

BLP Estimation using Laplace Transformation and Overlapping Simulation Draws *

Han Hong[†] Huiyu Li[‡] Jessie Li[§]

October 14, 2019

We derive the asymptotic distribution of the parameters of the [Berry et al. \(1995\)](#) (BLP) model in a many markets setting which takes into account simulation noise under the assumption of overlapping simulation draws. We show that as long as the number of simulation draws R and the number of markets T approach infinity, our estimator is $\sqrt{m} = \sqrt{\min(R, T)}$ consistent and asymptotically normal. We do not impose any relationship between the rates at which R and T go to infinity, thus allowing for the case of $R \ll T$. We provide a consistent estimate of the asymptotic variance which can be used to form asymptotically valid confidence intervals. Instead of directly minimizing the BLP GMM objective function, we propose using Hamiltonian Markov Chain Monte Carlo methods to implement a Laplace-type estimator which is asymptotically equivalent to the GMM estimator.

JEL CLASSIFICATION: C10; C11; C13; C15

KEYWORDS: BLP model, Simulation estimator, Laplace-type estimator

*We thank the participants of the Conference in Honor of Daniel McFadden for helpful comments. This research was supported by a Faculty Research Grant awarded by the Committee on Research from the University of California, Santa Cruz. Theodore Naff provided excellent research assistance. Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the Federal Reserve System.

[†]Department of Economics, Stanford University, 579 Serra Mall, Stanford, CA 94305. Email: hanhong@stanford.edu

[‡]Economic Research, Federal Reserve Bank of San Francisco, 101 Market St, San Francisco, CA 94105. Email: tohuiyu@gmail.com

[§]Department of Economics, University of California, Santa Cruz, 1156 High St, Santa Cruz, CA 95064. Email: jeqli@ucsc.edu(corresponding author)

1 Introduction

With the availability of larger datasets, estimation of the [Berry et al. \(1995\)](#) (BLP) model through minimization of the simulated GMM objective function has become increasingly more computationally intensive. For instance, the Nielsen RMS retail scanner dataset records sales information at the week-level for around 40,000 stores. By defining a market as a week/store combination and restricting attention to only California grocery stores that sold fresh strawberries in 2014, [Compiani \(2018\)](#) already has 86,562 markets. The traditional asymptotic theory for simulation estimators of nonlinear models under independent simulation draws either used a fixed number of simulation draws for each observation ([McFadden \(1989\)](#), [Pakes and Pollard \(1989\)](#)) or required that the number of simulation draws (R) increases faster than the than the square root of the number of markets (\sqrt{T}) in order to eliminate asymptotic bias ([Lee \(1995\)](#)). However, choosing a large number of simulation draws may be impractical when T is very large, which is becoming more common in the age of big data. It would be desirable to develop an asymptotic theory that eliminates asymptotic bias under the case of $R \ll T$. Unfortunately we are unaware of any such results in the case of independent simulation draws where different draws are used in different markets. We instead adopt the framework of overlapping simulation draws where the same draws are used in all markets, even though we acknowledge that there are efficiency gains from using independent simulation draws ([Lee \(1995\)](#), [Kristensen and Salanié \(2017\)](#)).

Our main contribution is to derive the asymptotic distribution of the BLP estimator in the case of overlapping simulation draws in a large R and large T setting. We show that our estimator scaled by \sqrt{m} for $m = \min(R, T)$ has zero asymptotic bias under the relatively weak conditions of $R \rightarrow \infty$ and $T \rightarrow \infty$, and it has an asymptotic variance term that consists of two parts, one for the sampling variance and another for the simulation variance.

An insightful paper that explicitly derives the asymptotic distribution of the BLP estimator in a many markets setting and accounting for simulation noise is [Freyberger \(2015\)](#). His main results focus on the case of independent simulation draws, but he also states without proof a related result under overlapping simulation draws. In order for the asymptotic bias to disappear in the case of overlapping simulation draws, he requires $\lim_{T, R \rightarrow \infty} \frac{T}{R} < \infty$. We show that this condition is not necessary in order for the asymptotic bias to disappear. By changing the rate of convergence from

\sqrt{T} to \sqrt{m} for $m = \min(R, T)$, we can remove the asymptotic bias and allow for the case of $R \ll T$, as long as $R \rightarrow \infty$ and $T \rightarrow \infty$.

On the computational front, we propose using Hamiltonian Markov Chain Monte Carlo (HMCMC) (originally developed by [Duane et al. \(1987\)](#) and discussed at length in [Neal et al. \(2011\)](#)) to implement the Laplace-type estimator of [Chernozhukov and Hong \(2003\)](#), which we show to be consistent for the true parameters and also asymptotically normal. We use HMCMC instead of standard MCMC because the former incorporates information about the gradient of the objective when searching for the optimum while the latter essentially uses a random walk. We compare the performance of our HMCMC routine to two other routines that seem to be commonly used: the nested fixed point (NFP) method originally proposed by [Berry et al. \(1995\)](#) and discussed further in e.g. [Berry \(1994\)](#) and [Nevo \(2000\)](#) and the mathematical programming with equilibrium constraints (MPEC) method discussed in e.g. [Dubé et al. \(2012\)](#) and [Su and Judd \(2012\)](#). We find that our HMCMC method generally gives less biased estimates and empirical coverage frequencies closer to the nominal level than both NFP and MPEC. We acknowledge that many other methods exist such as Bayesian Hierarchical Models ([Yang et al. \(2003\)](#)) or Nonparametric Instrumental Variables ([Compiani \(2018\)](#)), but a comparison with these methods is beyond the scope of the paper.

Section 2 reviews the BLP model in greater detail. Section 3 contains the main components of the asymptotic theory, demonstrating consistency of the BLP estimator and deriving the asymptotic distribution which is normal with mean zero and variance which takes into account simulation noise. Section 4 provides a way for applied researchers to obtain consistent estimates of the standard errors. Section 5 discusses consistency and asymptotic normality of the Laplace-type estimator. Section 6 compares our results with [Freyberger \(2015\)](#)'s results. Section 7 outlines the results of Monte Carlo studies that compare HMCMC to NFP and MPEC and illustrate the severe undercoverage of confidence intervals obtained using the typical GMM standard errors which do not take into account simulation noise. Section 8 concludes. Section 9 contains proofs and additional comparisons with [Freyberger \(2015\)](#).

2 BLP Model

We follow the standard setup in [Berry et al. \(1995\)](#). Each consumer r is choosing between J products and an outside good in T independent markets. Each consumer has her own individual taste parameters β_r for products with observed characteristics x_{jt} and unobserved characteristics ξ_{jt} . Note that the ξ_{jt} may be correlated among products in the same market. Each consumer also has an idiosyncratic horizontal preference component ϵ_{rjt} that is i.i.d. Type 1 extreme value. The price of product j in market t is p_{jt} .

The utility of consumer r choosing product j in market t is given by

$$u_{rjt} = \beta_r^0 + x'_{jt}\beta_r^x - \alpha p_{jt} + \xi_{jt} + \epsilon_{rjt}$$

The individual taste parameters $\beta_r \equiv (\beta_r^0, \beta_r^x)' \in \mathbb{R}^d$ have mean $E[\beta_r] = \beta = (\beta^0, \beta^x)'$ and variance $Var[\beta_r] = Var[\Sigma_0 v_r] = \Sigma_0 \Sigma_0'$, where Σ_0 is assumed to be a diagonal matrix. The randomness of the individual taste parameters comes from the consumer types v_r :

$$\beta_r = \beta + \Sigma_0 v_r, v_r \stackrel{iid}{\sim} F_0$$

Define $\theta_0 \equiv (\theta_{0,1}, \theta_{0,2})'$, where $\theta_{0,1} = (\beta, -\alpha)'$ and $\theta_{0,2} = \text{diag}(\Sigma_0)$ is the $d \times 1$ vector of the diagonal elements of Σ_0 . We rewrite the utility function using the mean utilities δ :

$$\begin{aligned} u_{rjt} &= \delta_{0jt} + \mu_{0rjt} + \epsilon_{rjt} \\ X'_{jt} &= [1, x'_{jt}, p_{jt}] \\ \delta_{0jt} &\equiv \delta_{jt}(\theta_0) = \beta^0 + x'_{jt}\beta^x - \alpha p_{jt} + \xi_{jt} \equiv X'_{jt}\theta_{0,1} + \xi_{jt} \\ \mu_{0rjt} &= \mu_{rjt}(\theta_{0,2}, v_r) = [1, x'_{jt}] \Sigma_0 v_r. \end{aligned}$$

Note that we will suppress dependence of $\delta_{jt}(\theta)$ on X_{jt} and ξ_{jt} to simplify notation. The observed market share for product j in market t is the probability that the utility from purchasing product j is greater than the utility from purchasing any other product in market t . Let $\mathcal{N}(t)$ denote the set of products in market t . Let δ_{0t} and X_t be vectors of δ_{0jt} and X_{jt} for $j = 1, \dots, J$. The observed

market shares are

$$S_{jt} \equiv s_{jt}(\delta_{0t}, X_t, F_0; \theta_0) = \int \frac{\exp(\delta_{0jt} + \mu_{0rjt})}{1 + \sum_{k \in \mathcal{N}(t)} \exp(\delta_{0kt} + \mu_{0rkt})} dF_0(v_r).$$

The BLP model is solved using a simulated GMM approach. Let $Z_t \in \mathbb{R}^{L \times J}$ be a matrix of instruments that are uncorrelated with the unobserved product characteristics ξ_t . The population moment conditions at the true parameters are $\gamma(\theta_0) = E[Z_t(\delta_{0t} - X_t' \theta_{0,1})] = E[Z_t \xi_t] = 0$. To form the sample moments, we need to obtain estimates of δ_t at arbitrary values of θ . We do so by solving for the fixed point from equating the simulated market shares $\hat{s}_{jt}(\delta_t, X_t, \hat{F}; \theta) = \frac{1}{R} \sum_{r=1}^R \frac{\exp(\delta_{jt} + \mu_{rjt}(\theta_2, v_r))}{1 + \sum_{k \in \mathcal{N}(t)} \exp(\delta_{kt} + \mu_{rkt}(\theta_2, v_r))}$ to the observed market shares S_{jt} for all products j in all markets t . [Berry et al. \(1995\)](#) prove that $\delta_t^{(k+1)} = \delta_t^{(k)} + \log(S_t) - \log(\hat{s}_t(\delta_t^{(k)}, X_t, \hat{F}; \theta))$ is a contraction mapping and that a unique fixed point $\hat{\delta}_t(\theta)$ exists. We will suppress the dependence of $\hat{\delta}_t(\theta)$ on X_t to simplify notation. Note that $\hat{\delta}_t(\theta)$ depends on $\theta_2 = \text{diag}(\Sigma)$ through μ_{rt} . The θ_1 are already absorbed into the δ and do not affect the solution of the fixed point algorithm. The GMM estimates of θ_0 are found by minimizing a quadratic form in the sample moment conditions $\hat{\gamma}(\theta) = \frac{1}{T} \sum_{t=1}^T Z_t (\hat{\delta}_t(\theta) - X_t' \theta_1)$ using a positive-definite weighting matrix W_T .

$$\hat{\theta} = \arg \min_{\theta} \hat{\gamma}(\theta)' W_T \hat{\gamma}(\theta)$$

3 Asymptotic Theory of Simulation Estimation

In order to derive the asymptotic distribution of $\hat{\theta}$, we need to first derive the asymptotic distribution of the sample moment conditions $\sqrt{m} \hat{\gamma}(\theta_0)$ scaled by $m = \min(R, T)$. Our strategy will be to take a first order Taylor expansion of the market shares and then invert the linearized market shares to obtain the linearized product qualities $\hat{\delta}(\theta_0)$ which will be shown to be \sqrt{R} consistent for δ_0 . Next, we will express the sample moment conditions as the sum of two terms. The first term is a sample average while the second term is a two-sample U-statistic in the sample of simulation draws v_r and the sample of covariates X_t , instruments Z_t , and product characteristics ξ_t . Using projection arguments for two-sample U-statistics, we show asymptotic normality of the sample moments. In order to show consistency of $\hat{\theta}$ for θ_0 , we will show that the sample moments converge

uniformly to the population moments over the parameter space Θ . Afterwards, we will show that $\hat{\theta}$ is \sqrt{m} consistent for θ_0 and derive the asymptotic distribution of $\sqrt{m}(\hat{\theta} - \theta_0)$. We also provide a consistent estimate of the asymptotic variance.

3.1 Nonsingularity of Jacobian Matrix

Recall that the true (observed) market shares are

$$s_{jt}(\delta_{0t}, X_t, F_0; \theta_0) = \int \frac{\exp(\delta_{0jt} + \mu_{0rjt})}{1 + \sum_{k \in \mathcal{N}(t)} \exp(\delta_{0kt} + \mu_{0rkt})} dF_0(v_r) \equiv \int g_{jt}(\delta_{0t}, X_t, v_r; \theta_0) dF_0(v_r).$$

The simulated market shares at any θ are

$$\hat{s}_{jt}(\hat{\delta}_t, X_t, \hat{F}; \theta) = \frac{1}{R} \sum_{r=1}^R \frac{\exp(\hat{\delta}_{jt}(\theta) + \mu_{rjt}(\theta_2, v_r))}{1 + \sum_{k \in \mathcal{N}(t)} \exp(\hat{\delta}_{kt}(\theta) + \mu_{rkt}(\theta_2, v_r))} \equiv \int g_{jt}(\hat{\delta}_t, X_t, v_r; \theta) d\hat{F}(v_r).$$

We can also define the market shares using arbitrary δ , X , θ , and F as

$$s_{jt}(\delta_t, X_t, F; \theta) = \int \frac{\exp(\delta_{jt}(\theta) + \mu_{rjt}(\theta_2, v_r))}{1 + \sum_{k \in \mathcal{N}(t)} \exp(\delta_{kt}(\theta) + \mu_{rkt}(\theta_2, v_r))} dF(v_r) \equiv \int g_{jt}(\delta_t, X_t, v_r; \theta) dF(v_r).$$

Let $g(\delta, X, v_r; \theta) \equiv \{g_{jt}(\delta, X, v_r; \theta)\}_{j,t=1}^{J,T}$. We now state a result that is proven in [Berry \(1994\)](#) and [Berry et al. \(2013\)](#) regarding the nonsingularity of the Jacobian matrix of the market shares with respect to δ .

Lemma 1. *For $G_\delta(\delta, X, v_r; \theta) \equiv \nabla_\delta g(\delta, X, v_r; \theta)$, $\int G_\delta(\delta, X, v_r; \theta) dF(v_r)$ is nonsingular for all X , θ , δ , and F .*

3.2 \sqrt{R} Consistency of $\hat{\delta}$ at θ_0

The next proposition proves \sqrt{R} consistency of $\hat{\delta}(\theta_0)$ and provides a linearization that will appear in the two sample U-statistic for the sample moments. The strategy is to take a first order Taylor expansion of $s(\hat{\delta}, X, \hat{F}; \theta_0) - s(\delta_0, X, F_0; \theta_0)$ with respect to both δ and F around some intermediate value between $\hat{\delta}(\theta_0)$ and δ_0 and between \hat{F} and F_0 . Since F is a function, we use the Intermediate Value Theorem for functionals.

Proposition 2. (\sqrt{R} consistency of $\hat{\delta}$ at θ_0) Suppose the following conditions are satisfied:

(i) Let $\hat{\delta}(\theta_0)$ and δ_0 lie in an open, bounded, and convex subset \mathbb{D} of \mathbb{R}^{JT} .

(ii) Let \hat{F} and F_0 lie in an open and convex subset of the Banach space \mathbb{F} of distribution functions $\mathbb{R}^d \mapsto [0, 1]$ equipped with the sup-norm.

Then, for $E_v[g(\delta_0, X, v; \theta_0)] \equiv \int g(\delta_0, X, v; \theta_0)dF_0(v)$, as $R \rightarrow \infty$ for any fixed T ,

$$\begin{aligned} \sqrt{R}(\hat{\delta}(\theta_0) - \delta_0) &= - \left(\int G_\delta(\delta_0, X, v_r; \theta_0)dF_0(v_r) \right)^{-1} \frac{1}{\sqrt{R}} \sum_{r=1}^R \{g(\delta_0, X, v_r; \theta_0) - E_v[g(\delta_0, X, v; \theta_0)]\} \\ &\quad + O_p\left(\frac{1}{\sqrt{R}}\right). \end{aligned}$$

PROOF. See appendix section 9.2. ■

3.3 Asymptotic Distribution of Sample Moment Conditions

Next we derive the asymptotic distribution of $\sqrt{m}\hat{\gamma}(\theta_0) = \sqrt{m}\frac{1}{T}\sum_{t=1}^T Z_t (\hat{\delta}_t(\theta_0) - X_t'\theta_1)$. Let $\tilde{g}(\theta_0) = \frac{1}{T}\sum_{t=1}^T Z_t (\delta_{0t} - X_t'\theta_{0,1})$. We will express $\hat{\gamma}(\theta_0) - \tilde{g}(\theta_0)$ as a two sample U-statistic in two i.i.d. samples $\{X_t, Z_t, \xi_t\}_{t=1}^T$ and $\{v_r\}_{r=1}^R$ which are independent of each other.

