BLP Estimation using Laplace Transformation and Overlapping Simulation Draws *

Han Hong[†] Huiyu Li[‡] Jessie Li[§]

May 2, 2018

Abstract

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.
[†]Department of Economics, Stanford University. Email: hanhong@stanford.edu
[‡]Federal Reserve Bank of San Francisco. E-mail: tohuiyu@gmail.com
[§]Department of Economics, UC Santa Cruz. E-mail: jessieqli@gmail.com (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. The traditional asymptotic theory for simulation estimators of nonlinear models under independent simulation draws requires 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 (Train (2009)), but 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) ,McFadden (1989), 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 $R \to \infty$ and $T \to \infty$, and it has an asymptotic variance term that consists of two parts, one for the sampling variance and another for the simulation variance. The approach we use to derive the asymptotic distribution of the BLP estimator is to write the moment condition as a two sample U-statistic and then apply the technical machinery of Van der Vaart (1998) and Neumeyer (2004) to show consistency and asymptotic normality.

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 $\sqrt{T}$ consistency, he requires $\underset{T,R\to\infty}{lim} \frac{\sqrt{T}}{R} < \infty$ for the case of independent simulation draws and $\underset{T,R\to\infty}{lim} \frac{T}{R} < \infty$ for the case of overlapping simulation draws. Our paper deals solely with the case of overlapping simulation draws. However, instead of assuming that R must grow as least as fast as T, we impose the weaker requirement that $R \to \infty$ and $T \to \infty$, thus allowing for the case where $R \ll T$. The rate of convergence of our estimator is $\sqrt{m}$ for $m = min(R, T)$, which coincides with the $\sqrt{T}$ rate under the assumptions given by Freyberger (2015).

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.

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 a Monte Carlo study that illustrates the severe undercoverage of confidence intervals obtained using the typical GMM standard errors which do not take into account simulation noise. Section 8 concludes.

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:

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

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

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

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 $j = 1, ..., J$. The observed market shares are

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

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 $\gamma\left(\theta_0\right) = \mathbb{E}\left[Z_t\left(\delta_{0t} - X_t'\theta_{0,1}\right)\right] = \mathbb{E}\left[Z_t \xi_t\right] = 0$. 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 $\hat{s}_{jt}\left(\delta_t, X_t, \hat{F}; \theta\right) = \frac{1}{R}\sum_{r=1}^{R} \frac{exp(\delta_{jt} + \mu_{rjt}(\theta_2, v_r))}{1 + \sum_{k \in \mathcal{N}(t)} exp(\delta_{kt} + \mu_{rkt}(\theta_2, v_r))}$ to the observed market shares S_{jt} for all products j in all markets t. Berry et al. (1995) prove that $\delta_t^{(k+1)} = \delta_t^{(k)} + log\left(S_t\right) - log\left(\hat{s}_t\left(\delta_t^{(k)}, X_t, \hat{F}; \theta\right)\right)$ is a contraction mapping and that a unique fixed point $\hat{\delta}_t\left(\theta\right)$ exists. We will suppress the dependence of $\hat{\delta}_t\left(\theta\right)$ on X_t to simplify notation. Note that $\hat{\delta}_t\left(\theta\right)$ depends on $\theta_2 = diag\left(\Sigma\right)$ 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 $\hat{\gamma}\left(\theta\right) = \frac{1}{T}\sum_{t=1}^{T} Z_t\left(\hat{\delta}_t\left(\theta\right) - X_t'\theta_1\right)$ using a positive-definite weighting matrix W_T.

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

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 $\hat{\delta}\left(\theta_0\right)$ 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. We will employ the central limit theorems of Neumeyer (2004) to obtain the asymptotic distribution of the U-statistic. 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 $\sqrt{m}\left(\hat{\theta}-\theta_0\right)$. We also provide a consistent estimate of the asymptotic variance.

3.1 Nonsingularity of Jacobian Matrix

Recall that the true (observed) market shares are

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

The simulated market shares at any θ are

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

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

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

Let $g(\delta, X, v_r; \theta) \equiv \{g_{jt}(\delta, X, v_r; \theta)\}_{j,t=1}^{J,T}$. We first need to show nonsingularity of the Jacobian matrix of the market shares with respect to δ because the matrix will appear later in the asymptotic distribution of the BLP estimator.

Lemma 1. *(Nonsingular Jacobian of Market Shares): 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.

PROOF. See appendix. ■

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 later be part of 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 that is stated in the appendix.

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 $\mathbb{E}_v[g(\delta_0, X, v; \theta_0)] \equiv \int g(\delta_0, X, v; \theta_0) dF_0(v)$,

$$\sqrt{R}(\hat{\delta}(\theta_0) - \delta_0) = -\left(\int G_\delta(\delta_0, X, v_r; \theta_0) dF_0(v_r)\right)^{-1} \frac{1}{\sqrt{R}} \sum_{r=1}^{R} \{g(\delta_0, X, v_r; \theta_0) - \mathbb{E}_v[g(\delta_0, X, v; \theta_0)]\}$$
$$+ o_p(1)$$

PROOF. See appendix. ■

3.3 Asymptotic Distribution of Sample Moment Conditions

Next we derive the asymptotic distribution of $\sqrt{m}\hat{\gamma}(\theta_0) = \sqrt{m}\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{J} Z_t\left(\hat{\delta}_t(\theta_0) - X_t'\theta_1\right)$. 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* $k = \underset{T\to\infty, R\to\infty}{lim} \frac{R}{T}$ *exists and the assumptions in proposition 2 are satisfied. Then for* $m = min(T, R)$*,*

