Supplementary appendix
Asymptotic theory for differentiated products demand models with many markets

Joachim Freyberger[‡]

October 12, 2014

A Proofs of lemmas

Proof of Lemma A1. Denote the expectation with respect to the distribution v_{rt} and conditional on x_t as E_t^*. Then we have to show that

$$\sup_{\theta\in\Theta}\max_{1\leq t\leq T}\left|\frac{1}{R}\sum_{r=1}^{R}\left(f(x_t,\theta,v_{rt})-E_t^*(f(x_t,\theta,v_{rt}))\right)\right|\xrightarrow{p}0$$

or that for any $\varepsilon>0$,

$$E_x\left(\Pr_t^*\left(\sup_{\theta\in\Theta}\max_{1\leq t\leq T}\left|\frac{1}{R}\sum_{r=1}^{R}\left(f(x_t,\theta,v_{rt})-E_t^*(f(x_t,\theta,v_{rt}))\right)\right|>\varepsilon\right)\right)\rightarrow 0 \text{ as } T\rightarrow\infty,$$

where $\Pr_t^*$ denotes the probability with respect to the distribution v and conditional on x.

Notice that

$$\sup_{\theta\in\Theta}\left|\frac{1}{R}\sum_{r=1}^{R}\left(f(x_t,\theta,v_{rt})-E_t^*(f(x_t,\theta,v_{rt}))\right)\right|\xrightarrow{p}0$$

simply follows from Jennrich's uniform law of large numbers (ULLN). The difference here is that we also have a maximum over t and this index affects both x_t (which is easy to deal with since $\mathcal{X}$ is compact) and v_{rt}. The proof uses similar arguments as the proof of Jennrich's ULLN. First define $\lambda=(x,\theta)$ and $\Lambda=\mathcal{X}\times\Theta$. Furthermore, denote $f(\lambda,v)=f(x,\theta,v)$. Now partition Λ in $\Lambda_1^n,\ldots,\Lambda_n^n$ such that the difference between any two elements in Λ_i^n goes to 0 as $n\rightarrow\infty$ for all i. Let λ_i^n be an arbitrary element from Λ_i^n for all i. Then

[‡]University of Wisconsin - Madison, Department of Economics. Email: jfreyberger@ssc.wisc.edu.

$$
\begin{aligned}
&\Pr_t^*\left(\sup_{\theta\in\Theta}\max_{1\le t\le T}\left|\frac{1}{R}\sum_{r=1}^{R}\left(f(x_t,\theta,v_{rt})-E_t^*(f(x_t,\theta,v_{rt}))\right)\right|>\varepsilon\right)\\
&\le\sum_{t=1}^{T}\Pr_t^*\left(\sup_{\lambda\in\Lambda}\left|\frac{1}{R}\sum_{r=1}^{R}\left(f(\lambda,v_{rt})-E_t^*(f(\lambda,v_{rt}))\right)\right|>\varepsilon\right)\\
&\le\sum_{t=1}^{T}\Pr_t^*\left(\bigcup_{i=1}^{n}\sup_{\lambda\in\Lambda_i^n}\left|\frac{1}{R}\sum_{r=1}^{R}\left(f(\lambda,v_{rt})-E_t^*(f(\lambda,v_{rt}))\right)\right|>\varepsilon\right)\\
&\le\sum_{t=1}^{T}\sum_{i=1}^{n}\Pr_t^*\left(\sup_{\lambda\in\Lambda_i^n}\left|\frac{1}{R}\sum_{r=1}^{R}\left(f(\lambda,v_{rt})-E_t^*(f(\lambda,v_{rt}))\right)\right|>\varepsilon\right)\\
&\le\sum_{t=1}^{T}\sum_{i=1}^{n}\Pr_t^*\left(\left|\frac{1}{R}\sum_{r=1}^{R}\left(f(\lambda_i^n,v_{rt})-E_t^*(f(\lambda_i^n,v_{rt}))\right)\right|>\varepsilon/2\right)\\
&\quad+\sum_{t=1}^{T}\sum_{i=1}^{n}\Pr_t^*\left(\frac{1}{R}\sum_{r=1}^{R}\sup_{\lambda\in\Lambda_i^n}\left|f(\lambda,v_{rt})-f(\lambda_i^n,v_{rt})+E_t^*(f(\lambda_i^n,v_{rt}))-E_t^*(f(\lambda,v_{rt}))\right|>\varepsilon/2\right)
\end{aligned}
$$