Theorem 3. (Asymptotic Distribution of sample moment conditions): Suppose the assumptions in proposition 2 are satisfied. Then for $m \equiv \min(T, R)$ and $k \equiv \lim_{T \rightarrow \infty, R \rightarrow \infty} \frac{R}{T}$,

$$\sqrt{m}\hat{\gamma}(\theta_0) \xrightarrow{d} N(0, \Sigma) \equiv N(0, (1 \wedge k)\Omega + (1 \wedge 1/k)\Sigma_h),$$

where $\Omega = \text{Var}(Z_t(\delta_{0t} - X_t'\theta_{0,1}))$, $\Sigma_h = \text{Var}[h(v_r; \theta_0)]$, and

$$h(v_r; \theta_0) = - \int \left\{ Z_t \left(\int \nabla_\delta g_t(\delta_{0t}, X_t, v; \theta_0)dF_0(v) \right)^{-1} (g_t(\delta_{0t}, X_t, v_r; \theta_0) - E_v[g_t(\delta_{0t}, X_t, v; \theta_0)]) \right\} dP(Z_t, X_t, \xi_t),$$

where $P(\cdot)$ is the joint distribution of Z_t, X_t, ξ_t .

PROOF. We give only a proof sketch. For a complete proof, see appendix section 9.3. The idea is to write the sample moment conditions scaled by \sqrt{m} as the sum of two terms which are independent

of each other and asymptotically normal:

$$\begin{aligned}\sqrt{m}\hat{\gamma}(\theta_0) &= \frac{\sqrt{m}}{\sqrt{R}} \frac{1}{\sqrt{R}} \sum_{r=1}^R h(v_r; \theta_0) + \frac{\sqrt{m}}{\sqrt{T}} \sqrt{T}\tilde{g}(\theta_0) + O_p\left(\frac{\sqrt{m}}{R}\right) + O_p\left(\frac{\sqrt{m}}{\sqrt{TR}}\right) \\ &\xrightarrow{d} N(0, (1 \wedge 1/k)\Sigma_h + (1 \wedge k)\Omega).\end{aligned}$$

The first term is one of the projection terms of a two-sample U-statistic and captures the simulation variance while the second term captures the sampling variance. ■

3.4 Uniform Consistency of $\hat{\gamma}(\theta)$ for $\gamma(\theta)$

Before we can show consistency of $\hat{\theta}$ for θ , we need to show that the simulated moment conditions $\hat{\gamma}(\theta)$ are consistent for the population moments $\gamma(\theta)$ uniformly over $\theta \in \Theta$. The approach we take is to first show stochastic equicontinuity and then appeal to the fact that pointwise convergence to a continuous function over a compact set in combination with stochastic equicontinuity implies uniform convergence.

Theorem 4. (*Uniform Consistency of $\hat{\gamma}(\theta)$ for $\gamma(\theta)$*): *Suppose the following conditions are satisfied.*

- (i) $\theta_0 \in \text{Interior}(\Theta)$, where Θ is a compact subset of \mathbb{R}^{2d+1} .
- (ii) $E \|Z_t\|_\infty < \infty$.
- (iii) $E \|Z_t X_t'\|_2 < \infty$.
- (iv) $E \left[\max_{j=1 \dots J} \left| \left[1, x'_{jt} \right]' \circ v_r \right| \right] < \infty$.

Then for any $\kappa_m \rightarrow 0$,

$$\begin{aligned}\sup_{\|\theta - \theta_0\| \leq \kappa_m} \sqrt{m} \|\hat{\gamma}(\theta) - \hat{\gamma}(\theta_0) - \gamma(\theta)\| &\xrightarrow{p} 0, \\ \sup_{\theta \in \Theta} \|\hat{\gamma}(\theta) - \gamma(\theta)\| &\xrightarrow{p} 0.\end{aligned}$$

PROOF. See appendix section 9.4. ■

3.5 Consistency of $\hat{\theta}$

Theorem 5. (*Consistency of $\hat{\theta}$*): Suppose the following assumptions and those in Theorem 4 are satisfied:

- (i) $\|\hat{\gamma}(\hat{\theta})\|_{W_T} \leq o_p(1) + \inf_{\theta \in \Theta} \|\hat{\gamma}(\theta)\|_{W_T}$.
- (ii) $W_T = W + o_p(1)$ where W is positive definite.
- (iii) For every open set G that contains θ_0 , $\inf_{\theta \notin G} \|\gamma(\theta)\|_W > \|\gamma(\theta_0)\|_W$.

Then $\hat{\theta} \xrightarrow{P} \theta_0$.

PROOF. The proof is a direct application of the argmax continuous mapping theorem in [van der Vaart and Wellner \(1996\)](#) (Corollary 3.2.3) since Theorem 4 and assumption (ii) imply that the sample objective $\|\hat{\gamma}(\theta)\|_{W_T}$ converges uniformly to the population objective $\|\gamma(\theta)\|_W$. ■

3.6 \sqrt{m} -Consistency of $\hat{\theta}$

Our final goal is to derive the asymptotic distribution of $\sqrt{m}(\hat{\theta} - \theta_0)$.

Theorem 6. (*\sqrt{m} -Consistency and Asymptotic Normality of $\hat{\theta}$*): Suppose the following assumptions and those in Theorem 5 are satisfied:

- (i) $\Gamma \equiv \left. \frac{\partial}{\partial \theta} \lim_{T, R \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E \left[Z_t \left(\hat{\delta}_t(\theta_2) - X_t' \theta_1 \right) \right] \right|_{\theta_0}$ exists, and $\Gamma' W \Gamma$ is nonsingular.
- (ii) $\|\hat{\gamma}(\hat{\theta})\|_{W_T} \leq \inf_{\theta} \|\hat{\gamma}(\theta)\|_{W_T} + o_p(m^{-1})$.

Then, for $\Sigma = (1 \wedge k)\Omega + (1 \wedge 1/k)\Sigma_h$,

$$\sqrt{m}(\hat{\theta} - \theta_0) \xrightarrow{d} N \left(0, (\Gamma' W \Gamma)^{-1} \Gamma' W \Sigma W \Gamma (\Gamma' W \Gamma)^{-1} \right).$$

PROOF. The first four conditions in Theorem 7.2 of [Newey and McFadden \(1994\)](#) are satisfied by assumption (i) in Theorem 4 and assumptions (i)-(iii) of the present theorem. The fifth condition follows from Theorem 4. ■

4 Consistent Estimation of Variance of $\hat{\theta}$

This section discusses how to compute standard errors that take into account simulation noise. Using the formula for the asymptotic variance of $\hat{\theta}$, we can calculate the standard errors as the square root of the diagonal of the matrix:

$$\frac{1}{m} \left(\hat{\Gamma}' W_T \hat{\Gamma} \right)^{-1} \hat{\Gamma}' W_T \hat{\Sigma} W_T \hat{\Gamma} \left(\hat{\Gamma}' W_T \hat{\Gamma} \right)^{-1}$$

Let us describe each of the components separately. The Jacobian of the sample moments can be estimated using

$$\hat{\Gamma} \equiv \hat{\Gamma}(\hat{\theta}) = \frac{\partial \hat{\gamma}(\hat{\theta})}{\partial \theta} = \left[-\frac{1}{T} Z' X, \quad \frac{1}{T} Z' \frac{\partial \hat{\delta}_t(\theta)}{\partial \theta_2} \Big|_{\hat{\theta}} \right].$$

In order to estimate $\frac{\partial \hat{\delta}_t(\theta)}{\partial \theta_2} \Big|_{\hat{\theta}}$, recall that the estimated market shares are

$$s_{jt}(\hat{\delta}_t, X_t, \hat{F}; \hat{\theta}) = \int \frac{\exp(\hat{\delta}_{jt}(\hat{\theta}_2) + \mu_{rjt}(\hat{\theta}_2, v_r))}{1 + \sum_{k \in \mathcal{N}(t)} \exp(\hat{\delta}_{kt}(\hat{\theta}_2) + \mu_{rkt}(\hat{\theta}_2, v_r))} d\hat{F}(v_r) \equiv \frac{1}{R} \sum_{r=1}^R g_{jt}(\hat{\delta}_t, X_t, v_r; \hat{\theta}).$$

The fixed point solution $\hat{\delta}$ at each θ is found by equating the estimated market shares to the observed market shares:

$$s(\hat{\delta}, X, \hat{F}; \theta) = S.$$

The Implicit Function Theorem implies that $\hat{\delta}$ is continuously differentiable in θ . Since $\hat{\delta}$ does not depend explicitly on θ_1 , it suffices to compute the derivative with respect to θ_2 , evaluated at $\hat{\theta}_2$:

$$\begin{aligned} \frac{\partial \hat{\delta}_t(\theta)}{\partial \theta_2} \Big|_{\hat{\theta}} &= - \left(\frac{\partial s(\hat{\delta}, X, \hat{F}; \hat{\theta})}{\partial \delta} \right)^{-1} \left(\frac{\partial s(\hat{\delta}, X, \hat{F}; \hat{\theta})}{\partial \theta_2} \right) \\ &= - \left(\int G_{\delta}(\hat{\delta}, X, v_r; \hat{\theta}) d\hat{F}(v_r) \right)^{-1} \int G_{\theta_2}(\hat{\delta}, X, v_r; \hat{\theta}) d\hat{F}(v_r) \\ &= - \left(\frac{1}{R} \sum_{r=1}^R G_{\delta}(\hat{\delta}, X, v_r; \hat{\theta}) \right)^{-1} \frac{1}{R} \sum_{r=1}^R G_{\theta_2}(\hat{\delta}, X, v_r; \hat{\theta}). \end{aligned}$$

$G_\delta(\delta, X, v_r; \theta)$ is a $JT \times JT$ block diagonal matrix with $\frac{\partial g_{jt}}{\partial \delta_{jt}} = g_{jt}(1 - g_{jt})$ and $\frac{\partial g_{jt}}{\partial \delta_{kt}} = -g_{jt}g_{kt}$ and $G_{\theta_2}(\delta, X, v_r; \theta)$ is the $JT \times d$ Jacobian matrix of $g(\delta, X, F; \theta)$ with respect to θ_2 . The jt^{th} row of $G_{\theta_2}(\hat{\delta}, X, v_r; \hat{\theta})$ is given by, for $\hat{\mu}_{rjt} \equiv \mu_{rjt}(\hat{\theta}_2, v_r)$,

$$\frac{\exp(\hat{\delta}_{jt}(\hat{\theta}_2) + \hat{\mu}_{rjt})}{1 + \sum_{k \in \mathcal{N}(t)} \exp(\hat{\delta}_{kt}(\hat{\theta}_2) + \hat{\mu}_{rkt})} \left([1, x'_{jt}]' \circ v_r - \frac{\sum_{k \in \mathcal{N}(t)} \exp(\hat{\delta}_{kt}(\hat{\theta}_2) + \hat{\mu}_{rkt}) [1, x'_{kt}] \circ v_r}{1 + \sum_{k \in \mathcal{N}(t)} \exp(\hat{\delta}_{kt}(\hat{\theta}_2) + \hat{\mu}_{rkt})} \right).$$

As noted by [Berry et al. \(1995\)](#), a large number of simulation draws is required for accuracy of the numerical integrals in the expression for the derivative of $\hat{\delta}$ with respect to θ_2 .

The variance covariance matrix of the sample moments is estimated as follows:

$$\begin{aligned} \hat{\Sigma} &= \min\left(1, \frac{R}{T}\right) \hat{\Omega} + \min\left(1, \frac{T}{R}\right) \hat{\Sigma}_h, \\ \hat{\Omega} &= \frac{1}{T} \sum_{t=1}^T \left(Z_t \left(\hat{\delta}_t(\hat{\theta}_2) - X'_t \hat{\theta}_1 \right) \right) \left(Z_t \left(\hat{\delta}_t(\hat{\theta}_2) - X'_t \hat{\theta}_1 \right) \right)', \\ \hat{\Sigma}_h &= \frac{1}{R} \sum_{r=1}^R \hat{h}(v_r; \hat{\theta}) \hat{h}(v_r; \hat{\theta})', \\ \hat{h}(v_r; \hat{\theta}) &= -\frac{1}{T} \sum_{t=1}^T Z_t \left(\frac{1}{R} \sum_{r'=1}^R \nabla_{\delta} g_t(\hat{\delta}_t, X_t, v_{r'}; \hat{\theta}) \right)^{-1} \left(g_t(\hat{\delta}_t, X_t, v_r; \hat{\theta}) - \frac{1}{R} \sum_{r'=1}^R g_t(\hat{\delta}_t, X_t, v_{r'}; \hat{\theta}) \right). \end{aligned}$$

The optimal weighting matrix is estimated as

$$W_T = \hat{\Sigma}^{-1}.$$

Theorem 7. (*Consistent Estimate of Asymptotic Variance*): Suppose the conditions in all of the previous theorems are satisfied. In addition, suppose

- (i) There exists $\kappa_m \downarrow 0$ such that $E \left[\sup_{\|\theta - \theta_0\| \leq \kappa_m} \left\| Z_t \left(\hat{\delta}_t(\theta) - X'_t \theta_1 \right) \right\| \right] < \infty$.
- (ii) $\max_{r=1 \dots R} \max_{t=1 \dots T} \left\| g_t(\hat{\delta}_t, X_t, v_r; \hat{\theta}) - g_t(\delta_{0t}, X_t, v_r; \theta_0) \right\|_{\infty} \xrightarrow{p} 0$.

Then,

$$\widehat{AsyVar}[\hat{\theta}] = \left(\hat{\Gamma}' W_T \hat{\Gamma} \right)^{-1} \hat{\Gamma}' W_T \hat{\Sigma} W_T \hat{\Gamma} \left(\hat{\Gamma}' W_T \hat{\Gamma} \right)^{-1} \xrightarrow{p} \left(\Gamma' W \Gamma \right)^{-1} \Gamma' W \Sigma W \Gamma \left(\Gamma' W \Gamma \right)^{-1}.$$

PROOF. See appendix 9.5. ■

5 Consistency of Laplace type Estimator

Laplace type estimators (LTEs) provide a computationally attractive alternative to directly minimizing the GMM objective, which is nonconvex and contains many local minima. LTEs are typically computed using Markov Chain Monte Carlo (MCMC) methods which generate a series of parameter draws such that the marginal distribution of this series is approximately the quasi-posterior distribution of the parameters. It is well-known (see e.g. Chernozhukov and Hong (2003)) that LTEs can be more generally expressed as the minimizer of a quasi-posterior risk function formed using a convex loss function.

Theorem 8. (*Asymptotic Normality of Laplace Estimator*): *Suppose the following conditions and those in Theorem 6 are satisfied.*

(i) Θ is a convex, compact subset of \mathbb{R}^{2d+1} .

(ii) The loss function $\rho_m : \mathbb{R}^{2d+1} \rightarrow \mathbb{R}_+$ satisfies (1) $\rho_m(u) = \rho(\sqrt{m}u)$ where $\rho(u) \geq 0$ and $\rho(u) = 0$ iff $u = 0$. (2) ρ is convex and $\rho(h) \leq 1 + |h|^p$ for some $p \geq 1$. (3) $\rho(u) = \rho(-u)$.

(iii) $\pi : \Theta \rightarrow \mathbb{R}_+$ is a continuous, uniformly positive density function.

Then, for $p_m(\theta) = \frac{\exp(-m\hat{\gamma}(\theta)'W_T\hat{\gamma}(\theta))\pi(\theta)}{\int_{\Theta} \exp(-m\hat{\gamma}(\theta)'W_T\hat{\gamma}(\theta))\pi(\theta)d\theta}$,

$$\tilde{\theta} = \arg \min_{\theta \in \Theta} \int_{\Theta} \rho_m(\theta' - \theta)p_m(\theta')d\theta'$$

is consistent for θ_0 and has the same asymptotic distribution as $\hat{\theta}$:

$$\sqrt{m}(\tilde{\theta} - \theta_0) \xrightarrow{d} N\left(0, (\Gamma'W\Gamma)^{-1} \Gamma'W\Sigma W\Gamma (\Gamma'W\Gamma)^{-1}\right).$$

PROOF. See appendix section 9.6. ■

Examples of LTEs include the quasi-posterior mean, which corresponds to $\rho(u) = \|u\|_2^2$ and the

quasi-posterior median, which corresponds to $\rho(u) = \|u\|_1$.

