_{s \neq t} W_s \cdot W_t} \cdot f\left(X_{k \ell}\right)-\sum_{k \neq \ell} \frac{1}{n(n-1)} \cdot f\left(X_{k \ell}\right)\right\} .
\end{aligned} \quad (37)$$

We have the following lemma for U-processes based on Arcones and Giné (1993) and Zhang (2001):

Lemma 3 (Weak convergence of empirical processes of U-processes)**.** *Let* $\tilde{\mathcal{F}} \subseteq (\mathcal{C}^2)^{\mathbb{R}}$ *be a measurable class of* symmetric *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*

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

Suppose that:

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

(ii) We have non-degeneracy, meaning that

$$\operatorname{Cov}\left(\tilde{f}_1(C_1, C_2), \tilde{f}_2(C_1, C_{2'})\right) > 0 \ \forall \tilde{f}_1, \tilde{f}_2 \in \tilde{\mathcal{F}}.$$

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

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

Then, defining the empirical processes

$$\tilde{\mathbb{G}}_n \tilde{f} = \sqrt{n}\left\{\sum_{k \neq \ell} \frac{1}{n(n-1)} \cdot \tilde{f}(C_k, C_\ell) - \mathbb{E}_{\mathbb{P}_C}\left[\tilde{f}(C_i, C_j)\right]\right\}$$
$$\tilde{\mathbb{G}}_n^* \tilde{f} = \sqrt{n}\left\{\sum_{k \neq \ell} W_k \cdot W_\ell \cdot \tilde{f}(C_k, C_\ell) - \sum_{k \neq \ell} \frac{1}{n(n-1)} \cdot \tilde{f}(C_k, C_\ell)\right\},$$

we have weak convergence over $\ell^\infty\left(\tilde{\mathcal{F}}\right)$ *of both* $\tilde{\mathbb{G}}_n$ *and* $\tilde{\mathbb{G}}_n^*$ *to the same centered Gaussian process* $\tilde{\mathbb{G}}$ *with covariance kernel*

$$\tilde{K}\left(\tilde{f}_1, \tilde{f}_2\right) = 4 \operatorname{Cov}\left(\tilde{f}_1(C_1, C_2), \tilde{f}_2(C_1, C_{2'})\right),$$

where the convergence of $\tilde{\mathbb{G}}_n^*$ *holds conditional on the data* $\{C_k\}$ *and outer almost surely.*

To use this lemma, first note that as $n \to \infty$, we can ignore the normalization term in the denominator of Equation (37), because

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

The convergence in probability follows since $\mathbb{E}\left[\sum_{s=1}^{n} W_s^2\right] = \frac{2}{n+1}$, and convergence in mean to zero for a non-negative random variable implies convergence in probability.

Next, note that we can always "symmetrize" a sum of non-symmetric functions since

$$\sum_{k \neq \ell} f\left(X_{k\ell}\right) = \sum_{k \neq \ell} \frac{f\left(X_{k\ell}\right) + f\left(X_{\ell k}\right)}{2}.$$

This symmetrization implies that the relevant covariance kernel is

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

Given these observations, it remains to check that Assumption 6 implies that we can apply Lemma 3 with $\tilde{\mathcal{F}} = \mathcal{F} \circ h$.

Towards that end, note that $\mathcal{F}$ being permissible implies $\mathcal{F} \circ h$ is permissible for measurable h. The existence of the positive integrable envelope function F for $\mathcal{F}$ implies the existence of a positive integrable envelope function $F \circ h$ for $\mathcal{F} \circ h$, since for any $(f \circ h) \in \mathcal{F} \circ h$ and $(c_1, c_2) \in \mathcal{C}^2$,

$$\begin{aligned}
&\left|\left(f \circ h\right)\left(c_1, c_2\right)\right| = \left|f\left(x_{12}\right)\right| \leq F\left(x_{12}\right) = \left(F \circ h\right)\left(c_1, c_2\right) \\
&\left(F \circ h\right)\left(c_1, c_2\right) = F\left(x_{12}\right) > 0 \\
&\mathbb{P}_C\left(F \circ h\right)^2 = \mathbb{P}_{X_{ij}} F^2 < \infty.
\end{aligned}$$