The first term converges to 0 if $\frac{\ln(T)}{R}\to 0$ because by the Bernstein inequality for bounded random variables, there exists a constant C such that for each fixed λ and t

$$
\Pr_t^*\left(\left|\sum_{r=1}^{R}\left(f(\lambda,v_{rt})-E_t^*(f(\lambda,v_{rt}))\right)\right|>R\varepsilon\right)\le 2\exp\left(-\frac{\varepsilon^2R^2}{CR}\right)=O\left(\exp(-R)\right).
$$

For the second term first assume that w_t has compact support. Then for all r and t, $v_{rt}\in\mathcal{V}$ where $\mathcal{V}$ is compact. Then, since f is a continuous function on a compact set, f is by the Heine Cantor theorem uniformly continuous. Hence, for n large enough (so large that $\sup_{\lambda\in\Lambda_i^n}||\lambda_i^n-\lambda||\le\delta$ for some small δ, but n is finite), $\sup_{v\in\mathcal{V}}\sup_{\lambda\in\Lambda_i^n}|f(\lambda,v)-f(\lambda_i^n,v)|\le\varepsilon/4$. Hence, also for all t,

$$
\sup_{\lambda\in\Lambda_i^n}|E_t^*(f(\lambda_i^n,v_{rt}))-E_t^*(f(\lambda,v_{rt}))|\le\varepsilon/4,
$$

which implies that

$$
\Pr_t^*\left(\frac{1}{R}\sum_{r=1}^{R}\sup_{\lambda\in\Lambda_i^n}|f(\lambda,v_{rt})-f(\lambda_i^n,v_{rt})+E_t^*(f(\lambda_i^n,v_{rt}))-E_t^*(f(\lambda,v_{rt}))|>\varepsilon/2\right)=0.
$$

Alternatively assume that for all l, $v_{rt}=g(a_t,w_{rt})$ where $w_{lrt}\sim\tilde{P}_t\in\{P^1,\dots,P^m\}$ where m is finite and $a_t\in\mathcal{A}$. With abuse of notation now define $\lambda=(x,\theta,a)$ and $\Lambda=\mathcal{X}\times\Theta\times\mathcal{A}$. Furthermore, denote $f(\lambda,w)=f(x,\theta,v)=f(x,\theta,g(a,w))$. Now partition Λ in $\Lambda_1^n,\dots,\Lambda_n^n$ such that the difference between any two elements in Λ_i^n goes to 0 as $n\to\infty$ for all i. Let λ_i^n be an arbitrary element from Λ_i^n for all i. Then

$$\begin{aligned}
&\Pr_t^*\left(\sup_{\theta\in\Theta}\max_{1\le t\le T}\left|\frac{1}{R}\sum_{r=1}^{R}(f(x_t,\theta,v_{rt})-E_t^*(f(x_t,\theta,v_{rt})))\right|>\varepsilon\right)\\
&\le\sum_{t=1}^{T}\Pr_t^*\left(\sup_{\theta\in\Theta}\sup_{x\in\mathcal{X}}\sup_{a\in\mathcal{A}}\left|\frac{1}{R}\sum_{r=1}^{R}(f(x,\theta,g(a,w_{rt}))-E_t^*(f(x,\theta,g(a,w_{rt}))))\right|>\varepsilon\right)\\
&=\sum_{t=1}^{T}\Pr_t^*\left(\sup_{\lambda\in\Lambda}\left|\frac{1}{R}\sum_{r=1}^{R}(f(\lambda,w_{rt})-E_t^*(f(\lambda,w_{rt})))\right|>\varepsilon\right)\\
&\le\sum_{t=1}^{T}\sum_{i=1}^{n}\Pr_t^*\left(\left|\frac{1}{R}\sum_{r=1}^{R}(f(\lambda_i^n,w_{rt})-E_t^*(f(\lambda_i^n,w_{rt})))\right|>\varepsilon/2\right)\\
&\quad+\sum_{t=1}^{T}\sum_{i=1}^{n}\Pr_t^*\left(\frac{1}{R}\sum_{r=1}^{R}\sup_{\lambda\in\Lambda_i^n}\left|f(\lambda,w_{rt})-f(\lambda_i^n,w_{rt})+E_t^*(f(\lambda_i^n,w_{rt}))-E_t^*(f(\lambda,w_{rt}))\right|>\varepsilon/2\right)
\end{aligned}$$