6 Comparison with Freyberger (2015)

We now discuss how our results relate to Freyberger (2015)'s results for overlapping simulation draws. Freyberger (2015) conjectures without proof the following asymptotic distribution under overlapping simulation draws:

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, V_1 + \tilde{\lambda}^2 V_2).$$

$$\tilde{\lambda} = \lim_{T \rightarrow \infty, R \rightarrow \infty} \frac{\sqrt{T}}{\sqrt{R}} < \infty, V_1 = (\Gamma'W\Gamma)^{-1}(\Gamma'W\Phi_1W\Gamma)(\Gamma'W\Gamma)^{-1}, V_2 = (\Gamma'W\Gamma)^{-1}(\Gamma'W\Phi_2W\Gamma)(\Gamma'W\Gamma)^{-1}.$$

We show in the appendix section 9.7 that $\Phi_1 = \Omega$, $\Phi_2 = \Sigma_h + o(1)$, $\hat{\Phi}_1 = \hat{\Omega}$, and $\hat{\Phi}_2 = \hat{\Sigma}_h$. It turns out that our finite sample variance estimate is numerically identical to Freyberger (2015)'s estimate, regardless of whether $R \ll T$, $R \gg T$, or $R \sim T$. Thus we are not doing any better than Freyberger (2015) in terms of approximating the estimator's distribution in finite sample, but rather we are simply making the point that we can do away with Freyberger (2015)'s rate condition $\lim_{T \rightarrow \infty, R \rightarrow \infty} \frac{\sqrt{T}}{\sqrt{R}} < \infty$. The main difference between Freyberger (2015)'s asymptotic distribution and our asymptotic distribution is the scaling of \sqrt{T} versus \sqrt{m} . The scaling of \sqrt{m} allows the asymptotic bias to disappear even without the rate condition $\lim_{T \rightarrow \infty, R \rightarrow \infty} \frac{\sqrt{T}}{\sqrt{R}} < \infty$ imposed by Freyberger (2015). Intuitively, the scaling of \sqrt{m} allows our asymptotic variance to put different weights on the sampling variation and the simulation noise depending on how R is related to T . More weight is placed on the simulation noise Σ_h when $R \ll T$, and more weight is placed on the sampling variation Ω when $R \gg T$ or $R \sim T$. Freyberger (2015) does not allow for the simulation noise to dominate the sampling variation because he uses a scaling of \sqrt{T} which implicitly always places more weight on the sampling variation. Freyberger (2015)'s Taylor expansion of $\sqrt{T}(\hat{\theta} - \theta_0)$ is

$$\sqrt{T}(\hat{\theta} - \theta_0) = \left((\Gamma'W\Gamma)^{-1} \Gamma'W + o_p(1) \right) \left(Q_{1,T} + \frac{\sqrt{T}}{\sqrt{R}} Q_{2,T,R} + \frac{\sqrt{T}}{R} C_{1,T,R} + o_p \left(\frac{\sqrt{T}}{R} \right) \right)$$

where $Q_{1,T} \xrightarrow{d} N(0, \Phi_1)$, $Q_{2,T,R} \xrightarrow{d} N(0, \Phi_2)$, and $C_{1,T,R} \xrightarrow{p} \bar{\mu}$, which represents the asymptotic bias. The assumption $\tilde{\lambda} = \lim_{T \rightarrow \infty, R \rightarrow \infty} \frac{\sqrt{T}}{\sqrt{R}} < \infty$ implies that $\frac{\sqrt{T}}{R} C_{1,T,R} \rightarrow 0$, which means the asymptotic bias disappears.

If we scale Freyberger (2015)'s Taylor expansion by $\sqrt{m} = \sqrt{\min(R, T)}$, for $k = \lim_{T \rightarrow \infty, R \rightarrow \infty} \frac{R}{T}$,

$$\begin{aligned}
& \sqrt{m}(\hat{\theta} - \theta_0) \\
&= \frac{\sqrt{\min(R, T)}}{\sqrt{T}} \sqrt{T}(\hat{\theta} - \theta_0) \\
&= ((\Gamma'W\Gamma)^{-1}\Gamma'W + o_p(1)) \left(\frac{\sqrt{\min(R, T)}}{\sqrt{T}} Q_{1,T} + \frac{\sqrt{\min(R, T)}}{\sqrt{R}} Q_{2,T,R} + \frac{\sqrt{\min(R, T)}}{R} C_{1,T,R} \right. \\
&\quad \left. + o_p\left(\frac{\sqrt{\min(R, T)}}{R}\right) \right) \\
&= ((\Gamma'W\Gamma)^{-1}\Gamma'W + o_p(1)) \left(\min\left(1, \sqrt{k}\right) Q_{1,T} + \min\left(1, \sqrt{\frac{1}{k}}\right) Q_{2,T,R} + \frac{1}{\sqrt{R}} \min\left(1, \sqrt{\frac{1}{k}}\right) C_{1,T,R} \right. \\
&\quad \left. + o_p\left(\frac{1}{\sqrt{R}} \min\left(1, \sqrt{\frac{1}{k}}\right)\right) \right).
\end{aligned}$$

Since $0 \leq \min\left(1, \sqrt{\frac{1}{k}}\right) \leq 1$, $\frac{1}{\sqrt{R}} \min\left(1, \sqrt{\frac{1}{k}}\right) C_{1,T,R} \rightarrow 0$ as $R \rightarrow \infty$ even if $R \ll T$.

Now we consider our estimator's Taylor expansion. It follows from theorems 3 and 5 that

$$\begin{aligned}
& \sqrt{m}(\hat{\theta} - \theta_0) \\
&= \left((\Gamma'W\Gamma)^{-1}\Gamma'W + o_p(1) \right) \sqrt{m}\hat{\gamma}(\theta_0) \\
&= \left((\Gamma'W\Gamma)^{-1}\Gamma'W + o_p(1) \right) \left(\frac{\sqrt{m}}{\sqrt{T}} \sqrt{T}\hat{g}(\theta_0) + \frac{\sqrt{m}}{\sqrt{R}} \frac{1}{\sqrt{R}} \sum_{r=1}^R h(v_r; \theta_0) + O_p\left(\frac{\sqrt{m}}{R}\right) + O_p\left(\frac{\sqrt{m}}{\sqrt{TR}}\right) \right).
\end{aligned}$$

Matching the terms in our expansion with those in Freyberger (2015)'s,

$$Q_{1,T} = \sqrt{T}\hat{g}(\theta_0) \xrightarrow{d} N(0, \Omega), \quad Q_{2,R} = \frac{1}{\sqrt{R}} \sum_{r=1}^R h(v_r; \theta_0) \xrightarrow{d} N(0, \Sigma_h).$$

In the case of $R \ll T$, the leading bias term is $O_p\left(\frac{\sqrt{m}}{R}\right)$, which converges in probability to zero.

7 Monte Carlo

We conduct Monte Carlo simulations using empirical moments from the automobiles dataset of [Berry et al. \(1995\)](#) that is posted as supplementary material to [Knittel and Metaxoglou \(2014\)](#). The dataset contains information on automobiles sold in the twenty year period between 1971 and 1990. Because some models enter and exist during that period, there is an unbalanced panel with a total of 2217 observations. The characteristics that enter the utility function are the price of the vehicle, the ratio of horsepower to weight (HPWT), whether or not the vehicle has air conditioning (AIR), the number of ten mile increments one could drive for one dollar's worth of gasoline (MPD), and the size of the vehicle (SIZE). The instruments for price are the characteristics of the vehicle, the sum of the characteristics of all other vehicles produced by the same firm, and the sum of the characteristics of all vehicles produced by rival firms.

We generate data using certain statistics of the automobiles data. The vector of observed product characteristics x is drawn from a multivariate normal distribution with a mean vector equal to the sample mean of the auto characteristics and a covariance matrix equal to the sample covariance matrix of the auto characteristics. The unobservable product characteristics ξ are generated as the sum of two independent mean zero normal random variables each with variance equal to half of the sample variance of ξ . The additional instruments besides x are generated as the sum of attributes across vehicles produced by rival firms. When forming the GMM objective, $E[x_{jt}\xi_{jt}] = 0$ are included in the moment conditions because x are generated independently of ξ . The unobservable cost characteristics η are generated as 0.001 times the sum of v_1 and v_3 , where v_3 are drawn from a mean zero normal random variable with variance equal to the absolute difference between the sample variance of ξ and the sample covariance of price and ξ . Similar to [Berry \(1994\)](#), we assume a linear structure for marginal costs which depends only on the characteristics of the products: $mc_j = x_j'\gamma_0 + \eta_j$. Using some assumed true values for the cost parameter γ_0 and the means $(\theta_{0,1})$ and standard deviations $(\theta_{0,2})$ of the demand parameters, we simulate the market shares of the products and the outside good and compute prices for all products in a given market as the firms' best response functions in a game of Bertrand competition. Specifically, for each market $t = 1 \dots T$,

prices are determined as

$$p^* = \underset{p \geq 0}{\operatorname{argmax}} (p - mc) \left(\sum_{q=1}^Q \omega_q \frac{\exp \left(\begin{bmatrix} 1 & x_{jt} & p_{jt} \end{bmatrix} \theta_{0,1} + \xi_{jt} + \mu_{rjt}(\theta_{0,2}, v_q) \right)}{1 + \sum_{k \in \mathcal{N}(t)} \exp \left(\begin{bmatrix} 1 & x_{kt} & p_{kt} \end{bmatrix} \theta_{0,1} + \xi_{kt} + \mu_{rkt}(\theta_{0,2}, v_q) \right)} \right)_{j=1}^J,$$

where ω_q are the Gauss-Hermite quadrature weights and the v_q are the quadrature nodes.

Using the generated data and the computed prices, we apply the Hamiltonian Markov Chain Monte Carlo (HMCMC) algorithm to compute the LTE estimates of θ_0 . The benefit of using the HMCMC algorithm as opposed to a traditional Metropolis Hastings MCMC algorithm is that the former uses the gradient of the objective function to guide the search for the true parameter values while the latter is essentially a random walk which can take extremely long to converge for parameters of even moderate dimensions.

The assumed true values of θ_0 and γ_0 , with the exception of the coefficient on price, are obtained from Table IV in [Berry et al. \(1995\)](#) and repeated in Table 1.

Table 1: Assumed True Values for θ_0 and γ_0

	Price	Constant	HPWT	AIR	MPD	SIZE
$\theta_{0,1}$	-0.2	-7.304	2.185	0.579	-0.049	2.604
$\theta_{0,2}$		2.009	1.586	1.215	0.670	1.510
γ_0		0.726	0.313	0.290	0.293	1.499

We compare the empirical coverage frequencies of two types of confidence intervals:

$$\left[\hat{\theta}_1 \pm \frac{1.96}{\sqrt{m}} \left(\hat{\Gamma}' W_T \hat{\Gamma} \right)^{-1} \hat{\Gamma}' W_T \hat{\Sigma} W_T \hat{\Gamma} \left(\hat{\Gamma}' W_T \hat{\Gamma} \right)^{-1} \right]$$

$$\left[\hat{\theta}_1 \pm \frac{1.96}{\sqrt{n}} \left(\hat{\Gamma}' W_T \hat{\Gamma} \right)^{-1} \hat{\Gamma}' W_T \hat{\Omega} W_T \hat{\Gamma} \left(\hat{\Gamma}' W_T \hat{\Gamma} \right)^{-1} \right]$$

We use the posterior mean as $\hat{\theta}$ and the formulas provided in section 4 to compute $\hat{\Gamma}$, $\hat{\Sigma}$, and $\hat{\Omega}$. Results using the posterior median as $\hat{\theta}$ are very similar and are available upon request.

Table 2 provides the average $\hat{\theta}_1$ computed using HMCMC, the average bias of $\hat{\theta}_1$, the average $\hat{\theta}_2$, the average bias of $\hat{\theta}_2$, the empirical coverage frequencies for θ_1 taking into account simulation noise, and the empirical coverage frequencies for θ_1 not taking into account simulation noise in settings

with either 500 or 1000 markets and 20 products each. We use 4000 Markov Chain iterations and burn in the first 2000 periods.

With the exception of the case of 500 markets and 50 simulation draws where there is overcoverage, the empirical coverage is fairly close to the nominal level of 95%. However, we would like to point out that the good coverage does not necessarily imply the parameter estimates have the correct sign or are close to the truth. For example, in the case of 500 markets and 100 simulation draws, the mean price coefficient is positive. In the case of 500 markets and 50 simulation draws, the mean price coefficient is more than three times too large. In the case of 1000 markets and 50 simulation draws, the mean price coefficient is more than three times too small.

Table 2: HMCMC results

500 markets, 20 products, 100 Monte Carlo Simulations, 50 Simulation Draws						
	Price	Constant	HPWT	AIR	MPD	SIZE
average $\hat{\theta}_1$	-0.052	-8.000	2.057	0.540	-0.163	2.322
average bias of $\hat{\theta}_1$	0.148	-0.696	-0.128	-0.039	-0.114	-0.282
average $\hat{\theta}_2$		1.640	1.878	2.085	1.171	1.790
average bias of $\hat{\theta}_2$		-0.370	0.292	0.870	0.501	0.280
Empirical Coverage w/ simulation noise	0.990	0.990	0.990	0.990	0.990	0.980
Empirical Coverage w/o simulation noise	0.280	0.280	0.190	0.290	0.200	0.290
500 markets, 20 products, 100 Monte Carlo Simulations, 100 Simulation Draws						
	Price	Constant	HPWT	AIR	MPD	SIZE
average $\hat{\theta}_1$	0.020	-8.429	2.050	0.506	-0.136	2.191
average bias of $\hat{\theta}_1$	0.220	-1.125	-0.135	-0.073	-0.087	-0.413
average $\hat{\theta}_2$		2.081	2.456	1.688	0.473	2.114
average bias of $\hat{\theta}_2$		0.072	0.870	0.473	-0.197	0.604
Empirical Coverage w/ simulation noise	0.941	0.941	0.952	0.941	0.941	0.941
Empirical Coverage w/o simulation noise	0.262	0.262	0.286	0.262	0.238	0.262
1000 markets, 20 products, 50 Monte Carlo Simulations, 50 Simulation Draws						
	Price	Constant	HPWT	AIR	MPD	SIZE
average $\hat{\theta}_1$	-0.703	-4.628	2.456	0.751	0.104	3.499
average bias of $\hat{\theta}_1$	-0.503	2.676	0.271	0.172	0.153	0.895
average $\hat{\theta}_2$		1.411	1.843	2.214	1.525	1.036
average bias of $\hat{\theta}_2$		-0.599	0.257	0.999	0.855	-0.474
Empirical Coverage w/ simulation noise	0.95	0.95	0.95	0.95	0.95	0.95
Empirical Coverage w/o simulation noise	0.10	0.10	0.10	0.10	0.10	0.10

Next we compare our HMCMC routine to two other commonly used methods for computing the parameter estimates: the nested fixed point (NFP) method as implemented in [Conlon \(2014\)](#) and the mathematical programming with equilibrium constraints (MPEC) method as implemented

in Dubé et al. (2012). We used KNITRO 11.0.0 with the options specified by Dubé et al. (2012). Table 3 provides the average $\hat{\theta}_1$ computed using MPEC, the average bias of $\hat{\theta}_1$, the average $\hat{\theta}_2$, the average bias of $\hat{\theta}_2$, the empirical coverage frequencies for θ_1 taking into account simulation noise, and the empirical coverage frequencies for θ_1 in settings with either 500 or 1000 markets and 20 products each. We can see that MPEC's coverage frequencies are far from the nominal level and the average bias for some estimates can be quite high.

Table 3: MPEC results

500 markets, 20 products, 100 Monte Carlo Simulations, 50 Simulation Draws						
	Price	Constant	HPWT	AIR	MPD	SIZE
average $\hat{\theta}_1$	-5.520	21.053	4.635	2.424	1.437	11.658
average bias of $\hat{\theta}_1$	-5.320	28.357	2.450	1.845	1.486	9.054
average $\hat{\theta}_2$		1.205	2.537	2.366	0.755	2.296
average bias of $\hat{\theta}_2$		-0.804	0.951	1.151	0.085	0.786
Empirical Coverage w/ simulation noise	0.47	0.47	0.45	0.44	0.52	0.53
Empirical Coverage w/o simulation noise	0.11	0.10	0.07	0.08	0.13	0.11
500 markets, 20 products, 100 Monte Carlo Simulations, 100 Simulation Draws						
	Price	Constant	HPWT	AIR	MPD	SIZE
average $\hat{\theta}_1$	-2.373	4.410	2.948	1.383	0.519	6.201
average bias of $\hat{\theta}_1$	-2.173	11.714	0.763	0.804	0.568	3.597
average $\hat{\theta}_2$		0.910	2.176	1.913	0.687	1.728
average bias of $\hat{\theta}_2$		-1.099	0.590	0.698	0.017	0.218
Empirical Coverage w/ simulation noise	0.36	0.35	0.33	0.37	0.37	0.36
Empirical Coverage w/o simulation noise	0.06	0.05	0.06	0.06	0.07	0.05
1000 markets, 20 products, 100 Monte Carlo Simulations, 50 Simulation Draws						
	Price	Constant	HPWT	AIR	MPD	SIZE
average $\hat{\theta}_1$	-0.675	-4.798	1.983	0.730	0.014	3.424
average bias of $\hat{\theta}_1$	-0.475	2.506	-0.202	0.151	0.063	0.820
average $\hat{\theta}_2$		1.345	2.100	1.857	0.693	1.397
average bias of $\hat{\theta}_2$		-0.664	0.514	0.642	0.023	-0.113
Empirical Coverage w/ simulation noise	0.60	0.59	0.59	0.58	0.57	0.59
Empirical Coverage w/o simulation noise	0.11	0.10	0.08	0.08	0.11	0.12

Table 4 provides the average $\hat{\theta}_1$ computed using NFP, the average bias of $\hat{\theta}_1$, the average $\hat{\theta}_2$, the average bias of $\hat{\theta}_2$, the empirical coverage frequencies for θ_1 taking into account simulation noise, and the empirical coverage frequencies for θ_1 in settings with either 500 or 1000 markets and 20 products each. We can see that NFP typically overcovers, and the average bias for some estimates can be quite high.