Non-degeneracy is satisfied because

$$\begin{aligned}
&\operatorname{Cov}\left(\left(f_1 \circ h\right)\left(C_1, C_2\right) + \left(f_1 \circ h\right)\left(C_2, C_1\right), \left(f_2 \circ h\right)\left(C_1, C_2'\right) + \left(f_2 \circ h\right)\left(C_2', C_1\right)\right) \\
&= \operatorname{Cov}\left(f_1\left(X_{12}\right) + f_1\left(X_{21}\right), f_2\left(X_{12'}\right) + f_2\left(X_{2'1}\right)\right) > 0.
\end{aligned}$$

Lastly, for each $\tilde{Q} \in \Delta\left(\mathcal{C}\right)$ such that $\tilde{Q}\left(F \circ h\right)^2 < \infty$, there exists a $Q \in \Delta\left(\mathcal{X}\right)$ such that $QF^2 < \infty$ and

$$\left\|F \circ h\right\|_{L_2\left(\tilde{Q}\right)} = \left\|F\right\|_{L_2(Q)}.$$

This implies we have

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

C.6 Proof of Lemma 3

The lemma follows almost directly from combining Theorem 4.9 in Arcones and Giné (1993) and Corollary 1 in Zhang (2001). Condition (iii) in Lemma 3 differs from Condition 2 in Zhang (2001), as the author assumes $\tilde{\mathcal{F}}$ has polynomial discrimination 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 bounded 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.

C.7 Proof of Theorem 5

From Theorem 4 we know that under Assumptions 6, $\mathbb{G}_n$ defined by $\mathbb{G}_n f=\sqrt{n}\left\{\mathbb{P}_{n,X_{ij}}f-\mathbb{P}_{X_{ij}}f\right\}$ converges unconditionally in distribution to a tight random element $\mathbb{G}$, and $\mathbb{G}_n^*$ defined by $\mathbb{G}_n^* f=\sqrt{n}\left\{\mathbb{P}_{n,X_{ij}}^* f-\mathbb{P}_{n,X_{ij}}f\right\}$ converges, conditionally given $\{X_{k\ell}\}_{k\neq\ell}$ and outer almost surely, to the same random element. This implies

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

for BL_1 the set of bounded Lipschitz functions from $\ell^\infty\left(\mathcal{F}\right)$ to $[0,1]$. Since $\hat{\theta}$ is assumed to be of the form $T\left(\mathbb{P}_{n,X_{ij}}\right)=\varphi\left(\mathbb{P}_{n,X_{ij}}f\right)$ for $f\in\mathcal{F}$, the result then follows by applying the functional delta method for the bootstrap, Theorem 23.9 in Van der Vaart (2000).

C.8 Proof of Corollary 2

The relevant function class is $\mathcal{F}_Z\equiv\{\nu_{\vartheta,\eta}:(\vartheta,\eta)\in\Theta\times\mathcal{H}\}$. Now, $\varphi:\ell^\infty\left(\mathcal{F}_Z\right)\mapsto\Theta$ is the map that extracts the zero from the estimating equation, so that we have,

$$\begin{aligned}
\theta&=T\left(\mathbb{P}_{X_{ij}}\right)=\varphi\left(\Psi\right)\equiv\varphi\left(\sup_{\eta\in\mathcal{H}}\left|\mathbb{P}_{X_{ij}}\nu_{\vartheta,\eta}\right|\right)\\
\hat{\theta}&=T\left(\mathbb{P}_{n,X_{ij}}\right)=\varphi\left(\Psi_n\right)\equiv\varphi\left(\sup_{\eta\in\mathcal{H}}\left|\mathbb{P}_{n,X_{ij}}\nu_{\vartheta,\eta}\right|\right)\\
\hat{\theta}^*&=T\left(\mathbb{P}_{n,X_{ij}}^*\right)=\varphi\left(\Psi_n^*\right)\equiv\varphi\left(\sup_{\eta\in\mathcal{H}}\left|\mathbb{P}_{n,X_{ij}}^*\nu_{\vartheta,\eta}\right|\right).
\end{aligned}$$

The proof follows Corollary 13.6 in Kosorok (2008). From Theorem 13.5 in Kosorok (2008), we know that conditions (i)-(iii) are sufficient conditions for Fréchet differentiability of φ. Conditions (iv) and (v) are regularity conditions. Condition (vi) guarantees convergence weak convergence of the empirical processes using Theorem 4. We can then apply Theorem 5 and the result follows.