Again, the first term converges to 0 if $\frac{\ln(T)}{R}\to 0$ by the Bernstein inequality for bounded random variables. For the second term define $h_t(\lambda,w_{rt})=f(\lambda,w_{rt})-E_t^*(f(\lambda,w_{rt}))$ where the expected value is with respect to w_{rt} which can only be drawn from a finite number of distributions for each t. Then for all t,

$$\lim_{n\to\infty}\sup_{\lambda\in\Lambda_i^n}|(h_t(\lambda,w_{rt})-h_t(\lambda_i^n,w_{rt}))|=0$$

uniformly over i. Thus, by Lebesgue's dominated convergence theorem,

$$\lim_{n\to\infty}E_t^*\sup_{\lambda\in\Lambda_i^n}|(h_t(\lambda,w_{rt})-h_t(\lambda_i^n,w_{rt}))|=0$$

uniformly over i. This function depends on t since the distribution might differ. However, since only a finite number of distributions are allowed, there exists an n such that for all t

$$E_t^*\sup_{\lambda\in\Lambda_i^n}|(h(\lambda,w_{rt})-h(\lambda_i^n,w_{rt}))|\le\varepsilon/4.$$

Then

$$\begin{aligned}
&\sum_{t=1}^{T}\sum_{i=1}^{n}\Pr^*\left(\frac{1}{R}\sum_{r=1}^{R}\sup_{\lambda\in\Lambda_i^n}|(h_t(\lambda,w_{rt})-h_t(\lambda_i^n,w_{rt}))|>\varepsilon/2\right)\\
\le\;&\sum_{t=1}^{T}\sum_{i=1}^{n}\Pr^*\left(\frac{1}{R}\sum_{r=1}^{R}\sup_{\lambda\in\Lambda_i^n}|(h_t(\lambda,w_{rt})-h_t(\lambda_i^n,w_{rt}))|-E_t^*\sup_{\lambda\in\Lambda_i^n}|(h_t(\lambda,w_{rt})-h_t(\lambda_i^n,w_{rt}))|>\varepsilon/4\right)
\end{aligned}$$

This probability converges to 0 using the Bernstein inequality as in the first part.

One can combine these results to show that

$$\sup_{\theta\in\Theta}\max_{1\le t\le T}\left|\frac{1}{R}\sum_{r=1}^{R} f(x_t,\theta,v_{rt}^1,v_{rt}^2)-\int f(x_t,\theta,v^1,v^2)dP_t^1(v_1)dP_t^2(v_2)\right|\xrightarrow{p}0 \qquad \text{as } T\to\infty$$

where v^1 satisfies condition $(iii-a)$ and v^2 satisfies condition $(iii-b)$. □

Proof of Lemma A3. See Lemma A in Serfling (1980). □

Proof of Lemma A2. The proof is very similar to the proof of Lemma A.2 in Lee (1995). By Hölder's inequality

$$E\left(y_{Rt}^2\right)\le\left(E\left(|s(x_t)|^{2a}\left[\frac{1}{R}\sum_{r=1}^{R}p\left(v_r^{(t)},x_t\right)\right]^{2am_1}\right)\right)^{1/a}\left(E\left(|s(x_t)|^{2b}\left[\frac{1}{R}\sum_{r=1}^{R}q\left(v_r^{(t)},x_t\right)\right]^{2bm_2}\right)\right)^{1/b}.$$