Table 5 shows the average across-market means, mins, maxes, and quantiles of the estimated

Table 4: NFP results

500 markets, 20 products, 100 Monte Carlo Simulations, 50 Simulation Draws						
	Price	Constant	HPWT	AIR	MPD	SIZE
average $\hat{\theta}_1$	2.175	0.078	-8.995	1.362	0.490	-0.185
average bias of $\hat{\theta}_1$	2.375	7.382	-11.180	0.783	0.539	-2.789
average $\hat{\theta}_2$		0.942	2.126	2.113	0.618	1.253
average bias of $\hat{\theta}_2$		-1.068	0.540	0.898	-0.052	-0.257
Empirical Coverage w/ simulation noise	1	1	1	1	1	1
Empirical Coverage w/o simulation noise	1	1	0.83	1	1	1
500 markets, 20 products, 100 Monte Carlo Simulations, 100 Simulation Draws						
	Price	Constant	HPWT	AIR	MPD	SIZE
average $\hat{\theta}_1$	1.645	0.373	-10.601	1.414	0.421	-0.251
average bias of $\hat{\theta}_1$	1.845	7.677	-12.786	0.835	0.470	-2.855
average $\hat{\theta}_2$		1.005	2.076	2.165	0.621	1.321
average bias of $\hat{\theta}_2$		-1.004	0.490	0.950	-0.049	-0.189
Empirical Coverage w/ simulation noise	1	1	0.98	1	1	1
Empirical Coverage w/o simulation noise	1	1	0.77	0.99	1	1
1000 markets, 20 products, 100 Monte Carlo Simulations, 50 Simulation Draws						
	Price	Constant	HPWT	AIR	MPD	SIZE
average $\hat{\theta}_1$	8.997	-3.934	12.595	3.456	1.944	1.005
average bias of $\hat{\theta}_1$	9.197	3.370	10.410	2.877	1.993	-1.599
average $\hat{\theta}_2$		1.370	2.681	1.896	0.709	1.995
average bias of $\hat{\theta}_2$		-0.639	1.095	0.681	0.039	0.485
Empirical Coverage w/ simulation noise	1	1	0.96	1	1	1
Empirical Coverage w/o simulation noise	1	0.98	0.70	0.98	1	0.98

own price elasticities for the first product. The elasticities of the other 19 products are very similar and available upon request. We compute the bias between the price elasticities computed using the estimated parameter values versus those computed using the true parameter values. We can see that the elasticities computed using the HMCMC parameter estimates typically have smaller bias than those computed using the MPEC or NFP parameter estimates. The one exception is the case of 1000 markets, 20 products, and 50 Simulation Draws, in which case HMCMC has similar bias as MPEC, but still smaller bias than NFP.

We acknowledge that improvements in the various computational methods are on-going (see e.g. [Conlon and Gortmaker \(2019\)](#)) and remain a good subject for further research.

Table 5: Means, Mins, Maxes, and Quantiles of Own Price Elasticities for 1st Product
500 markets, 20 products, 50 Simulation Draws

	Mean	Min	10th	25th	50th	75th	90th	Max
HMCMC	-0.408	-0.468	-0.442	-0.431	-0.418	-0.399	-0.362	-0.180
Bias	1.211	1.347	1.272	1.245	1.214	1.183	1.154	0.844
MPEC	-43.909	-51.601	-47.327	-46.180	-44.802	-42.922	-39.445	-21.412
Bias	-42.290	-49.785	-45.613	-44.503	-43.169	-41.339	-37.929	-20.388
NFP	17.063	5.227	14.699	16.679	17.624	18.238	18.729	20.049
Bias	18.682	7.042	16.413	18.355	19.256	19.820	20.245	21.073

500 markets, 20 products, 100 Simulation Draws

	Mean	Min	10th	25th	50th	75th	90th	Max
HMCMC	0.157	0.080	0.145	0.154	0.160	0.164	0.168	0.178
Bias	1.776	1.895	1.858	1.831	1.793	1.747	1.684	1.202
MPEC	-19.091	-21.664	-20.342	-19.892	-19.356	-18.688	-17.622	-10.923
Bias	-17.473	-19.849	-18.628	-18.215	-17.723	-17.106	-16.106	-9.899
NFP	12.876	4.319	10.940	12.506	13.310	13.821	14.228	15.337
Bias	14.495	6.135	12.663	14.183	14.943	15.403	15.744	16.361

1000 markets, 20 products, 50 Simulation Draws

	Mean	Min	10th	25th	50th	75th	90th	Max
HMCMC	-5.536	-6.516	-6.023	-5.876	-5.691	-5.415	-4.852	-2.030
Bias	-3.917	-4.685	-4.309	-4.199	-4.058	-3.833	-3.336	-1.056
MPEC	-5.449	-6.218	-5.800	-5.673	-5.521	-5.336	-5.057	-2.799
Bias	-3.830	-4.388	-4.087	-3.996	-3.888	-3.753	-3.540	-1.824
NFP	72.013	0.034	66.654	73.267	75.728	77.727	79.357	84.910
Bias	73.632	1.865	68.367	74.944	77.361	79.309	80.874	85.884

8 Conclusion

This paper has derived the asymptotic distribution of the parameters of the BLP model in the case of overlapping simulation draws. Asymptotics have been performed by sending the number of simulation draws and the number of markets to infinity but keeping the number of products in each market fixed. By writing the sample moment condition as a two-sample U-statistic, we have shown that the simulated GMM estimator is asymptotically normal. Our results have allowed for the case where $R \ll T$ as long as both $R \rightarrow \infty$ and $T \rightarrow \infty$. We have derived the form of the asymptotic variance that accounts for both simulation variance and sampling variance and have also provided a consistent estimate which can be used to form asymptotically valid confidence intervals. To improve the computational performance, we have proposed using a Laplace-type estimator implemented using Hamiltonian Markov Chain Monte Carlo (HMCMC). We have demonstrated consistency of the Laplace-type estimator and have illustrated through Monte Carlo simulations

the better performance of our estimator compared to Nested Fixed Point and MPEC in terms of coverage and bias.

9 Appendix

9.1 Definitions and Theorems about Functionals

Definition. Suppose $f : U \rightarrow Y$ is a mapping from an open subset $U \subset X$ of a Banach space to another Banach space Y . Then, f is **Fréchet Differentiable** at $u_0 \in U$ if there is a bounded linear map $Df(u_0) : X \rightarrow Y$ such that for every $\epsilon > 0$, there is a $\delta > 0$ such that whenever $0 < \|u - u_0\| < \delta$, we have

$$\frac{\|f(u) - f(u_0) - Df(u_0) \cdot (u - u_0)\|}{\|u - u_0\|} < \epsilon.$$

The **Fréchet Derivative** of f at u_0 , $Df(u_0)$, is related to the directional derivative (sometimes called the Gateaux Derivative) of f at u_0 in the direction h :

$$Df(u_0) \cdot h = \lim_{t \rightarrow 0} \frac{f(u_0 + th) - f(u_0)}{t} \equiv f'_{u_0}(h).$$

The Mean Value Theorem can be extended to Fréchet differentiable functionals.

Theorem. (*Mean Value Theorem*) Let $U \subset X$ be an open and convex subset of a Banach space X and let $f : U \rightarrow Y$ be a C^1 mapping from U to a Banach space Y . For $u, v \in U$, assume $\{(1-t)u + tv | t \in [0, 1]\} \subset U$. Then,

$$\begin{aligned} f(v) - f(u) &= \int_0^1 Df((1-t)u + tv) dt \cdot (v - u) \\ &= Df(u) \cdot (v - u) + \int_0^1 (Df((1-t)u + tv) - Df(u)) dt \cdot (v - u). \end{aligned}$$

Corollary. (*Intermediate Value Theorem*) Let $U \subset X$ be an open convex subset of a Banach space X and let $f : U \rightarrow \mathbb{R}$ be C^1 map. For all $u, v \in U$, there exists a $c = (1-t)u + tv$ for some $t \in [0, 1]$ such that $f(v) - f(u) = Df(c) \cdot (v - u)$.

9.2 Proof of Proposition 2

PROOF. We will show asymptotic normality of $\sqrt{R}(\hat{\delta}(\theta_0) - \delta_0)$ by applying a Taylor expansion to $\sqrt{R}(s(\hat{\delta}, X, \hat{F}; \theta_0) - s(\delta_0, X, F_0; \theta_0))$ around some $\delta^* = t_\delta \delta_0 + (1 - t_\delta) \hat{\delta}(\theta_0)$ and $F^* = t_F F_0 + (1 - t_F) \hat{F}$ for $t_F \in [0, 1]$ and $t_\delta \in [0, 1]$:

$$D_\delta s(\delta^*, X, F^*; \theta_0) \cdot (\hat{\delta}(\theta_0) - \delta_0) + D_F s(\delta^*, X, F^*; \theta_0) \cdot (\hat{F} - F_0).$$

In order to apply the Intermediate Value Theorem for functionals, we need to show that $s(\delta, X, F; \theta_0)$ is Fréchet differentiable in δ and F , that $D_\delta s(\delta, X, F; \theta_0)$ is continuous in δ , and that $D_F s(\delta, X, F; \theta_0)$ is continuous in F .

First we compute the directional derivative of $s(\delta, X, F; \theta_0)$ in the direction $\hat{\delta}(\theta_0) - \delta_0$ evaluated at δ^* and F^* . Since $g(\delta, X, v_r; \theta_0) \in (0, 1)$, we can use Lebesgue's Bounded Convergence Theorem to interchange integration and differentiation.

$$\begin{aligned} s'_{\delta^*}(\hat{\delta}(\theta_0) - \delta_0) &= \lim_{t \rightarrow 0} \frac{\int g(\delta^* + t(\hat{\delta}(\theta_0) - \delta_0), X, v_r; \theta_0) dF^*(v_r) - \int g(\delta^*, X, v_r; \theta_0) dF^*(v_r)}{t} \\ &= \int \lim_{t \rightarrow 0} \frac{g(\delta^* + t(\hat{\delta}(\theta_0) - \delta_0), X, v_r; \theta_0) - g(\delta^*, X, v_r; \theta_0)}{t} dF^*(v_r) \\ &= \int G_\delta(\delta^*, X, v_r; \theta_0) \cdot (\hat{\delta}(\theta_0) - \delta_0) dF^*(v_r) \\ &= \int G_\delta(\delta^*, X, v_r; \theta_0) dF^*(v_r) \cdot (\hat{\delta}(\theta_0) - \delta_0). \end{aligned}$$

The third equality follows from the fact that $g(\delta, X, v_r; \theta)$ is differentiable in δ . Note that $s'_\delta(\cdot)$ is a linear map for all δ and F because for all $\lambda_1, \lambda_2 \in \mathbb{R}$, $h_1, h_2 \in \mathbb{R}^{JT}$,

$$\begin{aligned} s'_\delta(\lambda_1 h_1 + \lambda_2 h_2) &= \int G_\delta(\delta, X, v_r; \theta_0) dF(v_r) \cdot (\lambda_1 h_1 + \lambda_2 h_2) \\ &= \lambda_1 \int G_\delta(\delta, X, v_r; \theta_0) dF(v_r) \cdot h_1 + \lambda_2 \int G_\delta(\delta, X, v_r; \theta_0) dF(v_r) \cdot h_2 \\ &= \lambda_1 s'_\delta(h_1) + \lambda_2 s'_\delta(h_2). \end{aligned}$$

$s'_\delta(\cdot)$ is also a bounded map for all δ because the elements of $G_\delta(\delta, X, v_r; \theta_0)$ lie in $(-1, 0) \cup (0, 1)$

for all δ , X , and v_r :

$$\|s'_\delta(h_1)\| \leq \left\| \int G_\delta(\delta, X, v_r; \theta_0) dF(v_r) \right\| \|h_1\| \leq \|h_1\|.$$

Therefore we have shown that $s(\delta, X, F; \theta_0)$ is Fréchet differentiable in δ and we can write $s'_{\delta^*}(\hat{\delta}(\theta_0) - \delta_0) = D_\delta s(\delta^*, X, F; \theta_0) \cdot (\hat{\delta}(\theta_0) - \delta_0)$.

We can show that $D_\delta s(\delta, X, F; \theta_0)$ is continuous in δ by noting that $G_\delta(\delta, X, v_r; \theta_0)$ is continuous in δ . For all $\epsilon > 0$, there exists $\nu > 0$ such that $\|\delta - \delta'\| < \nu \implies \|G_\delta(\delta, X, v_r; \theta_0) - G_\delta(\delta', X, v_r; \theta_0)\| < \epsilon$. Then,

$$\begin{aligned} \|D_\delta s(\delta, X, F; \theta_0) - D_\delta s(\delta', X, F; \theta_0)\| &= \left\| \int G_\delta(\delta, X, v_r; \theta_0) dF(v_r) - \int G_\delta(\delta', X, v_r; \theta_0) dF(v_r) \right\| \\ &\leq \int \|G_\delta(\delta, X, v_r; \theta_0) - G_\delta(\delta', X, v_r; \theta_0)\| dF(v_r) \\ &\leq \int \epsilon dF(v_r) = \epsilon. \end{aligned}$$

The directional derivative of $s(\delta, X, F; \theta_0)$ in the direction $\hat{F} - F_0$ evaluated at δ^* and F^* is

$$\begin{aligned} s'_{F^*}(\hat{F} - F_0) &= \lim_{t \rightarrow 0} \left[\frac{1}{t} \left(s(\delta^*, X, F^* + t(\hat{F} - F_0); \theta_0) - s(\delta^*, X, F^*; \theta_0) \right) \right] \\ &= \lim_{t \rightarrow 0} \left[\frac{1}{t} \left(\int g(\delta^*, X, v_r; \theta_0) d(F^* + t(\hat{F} - F_0))(v_r) - \int g(\delta^*, X, v_r; \theta_0) dF^*(v_r) \right) \right] \\ &= \lim_{t \rightarrow 0} \left[\frac{1}{t} \left(\int g(\delta^*, X, v_r; \theta_0) d(t(\hat{F} - F_0))(v_r) \right) \right] \\ &= \int g(\delta^*, X, v_r; \theta_0) d(\hat{F} - F_0)(v_r). \end{aligned}$$

Note that $s'_F(\cdot)$ is a linear map for all δ since for all $\lambda_1, \lambda_2 \in \mathbb{R}$, $F_1, F_2 \in \mathbb{F}$,

$$\begin{aligned} s'_F(\lambda_1 F_1 + \lambda_2 F_2) &= \int g(\delta, X, v_r; \theta_0) d(\lambda_1 F_1 + \lambda_2 F_2)(v_r) \\ &= \lambda_1 \int g(\delta, X, v_r; \theta_0) dF_1(v_r) + \lambda_2 \int g(\delta, X, v_r; \theta_0) dF_2(v_r) \\ &= \lambda_1 s'_F(F_1) + \lambda_2 s'_F(F_2). \end{aligned}$$

$s'_F(\cdot)$ is also a bounded map for all δ and F because $g(\delta, X, v_r; \theta_0) \in (0, 1)$ for all δ , X , and v_r . For all $F_1 \in \mathbb{F}$,

$$\|s'_F(F_1)\| = \left\| \int g(\delta, X, v_r; \theta_0) dF_1(v_r) \right\| \leq \left\| \int dF_1(v_r) \right\| = \|F_1\|.$$

Therefore we have shown that $s(\delta, X, F; \theta_0)$ is Fréchet differentiable in F and we can write $s'_{F^*}(\hat{F} - F_0) \equiv D_F s(\delta^*, X, F^*; \theta_0) \cdot (\hat{F} - F_0)$.

To show that $D_F s(\delta, X, F; \theta_0)$ is continuous in F , we will show that for all $\epsilon > 0$, there exists $0 < \nu < \infty$ such that $\|F - F'\| < \nu \implies \|\int g(\delta, X, v_r; \theta_0) d(F' - F)(v_r)\| < \epsilon$. Since $\|\int g(\delta, X, v_r; \theta_0) d(F' - F)(v_r)\| \leq \|\int |g(\delta, X, v_r; \theta_0)| d(F' - F)(v_r)\| \leq \|\int d(F' - F)(v_r)\| = \|F - F'\|$, we can take $\epsilon = \nu$.