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

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

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

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

Proof. From proposition 2, for all $t = 1...T$,

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

The sample and population moments are

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

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

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

$$\Omega = \underset{T\to\infty}{lim}\frac{1}{T}\sum_{t=1}^T Var\left(Z_t(\delta_{0t} - X_t'\theta_{0,1})\right) = Var\left(Z_t(\delta_{0t} - X_t'\theta_{0,1})\right)$$

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

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

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

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

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.

$$
\frac{1}{TR}S_{TR}(\theta_0) = \frac{1}{T}\sum_{t=1}^{T} f(Z_t, \delta_{0t}, X_t; \theta_0) + \frac{1}{R}\sum_{r=1}^{R} h\left(v_r; \theta_0\right) + o_p\left(\frac{1}{\sqrt{m}}\right)
$$

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

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

Using Theorem 2.7 in Neumeyer (2004),

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

where $k = \underset{T \to \infty, R \to \infty}{lim} \frac{R}{T}$, $\Sigma_f = Var(f(Z_t, \delta_{0t}, X_t; \theta_0)) = 0$, and $\Sigma_h = Var\left(h\left(v_r; \theta_0\right)\right)$.
Since $\{X_t, Z_t, \xi_t\}_{t=1}^{T}$ are independent of $\{v_r\}_{r=1}^{R}$,

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

■

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) $\theta_0 \in Interior\,(\Theta)$, *where* Θ *is a compact subset of* $\mathbb{R}^{2d+1}$.

(ii) $\mathbb{E}\left\|Z_t\right\|_\infty < \infty$

(iii) $\mathbb{E}\left\|Z_t X_t'\right\|_2 < \infty$

(iv) $\mathbb{E}\left[\max_{j=1\ldots J}\left|\left[1, x_{jt}'\right]' \circ v_r\right|\right] < \infty.$

Then for any $\kappa_m \to 0$,

$$\sup_{\|\theta-\theta_0\|\leq\kappa_m} \sqrt{m}\left\|\hat{\gamma}\left(\theta\right) - \hat{\gamma}\left(\theta_0\right) - \gamma\left(\theta\right)\right\| \xrightarrow{p} 0$$

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

Proof. See appendix. ■

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) after we show that the sample objective $\|\hat{\gamma}(\theta)\|_{W_T}$ converges uniformly to the population objective $\|\gamma(\theta)\|_W$. Theorem 4 states that $\underset{\theta\in\Theta}{sup}\|\hat{\gamma}(\theta)-\gamma(\theta)\| = o_p(1)$, which implies that $\underset{\theta\in\Theta}{sup}|\hat{\gamma}(\theta)-\gamma(\theta)| = o_p(1)$. Therefore,

$$
\begin{aligned}
&\underset{\theta\in\Theta}{sup}\left|\|\hat{\gamma}(\theta)\|_{W_T}-\|\gamma(\theta)\|_W\right| \\
&\leq \underset{\theta\in\Theta}{sup}\left|\|\hat{\gamma}(\theta)\|_W-\|\gamma(\theta)\|_W\right| + \underset{\theta\in\Theta}{sup}\left|\|\hat{\gamma}(\theta)\|_{W_T}-\|\hat{\gamma}(\theta)\|_W\right| \\
&\leq \underset{\theta\in\Theta}{sup}\|\hat{\gamma}(\theta)-\gamma(\theta)\|_W + \underset{\theta\in\Theta}{sup}|\hat{\gamma}(\theta)'W_T\hat{\gamma}(\theta)-\hat{\gamma}(\theta)'W\hat{\gamma}(\theta)| \\
&= \underset{\theta\in\Theta}{sup}\left(\hat{\gamma}(\theta)-\gamma(\theta)\right)'W\left(\hat{\gamma}(\theta)-\gamma(\theta)\right) + \underset{\theta\in\Theta}{sup}|\hat{\gamma}(\theta)'\left(W_T-W\right)\hat{\gamma}(\theta)| \\
&\leq \left(\underset{\theta\in\Theta}{sup}|\hat{\gamma}(\theta)-\gamma(\theta)|\right)'W\left(\underset{\theta\in\Theta}{sup}|\hat{\gamma}(\theta)-\gamma(\theta)|\right) + \underset{\theta\in\Theta}{sup}|\hat{\gamma}(\theta)'\left(W_T-W\right)\hat{\gamma}(\theta)| \\
&= o_p(1)
\end{aligned}
$$

■

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

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

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

(i) $\Gamma \equiv \frac{\partial}{\partial\theta}\underset{T,R\to\infty}{lim}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[Z_t\left(\hat{\delta}_t\left(\theta_2\right)-X_t'\theta_1\right)\right]\Big|_{\theta_0}$ *exists, and* $\Gamma'W\Gamma$ *is nonsingular.*

(ii) $\|\hat{\gamma}(\hat{\theta})\|_{W_T} \leq \underset{\theta}{inf}\|\hat{\gamma}(\theta)\|_{W_T} + o_p\left(m^{-1}\right)$.