C.9 Proof of Theorem 6

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 (28), 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.

D Extra Results for Limiting Marginal Prior

Since the Dirichlet process prior in Equation (24) is only supported on $\{C_k\}$, we have

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

which collapses to the Bayesian bootstrap *posterior* in Equation (23) 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 $\left\{\hat{\theta}^{*,(1)}, ..., \hat{\theta}^{*,(B)}\right\}$.

E Other Extensions

E.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

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

with corresponding priors

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

E.2 Conditional Exchangeability

The key underlying model assumption, as discussed in Section 3.1, 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

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

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

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

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 3.1.1, we can again motivate the model using an Aldous-Hoover representation. Now, $\left\{X_{ij}\right\}_{i,j\in\mathbb{N},i\neq j}$ is assumed to be *relatively exchangeable* with respect to "types" R, which means that there exist subpopulations within which agents are exchangeable. Then,

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

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

$$X_{ij}^* = \tilde{h}^{AH}\left(U, R_i, R_j, C_i, C_j, D_{ij}\right), \tag{38}$$

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

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

F Extra Results for Applications

F.1 Extra Results for Application 1: Caliendo and Parro (2015)

F.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. I also follow the authors and replace the elasticity for sectors "Auto" and "Other Transport" by the average elasticity of the other sectors.

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

We can use Theorem 3 with

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

and continuity of χ is satisfied. Figure 10 plots the bootstrap posterior and the limiting marginal prior using Theorem 3. 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.

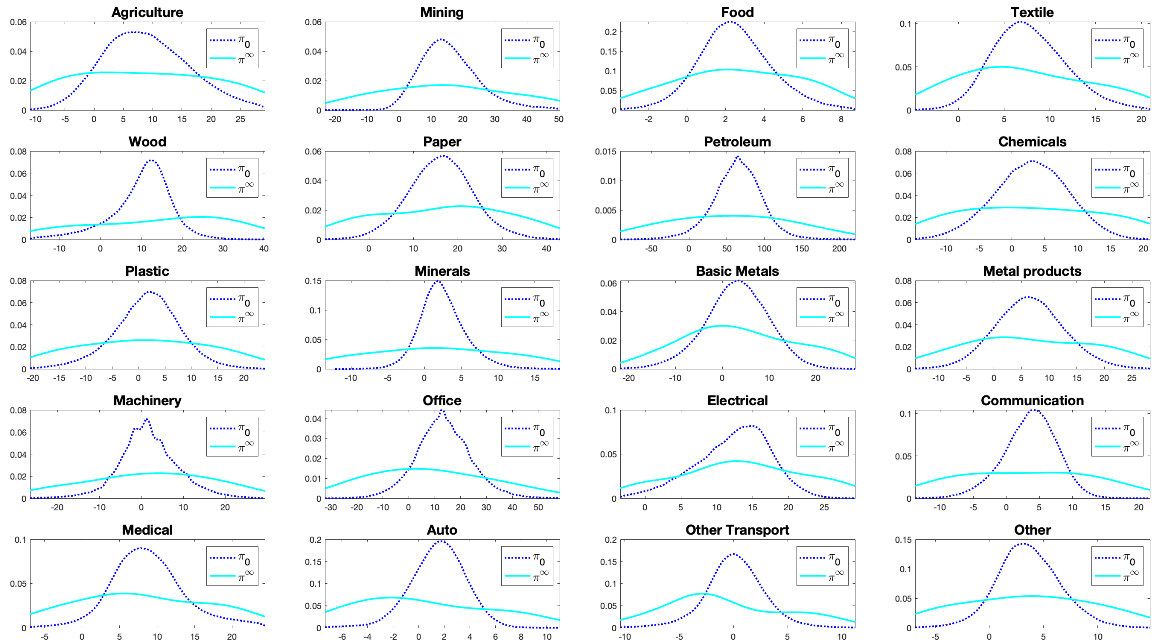

Figure 10: Limiting marginal prior for trade elasticities as in Caliendo and Parro (2015).