Now that we have checked that the Fréchet derivatives of $s(\delta, X, F; \theta_0)$ are continuous with respect to δ and F , we can apply the Intermediate Value Theorem to $s(\hat{\delta}, X, \hat{F}; \theta_0) - s(\delta_0, X, F_0; \theta_0)$. Furthermore, [Berry \(1994\)](#) showed that for all $X \in \mathbb{R}^{d+1}$, there exists a unique $\hat{\delta}(\theta_0)$ that solves $s(\hat{\delta}, X, \hat{F}; \theta_0) - s(\delta_0, X, F_0; \theta_0) = 0$. Therefore,

$$\begin{aligned} 0 &= \sqrt{R} \left(s(\hat{\delta}, X, \hat{F}; \theta_0) - s(\delta_0, X, F_0; \theta_0) \right) \\ &= \sqrt{R} \left(D_\delta s(\delta^*, X, F^*; \theta_0) \cdot (\hat{\delta}(\theta_0) - \delta_0) + D_F s(\delta^*, X, F^*; \theta_0) \cdot (\hat{F} - F_0) \right) \\ &= \sqrt{R} \left((\hat{\delta}(\theta_0) - \delta_0) \left(\int G_\delta(\delta^*, X, v_r; \theta_0) dF^*(v_r) \right) + \int g(\delta^*, X, v_r; \theta_0) d(\hat{F} - F_0)(v_r) \right). \end{aligned}$$

Rearranging,

$$\begin{aligned} \sqrt{R} (\hat{\delta}(\theta_0) - \delta_0) &= - \left(\int G_\delta(\delta^*, X, v_r; \theta_0) dF^*(v_r) \right)^{-1} \sqrt{R} \int g(\delta^*, X, v_r; \theta_0) d(\hat{F} - F_0)(v_r) \\ &= - \left(\int G_\delta(\delta^*, X, v_r; \theta_0) dF^*(v_r) \right)^{-1} \frac{1}{\sqrt{R}} \sum_{r=1}^R \{g(\delta^*, X, v_r; \theta_0) - E_v[g(\delta^*, X, v; \theta_0)]\}. \end{aligned}$$

We will show that $\sqrt{R} (\hat{\delta}(\theta_0) - \delta_0)$ converges in distribution to a Gaussian random variable. To do so, we first need to show that $\mathcal{G} = \{g(\delta, X, v_r; \theta_0) : \delta \in \mathbb{D}\}$ is a Donsker class. We will show that $g(\delta, X, v_r; \theta_0)$ is Lipschitz in δ with a uniformly bounded Lipschitz constant, which is an example of a parametric class ([van der Vaart \(1998\)](#)). Since $g(\delta, X, v_r; \theta_0)$ is continuously differentiable in

δ , the intermediate value theorem implies that for all $\delta_1, \delta_2 \in \mathbb{D}$, there exists $\tilde{\delta} \in [\delta_1, \delta_2]$ such that

$$g(\delta_1, X, v_r; \theta_0) - g(\delta_2, X, v_r; \theta_0) = G_\delta(\tilde{\delta}, X, v_r; \theta_0)(\delta_1 - \delta_2).$$

Recall that all of the elements of $G_\delta(\delta, X, v_r; \theta_0)$ lie in $(-1, 0) \cup (0, 1)$ since $\frac{\partial g_{jt}}{\partial \delta_{jt}} = g_{jt}(1 - g_{jt})$ for all $j = 1 \dots J, t = 1 \dots T$, and $\frac{\partial g_{jt}}{\partial \delta_{kt}} = -g_{jt}g_{kt}$ for all $k \neq j, t = 1 \dots T$. Therefore, $\sup_{\delta \in \mathbb{D}} \|G_\delta(\delta, X, v_r; \theta_0)\| < \infty$ and $g(\delta, X, v_r; \theta_0)$ is Lipschitz in δ with uniformly bounded Lipschitz constant $\|G_\delta(\delta, X, v_r; \theta_0)\|$. Since we also showed in lemma 1 that $\int G_\delta(\delta, X, v; \theta) dF(v)$ is nonsingular for all δ, X, θ , and F , it follows that $\sqrt{R}(\hat{\delta}(\theta_0) - \delta_0) = O_p(1)$.

Next, note that the intermediate value theorem implies there exists $\tilde{\delta} \in [\delta^*, \delta_0]$ such that

$$\begin{aligned} & \frac{1}{\sqrt{R}} \sum_{r=1}^R \{g(\delta^*, X, v_r; \theta_0) - E_v[g(\delta^*, X, v; \theta_0)]\} \\ &= \frac{1}{\sqrt{R}} \sum_{r=1}^R \{g(\delta_0, X, v_r; \theta_0) - E_v[g(\delta_0, X, v; \theta_0)]\} \\ &+ \frac{1}{R} \sum_{r=1}^R \left(G_\delta(\tilde{\delta}, X, v_r; \theta_0) - E_v[G_\delta(\tilde{\delta}, X, v; \theta_0)] \right) \sqrt{R}(\delta^* - \delta_0). \end{aligned}$$

Note that all of the elements of $\nabla_\delta^2 g(\delta, X, v_r; \theta_0)$ lie in $(-2, 0) \cup (0, 2)$ since for all $t = 1 \dots T$, $\frac{\partial^2 g_{jt}}{\partial \delta_{jt}^2} = g_{jt}(1 - g_{jt})^2 - 2g_{jt}^2(1 - g_{jt})$ for all $j = 1 \dots J$, $\frac{\partial^2 g_{jt}}{\partial \delta_{kt}^2} = -g_{jt}(1 - g_{kt})g_{kt} + g_{jt}g_{kt}^2$ for all $k \neq j$, $\frac{\partial^2 g_{jt}}{\partial \delta_{kt} \partial \delta_{jt}} = -g_{jt}(1 - g_{jt})g_{kt} + g_{jt}^2g_{kt}$ for all $k \neq j$, and $\frac{\partial^2 g_{jt}}{\partial \delta_{kt} \partial \delta_{ht}} = -2g_{jt}g_{kt}g_{ht}$ for all $k \neq j, h \neq j, k \neq h$. Therefore, $\sup_{\delta \in \mathbb{D}} \|\nabla_\delta^2 g(\delta, X, v_r; \theta_0)\| < \infty$ and $\nabla \mathcal{G} = \{G_\delta(\delta, X, v_r; \theta_0) : \delta \in \mathbb{D}\}$ is a parametric class and therefore a Donsker class. It follows then that

$$\frac{1}{R} \sum_{r=1}^R \left(G_\delta(\tilde{\delta}, X, v_r; \theta_0) - E_v[G_\delta(\tilde{\delta}, X, v; \theta_0)] \right) = O_p\left(\frac{1}{\sqrt{R}}\right).$$

Note that δ^* is also \sqrt{R} -consistent for δ_0 since it lies between the \sqrt{R} -consistent estimator $\hat{\delta}(\theta_0)$ and δ_0 . Since \hat{F} is consistent for F_0 , and F^* lies between \hat{F} and F_0 , F^* is consistent for F_0 . The continuous mapping theorem implies that

$$\left(\int G_\delta(\delta^*, X, v_r; \theta_0) dF^*(v_r) \right)^{-1} = \left(\int G_\delta(\delta_0, X, v_r; \theta_0) dF_0(v_r) \right)^{-1} + o_p(1).$$

Therefore,

$$\begin{aligned} & \sqrt{R} \left(\hat{\delta}(\theta_0) - \delta_0 \right) \\ &= - \left(\int G_{\delta}(\delta_0, X, v_r; \theta_0) dF_0(v_r) \right)^{-1} \frac{1}{\sqrt{R}} \sum_{r=1}^R \{g(\delta_0, X, v_r; \theta_0) - E_v[g(\delta_0, X, v; \theta_0)]\} + O_p \left(\frac{1}{\sqrt{R}} \right). \end{aligned}$$

■

9.3 Proof of Theorem 3

PROOF. From proposition 2, for all $t = 1 \dots T$,

$$\begin{aligned} & \sqrt{R} \left(\hat{\delta}_t(\theta_0) - \delta_{0t} \right) \\ &= - \left(\int \nabla_{\delta} g_t(\delta_{0t}, X_t, v; \theta_0) dF_0(v) \right)^{-1} \frac{1}{\sqrt{R}} \sum_{r=1}^R (g_t(\delta_{0t}, X_t, v_r; \theta_0) - E_v[g_t(\delta_{0t}, X_t, v; \theta_0)]) + O_p \left(\frac{1}{\sqrt{R}} \right). \end{aligned}$$

The sample and population moments are

$$\begin{aligned} \hat{\gamma}(\theta) &= \frac{1}{T} \sum_{t=1}^T Z_t \left(\hat{\delta}_t(\theta) - X_t' \theta_1 \right) \\ \gamma(\theta) &= \lim_{T, R \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E \left[Z_t \left(\hat{\delta}_t(\theta) - X_t' \theta_1 \right) \right]. \end{aligned}$$

Note that $\gamma(\theta_0) = E[Z_t(\delta_{0t} - X_t' \theta_{0,1})] = E[Z_t \xi_t] = 0$.

Let $\tilde{g}(\theta_0) = \frac{1}{T} \sum_{t=1}^T Z_t (\delta_{0t} - X_t' \theta_{0,1})$. Then $\sqrt{T} \tilde{g}(\theta_0) \xrightarrow{d} N(0, \Omega)$, where

$$\Omega = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \text{Var} (Z_t (\delta_{0t} - X_t' \theta_{0,1})) = \text{Var} (Z_t (\delta_{0t} - X_t' \theta_{0,1})).$$

Our sample moments scaled by \sqrt{m} , where $m = \min(R, T)$, can be rewritten as

$$\begin{aligned}
& \sqrt{m}(\hat{\gamma}(\theta_0) - \tilde{g}(\theta_0) + \tilde{g}(\theta_0)) \\
&= \frac{\sqrt{m}}{\sqrt{R}} \sqrt{R} \frac{1}{T} \sum_{t=1}^T Z_t (\hat{\delta}_t(\theta_0) - \delta_{0t}) + \frac{\sqrt{m}}{\sqrt{T}} \sqrt{T} \tilde{g}(\theta_0) \\
&= -\frac{\sqrt{m}}{\sqrt{R}} \frac{1}{T} \sum_{t=1}^T Z_t \left(\int \nabla_{\delta} g_t(\delta_{0t}, X_t, v; \theta_0) dF_0(v) \right)^{-1} \frac{1}{\sqrt{R}} \sum_{r=1}^R (g_t(\delta_{0t}, X_t, v_r; \theta_0) - E_v[g_t(\delta_{0t}, X_t, v; \theta_0)]) \\
&+ O_p\left(\frac{\sqrt{m}}{R}\right) + \frac{\sqrt{m}}{\sqrt{T}} \sqrt{T} \tilde{g}(\theta_0) \\
&= \sqrt{m} \frac{1}{TR} S_{TR}(\theta_0) + O_p\left(\frac{\sqrt{m}}{R}\right) + \frac{\sqrt{m}}{\sqrt{T}} \sqrt{T} \tilde{g}(\theta_0).
\end{aligned}$$

Since $X_1, \dots, X_T, Z_1, \dots, Z_T, \xi_1, \dots, \xi_T$ and v_1, \dots, v_R are drawn i.i.d. from two different independent distributions, the first term is a two-sample U-statistic:

$$\begin{aligned}
\frac{1}{TR} S_{TR}(\theta_0) &= \frac{1}{TR} \sum_{t=1}^T \sum_{r=1}^R q(Z_t, X_t, v_r; \theta_0, \delta_{0t}) \\
q(Z_t, X_t, v_r; \theta_0, \delta_{0t}) &= -Z_t \left(\int \nabla_{\delta} g_t(\delta_{0t}, X_t, v; \theta_0) dF_0(v) \right)^{-1} (g_t(\delta_{0t}, X_t, v_r; \theta_0) - E_v[g_t(\delta_{0t}, X_t, v; \theta_0)]).
\end{aligned}$$

We can decompose the two sample U-statistic into the sum of two projection terms and a remainder term. The first projection term integrates out the v_r while the second term integrates out the Z_t , X_t , and ξ_t .

$$\frac{1}{TR} S_{TR}(\theta_0) = \frac{1}{T} \sum_{t=1}^T f(Z_t, \delta_{0t}, X_t; \theta_0) + \frac{1}{R} \sum_{r=1}^R h(v_r; \theta_0) + \frac{1}{TR} \sum_{t=1}^T \sum_{r=1}^R \tilde{q}(Z_t, X_t, v_r; \theta_0, \delta_{0t}).$$

$$\begin{aligned}
& f(Z_t, \delta_{0t}, X_t; \theta_0) \\
&= \int q(Z_t, X_t, v_r; \theta_0, \delta_{0t}) dF_0(v_r) \\
&= -Z_t \int \left(\int \nabla_{\delta} g_t(\delta_{0t}, X_t, v; \theta_0) dF_0(v) \right)^{-1} \int (g_t(\delta_{0t}, X_t, v_r; \theta_0) - E_v[g_t(\delta_{0t}, X_t, v; \theta_0)]) dF_0(v_r) \\
&= 0.
\end{aligned}$$

$$\begin{aligned}
& h(v_r; \theta_0) \\
&= \int q(Z_t, X_t, v_r; \theta_0, \delta_{0t}) dP(Z_t, X_t, \xi_t) \\
&= - \int \left\{ Z_t \left(\int \nabla_{\delta} g_t(\delta_{0t}, X_t, v; \theta_0) dF_0(v) \right)^{-1} (g_t(\delta_{0t}, X_t, v_r; \theta_0) - E_v[g_t(\delta_{0t}, X_t, v; \theta_0)]) \right\} dP(Z_t, X_t, \xi_t).
\end{aligned}$$

Neumeyer (2004) showed that

$$\frac{1}{TR} \sum_{t=1}^T \sum_{r=1}^R \tilde{q}(Z_t, X_t, v_r; \theta_0, \delta_{0t}) = O_p \left(\frac{1}{\sqrt{TR}} \right).$$

It follows from the Lindeberg-Levy central limit theorem that

$$\begin{aligned}
\sqrt{m} \frac{1}{TR} S_{TR}(\theta_0) &= \frac{\sqrt{m}}{\sqrt{R}} \frac{1}{\sqrt{R}} \sum_{r=1}^R h(v_r; \theta_0) + O_p \left(\frac{\sqrt{m}}{\sqrt{TR}} \right) \\
&\xrightarrow{d} N(0, (1 \wedge k) \Sigma_f + (1 \wedge 1/k) \Sigma_h).
\end{aligned}$$

where $k = \lim_{T \rightarrow \infty, R \rightarrow \infty} \frac{R}{T}$, $\Sigma_f = \text{Var}(f(Z_t, \delta_{0t}, X_t; \theta_0)) = 0$, and $\Sigma_h = \text{Var}(h(v_r; \theta_0))$.