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

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

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. We showed in Theorem 4 $\underset{\|\theta-\theta_0\|\leq\kappa_m}{sup} \frac{\sqrt{m}\|\hat{\gamma}(\theta)-\hat{\gamma}(\theta_0)-\gamma(\theta)\|}{1+\sqrt{m}\|\theta-\theta_0\|} \xrightarrow{p} 0$ for any $\kappa_m \to 0$. This is condition (v) in Theorem 7.2 of Newey and McFadden (1994), from which $\sqrt{m}$ consistency of $\hat{\theta}$ then follows. Since we showed in theorem 3 that $\sqrt{m}\hat{\gamma}(\theta_0) \xrightarrow{d} N(0,\Sigma) = N\left(0,(1\wedge k)\Omega + (1\wedge 1/k)\Sigma_h\right)$, $\sqrt{m}(\hat{\theta}-\theta_0) \xrightarrow{d} N\left(0, (\Gamma'W\Gamma)^{-1}\Gamma'W\Sigma W\Gamma\,(\Gamma'W\Gamma)^{-1}\right)$. ■

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

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

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

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

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 $\hat{\theta}_2$:

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

$G_\delta(\delta, X, v_r; \theta)$ 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 $\hat{\mu}_{rjt} \equiv \mu_{rjt}\left(\hat{\theta}_2, v_r\right)$,

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

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

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

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* $\mathbb{E}\left[\sup_{\|\theta-\theta_0\|\leq\kappa_m}\left\|Z_t\left(\hat{\delta}_t(\theta) - X_t'\theta_1\right)\right\|\right] < \infty$

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

Then,

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

Proof. See appendix. ■

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 $\rho_m : \mathbb{R}^{2d+1} \to \mathbb{R}_+$ *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)* $\rho(u) = \rho(-u)$.

(ii) $\pi : \Theta \to \mathbb{R}_+$ *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* $\hat{\theta}$*:*

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

Proof. See appendix. ∎

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.

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

$\tilde{\lambda}=\underset{T\to\infty,R\to\infty}{lim}\frac{\sqrt{T}}{\sqrt{R}}<\infty$, $V_1=(\Gamma'W\Gamma)^{-1}(\Gamma'W\Phi_1W\Gamma)(\Gamma'W\Gamma)^{-1}$, $V_2=(\Gamma'W\Gamma)^{-1}(\Gamma'W\Phi_2W\Gamma)(\Gamma'W\Gamma)^{-1}$. The Taylor expansion of $\sqrt{T}(\hat{\theta}-\theta_0)$ is

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

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

If we scale by $\sqrt{m}=\sqrt{min(R,T)}$, our Taylor expansion becomes, for $k=\underset{T\to\infty,R\to\infty}{lim}\frac{R}{T}$,

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

Since $0 \leq min\left(1, \sqrt{\frac{1}{k}}\right) \leq 1$ and $0 \leq min\left(1, \sqrt{\frac{1}{k}}\right) \leq 1$, $\frac{1}{R} min\left(1, \sqrt{\frac{1}{k}}\right) C_{1,T,R} \to 0$ as $R \to \infty$ even if $R \ll T$. The asymptotic bias disappears even if the number of simulation draws is much smaller than the number of markets.

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 v_1 plus v_2, where v_1 are draws from a mean zero normal random variable with variance equal to the absolute difference between the sample variance of price and the sample variance of ξ, and v_2 are draws from a multivariate normal random variable with mean and covariance equal to the sample mean and sample covariance of the instruments used in the Berry et al. (1995) paper. When forming the GMM objective, $\mathbb{E}\left[x_{jt}\xi_{jt}\right] = 0$ 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 $t = 1...T$, prices are determined as

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

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.

We first consider a simplified version of the Berry et al. (1995) model with only three covariates: Price, HPWT, and SIZE. The assumed true values of θ_0 and γ_0 are obtained from Table IV in Berry et al. (1995). We compare the empirical coverage frequencies of the following two types of confidence intervals:

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

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

Table 1 provides the average $\hat{\beta}$, empirical coverage frequencies taking into account simulation noise, and empirical coverage frequencies not taking into account simulation noise across with 2 products. We use 2800 markov chain iterations and burn in the first 1400 periods.

[Table 1 about here.]

Table 2 provides the average $\hat{\beta}$, empirical coverage frequencies taking into account simulation noise, and empirical coverage frequencies not taking into account simulation noise across 100 Monte Carlo Simulations, 100 Simulation Draws, and 100 markets with 2 products. We use 2800 markov chain iterations and burn in the first 1400 periods.

[Table 2 about here.]

Table 3 provides the average $\hat{\beta}$, empirical coverage frequencies taking into account simulation noise, and empirical coverage frequencies not taking into account simulation noise across 200 Monte Carlo Simulations, 200 Simulation Draws, and 50 markets with 2 products. We use 5600 markov chain iterations and burn in the first 1400 periods.

[Table 3 about here.]

As we can see, the empirical coverage frequencies for the confidence intervals taking into account simulation noise are very close to the nominal level of 95% while the confidence intervals which do not take into account simulation noise undercover.

In the second set of Monte Carlo simulations, we use all of the covariates in the Berry et al. (1995) model.

Table 4 provides the empirical coverage frequencies, average $\hat{\beta}$, average standard errors taking into account simulation noise, and average standard errors not taking into account simulation noise for 100 Monte Carlo Simulations, 50 Simulation Draws, and 100 markets with 2 products. We use 8800 markov chain iterations and burn in the first 2200 periods.

[Table 4 about here.]

Table 5 provides the empirical coverage frequencies, average $\hat{\beta}$, average standard errors taking into account simulation noise, and average standard errors not taking into account simulation noise for 100 Monte Carlo Simulations, 100 Simulation Draws, and 100 markets with 2 products. We use 5600 markov chain iterations and burn in the first 1400 periods.

[Table 5 about here.]

Table 6 provides the empirical coverage frequencies, average $\hat{\beta}$, average standard errors taking into account simulation noise, and average standard errors not taking into account simulation noise for 200 Monte Carlo Simulations, 200 Simulation Draws, and 50 markets with 2 products. We use 5600 markov chain iterations and burn in the first 1400 periods.

[Table 6 about here.]