F.1.3 Alternative Methods

Table 13, Figure 11 and Table 14 reproduce Table 4, Figure 3 and Table 5, respectively, but add the results corresponding to the pigeonhole bootstrap from Section 7.1. The confidence intervals for the sectoral elasticities constructed using the pigeonhole bootstrap are consistently larger and all include zero. This results also in that 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 7.2 cannot be applied here because data are triadic and a small share of observations are missing.

F.2 Extra Results for Application 2: Artuç, Chaudhuri, and McLaren (2010)

F.2.1 Details for Implementation

In a small number of counterfactual draws, the welfare effects are complex numbers. I omit these draws.

	Point estimate	As in paper	Bayesian bootstrap	Pigeonhole bootstrap
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 13: 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).

F.2.2 Iterated GMM and Different Weight Matrix

In Artuç, Chaudhuri, and McLaren (2010), the authors use a different weight matrix, namely

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

for the residuals

$$e_{ij,t}(\theta) = \left(Y_{ij,t} - \begin{pmatrix} \frac{\zeta-1}{\sigma^2}\mu & \frac{\zeta}{\sigma^2} & \zeta \end{pmatrix} R_{ij,t} \right).$$

They also use iterated GMM rather than two-step GMM. Using a different weight matrix and a different GMM estimator results in slightly different point estimates. However, the iterated GMM estimator satisfies Assumption 1 and, analogous to the discussion in Section 4.1.3, Imbens (1997) shows it can be combined into a single just-identified system, so that the frequentist guarantees also hold. The resulting Bayesian bootstrap algorithm that only

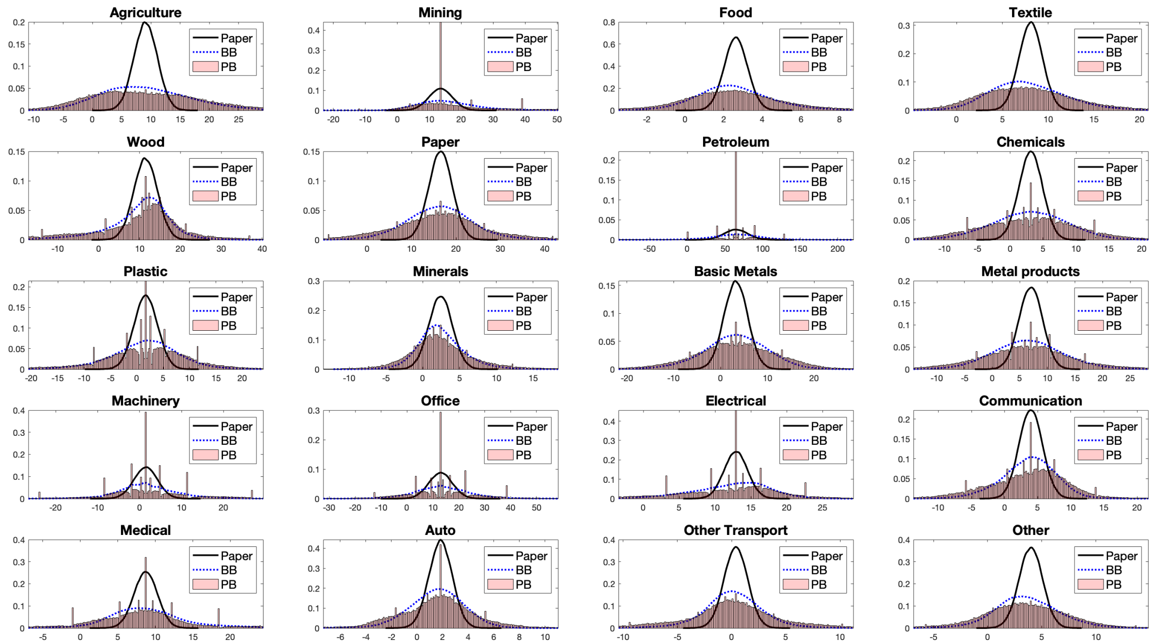

Figure 11: Distributions of the benchmark estimates (which remove the countries with the lowest 1% share of trade for each sector) in Table 1 of Caliendo and Parro (2015). "Paper" corresponds to the normal approximation as implied by the standard errors reported in the paper. "BB" and "PB" correspond to the Bayesian bootstrap posterior and pigeonhole bootstrap distribution, respectively.

assumes exchangeability across industries is summarized in Algorithm 6.