Since $\{X_t, Z_t, \xi_t\}_{t=1}^T$ are independent of $\{v_r\}_{r=1}^R$,

$$\begin{aligned}
\sqrt{m} \hat{\gamma}(\theta_0) &= \frac{\sqrt{m}}{\sqrt{R}} \frac{1}{\sqrt{R}} \sum_{r=1}^R h(v_r; \theta_0) + \frac{\sqrt{m}}{\sqrt{T}} \sqrt{T} \tilde{g}(\theta_0) + O_p \left(\frac{\sqrt{m}}{R} \right) + O_p \left(\frac{\sqrt{m}}{\sqrt{TR}} \right) \\
&\xrightarrow{d} N(0, (1 \wedge 1/k) \Sigma_h + (1 \wedge k) \Omega).
\end{aligned}$$

■

9.4 Proof of Theorem 4

PROOF. First we will show stochastic equicontinuity. Recall that the implicit function theorem

applied to $s(\hat{\delta}, X, \hat{F}; \theta) = S$ implies that $\hat{\delta}(\theta)$ is a continuously differentiable function of θ . By the intermediate value theorem, there exists $\theta^* \in [\theta, \theta_0]$ such that $\hat{\delta}_t(\theta) - \hat{\delta}_t(\theta_0) = \frac{\partial \hat{\delta}_t(\theta^*)}{\partial \theta}(\theta - \theta_0)$. It follows that

$$\begin{aligned} & \|\hat{\gamma}(\theta) - \hat{\gamma}(\theta_0) - (\gamma(\theta) - \gamma(\theta_0))\| \\ &= \left\| \frac{1}{T} \sum_{t=1}^T \left\{ Z_t \left(\hat{\delta}_t(\theta) - \hat{\delta}_t(\theta_0) - (X_t' \theta_1 - X_t' \theta_{0,1}) \right) \right\} - \lim_{T, R \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E \left[Z_t \left(\hat{\delta}_t(\theta) - \hat{\delta}_t(\theta_0) - (X_t' \theta_1 - X_t' \theta_{0,1}) \right) \right] \right\| \\ &\leq \left\| \frac{1}{T} \sum_{t=1}^T Z_t \frac{\partial \hat{\delta}_t(\theta^*)}{\partial \theta_2} - \lim_{T, R \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E \left[Z_t \frac{\partial \hat{\delta}_t(\theta^*)}{\partial \theta_2} \right] \right\|_2 \|\theta_2 - \theta_{0,2}\| + \left\| \frac{1}{T} \sum_{t=1}^T Z_t X_t' - E[Z_t X_t'] \right\|_2 \|\theta_1 - \theta_{0,1}\| \\ &\leq \sup_{\theta \in \Theta} \left\| \frac{1}{T} \sum_{t=1}^T Z_t \frac{\partial \hat{\delta}_t(\theta)}{\partial \theta_2} - \lim_{T, R \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E \left[Z_t \frac{\partial \hat{\delta}_t(\theta)}{\partial \theta_2} \right] \right\|_2 \|\theta_2 - \theta_{0,2}\| + \left\| \frac{1}{T} \sum_{t=1}^T Z_t X_t' - E[Z_t X_t'] \right\|_2 \|\theta_1 - \theta_{0,1}\|. \end{aligned}$$

Recall that $\frac{\partial \hat{\delta}(\theta)}{\partial \theta_1} = 0$ and $\frac{\partial \hat{\delta}(\theta)}{\partial \theta_2} = - \left(\frac{1}{R} \sum_{r=1}^R G_{\delta}(\hat{\delta}, X, v_r; \theta) \right)^{-1} \frac{1}{R} \sum_{r=1}^R G_{\theta_2}(\hat{\delta}, X, v_r; \theta)$. If we can show that $E \left[\sup_{\theta \in \Theta} \left\| Z_t \frac{\partial \hat{\delta}_t(\theta)}{\partial \theta_2} \right\|_2 \right] < \infty$, then since $\frac{\partial \hat{\delta}(\theta)}{\partial \theta_2}$ is continuous in θ and Θ is a compact set, we will have that the uniform law of large numbers holds (see e.g. Lemma 2.4 in [Newey and McFadden \(1994\)](#)):

$$\sup_{\theta \in \Theta} \left\| \frac{1}{T} \sum_{t=1}^T Z_t \frac{\partial \hat{\delta}_t(\theta)}{\partial \theta_2} - \lim_{T, R \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E \left[Z_t \frac{\partial \hat{\delta}_t(\theta)}{\partial \theta_2} \right] \right\|_2 = o_p(1).$$

Recall that $Z_t \in \mathbb{R}^{L \times J}$ for finite L and that $E \left[\sup_{\theta \in \Theta} \left\| Z_t \frac{\partial \hat{\delta}_t(\theta)}{\partial \theta_2} \right\|_2 \right] < \sqrt{L} E \left[\sup_{\theta \in \Theta} \left\| Z_t \frac{\partial \hat{\delta}_t(\theta)}{\partial \theta_2} \right\|_{\infty} \right]$. It therefore suffices to show that $E \left[\sup_{\theta \in \Theta} \left\| Z_t \frac{\partial \hat{\delta}_t(\theta)}{\partial \theta_2} \right\|_{\infty} \right] < \infty$. Note that

$$\begin{aligned} & E \left[\sup_{\theta \in \Theta} \left\| Z_t \frac{\partial \hat{\delta}_t(\theta)}{\partial \theta_2} \right\|_{\infty} \right] \\ &\leq E \left[\|Z_t\|_{\infty} \right] + E \left[\sup_{\theta \in \Theta} \left\| \frac{\partial \hat{\delta}_t(\theta)}{\partial \theta_2} \right\|_{\infty} \right] \\ &\leq E \left[\|Z_t\|_{\infty} \right] + E \left[\sup_{\theta \in \Theta} \left\| \left(\frac{1}{R} \sum_{r=1}^R \nabla_{\delta} g_t(\hat{\delta}_t, X_t, v_r; \theta) \right)^{-1} \sup_{\theta \in \Theta} \left\| \frac{1}{R} \sum_{r=1}^R \nabla_{\theta_2} g_t(\hat{\delta}_t, X_t, v_r; \theta) \right\|_{\infty} \right\| \right]. \end{aligned}$$

We showed in lemma 1 that $\int \nabla_{\delta} g_t(\delta_t, X_t, v_r; \theta) dF(v_r)$ is strictly diagonally dominant for all θ , δ_t , X_t , and F , which implies $\frac{1}{R} \sum_{r=1}^R \nabla_{\delta} g_t(\delta_t, X_t, v_r; \theta)$ is strictly diagonally dominant for all θ , δ_t , and X_t . The Ahlberg-Nilson-Varah bound ([Ahlberg and Nilson \(1963\)](#); [Varah \(1975\)](#)) states that for all

$t = 1 \dots T$,

$$\sup_{\theta \in \Theta} \left\| \left(\frac{1}{R} \sum_{r=1}^R \nabla_{\delta} g_t(\hat{\delta}_t, X_t, v_r; \theta) \right)^{-1} \right\|_{\infty} \leq \sup_{\theta \in \Theta} \frac{1}{\min_{1 \leq i \leq JT} (|a_t^{ii}(\theta)| - \sum_{j \neq i} |a_t^{ij}(\theta)|)},$$

where $a_t^{ij}(\theta)$ is the i, j th element of $\frac{1}{R} \sum_{r=1}^R \nabla_{\delta} g_t(\hat{\delta}_t, X_t, v_r; \theta)$. Since $a_t^{ij}(\theta) \in (-1, 0) \cup (0, 1)$ for all θ , there exists a constant C such that $\max_{t=1 \dots T} \sup_{\theta \in \Theta} \left\| \left(\frac{1}{R} \sum_{r=1}^R \nabla_{\delta} g_t(\hat{\delta}_t, X_t, v_r; \theta) \right)^{-1} \right\|_{\infty} < C$. Next we show $E \left[\sup_{\theta \in \Theta} \left\| \frac{1}{R} \sum_{r=1}^R \nabla_{\theta_2} g_t(\hat{\delta}_t, X_t, v_r; \theta) \right\|_{\infty} \right] < \infty$ by showing that the vector $E \left[\sup_{\theta \in \Theta} \frac{1}{R} \sum_{r=1}^R \left| \frac{\partial g_{jt}(\hat{\delta}_t, X_t, v_r; \theta)}{\partial \theta_2} \right| \right] < \infty$ for all $j = 1 \dots J$. Note that for all $t = 1 \dots T$, $j = 1 \dots J$, and $r = 1 \dots R$,

$$\begin{aligned} & \sup_{\theta \in \Theta} \left| \frac{\partial g_{jt}(\hat{\delta}_t, X_t, v_r; \theta)}{\partial \theta_2} \right| \\ &= \sup_{\theta \in \Theta} \left| \frac{\exp(\hat{\delta}_{jt} + \mu_{rjt})}{1 + \sum_{k \in \mathcal{N}(t)} \exp(\hat{\delta}_{kt} + \mu_{rkt})} \left([1, x'_{jt}]' \circ v_r - \frac{\sum_{k \in \mathcal{N}(t)} \exp(\hat{\delta}_{kt} + \mu_{rkt}) x_{kt} \circ v_r}{1 + \sum_{k \in \mathcal{N}(t)} \exp(\hat{\delta}_{kt} + \mu_{rkt})} \right) \right| \\ &\leq \max_{k=1 \dots J} \left| [1, x'_{kt}]' \circ v_r \right| \sup_{\theta \in \Theta} \left| \left(\frac{\exp(\hat{\delta}_{jt} + \mu_{rjt})}{1 + \sum_{k \in \mathcal{N}(t)} \exp(\hat{\delta}_{kt} + \mu_{rkt})} \right) \left(1 + \frac{\sum_{k \in \mathcal{N}(t)} \exp(\hat{\delta}_{kt} + \mu_{rkt})}{1 + \sum_{k \in \mathcal{N}(t)} \exp(\hat{\delta}_{kt} + \mu_{rkt})} \right) \right| \\ &\leq 2 \max_{k=1 \dots J} \left| [1, x'_{kt}]' \circ v_r \right|. \end{aligned}$$

Since $E \left[\max_{j=1 \dots J} \left| [1, x'_{jt}]' \circ v_r \right| \right] < \infty$ by assumption, $E \left[\sup_{\theta \in \Theta} \frac{1}{R} \sum_{r=1}^R \left| \frac{\partial g_{jt}(\hat{\delta}_t, X_t, v_r; \theta)}{\partial \theta_2} \right| \right] \leq E \left[\frac{1}{R} \sum_{r=1}^R \max_{j=1 \dots J} \left| [1, x'_{jt}]' \circ v_r \right| \right] < \infty$ and $E \left[\sup_{\theta \in \Theta} \left\| \frac{\partial \hat{\delta}_t(\theta)}{\partial \theta_2} \right\|_{\infty} \right] \leq CE \left[\sup_{\theta \in \Theta} \left\| \frac{1}{R} \sum_{r=1}^R \nabla_{\theta_2} g_t(\hat{\delta}_t, X_t, v_r; \theta) \right\|_{\infty} \right] < \infty$. This combined with $E[\|Z_t\|_{\infty}] < \infty$ implies that $E \left[\sup_{\theta \in \Theta} \left\| Z_t \frac{\partial \hat{\delta}_t(\theta)}{\partial \theta_2} \right\|_{\infty} \right] < \infty$ which implies that $E \left[\sup_{\theta \in \Theta} \left\| Z_t \frac{\partial \hat{\delta}_t(\theta)}{\partial \theta_2} \right\|_2 \right] < \infty$. It follows that

$$\sup_{\theta \in \Theta} \left\| \frac{1}{T} \sum_{t=1}^T Z_t \frac{\partial \hat{\delta}_t(\theta)}{\partial \theta_2} - \lim_{T, R \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E \left[Z_t \frac{\partial \hat{\delta}_t(\theta)}{\partial \theta_2} \right] \right\|_2 = o_p(1).$$

Additionally, since $E[\|Z_t X'_t\|_2] < \infty$, the weak law of large numbers implies that

$$\left\| \frac{1}{T} \sum_{t=1}^T Z_t X'_t - E[Z_t X'_t] \right\|_2 = o_p(1).$$

Therefore, stochastic equicontinuity holds:

$$\begin{aligned}
& \sup_{\|\theta - \theta_0\| \leq \kappa_m} \sqrt{m} \|\hat{\gamma}(\theta) - \hat{\gamma}(\theta_0) - (\gamma(\theta) - \gamma(\theta_0))\| / (1 + \sqrt{m} \|\theta - \theta_0\|) \\
& \leq \sup_{\|\theta - \theta_0\| \leq \kappa_m} \|\hat{\gamma}(\theta) - \hat{\gamma}(\theta_0) - (\gamma(\theta) - \gamma(\theta_0))\| / \|\theta - \theta_0\| \\
& \leq \sup_{\theta \in \Theta} \left\| \frac{1}{T} \sum_{t=1}^T Z_t \frac{\partial \hat{\delta}_t(\theta)}{\partial \theta_2} - \lim_{T, R \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E \left[Z_t \frac{\partial \hat{\delta}_t(\theta)}{\partial \theta_2} \right] \right\|_2 + \left\| \frac{1}{T} \sum_{t=1}^T Z_t X_t' - E[Z_t X_t'] \right\|_2 \\
& = o_p(1).
\end{aligned}$$

Using similar arguments, we can show that for all $\theta', \theta'' \in \Theta$,

$$\begin{aligned}
\|\hat{\gamma}(\theta') - \hat{\gamma}(\theta'')\| & \leq \sup_{\theta \in \Theta} \left\| \frac{1}{T} \sum_{t=1}^T Z_t \frac{\partial \hat{\delta}_t(\theta)}{\partial \theta_2} \right\|_2 \|\theta_2' - \theta_2''\| + \left\| \frac{1}{T} \sum_{t=1}^T Z_t X_t' \right\|_2 \|\theta_1' - \theta_1''\| \\
& \leq B_T \|\theta' - \theta''\|
\end{aligned}$$

for $B_T = \sup_{\theta \in \Theta} \left\| \frac{1}{T} \sum_{t=1}^T Z_t \frac{\partial \hat{\delta}_t(\theta)}{\partial \theta_2} \right\|_2 + \left\| \frac{1}{T} \sum_{t=1}^T Z_t X_t' \right\|_2 \leq \frac{1}{T} \sum_{t=1}^T \left(\sup_{\theta \in \Theta} \left\| Z_t \frac{\partial \hat{\delta}_t(\theta)}{\partial \theta_2} \right\|_2 + \|Z_t X_t'\|_2 \right) = O_p(1)$ since $E \left[\sup_{\theta \in \Theta} \left\| Z_t \frac{\partial \hat{\delta}_t(\theta)}{\partial \theta_2} \right\|_2 \right] < \infty$ and $E[\|Z_t X_t'\|_2] < \infty$. Since $\gamma(\theta) = \lim_{T, R \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E \left[Z_t \left(\hat{\delta}_t(\theta) - X_t' \theta_1 \right) \right]$ is continuous in θ , Θ is a compact set, and $\|\hat{\gamma}(\theta) - \gamma(\theta)\| \xrightarrow{p} 0$ for each θ , Lemma 2.9 in [Newey and McFadden \(1994\)](#) implies that

$$\sup_{\theta \in \Theta} \|\hat{\gamma}(\theta) - \gamma(\theta)\| \xrightarrow{p} 0.$$

■

9.5 Proof of Theorem 7

PROOF. Recall that

$$\frac{\partial \hat{\delta}(\theta)}{\partial \theta_2} = - \left(\frac{1}{R} \sum_{r=1}^R G_{\delta}(\hat{\delta}, X, v_r; \theta) \right)^{-1} \frac{1}{R} \sum_{r=1}^R G_{\theta_2}(\hat{\delta}, X, v_r; \theta).$$

We showed in theorem 4 that $E \left[\sup_{\theta \in \Theta} \left\| Z_t \frac{\partial \hat{\delta}_t(\theta)}{\partial \theta_2} \right\|_2 \right] < \infty$. Since $\hat{\theta} \xrightarrow{p} \theta_0$ and $\frac{\partial \hat{\delta}(\theta)}{\partial \theta_2}$ is continuous in θ ,

by Lemma 4.3 of [Newey and McFadden \(1994\)](#) and the weak law of large numbers,

$$\hat{\Gamma} = \left[-\frac{1}{T} \sum_{t=1}^T Z_t X_t', \quad \frac{1}{T} \sum_{t=1}^T Z_t \frac{\partial \hat{\delta}_t(\theta)}{\partial \theta_2} \Big|_{\hat{\theta}} \right] \xrightarrow{p} \left[-E[Z_t X_t'], \quad \lim_{T, R \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E \left[Z_t \frac{\partial \hat{\delta}_t(\theta)}{\partial \theta_2} \Big|_{\theta_0} \right] \right].$$

Since we also assumed $E \|Z_t X_t'\|_2 < \infty$ and $Z_t (\hat{\delta}_t(\theta) - X_t' \theta_1)$ is integrable for θ in a neighborhood of θ_0 , we can interchange differentiation and expectation so that $plim \hat{\Gamma} = \Gamma$.

Furthermore, $W_T = \hat{\Sigma}^{-1} \xrightarrow{p} W = \Sigma^{-1}$. Therefore, $\hat{\Gamma}' W_T \hat{\Gamma} \xrightarrow{p} \Gamma' W \Gamma$.