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 $R \to \infty$ and $T \to \infty$. 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 reduce the computational burden of solving the simulated GMM optimization problem, we have proposed using a Laplace-type estimator implemented using Hamiltonian Markov Chain Monte Carlo. We have demonstrated consistency of the Laplace-type estimator and have illustrated through Monte Carlo simulations the similarity between the empirical and nominal coverage frequencies of confidence intervals formed using the consistent estimate of the asymptotic variance.

9 Appendix

9.1 Definitions and Theorems about Functionals

Definition. Suppose $f : U \to Y$ 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 $Df(u_0) : X \to Y$ 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:

$$Df(u_0) \cdot h = \underset{t \to 0}{lim} \frac{f(u_0 + th) - f(u_0)}{t} \equiv f'_{u_0}(h)$$

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 $f : U \to Y$ 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,*

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

PROOF. Define $c(t)=(1-t)u+tv$ and $\gamma(t)=(f\circ c)(t)$. Using the chain rule for derivatives of functionals, $\gamma'(t)=Df(c(t))\cdot c'(t)$. By the fundamental theorem of calculus,

$$f(v)-f(u)=\gamma(1)-\gamma(0)=\int_0^1\gamma'(t)dt=\int_0^1 Df(c(t))\cdot c'(t)dt=\int_0^1 Df((1-t)u+tv)\cdot(v-u)dt$$

If U is convex, $\int_0^1 Df((1-t)u+tv)\cdot(v-u)dt=\int_0^1 Df((1-t)u+tv)dt\cdot(v-u)=Df(u)\cdot(v-u)+(\int_0^1(Df((1-t)u+tv)-Df(u))dt)\cdot(v-u)$. ■

Corollary. *(Intermediate Value Theorem) Let* $U\subset X$ *be an open convex subset of a Banach space* X *and let* $f:U\to\mathbb{R}$ *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* $f(v)-f(u)=Df(c)\cdot(v-u)$.

PROOF. Using the Intermediate Value Theorem in $\mathbb{R}$, $\int_0^1 Df((1-t)u+tv)dt=Df(c)(1-0)=Df(c)$ for some c on the line segment $\{(1-t)u+tv|t\in[0,1]\}$. Using the Mean Value Theorem, $f(v)-f(u)=\int_0^1 Df((1-t)u+tv)dt\cdot(v-u)=Df(c)\cdot(v-u)$. ■

9.2 Proof of Lemma 1

In order to prove nonsingularity, we will use a theorem from Mas-Colell et al. (1995) (MWG) which states that a matrix with a dominant diagonal is nonsingular. This approach has been used in Berry et al. (1995) and Berry (1994).

Definition. (MWG Definition M.D.2): A $N \times N$ matrix M with generic entry a_{ij} has a dominant diagonal if there is $(p_1, ..., p_N) \gg 0$ such that for every $i = 1...N$, $|p_i a_{ii}| > \sum_{j \neq i} |p_j a_{ij}|$. Such a matrix is also called **strictly diagonally dominant**.

Theorem. *(MWG Theorem M.D.5): Let M be a $N \times N$ matrix with a dominant diagonal. Then M is nonsingular. Furthermore, if M is symmetric and has a positive diagonal, then M is positive definite.*

Proof. The partial derivatives of $g_{jt}(\delta_t, X_t, v_r; \theta) = \frac{exp(\delta_{jt}+\mu_{rjt})}{1+\sum_{k \in \mathcal{N}(t)} exp(\delta_{kt}+\mu_{rkt})}$ with respect to δ_{jt} and δ_{ht} for all $h \neq j$ are

$$\begin{aligned}\frac{\partial g_{jt}(\delta_t, X_t, v_r; \theta)}{\partial \delta_{jt}} &= \left(\frac{exp(\delta_{jt} + \mu_{rjt})}{1 + \sum_{k \in \mathcal{N}(t)} exp(\delta_{kt} + \mu_{rkt})}\right)\left(1 - \frac{exp(\delta_{jt} + \mu_{rjt})}{1 + \sum_{k \in \mathcal{N}(t)} exp(\delta_{kt} + \mu_{rkt})}\right) \\ &= g_{jt}(\delta_t, X_t, v_r; \theta)(1 - g_{jt}(\delta_t, X_t, v_r; \theta))\end{aligned}$$

$$\begin{aligned}\frac{\partial g_{jt}(\delta_t, X_t, v_r; \theta)}{\partial \delta_{ht}} &= -\left(\frac{exp(\delta_{jt} + \mu_{rjt})}{1 + \sum_{k \in \mathcal{N}(t)} exp(\delta_{kt} + \mu_{rkt})}\right)\left(\frac{exp(\delta_{ht} + \mu_{rht})}{1 + \sum_{k \in \mathcal{N}(t)} exp(\delta_{kt} + \mu_{rkt})}\right) \\ &= -g_{jt}(\delta_t, X_t, v_r; \theta) g_{ht}(\delta_t, X_t, v_r; \theta)\end{aligned}$$

Note that $\frac{\partial g_{jt}(\delta_t, X_t, v_r; \theta)}{\partial \delta_{jt}} \in (0, 1)$ because the numerator of the first term is strictly less than the denominator. Similarly, $\frac{\partial g_{jt}(\delta_t, X_t, v_r; \theta)}{\partial \delta_{ht}} \in (-1, 0)$ because the first term is between -1 and 0 and the second term is between 0 and 1.

Define the Jacobian matrix of the market shares with respect to δ as

$$\int G_\delta(\delta, X, v_r; \theta) dF(v_r) = \left\{\int \nabla_\delta g_t(\delta_t, X_t, v_r; \theta) dF(v_r)\right\}_{t=1}^{T}$$

Notice that $\int G_\delta(\delta, X, v_r; \theta) dF(v_r)$ is a block diagonal matrix due to the assumption of independent markets. Since the inverse of a block diagonal matrix is another block diagonal matrix with each block the inverse of the block in the original matrix, we can prove that

$\int G_\delta(\delta, X, v_r; \theta)dF(v_r)$ is nonsingular by showing that $\int \nabla_\delta g_t(\delta_t, X_t, v_r; \theta)dF(v_r)$ is nonsingular for all t.