Table 15 reports the resulting coverage and credible intervals. Figure 12 plots the implied normal distributions by the point estimate and t-statistic reported in the paper, and the posteriors obtained by Algorithm 6.

F.2.3 Alternative Methods

In principle one could apply the approach using analytic standard errors from Section 7.2 by interpreting the two-step GMM estimator as a just-identified GMM estimator as per the discussion in Section 4.1.3. However, since this amounts to taking many numerical derivatives, the procedure is unstable.

	Point estimate	Bayesian bootstrap	Bayesian bootstrap
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 14: Bayesian uncertainty quantification for welfare effects as in Table 2 of Caliendo and Parro (2015).

Algorithm 6 Bayesian bootstrap procedure for iterated GMM and different weight matrix

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 $\omega_{k\ell,s}^{(b)} = V_k^{(b)} \cdot V_\ell^{(b)} / \left(\sum_{s=1}^{T} \sum_{u \neq v} V_u^{(b)} \cdot V_v^{(b)}\right)$ for $k, \ell = 1, ..., n$ and $s = 1, ..., T$.
   (c) Denote
   $$\mathbb{P}_{n,X_{ij,t}}^{*,(b)} = \sum_{k \neq \ell, s} \omega_{k\ell,s}^{(b)} \cdot \delta_{X_{k\ell,s}}.$$
   (d) Set $\hat{\Omega}_{(0)}^{(b)} = I_3$.
   (e) Until convergence, compute
      i. $\hat{\theta}_{(w+1)}^{(b)} = \underset{\vartheta \in \Theta}{\arg\min} \mathbb{E}_{\mathbb{P}_{n,X_{ij,t}}^{*,(b)}} \left[e_{ij,t}\,(\vartheta)\, Z_{ij,t}\right]' \hat{\Omega}_{(w)}^{(b)} \mathbb{E}_{\mathbb{P}_{n,X_{ij,t}}^{*,(b)}} \left[e_{ij,t}\,(\vartheta)\, Z_{ij,t}\right]$
      ii. $\hat{\Omega}_{(w+1)}^{(b)} = \mathbb{E}_{\mathbb{P}_{n,X_{ij,t}}^{*,(b)}} \left[e_{ij,t}\left(\hat{\theta}_{(w+1)}^{(b)}\right)^2\right] \mathbb{E}_{\mathbb{P}_{n,X_{ij,t}}^{*,(b)}} \left[Z_{ij,t} Z_{ij,t}'\right]$.
2. Report the quantiles of interest of $\left\{\hat{\theta}^{*,(1)}, ..., \hat{\theta}^{*,(B)}\right\}$.

G 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

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

	Mean	Variance
Point estimate	6.56	1.88
As in paper: analytic errors, exchangeability across all observations	[3.06, 10.07]	[1.04, 2.72]
Preferred approach: Bayesian bootstrap, exchangeability across industries	[4.47, 10.09]	[1.35, 2.83]

Table 15: Uncertainty quantification for Panel IV in Table 3 in Artuç, Chaudhuri, and McLaren (2010) for $\zeta = 0.97$.

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 13 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 yield comparable results for uncertainty quantification.