Lemma A2 implies that for some constant c,

$$E\left(\left[\frac{1}{R}\sum_{r=1}^{R}p\left(v_r^{(t)},x_t\right)\right]^{2am_1}\middle|\,x_t\right)\le\frac{c}{R^{m_1a}}E\left(p\left(v_r^{(t)},x_t\right)^{2am_1}\middle|\,x_t\right)$$

and

$$E\left(\left[\frac{1}{R}\sum_{r=1}^{R}q\left(v_r^{(t)},x_t\right)\right]^{2bm_2}\middle|\,x_t\right)\le\frac{c}{R^{m_2b}}E\left(q\left(v_r^{(t)},x_t\right)^{2bm_2}\middle|\,x_t\right).$$

It follows that

$$\begin{aligned}E\left(y_{Rt}^2\right)&\le\frac{c}{R^{m_1+m_2}}\left(E\left(|s(x_t)|^{2a}p\left(v_r^{(t)},x_t\right)^{2am_1}\right)\right)^{1/a}\left(E\left(|s(x_t)|^{2b}p\left(v_r^{(t)},x_t\right)^{2bm_2}\right)\right)^{1/b}\\&\le\frac{cM}{R^{m_1+m_2}}.\end{aligned}$$

By Markov's inequality it now follows that

$$\begin{aligned}P\left(R^{(m_1+m_2)/2}\frac{1}{T}\sum_{t=1}^{T}|y_{Rt}|\ge\varepsilon\right)&\le\frac{R^{(m_1+m_2)/2}}{\varepsilon}\frac{1}{T}\sum_{t=1}^{T}E\left(|y_{Rt}|\right)\\&\le\frac{R^{(m_1+m_2)/2}}{\varepsilon}\frac{1}{T}\sum_{t=1}^{T}\left(E\left(|y_{Rt}|^2\right)\right)^{1/2}\\&\le\frac{R^{(m_1+m_2)/2}}{\varepsilon}\frac{1}{T}\sum_{t=1}^{T}\left(\frac{cM}{R^{m_1+m_2}}\right)^{1/2}=\frac{(cM)^{1/2}}{\varepsilon},\end{aligned}$$

which means that $R^{(m_1+m_2)/2}\frac{1}{T}\sum_{t=1}^{T}|y_{Rt}|$ is $O_p(1)$. □

Proof of Lemma A4. This is a standard result. □

B Monte Carlo simulations with same draws

Table A1: Coverage rates of 95% confidence intervals for β^p and ratio of mean length

	50 draws	100 draws	200 draws	400 draws
	GMM point estimates and GMM standard errors			
50 markets	0.817	0.855	0.845	0.863
100 markets	0.772	0.832	0.864	0.875
200 markets	0.707	0.803	0.849	0.894
400 markets	0.640	0.767	0.801	0.871
	Bias corrected point estimates and GMM standard errors			
50 markets	0.876	0.891	0.892	0.907
100 markets	0.839	0.898	0.905	0.918
200 markets	0.825	0.885	0.908	0.946
400 markets	0.754	0.831	0.885	0.925
	GMM point estimates and adjusted standard errors			
50 markets	0.933	0.907	0.901	0.907
100 markets	0.912	0.924	0.913	0.911
200 markets	0.906	0.915	0.914	0.934
400 markets	0.890	0.895	0.891	0.915
	Bias corrected point estimates and adjusted standard errors			
50 markets	0.950	0.940	0.929	0.926
100 markets	0.952	0.952	0.940	0.944
200 markets	0.951	0.948	0.953	0.971
400 markets	0.947	0.953	0.948	0.955
	Ratio of mean length of nonadjusted and adjusted confidence intervals			
50 markets	1.288	1.178	1.126	1.102
100 markets	1.365	1.224	1.150	1.118
200 markets	1.492	1.307	1.194	1.137
400 markets	1.724	1.450	1.275	1.182
	Ratio of mean length of confidence intervals with same and different draws			
50 markets	1.084	1.051	1.046	1.025
100 markets	1.122	1.067	1.014	1.006
200 markets	1.167	1.049	1.031	1.031
400 markets	1.265	1.091	1.044	1.002

The DGP is as in Section 4, but now the same draws are used in all markets. Again, the nominal coverage rate is 0.95 and the number of Monte Carlo iterations is $1,000$.

References

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

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* New York: Wiley.