Note that due to the presence of the outside good, the sum of the shares of the products in each market is less than one:

$$\sum_{j=1}^{J} g_{jt}(\delta_t, X_t, v_r; \theta) = \sum_{j=1}^{J} \frac{exp(\delta_{jt} + \mu_{rjt})}{1 + \sum_{k\in\mathcal{N}(t)} exp(\delta_{kt} + \mu_{rkt})} = \frac{\sum_{j\in\mathcal{N}(t)} exp(\delta_{jt} + \mu_{rjt})}{1 + \sum_{k\in\mathcal{N}(t)} exp(\delta_{kt} + \mu_{rkt})} < 1$$

Define $g_{jt} \equiv g_{jt}(\delta_t, X_t, v_r; \theta)$ and rearranging, $g_{jt}(1 - g_{jt}) > g_{jt} \sum_{k\neq j} g_{kt} \forall j = 1...J$.

Recall that $\frac{\partial g_{jt}}{\partial \delta_{jt}} = g_{jt}(1 - g_{jt})$ and $\frac{\partial g_{jt}}{\partial \delta_{kt}} = -g_{jt}g_{kt}$. Define $(p_{1t}, .., p_{Jt}) = (1, ..., 1)$. Taking the integral of both sides, $\int \left|p_{jt}\frac{\partial g_{jt}}{\partial \delta_{jt}}\right| dF(v_r) > \sum_{k\neq j} \int \left|p_{kt}\frac{\partial g_{jt}}{\partial \delta_{kt}}\right| dF(v_r) \forall j = 1...J$.

Using $\frac{\partial g_{jt}}{\partial \delta_{jt}} > 0$ and the triangle inequality, $\forall j = 1...J$,

$$\begin{aligned} \left|p_{jt} \int \frac{\partial g_{jt}}{\partial \delta_{jt}} dF(v_r)\right| &= \int \left|p_{jt}\frac{\partial g_{jt}}{\partial \delta_{jt}}\right| dF(v_r) \\ &> \sum_{k\neq j} \int \left|p_{kt}\frac{\partial g_{jt}}{\partial \delta_{kt}}\right| dF(v_r) \\ &\geq \sum_{k\neq j} \left|p_{kt} \int \frac{\partial g_{jt}}{\partial \delta_{kt}} dF(v_r)\right| \end{aligned}$$

Therefore, $\int \nabla_\delta g_t(\delta_t, X_t, v_r; \theta)dF(v_r)$ has a dominant diagonal for all values of X_t, δ_t, θ, and F, which implies that $\int \nabla_\delta g_t(\delta_t, X_t, v_r; \theta)dF(v_r)$ is nonsingular for all values of X_t, δ_t, θ, and F. Therefore, $\int G_\delta(\delta, X, v_r; \theta)dF(v_r)$ is nonsingular for all values of X, δ, θ, and F. Furthermore, since $\frac{\partial g_{jt}}{\partial \delta_{jt}} > 0$ and $\frac{\partial g_{jt}}{\partial \delta_{kt}} = \frac{\partial g_{kt}}{\partial \delta_{jt}}$ for all j, k, t, $\int G_\delta(\delta, X, v_r; \theta)dF(v_r)$ has a positive diagonal and is symmetric, which implies that $\int G_\delta(\delta, X, v_r; \theta)dF(v_r)$ is positive definite by Theorem M.D.5 in Mas-Colell et al. (1995). ■

9.3 Proof of Proposition 2

Proof. We will show asymptotic normality of $\sqrt{R}(\hat{\delta}\left(\theta_0\right) - \delta_0)$ 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 $\delta^* = t_\delta \delta_0 + (1-t_\delta)\hat{\delta}\left(\theta_0\right)$ and $F^* = t_F F_0 + (1-t_F)\hat{F}$ for $t_F \in [0,1]$ and $t_\delta \in [0,1]$:

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

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 $\hat{\delta}\left(\theta_0\right) - \delta_0$ 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.

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

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

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

$s'_{\delta}(\cdot)$ 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 $s'_{\delta^*}\left(\hat{\delta}(\theta_0) - \delta_0\right) \equiv D_{\delta} s(\delta^*, X, F; \theta_0) \cdot \left(\hat{\delta}(\theta_0) - \delta_0\right)$.

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

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

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 $s'_{F^*}\left(\hat{F} - F_0\right) \equiv D_F s(\delta^*, X, F^*; \theta_0) \cdot \left(\hat{F} - F_0\right)$.

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 $\|F - F'\| < \nu \implies \left\| \int g(\delta, X, v_r; \theta_0) d(F' - F)(v_r) \right\| < \epsilon$. Since $\left\| \int g(\delta, X, v_r; \theta_0) d(F' - F)(v_r) \right\| \leq \left\| \int |g(\delta, X, v_r; \theta_0)| \, d(F' - F)(v_r) \right\| \leq \left\| \int d(F' - F)(v_r) \right\| = \|F - F'\|$, 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,

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

Rearranging,

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

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

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

Recall that all of the elements of $G_\delta\left(\delta, X, v_r; \theta_0\right)$ lie in $(-1,0)\cup(0,1)$ since $\frac{\partial g_{jt}}{\partial \delta_{jt}} = g_{jt}(1-g_{jt})$ for all $j=1...J$, $t=1...T$, and $\frac{\partial g_{jt}}{\partial \delta_{kt}} = -g_{jt}g_{kt}$ for all $k \neq j$, $t=1...T$. Therefore, $\underset{\delta\in\mathbb{D}}{sup}\left\|G_\delta\left(\delta, X, v_r; \theta_0\right)\right\| < \infty$ and $g\left(\delta, X, v_r; \theta_0\right)$ is Lipschitz in δ with uniformly bounded Lipschitz constant $\left\|G_\delta\left(\delta, X, v_r; \theta_0\right)\right\|$. 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 $\sqrt{R}\left(\hat{\delta}\left(\theta_0\right)-\delta_0\right) = O_p\left(1\right)$.