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

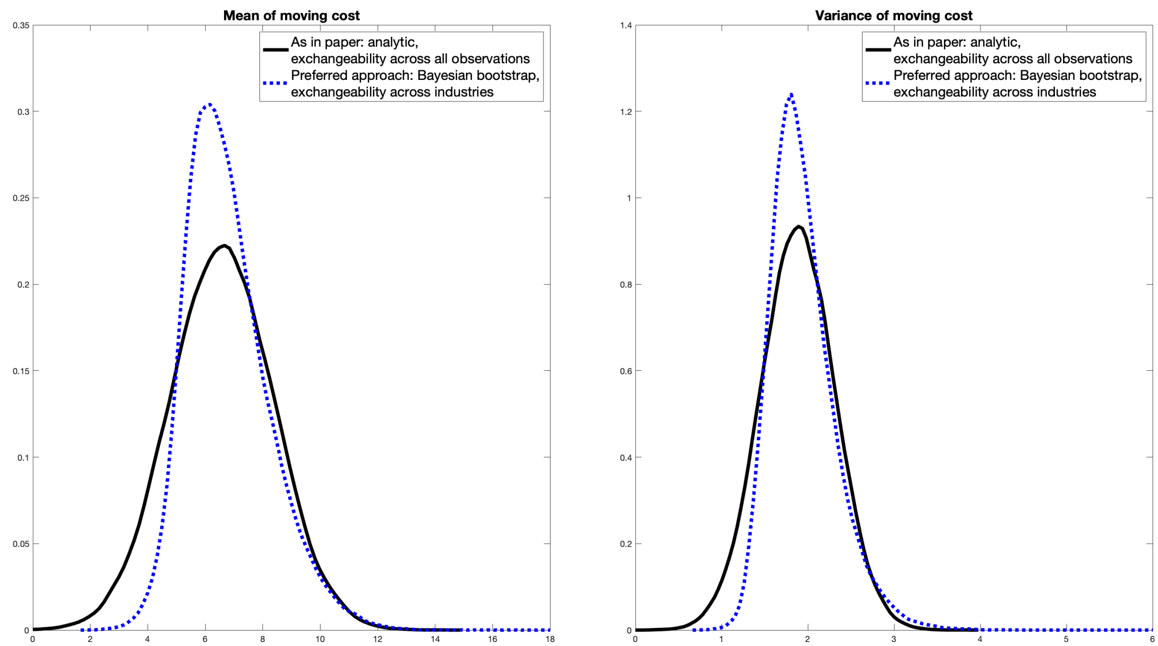

Figure 12: Distribution of estimators for Panel IV in Table 3 in Artuç, Chaudhuri, and McLaren (2010) for $\zeta = 0.97$.

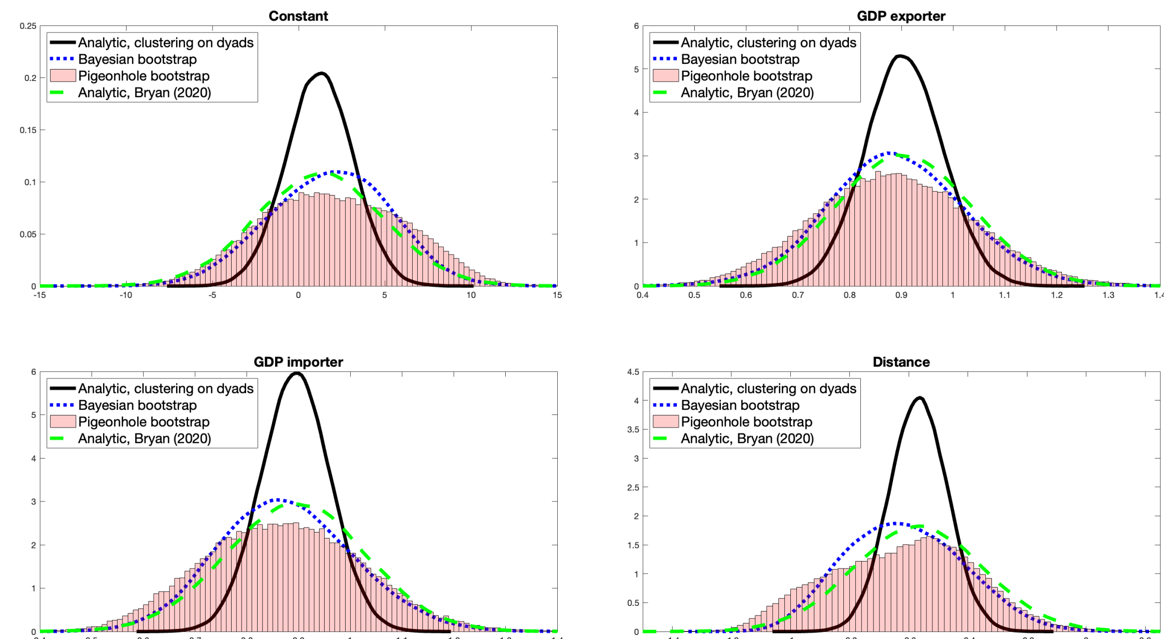

Figure 13: Distribution of estimators when regressing bilateral trade flows on a constant, log exporter and importer GDP, and log distance using PPML.