To show that $\hat{\Omega} \xrightarrow{p} \Omega$, note that since $\hat{\delta}(\hat{\theta}_2) \xrightarrow{p} \delta_0$, $\hat{\theta} \xrightarrow{p} \theta_0$, and there exists $\kappa_m \downarrow 0$ such that $E \left[\sup_{\|\theta - \theta_0\| \leq \kappa_m} \left\| Z_t (\hat{\delta}_t(\theta) - X_t' \theta_1) \right\| \right] < \infty$, by Lemma 4.3 of [Newey and McFadden \(1994\)](#),

$$\frac{1}{T} \sum_{t=1}^T \left(Z_t (\hat{\delta}_t(\hat{\theta}_2) - X_t' \hat{\theta}_1) \right) \left(Z_t (\hat{\delta}_t(\hat{\theta}_2) - X_t' \hat{\theta}_1) \right)' - E \left[\left(Z_t (\delta_{0t} - X_t' \theta_{0,1}) \right) \left(Z_t (\delta_{0t} - X_t' \theta_{0,1}) \right)' \right] \xrightarrow{p} 0.$$

To show that $\hat{\Sigma}_h \xrightarrow{p} \Sigma_h$, we first show that $\max_{r=1 \dots R} \left\| \hat{h}(v_r; \hat{\theta}) - \tilde{h}(v_r; \theta_0) \right\|_{\infty} \xrightarrow{p} 0$, where

$$\begin{aligned} \hat{h}(v_r; \hat{\theta}) &= -\frac{1}{T} \sum_{t=1}^T Z_t \left(\frac{1}{R} \sum_{r'=1}^R \nabla_{\delta} g_t(\hat{\delta}_t, X_t, v_{r'}; \hat{\theta}) \right)^{-1} \left(g_t(\hat{\delta}_t, X_t, v_r; \hat{\theta}) - \frac{1}{R} \sum_{r'=1}^R g_t(\hat{\delta}_t, X_t, v_{r'}; \hat{\theta}) \right) \\ \tilde{h}(v_r; \theta_0) &= -\frac{1}{T} \sum_{t=1}^T Z_t \left(\int \nabla_{\delta} g_t(\delta_{0t}, X_t, v; \theta_0) dF_0(v) \right)^{-1} \left(g_t(\delta_{0t}, X_t, v_r; \theta_0) - E_v[g_t(\delta_{0t}, X_t, v; \theta_0)] \right). \end{aligned}$$

Note that for all $t = 1 \dots T$, $\nabla_{\delta} g_t(\delta_t, X_t, v_r; \theta)$ is continuous in δ_t and θ , and $\nabla_{\delta} g_t(\delta_t, X_t, v_r; \theta) \in (-1, 0) \cup (0, 1)$ for all δ_t, X_t, v_r , and θ . Since $\hat{\theta} \xrightarrow{p} \theta_0$ and $\hat{\delta}(\hat{\theta}_2) \xrightarrow{p} \delta_0$, by Lemma 4.3 of [Newey and McFadden \(1994\)](#),

$$\max_{t=1 \dots T} \left\| \frac{1}{R} \sum_{r=1}^R \nabla_{\delta} g_t(\hat{\delta}_t, X_t, v_r; \hat{\theta}) - \int \nabla_{\delta} g_t(\delta_{0t}, X_t, v_r; \theta_0) dF_0(v_r) \right\|_{\infty} \xrightarrow{p} 0.$$

By the Continuous Mapping Theorem,

$$\max_{t=1 \dots T} \left\| \left(\frac{1}{R} \sum_{r=1}^R \nabla_{\delta} g_t(\hat{\delta}_t, X_t, v_r; \hat{\theta}) \right)^{-1} - \left(\int \nabla_{\delta} g_t(\delta_{0t}, X_t, v; \theta_0) dF_0(v) \right)^{-1} \right\|_{\infty} \xrightarrow{p} 0.$$

Similarly, note that for all $t = 1 \dots T$, $g_t(\delta_t, X_t, v_r; \theta)$ is continuous in δ_t and θ , and $g_t(\delta_t, X_t, v_r; \theta) \in$

(0, 1) for all δ_t, X_t, v_r , and θ . Since $\hat{\theta} \xrightarrow{p} \theta_0$ and $\hat{\delta}(\hat{\theta}_2) \xrightarrow{p} \delta_0$, by Lemma 4.3 of [Newey and McFadden \(1994\)](#),

$$\max_{t=1\dots T} \left\| \frac{1}{R} \sum_{r=1}^R g_t(\hat{\delta}_t, X_t, v_r; \hat{\theta}) - E_v [g_t(\delta_{0t}, X_t, v; \theta_0)] \right\|_{\infty} \xrightarrow{p} 0.$$

Note that the Ahlberg-Nilson-Varah ([Ahlberg and Nilson \(1963\)](#); [Varah \(1975\)](#)) bound on the strictly diagonally dominant matrices $\int \nabla_{\delta} g_t(\delta_{0t}, X_t, v_r; \theta_0) dF(v_r)$ implies that there exists a constant C such that

$$\max_{t=1\dots T} \left\| \left(\int \nabla_{\delta} g_t(\delta_{0t}, X_t, v_r; \theta_0) dF_0(v_r) \right)^{-1} \right\|_{\infty} < C.$$

Also, since $g_t(\delta_t, X_t, v_r; \theta) \in (0, 1)$ for all δ_t, X_t, v_r , and θ , there exists B such that

$$\max_{t=1\dots T} \max_{r=1\dots R} \left\| g_t(\hat{\delta}_t, X_t, v_r; \hat{\theta}) - \frac{1}{R} \sum_{r'=1}^R g_t(\hat{\delta}_t, X_t, v_{r'}; \hat{\theta}) \right\|_{\infty} < B.$$

We assumed in [theorem 4](#) that $E \|Z_t\|_{\infty} < \infty$, which implies that $\|Z_t\|_{\infty} = O_p(1)$.

Furthermore, we assumed

$$\max_{r=1\dots R} \max_{t=1\dots T} \left\| g_t(\hat{\delta}_t, X_t, v_r; \hat{\theta}) - g_t(\delta_{0t}, X_t, v_r; \theta_0) \right\|_{\infty} \xrightarrow{p} 0.$$

Therefore,

$$\begin{aligned}
& \max_{r=1\dots R} \left\| \hat{h}(v_r; \hat{\theta}) - \tilde{h}(v_r; \theta_0) \right\|_{\infty} \\
& \leq \max_{r=1\dots R} \left\| \frac{1}{T} \sum_{t=1}^T Z_t \left(\left(\frac{1}{R} \sum_{r'=1}^R \nabla_{\delta} g_t(\hat{\delta}_t, X_t, v_{r'}; \hat{\theta}) \right)^{-1} - \left(\int \nabla_{\delta} g_t(\delta_{0t}, X_t, v_{r'}; \theta_0) dF_0(v_{r'}) \right)^{-1} \right) \right. \\
& \quad \left. \left(g_t(\hat{\delta}_t, X_t, v_r; \hat{\theta}) - \frac{1}{R} \sum_{r'=1}^R g_t(\hat{\delta}_t, X_t, v_{r'}; \hat{\theta}) \right) \right\|_{\infty} \\
& \quad + \max_{r=1\dots R} \left\| \frac{1}{T} \sum_{t=1}^T Z_t \left(\int \nabla_{\delta} g_t(\delta_{0t}, X_t, v_{r'}; \theta_0) dF_0(v_{r'}) \right)^{-1} \left(g_t(\hat{\delta}_t, X_t, v_r; \hat{\theta}) - g_t(\delta_{0t}, X_t, v_r; \theta_0) \right) \right\|_{\infty} \\
& \quad + \left\| \frac{1}{T} \sum_{t=1}^T Z_t \left(\int \nabla_{\delta} g_t(\delta_{0t}, X_t, v_{r'}; \theta_0) dF_0(v_{r'}) \right)^{-1} \left(\frac{1}{R} \sum_{r'=1}^R g_t(\hat{\delta}_t, X_t, v_{r'}; \hat{\theta}) - E_v[g_t(\delta_{0t}, X_t, v; \theta_0)] \right) \right\|_{\infty} \\
& \leq \max_{r=1\dots R} \max_{t=1\dots T} \left\| Z_t \left(\left(\frac{1}{R} \sum_{r'=1}^R \nabla_{\delta} g_t(\hat{\delta}_t, X_t, v_{r'}; \hat{\theta}) \right)^{-1} - \left(\int \nabla_{\delta} g_t(\delta_{0t}, X_t, v_{r'}; \theta_0) dF_0(v_{r'}) \right)^{-1} \right) \right. \\
& \quad \left. \left(g_t(\hat{\delta}_t, X_t, v_r; \hat{\theta}) - \frac{1}{R} \sum_{r'=1}^R g_t(\hat{\delta}_t, X_t, v_{r'}; \hat{\theta}) \right) \right\|_{\infty} \\
& \quad + \max_{r=1\dots R} \max_{t=1\dots T} \left\| Z_t \left(\int \nabla_{\delta} g_t(\delta_{0t}, X_t, v_{r'}; \theta_0) dF_0(v_{r'}) \right)^{-1} \left(g_t(\hat{\delta}_t, X_t, v_r; \hat{\theta}) - g_t(\delta_{0t}, X_t, v_r; \theta_0) \right) \right\|_{\infty} \\
& \quad + \max_{t=1\dots T} \left\| Z_t \left(\int \nabla_{\delta} g_t(\delta_{0t}, X_t, v_{r'}; \theta_0) dF_0(v_{r'}) \right)^{-1} \left(\frac{1}{R} \sum_{r'=1}^R g_t(\hat{\delta}_t, X_t, v_{r'}; \hat{\theta}) - E_v[g_t(\delta_{0t}, X_t, v; \theta_0)] \right) \right\|_{\infty} \\
& \leq \max_{t=1\dots T} \|Z_t\|_{\infty} \max_{t=1\dots T} \left\| \left(\frac{1}{R} \sum_{r'=1}^R \nabla_{\delta} g_t(\hat{\delta}_t, X_t, v_{r'}; \hat{\theta}) \right)^{-1} - \left(\int \nabla_{\delta} g_t(\delta_{0t}, X_t, v_{r'}; \theta_0) dF_0(v_{r'}) \right)^{-1} \right\|_{\infty} \\
& \quad \max_{r=1\dots R} \max_{t=1\dots T} \left\| g_t(\hat{\delta}_t, X_t, v_r; \hat{\theta}) - \frac{1}{R} \sum_{r'=1}^R g_t(\hat{\delta}_t, X_t, v_{r'}; \hat{\theta}) \right\|_{\infty} \\
& \quad + \max_{t=1\dots T} \|Z_t\|_{\infty} \max_{t=1\dots T} \left\| \left(\int \nabla_{\delta} g_t(\delta_{0t}, X_t, v_{r'}; \theta_0) dF_0(v_{r'}) \right)^{-1} \right\|_{\infty} \\
& \quad \left\{ \max_{r=1\dots R} \max_{t=1\dots T} \left\| g_t(\hat{\delta}_t, X_t, v_r; \hat{\theta}) - g_t(\delta_{0t}, X_t, v_r; \theta_0) \right\|_{\infty} \right. \\
& \quad \left. + \max_{t=1\dots T} \left\| \frac{1}{R} \sum_{r'=1}^R g_t(\hat{\delta}_t, X_t, v_{r'}; \hat{\theta}) - E_v[g_t(\delta_{0t}, X_t, v; \theta_0)] \right\|_{\infty} \right\} \\
& = o_p(1).
\end{aligned}$$

Then it follows that

$$\begin{aligned}
\hat{\Sigma}_h &= \frac{1}{R} \sum_{r=1}^R \hat{h}(v_r; \hat{\theta}) \hat{h}(v_r; \hat{\theta})' \\
&= \frac{1}{R} \sum_{r=1}^R \left(\tilde{h}(v_r; \theta_0) + o_p(1) \right) \left(\tilde{h}(v_r; \theta_0) + o_p(1) \right)' \\
&= \frac{1}{R} \sum_{r=1}^R \tilde{h}(v_r; \theta_0) \tilde{h}(v_r; \theta_0)' + o_p(1).
\end{aligned}$$

Note that $\frac{1}{R} \sum_{r=1}^R \tilde{h}(v_r; \theta_0) \tilde{h}(v_r; \theta_0)'$ is a two-sample V-statistic, which is known to satisfy a weak law of large numbers (see e.g. [van der Vaart \(1998\)](#)). Therefore, for $k(\delta_{0t}, X_t, v_r; \theta_0) = -Z_t \left(\int \nabla_{\delta} g_t(\delta_{0t}, X_t, v; \theta_0) dF_0(v) \right)^{-1} (g_t(\delta_{0t}, X_t, v_r; \theta_0) - E_v[g_t(\delta_{0t}, X_t, v; \theta_0)])$,

$$\frac{1}{R} \sum_{r=1}^R \tilde{h}(v_r; \theta_0) \tilde{h}(v_r; \theta_0)' \xrightarrow{p} E \left[k(\delta_{0t}, X_t, v_r; \theta_0) k(\delta_{0t}, X_t, v_r; \theta_0)' \right] = \Sigma_h.$$

We have shown that $\hat{\Sigma} = \min(1, \frac{R}{T}) \hat{\Omega} + \min(1, \frac{T}{R}) \hat{\Sigma}_h \xrightarrow{p} \Sigma = (1 \wedge k)\Omega + (1 \wedge 1/k)\Sigma_h$. Therefore,

$$\widehat{AsyVar}[\hat{\theta}] = \left(\hat{\Gamma}' W_T \hat{\Gamma} \right)^{-1} \hat{\Gamma}' W_T \hat{\Sigma} W_T \hat{\Gamma} \left(\hat{\Gamma}' W_T \hat{\Gamma} \right)^{-1} \xrightarrow{p} (\Gamma' W \Gamma)^{-1} \Gamma' W \Sigma W \Gamma (\Gamma' W \Gamma)^{-1}.$$

■

9.6 Proof of Theorem 8

PROOF. We first check that assumptions 1-4 of [Chernozhukov and Hong \(2003\)](#) are satisfied. Condition (i) of Theorem 4 in combination with condition (i) of the present theorem is assumption 1 of [Chernozhukov and Hong \(2003\)](#). Condition (ii) of the present theorem is assumption 2 of [Chernozhukov and Hong \(2003\)](#). We assumed in Theorem 5 that $\inf_{\theta \notin G} \|\gamma(\theta)\|_W > \|\gamma(\theta_0)\|_W$ for every open set G that contains θ_0 , which is equivalent to saying that for any sequence $\{\theta_m\} \in \Theta$, $\limsup_{m \rightarrow \infty} \|\gamma(\theta_m)\|_W \leq \|\gamma(\theta_0)\|_W$ implies $\|\theta_m - \theta_0\| \rightarrow 0$. Since $\gamma(\theta)$ is a continuous function of θ and theorem 4 implies that $\|\hat{\gamma}(\hat{\theta})\|_{W_T}$ converges in probability to $\|\gamma(\theta)\|_W$ uniformly over Θ , it follows by Lemma 1 of [Chernozhukov and Hong \(2003\)](#) that assumption 3 is satisfied: for any $\kappa > 0$, there exists

ϵ such that $\liminf_{m \rightarrow \infty} P \left(\inf_{\|\theta - \theta_0\| \geq \kappa} (\|\hat{\gamma}(\theta)\|_{W_T} - \|\hat{\gamma}(\theta_0)\|_{W_T}) \geq \epsilon \right) = 1$. Assumption 4(i) is a quadratic expansion of the sample objective function around θ_0 , which we know to exist from differentiability of $\hat{\gamma}(\theta)$ at θ_0 (condition (i) in Theorem 6). Assumption 4(ii) is the asymptotic normality of $\Gamma'W\sqrt{m}\hat{\gamma}(\theta_0)$, which we showed in Theorem 6. Assumption 4(iii) follows from nonsingularity of $\Gamma'W\Gamma$.

We show assumption 4(iv) by showing that the remainder term in the quadratic expansion of the sample objective function around θ_0 converges in probability to zero uniformly over θ in a δ_m ball around θ_0 . Define

$$\begin{aligned} R_m(\theta) &= \frac{m}{2}\hat{\gamma}(\theta)'W_T\hat{\gamma}(\theta) - \frac{m}{2}\hat{\gamma}(\theta_0)'W_T\hat{\gamma}(\theta_0) \\ &\quad - \sqrt{m}(\theta - \theta_0)\Gamma'W\sqrt{m}\hat{\gamma}(\theta_0) - \frac{1}{2}\sqrt{m}(\theta - \theta_0)\Gamma'W\Gamma\sqrt{m}(\theta - \theta_0), \end{aligned}$$

where we can write $R_m(\theta) = R_{1m}(\theta) + R_{2m}(\theta)$ for

$$\begin{aligned} R_{1m}(\theta) &= m \left(\frac{1}{2}\hat{\gamma}(\theta)'W_T\hat{\gamma}(\theta) - \frac{1}{2}\hat{\gamma}(\theta_0)'W_T\hat{\gamma}(\theta_0) - (\theta - \theta_0)\Gamma'W_T\hat{\gamma}(\theta_0) - \frac{1}{2}(\theta - \theta_0)\Gamma'W\Gamma(\theta - \theta_0) \right), \\ R_{2m}(\theta) &= m(\theta - \theta_0)\Gamma'(W_T - W)\hat{\gamma}(\theta_0). \end{aligned}$$

We can show that for any $\kappa_m \rightarrow 0$,

$$\sup_{\|\theta - \theta_0\| \leq \kappa_m} \frac{|R_{2m}(\theta)|}{1 + m\|\theta - \theta_0\|^2} = o_p(1),$$

due to consistency of W_T for W and $\sqrt{m}\hat{\gamma}(\theta_0) = O_p(1)$. The more difficult term is $R_{1m}(\theta)$, which

can be decomposed into six terms expressed in terms of $\epsilon(\theta) \equiv \frac{\hat{\gamma}(\theta) - \hat{\gamma}(\theta_0) - \gamma(\theta)}{1 + \sqrt{m}\|\theta - \theta_0\|}$:

$$\begin{aligned}
\frac{1}{m}R_{1m}(\theta) &= \frac{1}{2} \underbrace{(1 + \sqrt{m}\|\theta - \theta_0\|)^2 \epsilon(\theta)' W_T \epsilon(\theta)}_{r_1(\theta)} + \underbrace{\hat{\gamma}(\theta_0)' W_T (\gamma(\theta) - \Gamma(\theta - \theta_0))}_{r_2(\theta)} \\
&+ \underbrace{(1 + \sqrt{m}\|\theta - \theta_0\|) \epsilon(\theta)' W_T \hat{\gamma}(\theta_0)}_{r_3(\theta)} + \underbrace{(1 + \sqrt{m}\|\theta - \theta_0\|) \epsilon(\theta)' W_T \gamma(\theta)}_{r_4(\theta)} \\
&+ \underbrace{\frac{1}{2} \gamma(\theta)' (W_T - W) \gamma(\theta)}_{r_5(\theta)} \\
&+ \underbrace{\frac{1}{2} \gamma(\theta)' W \gamma(\theta) - \frac{1}{2} (\theta - \theta_0)' \Gamma' W \Gamma (\theta - \theta_0)}_{r_6(\theta)}.
\end{aligned}$$