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

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

Note that all of the elements of $\nabla^2_\delta g\left(\delta, X, v_r; \theta_0\right)$ lie in $(-2,0)\cup(0,2)$ since for all $t=1...T$,

$\frac{\partial^2 g_{jt}}{\partial \delta_{jt}^2} = g_{jt}(1-g_{jt})^2 - 2g_{jt}^2(1-g_{jt})$ for all $j = 1...J$, $\frac{\partial^2 g_{jt}}{\partial \delta_{kt}^2} = -g_{jt}(1-g_{kt})g_{kt} + g_{jt}g_{kt}^2$ for all $k \neq j$ $\frac{\partial^2 g_{jt}}{\partial \delta_{kt}\partial \delta_{jt}} = -g_{jt}(1-g_{jt})g_{kt} + g_{jt}^2 g_{kt}$ 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, $\underset{\delta \in \mathbb{D}}{sup} \left\| \nabla_\delta^2 g\left(\delta, X, v_r; \theta_0\right) \right\| < \infty$ and $\nabla \mathcal{G} = \left\{ G_\delta\left(\delta, X, v_r; \theta_0\right) : \delta \in \mathbb{D} \right\}$ is a parametric class and therefore a Donsker class. It follows then that

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

Note that δ^* is also $\sqrt{R}$-consistent for δ_0 since it lies between the $\sqrt{R}$-consistent estimator $\hat{\delta}\left(\theta_0\right)$ 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,

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

■

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 $\hat{\delta}\left(\theta\right)$ is a continuously differentiable function of θ. By the intermediate value theorem, there exists $\theta^* \in [\theta, \theta_0]$ such that

$\hat{\delta}_t(\theta) - \hat{\delta}_t(\theta_0) = \frac{\partial \hat{\delta}_t(\theta^*)}{\partial \theta}(\theta - \theta_0)$.

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

Recall that $\frac{\partial \hat{\delta}(\theta)}{\partial \theta_1} = 0$ and $\frac{\partial \hat{\delta}(\theta)}{\partial \theta_2} = -\left(\frac{1}{R}\sum_{r=1}^{R} G_\delta\left(\hat{\delta}, X, v_r; \theta\right)\right)^{-1}\frac{1}{R}\sum_{r=1}^{R} G_{\theta_2}\left(\hat{\delta}, X, v_r; \theta\right)$. If we can show that $\mathbb{E}\left[\underset{\theta\in\Theta}{sup}\left\|Z_t\frac{\partial \hat{\delta}_t(\theta)}{\partial \theta_2}\right\|_2\right] < \infty$, 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)):

$$
\underset{\theta\in\Theta}{sup}\left\| \frac{1}{T}\sum_{t=1}^{T} Z_t\frac{\partial \hat{\delta}_t(\theta)}{\partial \theta_2} - \underset{T,R\to\infty}{lim}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[Z_t\frac{\partial \hat{\delta}_t(\theta)}{\partial \theta_2}\right]\right\|_2 = o_p(1)
$$

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

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

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 $t = 1...T$,

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

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

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

Since $\mathbb{E}\left[\underset{j=1...J}{max}\left|[1,x'_{jt}]'\circ v_r\right|\right] < \infty$ by assumption, $\mathbb{E}\left[\underset{\theta\in\Theta}{sup}\frac{1}{R}\sum_{r=1}^{R}\left|\frac{\partial g_{jt}\left(\hat{\delta}_t, X_t, v_r; \theta\right)}{\partial\theta_2}\right|\right] \leq \mathbb{E}\left[\frac{1}{R}\sum_{r=1}^{R}\underset{j=1...J}{max}\left|[1,x'_{jt}]'\circ v_r\right|\right] < \infty$ and $\mathbb{E}\left[\underset{\theta\in\Theta}{sup}\left\|\frac{\partial\hat{\delta}_t(\theta)}{\partial\theta_2}\right\|_{\infty}\right] \leq C\mathbb{E}\left[\underset{\theta\in\Theta}{sup}\left\|\frac{1}{R}\sum_{r=1}^{R}\nabla_{\theta_2}g_t\left(\hat{\delta}_t, X_t, v_r; \theta\right)\right\|_{\infty}\right] < \infty$. This combined with $\mathbb{E}\left[\|Z_t\|_{\infty}\right] < \infty$ implies that $\mathbb{E}\left[\underset{\theta\in\Theta}{sup}\left\|Z_t\frac{\partial\hat{\delta}_t(\theta)}{\partial\theta_2}\right\|_{\infty}\right] < \infty$ which implies that $\mathbb{E}\left[\underset{\theta\in\Theta}{sup}\left\|Z_t\frac{\partial\hat{\delta}_t(\theta)}{\partial\theta_2}\right\|_{2}\right] < \infty$. It follows that

$$\underset{\theta\in\Theta}{sup}\left\|\frac{1}{T}\sum_{t=1}^{T}Z_t\frac{\partial\hat{\delta}_t(\theta)}{\partial\theta_2} - \underset{T,R\to\infty}{lim}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[Z_t\frac{\partial\hat{\delta}_t(\theta)}{\partial\theta_2}\right]\right\|_2 = o_p(1)$$

Additionally, since $\mathbb{E}\left[\|Z_t X_t'\|_2\right] < \infty$, the weak law of large numbers implies that

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

Therefore, stochastic equicontinuity holds:

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

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

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

for $B_T = \underset{\theta\in\Theta}{sup} \left\| \frac{1}{T} \sum_{t=1}^{T} Z_t \frac{\partial \hat{\delta}_t(\theta)}{\partial \theta_2} \right\|_2 + \left\| \frac{1}{T} \sum_{t=1}^{T} Z_t X_t' \right\|_2 \leq \frac{1}{T} \sum_{t=1}^{T} \left(\underset{\theta\in\Theta}{sup} \left\| Z_t \frac{\partial \hat{\delta}_t(\theta)}{\partial \theta_2} \right\|_2 + \|Z_t X_t'\|_2 \right) = O_p(1)$ since $\mathbb{E}\left[\underset{\theta\in\Theta}{sup} \left\| Z_t \frac{\partial \hat{\delta}_t(\theta)}{\partial \theta_2} \right\|_2 \right] < \infty$ and $\mathbb{E}\left[\|Z_t X_t'\|_2\right] < \infty$.
Since $\gamma(\theta) = \underset{T,R\to\infty}{lim} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\left[Z_t \left(\hat{\delta}_t(\theta) - X_t'\theta_1 \right) \right]$ 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

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

■

9.5 Proof of Theorem 7

PROOF. Recall that

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

We showed in theorem 4 that $\mathbb{E}\left[\underset{\theta\in\Theta}{sup}\left\|Z_t\frac{\partial\hat{\delta}_t(\theta)}{\partial\theta_2}\right\|_2\right]<\infty$. 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,

$$\hat{\Gamma}=\left[\begin{array}{cc}-\frac{1}{T}\sum_{t=1}^{T}Z_tX_t', & \frac{1}{T}\sum_{t=1}^{T}Z_t\left.\frac{\partial\hat{\delta}_t(\theta)}{\partial\theta_2}\right|_{\hat{\theta}}\end{array}\right]\xrightarrow{p}\left[\begin{array}{cc}-\mathbb{E}\left[Z_tX_t'\right], & \underset{T,R\to\infty}{lim}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[Z_t\left.\frac{\partial\hat{\delta}_t(\theta)}{\partial\theta_2}\right|_{\theta_0}\right]\end{array}\right]$$

Since we also assumed $\mathbb{E}\left\|Z_tX_t'\right\|_2<\infty$ and $Z_t\left(\hat{\delta}_t\left(\theta\right)-X_t'\theta_1\right)$ 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 $\hat{\delta}\left(\hat{\theta}_2\right)\xrightarrow{p}\delta_0$, $\hat{\theta}\xrightarrow{p}\theta_0$, and there exists $\kappa_m\downarrow 0$ such that $\mathbb{E}\left[\underset{\|\theta-\theta_0\|\leq\kappa_m}{sup}\left\|Z_t\left(\hat{\delta}_t\left(\theta\right)-X_t'\theta_1\right)\right\|\right]<\infty$, by Lemma 4.3 of Newey and McFadden (1994),

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

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

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

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

Newey and McFadden (1994),

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

By the Continuous Mapping Theorem,

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

Similarly, note that for all $t=1...T$, $g_t\left(\delta_t,X_t,v_r;\theta\right)$ is continuous in δ_t and θ, and $g_t\left(\delta_t,X_t,v_r;\theta\right)\in(0,1)$ for all δ_t, X_t, v_r, and θ. Since $\hat{\theta}\xrightarrow{p}\theta_0$ and $\hat{\delta}\left(\hat{\theta}_2\right)\xrightarrow{p}\delta_0$, by Lemma 4.3 of Newey and McFadden (1994),

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

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

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

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

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

We assumed in theorem 4 that $\mathbb{E}\left\|Z_t\right\|_{\infty}<\infty$, which implies that $\left\|Z_t\right\|_{\infty}=O_p\left(1\right)$. Furthermore, we assumed

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

Therefore,

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

Then it follows that

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

Note that $\frac{1}{R}\sum_{r=1}^{R}\tilde{h}\left(v_r;\theta_0\right)\tilde{h}\left(v_r;\theta_0\right)'$ 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 $k\left(\delta_{0t},X_t,v_r;\theta_0\right)=-Z_t\left(\int\nabla_\delta g_t(\delta_{0t},X_t,v_r;\theta_0)dF_0(v_r)\right)^{-1}\left(g_t(\delta_{0t},X_t,v_r;\theta_0)-\mathbb{E}_v[g_t(\delta_{0t},X_t,v;\theta_0)]\right)$,

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

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

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

■

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 $\{\theta_m\} \in \Theta$, $\limsup_{m\to\infty} \|\gamma(\theta_m)\|_W \leq \|\gamma(\theta_0)\|_W$ implies $\|\theta_m - \theta_0\| \to 0$. We also showed in theorem 4 that $\gamma(\theta)$ is a continuous function of θ and that $\|\hat{\gamma}(\hat{\theta})\|_{W_T}$ converges in probability to $\|\gamma(\theta)\|_W$ uniformly over Θ. It then follows by Lemma 1 of Chernozhukov and Hong (2003) that assumption 3 is satisfied: for any $\kappa > 0$, there exists ϵ such that $\liminf_{m\to\infty} \mathbb{P}\left(\inf_{\|\theta-\theta_0\|\geq\kappa} (\|\hat{\gamma}(\theta)\|_{W_T} - \|\hat{\gamma}(\theta_0)\|_{W_T}) \geq \epsilon\right) = 1$. 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