It remains to show that for any $\kappa_m \rightarrow 0$, $\sup_{\|\theta - \theta_0\| \leq \kappa_m} \frac{m|r_j(\theta)|}{1 + m\|\theta - \theta_0\|^2} = o_p(1)$ for all $j = 1, \dots, 6$. Using the Taylor expansion $\gamma(\theta) = \Gamma(\theta - \theta_0) + o(\|\theta - \theta_0\|)$, $\sup_{\|\theta - \theta_0\| \leq \kappa_m} \sqrt{m} \|\epsilon(\theta)\| \xrightarrow{p} 0$ (the stochastic equicontinuity result shown in Theorem 4), and consistency of W_T for W ,

$$\begin{aligned}
\sup_{\|\theta - \theta_0\| \leq \kappa_m} \frac{m|r_1(\theta)|}{1 + m\|\theta - \theta_0\|^2} &\leq \sup_{\|\theta - \theta_0\| \leq \kappa_m} m \epsilon(\theta)' W_T \epsilon(\theta) = o_p(1), \\
\sup_{\|\theta - \theta_0\| \leq \kappa_m} \frac{m|r_2(\theta)|}{1 + m\|\theta - \theta_0\|^2} &\leq \sup_{\|\theta - \theta_0\| \leq \kappa_m} \frac{o(\sqrt{m}\|\theta - \theta_0\|)'}{1 + m\|\theta - \theta_0\|^2} |W_T \sqrt{m} \hat{\gamma}(\theta_0)| = o_p(1), \\
\sup_{\|\theta - \theta_0\| \leq \kappa_m} \frac{m|r_3(\theta)|}{1 + m\|\theta - \theta_0\|^2} &\leq \sup_{\|\theta - \theta_0\| \leq \kappa_m} 2\sqrt{m} \left| \epsilon(\theta)' W_T \frac{\sqrt{m} \hat{\gamma}(\theta_0)}{\sqrt{m}\|\theta - \theta_0\|} \right| = o_p(1), \\
\sup_{\|\theta - \theta_0\| \leq \kappa_m} \frac{m|r_4(\theta)|}{1 + m\|\theta - \theta_0\|^2} &\leq \sup_{\|\theta - \theta_0\| \leq \kappa_m} 2\sqrt{m} \left| \epsilon(\theta)' W_T \frac{\sqrt{m} \gamma(\theta)}{\sqrt{m}\|\theta - \theta_0\|} \right| = o_p(1), \\
\sup_{\|\theta - \theta_0\| \leq \kappa_m} \frac{m|r_5(\theta)|}{1 + m\|\theta - \theta_0\|^2} &\leq \sup_{\|\theta - \theta_0\| \leq \kappa_m} \left(\frac{\sqrt{m}\|\gamma(\theta)\|}{\sqrt{m}\|\theta - \theta_0\|} \right)^2 \|W_T - W\| = o_p(1), \\
\sup_{\|\theta - \theta_0\| \leq \kappa_m} \frac{m|r_6(\theta)|}{1 + m\|\theta - \theta_0\|^2} &\leq \sup_{\|\theta - \theta_0\| \leq \kappa_m} \frac{o(\|\theta - \theta_0\|^2 \|W\|)}{\|\theta - \theta_0\|^2} = o_p(1).
\end{aligned}$$

■

9.7 Comparison with Freyberger (2015)

We reproduce the formulas for Φ_1 and Φ_2 in Freyberger (2015) and compare them to our Ω and Σ_h . In the following expressions, we reproduce Freyberger (2015)'s notation under the assumption of overlapping simulation draws: $v_{rt} = v_r$ for all t . First note that Freyberger (2015) uses ν_{jt} when defining the market shares while we use g_{jt} .

$$\begin{aligned}\nu_{jt}(\theta, x_t, \xi_t, v_r) &= \frac{\exp\left(X'_{jt}\theta_1 + \xi_{jt} + \mu_{rjt}\right)}{1 + \sum_{k \in \mathcal{N}(t)} \exp\left(X'_{kt}\theta_1 + \xi_{kt} + \mu_{rkt}\right)} = g_{jt}(\delta_t, X_t, v_r; \theta) \\ \sigma_t(\theta, x_t, \xi_t, P_t) &= \int \nu_t(\theta, x_t, \xi_t, v) dP_t(v) = \int g_t(\delta_t, X_t, v; \theta) dF(v).\end{aligned}$$

Freyberger (2015) defines H_{0t} as the Jacobian matrix of the true market shares with respect to ξ_t , which is the same as our Jacobian matrix of the true market shares with respect to $\delta_t = X'_t\theta_1 + \xi_t$.

$$H_{0t} = \frac{\partial \sigma_t(\theta_0, \xi_t(\theta_0, P_{0t}), P_{0t})}{\partial \xi} = \int \nabla_{\delta} g_t(\delta_{0t}, X_t, v_r; \theta_0) dF_0(v_r).$$

Freyberger (2015) also defines

$$\epsilon_{0rt} = \nu_t(\theta_0, x_t, \xi_t(\theta_0, P_{0t}, s_t), v_r) - \int \nu_t(\theta_0, x_t, \xi_t(\theta_0, P_{0t}, s_t), v) dP_{0t}(v).$$

Freyberger (2015) uses $z_t \in \mathbb{R}^{J \times p}$ to denote the matrix of instruments for market t while we use $Z_t \in \mathbb{R}^{p \times J}$. His expression for Φ_1 coincides with our Ω under our assumption that the data are i.i.d. across markets.

$$\begin{aligned}\Phi_1 &= \lim_{T \rightarrow \infty} E \left[\frac{1}{T} \sum_{t=1}^T z'_t \xi_t(\theta_0, P_{0t}, s_t) \xi_t(\theta_0, P_{0t}, s_t)' z_t \right] \\ &= \lim_{T \rightarrow \infty} E \left[\frac{1}{T} \sum_{t=1}^T Z_t (\delta_{0t} - X'_t \theta_{01}) (\delta_{0t} - X'_t \theta_{01})' Z'_t \right] \\ &= E \left[Z_t (\delta_{0t} - X'_t \theta_{01}) (\delta_{0t} - X'_t \theta_{01})' Z'_t \right] = \Omega.\end{aligned}$$

Freyberger (2015)'s Φ_2 is not exactly the same as our Σ_h , but only differs by a $o(1)$ term. To see this, note that $\Phi_2 = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \text{Var} [z'_t H_{0t}^{-1} \epsilon_{0rt}]$ can be rewritten using the fact that v_r are i.i.d.

as $\lim_{T \rightarrow \infty} \text{Var} \left[\frac{1}{TR} \sum_{t=1}^T \sum_{r=1}^R q(Z_t, X_t, v_r; \theta_0, \delta_{0t}) \right]$, where

$$q(Z_t, X_t, v_r; \theta_0, \delta_{0t}) = -Z_t \left(\int \nabla_{\delta} g_t(\delta_{0t}, X_t, v; \theta_0) dF_0(v) \right)^{-1} \left(g_t(\delta_{0t}, X_t, v_r; \theta_0) - \int g_t(\delta_{0t}, X_t, v_r; \theta_0) dF_0(v_r) \right).$$

Recall that $\frac{1}{TR} \sum_{t=1}^T \sum_{r=1}^R q(Z_t, X_t, v_r; \theta_0, \delta_{0t})$ is a two-sample U-statistic whose decomposition is

$$\frac{1}{TR} \sum_{t=1}^T \sum_{r=1}^R q(Z_t, X_t, v_r; \theta_0, \delta_{0t}) = \frac{1}{R} \sum_{r=1}^R h(v_r; \theta_0) + o_p(1).$$

$$h(v_r; \theta_0) = - \int \left\{ Z_t \left(\int \nabla_{\delta} g_t(\delta_{0t}, X_t, v; \theta_0) dF_0(v) \right)^{-1} \left(g_t(\delta_{0t}, X_t, v_r; \theta_0) - \int g_t(\delta_{0t}, X_t, v_r; \theta_0) dF_0(v_r) \right) \right\} dP(Z_t, X_t, \xi_t).$$

It follows that $\Phi_2 = \lim_{T \rightarrow \infty} \text{Var} [h(v_r; \theta_0) + o_p(1)] = \Sigma_h + o(1)$. Next we explain why $\hat{\Phi}_1 = \hat{\Omega}$ and $\hat{\Phi}_2 = \hat{\Sigma}_h$. Note that

$$\begin{aligned} \hat{H}_t &= \frac{\partial \sigma_t(\hat{\theta}, x_t, \hat{\xi}_t, P_{rt})}{\partial \xi} = \frac{1}{R} \sum_{r=1}^R \text{diag} \left(\nu_t(\hat{\theta}, x_t, \hat{\xi}_t, v_r) \right) - \nu_t(\hat{\theta}, x_t, \hat{\xi}_t, v_r) \nu_t(\hat{\theta}, x_t, \hat{\xi}_t, v_r)' \\ &= \frac{1}{R} \sum_{r'=1}^R \nabla_{\delta} g_t(\hat{\delta}_t, X_t, v_{r'}; \hat{\theta}). \end{aligned}$$

$$\begin{aligned} \bar{\nu}_t(\hat{\theta}, x_t, \hat{\xi}_t, v_r) &= \nu_t(\hat{\theta}, x_t, \hat{\xi}_t, v_r) - \sigma_t(\hat{\theta}, x_t, \hat{\xi}_t, P_{rt}) \\ &= g_t(\hat{\delta}_t, X_t, v_r; \hat{\theta}) - \frac{1}{R} \sum_{r'=1}^R g_t(\hat{\delta}_t, X_t, v_{r'}; \hat{\theta}). \end{aligned}$$

It follows that

$$\begin{aligned} \hat{\Phi}_1 &= \frac{1}{T} \sum_{t=1}^T z_t' \hat{\xi}_t \hat{\xi}_t' z_t = \frac{1}{T} \sum_{t=1}^T Z_t (\hat{\delta}_t - X_t' \hat{\theta}_1) (\hat{\delta}_t - X_t' \hat{\theta}_1)' Z_t' = \hat{\Omega}. \\ \hat{\Phi}_2 &= \frac{1}{RT} \sum_{r=1}^R \sum_{t=1}^T z_t' \hat{H}_t^{-1} \bar{\nu}_t(\hat{\theta}, x_t, \hat{\xi}_t, v_r) \bar{\nu}_t(\hat{\theta}, x_t, \hat{\xi}_t, v_r)' (\hat{H}_t^{-1})' z_t \\ &= \frac{1}{R} \sum_{r=1}^R \hat{h}(v_r; \hat{\theta}) \hat{h}(v_r; \hat{\theta})' = \hat{\Sigma}_h. \end{aligned}$$

$$\hat{h}(v_r; \hat{\theta}) = -\frac{1}{T} \sum_{t=1}^T Z_t \left(\frac{1}{R} \sum_{r'=1}^R \nabla_{\delta} g_t(\hat{\delta}_t, X_t, v_{r'}; \hat{\theta}) \right)^{-1} \left(g_t(\hat{\delta}_t, X_t, v_r; \hat{\theta}) - \frac{1}{R} \sum_{r'=1}^R g_t(\hat{\delta}_t, X_t, v_{r'}; \hat{\theta}) \right).$$

We can also show that our standard errors are the same as [Freyberger \(2015\)](#)'s standard errors.

Our estimate of the finite sample variance of $\hat{\theta}$ is $\frac{1}{m} \left(\hat{\Gamma}' W_T \hat{\Gamma} \right)^{-1} \hat{\Gamma}' W_T \hat{\Sigma} W_T \hat{\Gamma} \left(\hat{\Gamma}' W_T \hat{\Gamma} \right)^{-1}$ where

$$\begin{aligned} \frac{1}{m} \hat{\Sigma} &= \frac{1}{m} \min \left(1, \frac{R}{T} \right) \hat{\Omega} + \frac{1}{m} \min \left(1, \frac{T}{R} \right) \hat{\Sigma}_h \\ &= \frac{1}{T} \hat{\Omega} + \frac{1}{R} \hat{\Sigma}_h \\ &= \frac{1}{T} \left(\hat{\Phi}_1 + \frac{T}{R} \hat{\Phi}_2 \right). \end{aligned}$$

$\frac{1}{T} \left(\hat{\Gamma}' W_T \hat{\Gamma} \right)^{-1} \hat{\Gamma}' W_T \left(\hat{\Phi}_1 + \frac{T}{R} \hat{\Phi}_2 \right) W_T \hat{\Gamma} \left(\hat{\Gamma}' W_T \hat{\Gamma} \right)^{-1}$ is [Freyberger \(2015\)](#)'s estimate of the finite sample variance of $\hat{\theta}$.

References

- Ahlberg, JH and EN Nilson**, "Convergence properties of the spline fit," *Journal of the Society for Industrial and Applied Mathematics*, 1963, 11 (1), 95–104.
- Berry, Steven**, "Estimating Discrete-Choice Models of Product Differentiation," *RAND Journal of Economics*, 1994, 25 (2), 242–262.
- , **Amit Gandhi, and Philip Haile**, "Connected substitutes and invertibility of demand," *Econometrica*, 2013, 81 (5), 2087–2111.
- , **James Levinsohn, and Ariel Pakes**, "Automobile Prices in Market Equilibrium," *Econometrica*, 1995, 63 (4), 841–890.
- Chernozhukov, Victor and Han Hong**, "A MCMC Approach to Classical Estimation," *Journal of Econometrics*, 2003, 115 (2), 293–346.
- Compiani, Giovanni**, "Nonparametric Demand Estimation in Differentiated Products Markets," *working paper*, 2018.

- Conlon, Christopher, “BLP-Demand repository,” [https://github.com/chrisconlon/](https://github.com/chrisconlon/blp-demand)
[blp-demand](#) 2014.
- and Jeff Gortmaker, “Best practices for differentiated products demand estimation with pyblp,” Technical Report, Working paper. url: <https://chrisconlon.github.io/site/pyblp.pdf> 2019.
- Duane, Simon, Anthony D Kennedy, Brian J Pendleton, and Duncan Roweth, “Hybrid monte carlo,” *Physics letters B*, 1987, 195 (2), 216–222.
- Dubé, Jean-Pierre, Jeremy T Fox, and Che-Lin Su, “Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation,” *Econometrica*, 2012, 80 (5), 2231–2267.
- Dubé, Jean-Pierre, Jeremy Fox, and Che-Lin Su, “MATLAB Code for BLP GMM estimator of the Random Coefficients Logit using the MPEC algorithm,” <http://faculty.chicagobooth.edu/jean-pierre.dube/research/MPECcode.html> 2012.
- Freyberger, Joachim, “Asymptotic theory for differentiated products demand models with many markets,” *Journal of Econometrics*, 2015, 185 (1), 162–181.
- Knittel, Christopher R and Konstantinos Metaxoglou, “Estimation of random-coefficient demand models: Two empiricists’ perspective,” *Review of Economics and Statistics*, 2014, 96 (1), 34–59.
- Kristensen, Dennis and Bernard Salanié, “Higher-order properties of approximate estimators,” *Journal of Econometrics*, 2017, 198 (2), 189–208.
- Lee, Lung-Fei, “Asymptotic bias in simulated maximum likelihood estimation of discrete choice models,” *Econometric Theory*, 1995, 11 (3), 437–483.
- McFadden, Daniel, “A method of simulated moments for estimation of discrete response models without numerical integration,” *Econometrica: Journal of the Econometric Society*, 1989, pp. 995–1026.

- Neal, Radford M et al.**, “MCMC using Hamiltonian dynamics,” *Handbook of Markov Chain Monte Carlo*, 2011, 2 (11).
- Neumeyer, Natalie**, “A central limit theorem for two-sample U-processes,” *Statistics and Probability Letters*, 2004, 67, 73–85.
- Nevo, Aviv**, “A practitioner’s guide to estimation of random-coefficients logit models of demand,” *Journal of economics & management strategy*, 2000, 9 (4), 513–548.
- Newey, Whitney K and Daniel McFadden**, “Large sample estimation and hypothesis testing,” *Handbook of econometrics*, 1994, 4, 2111–2245.
- Pakes, Ariel and David Pollard**, “Simulation and Asymptotics of Optimization Estimators,” *Econometrica*, 1989, 57, 1027–1057.
- Su, Che-Lin and Kenneth L Judd**, “Constrained optimization approaches to estimation of structural models,” *Econometrica*, 2012, 80 (5), 2213–2230.
- van der Vaart, Aad W**, *Asymptotic statistics*, Vol. 3, Cambridge university press, 1998.
- **and Jon Wellner**, *Weak Convergence and Empirical Processes*, Springer-Verlag, 1996.
- Varah, James M**, “A lower bound for the smallest singular value of a matrix,” *Linear Algebra and its Applications*, 1975, 11 (1), 3–5.
- Yang, Sha, Yuxin Chen, and Greg M Allenby**, “Bayesian analysis of simultaneous demand and supply,” *Quantitative marketing and economics*, 2003, 1 (3), 251–275.