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

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 $\kappa_m \to 0$,

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

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 $\epsilon(\theta) \equiv \frac{\hat{\gamma}(\theta)-\hat{\gamma}(\theta_0)-\gamma(\theta)}{1+\sqrt{m}\|\theta-\theta_0\|}$:

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

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

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

■

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. 32, 35

Berry, Steven, "Estimating Discrete-Choice Models of Product Differentiation," *RAND Journal of Economics*, 1994, *25* (2), 242–262. 19, 23, 28

_ **, James Levinsohn, and Ariel Pakes**, "Automobile Prices in Market Equilibrium," *Econometrica*, 1995, *63* (4), 841–890. 1, 2, 3, 5, 18, 19, 21, 23

Chernozhukov, Victor and Han Hong, "A MCMC Approach to Classical Estimation," *Journal of Econometrics*, 2003, *115* (2), 293–346. 3, 16, 37, 38

der Vaart, Aad W Van, *Asymptotic statistics*, Vol. 3, Cambridge university press, 1998. 2, 29, 37

Duane, Simon, Anthony D Kennedy, Brian J Pendleton, and Duncan Roweth, "Hybrid monte carlo," *Physics letters B*, 1987, *195* (2), 216–222. 3

Freyberger, Joachim, "Asymptotic theory for differentiated products demand models with many markets," *Journal of Econometrics*, 2015, *185* (1), 162–181. 2, 3, 17

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. 18

Kristensen, Dennis and Bernard Salanié, "Higher-order properties of approximate estimators," *Journal of Econometrics*, 2017, *198* (2), 189–208. 2

Lee, Lung-Fei, "Asymptotic bias in simulated maximum likelihood estimation of discrete choice models," *Econometric Theory*, 1995, *11* (3), 437–483. 2

Mas-Colell, Andreu, Michael Whinston, and Jerry Green, *Microeconomic Theory*, Oxford University Press, 1995. 23, 25

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. 2

Neal, Radford M et al., "MCMC using Hamiltonian dynamics," *Handbook of Markov Chain Monte Carlo*, 2011, *2* (11). 3

Neumeyer, Natalie, "A central limit theorem for two-sample U-processes," *Statistics and Probability Letters*, 2004, *67*, 73–85. 2, 6, 10

Newey, Whitney K and Daniel McFadden, "Large sample estimation and hypothesis testing," *Handbook of econometrics*, 1994, *4*, 2111–2245. 13, 31, 33, 34, 35

Train, Kenneth E, *Discrete choice methods with simulation*, Cambridge university press, 2009. 2

van der Vaart, Aad W. and Jon Wellner, *Weak Convergence and Empirical Processes*, Springer-Verlag, 1996. 12

Varah, James M, "A lower bound for the smallest singular value of a matrix," *Linear Algebra and its Applications*, 1975, *11* (1), 3–5. 32, 35

List of Tables

Table 1: Results for 100 Monte Carlo Simulations, 50 Simulation Draws, and 100 markets with 2 products each

	Price	Constant	HPWT	SIZE
average $\hat{\beta}$	-0.134	-7.745	2.181	2.390
Empirical Coverage w/ simulation noise	0.960	0.970	0.960	0.950
Empirical Coverage w/o simulation noise	0.860	0.850	0.820	0.830

Table 2: Results for 100 Monte Carlo Simulations, 100 Simulation Draws, and 100 markets with 2 products each

	Price	Constant	HPWT	SIZE
average $\hat{\beta}$	-0.105	-7.670	2.054	2.231
Empirical Coverage w/ simulation noise	0.940	0.960	0.940	0.940
Empirical Coverage w/o simulation noise	0.830	0.850	0.840	0.840

Table 3: Results for 200 Monte Carlo Simulations, 200 Simulation Draws, and 50 markets with 2 products each

	Price	Constant	HPWT	SIZE
average $\hat{\beta}$	-0.197	-7.192	1.972	2.469
Empirical Coverage w/ simulation noise	0.955	0.960	0.950	0.955
Empirical Coverage w/o simulation noise	0.845	0.835	0.815	0.825

Table 4: Results for 100 Monte Carlo Simulations, 50 Simulation Draws, and 100 markets with 2 products each

	Price	Constant	HPWT	AIR	MPD	SIZE
average $\hat{\beta}$	-0.323	-6.646	2.310	0.612	-0.098	2.797
Empirical Coverage w/ simulation noise	0.95	0.94	0.94	0.96	0.97	0.96
Empirical Coverage w/o simulation noise	0.80	0.80	0.80	0.81	0.76	0.79

Table 5: Results for 100 Monte Carlo Simulations, 100 Simulation Draws, and 100 markets with 2 products each

	Price	Constant	HPWT	AIR	MPD	SIZE
average $\hat{\beta}$	-0.169	-7.627	2.274	0.531	-0.096	2.441
Empirical Coverage w/ simulation noise	0.94	0.96	0.97	0.97	0.95	0.96
Empirical Coverage w/o simulation noise	0.82	0.82	0.84	0.86	0.85	0.87

Table 6: Results for 200 Monte Carlo Simulations, 200 Simulation Draws, and 50 markets with 2 products each

	Price	Constant	HPWT	AIR	MPD	SIZE
average $\hat{\beta}$	-0.126	-7.653	2.166	0.509	-0.149	2.281
Empirical Coverage w/ simulation noise	0.950	0.970	0.930	0.965	0.940	0.925
Empirical Coverage w/o simulation noise	0.830	0.835	0.800	0.820	0.825	0.835