SUPPLEMENT TO "INFERENCE IN DYNAMIC DISCRETE CHOICE MODELS WITH SERIALLY CORRELATED UNOBSERVED STATE VARIABLES"

(*Econometrica*, Vol. 77, No. 5, September 2009, 1665–1682)

BY ANDRIY NORETS

THIS SUPPLEMENT is organized as follows. In Section S1, nonuniform convergence results for the DP solution algorithm are presented. In Section S2, these results are extended to the uniform convergence. In Section S3, convergence of posterior expectations is proved. The final section of the supplement presents auxiliary results referred to in proofs and the main paper.

S1. NONUNIFORM CONVERGENCE

THEOREM 4: *Under Assumptions* 1–6, *the approximation to the expected value function in* (12) *converges completely* (*and thus a.s.*) *to the true value with probability bounds that are uniform over parameter and state spaces*; *that is*, *for any* $\tilde{\epsilon} > 0$, *there exists a sequence* $\{z_t\}$ *such that* $\sum_{t=0}^{\infty} z_t < \infty$, *and for any* $\theta \in \Theta$, $s \in S$, *and* $d \in D$,

$$P\Big(\Big|\hat{E}^{(t)}[V(s';\theta) \mid s, d; \theta] - E[V(s';\theta) \mid s, d; \theta]\Big| > \tilde{\epsilon}\Big) \le z_t.$$

PROOF: Let us decompose the error of approximation into three parts:

$$\begin{aligned}
&\Big|\hat{E}^{(t)}[V(s';\theta) \mid s, d; \theta] - E[V(s';\theta) \mid s, d; \theta]\Big| \\
&\quad = \left|\sum_{i=1}^{\tilde{N}(t)} \sum_{j=1}^{\hat{N}(k_i)} V^{k_i}(s^{k_i,j};\theta^{k_i}) W_{k_i,j,t}(s,d,\theta) - E[V(s';\theta) \mid s, d; \theta]\right| \\
&\quad \le \left|\sum_{i=1}^{\tilde{N}(t)} \sum_{j=1}^{\hat{N}(k_i)} V(s^{k_i,j};\theta) W_{k_i,j,t}(s,d,\theta) - E[V(s';\theta) \mid s, d; \theta]\right| \\
&\qquad + \left|\sum_{i=1}^{\tilde{N}(t)} \sum_{j=1}^{\hat{N}(k_i)} (V(s^{k_i,j};\theta^{k_i}) - V(s^{k_i,j};\theta)) W_{k_i,j,t}(s,d,\theta)\right| \\
&\qquad + \left|\sum_{i=1}^{\tilde{N}(t)} \sum_{j=1}^{\hat{N}(k_i)} (V^{k_i}(s^{k_i,j};\theta^{k_i}) - V(s^{k_i,j};\theta^{k_i})) W_{k_i,j,t}(s,d,\theta)\right| \\
&\quad = A_1^t(\theta,s,d) + A_2^t(\theta,s,d) + A_3^t(\theta,s,d) \\
&\quad \le \max_d A_1^t(\theta,s,d) + \max_d A_2^t(\theta,s,d) + \max_d A_3^t(\theta,s,d)
\end{aligned}$$

 DOI: 10.3982/ECTA7292

$$= A_1^t(\theta, s) + A_2^t(\theta, s) + A_3^t(\theta, s).$$

In Lemma 1, I show that $A_1^t(\theta, s)$ converges to zero completely with bounds on probabilities that are independent of θ and s. The proof uses Hoeffding's inequality, implying a strong law of large numbers (SLLN) for bounded random variables. However, some additional work is required since $s^{k_i,j}$ do not constitute a random sample. Using the continuity of the value function $V(\cdot)$, the compactness of the parameter space Θ, and the assumption that each parameter draw can get into any point in Θ (Assumption 5), I show analogous result for $A_2^t(\theta, s)$ in Lemma 2. In Lemma 3, I bound $A_3^t(\theta, s)$ by a weighted sum of $A_1^t(\theta, s)$ and $A_2^t(\theta, s)$ from previous iterations. Due to very fast convergence of $A_1^t(\theta, s)$ and $A_2^t(\theta, s)$, $A_3^t(\theta, s)$ also converges to zero completely. Thus, from the three lemmas the result follows. Formally, according to Lemmas 1, 2, and 3, there exist $\delta_1 > 0$, $\delta_2 > 0$, $\delta_3 > 0$, and T such that $\forall\theta \in \Theta$, $\forall s \in S$, and $\forall t > T$,

$$P\big(|A_1^t(\theta, s)| > \tilde{\epsilon}/3\big) \le e^{-0.5\delta_1 t^{\gamma_1}},$$
$$P\big(|A_2^t(\theta, s)| > \tilde{\epsilon}/3\big) \le e^{-0.5\delta_2 t^{\gamma_1}},$$
$$P\big(|A_3^t(\theta, s)| > \tilde{\epsilon}/3\big) \le e^{-\delta_3 t^{\gamma_0\gamma_1}}.$$

Combining the above equations gives

$$\begin{aligned} &P\Big(\big|\hat{E}^{(t)}[V(s'; \theta) \mid s, d; \theta] - E[V(s'; \theta) \mid s, d; \theta]\big| > \tilde{\epsilon}\Big) \\ &\quad \le P(A_1^t(\theta, s) + A_2^t(\theta, s) + A_3^t(\theta, s) > \tilde{\epsilon}) \\ &\quad \le P\big(|A_1^t(\theta, s)| > \tilde{\epsilon}/3\big) + P\big(|A_2^t(\theta, s)| > \tilde{\epsilon}/3\big) + P\big(|A_3^t(\theta, s)| > \tilde{\epsilon}/3\big) \\ &\quad \le e^{-0.5\delta_1 t^{\gamma_1}} + e^{-0.5\delta_2 t^{\gamma_1}} + e^{-\delta_3 t^{\gamma_0\gamma_1}} \quad (\forall t > T) \\ &\quad = z_t \quad (\forall t > T). \end{aligned}$$

For $t \le T$, set $z_t = 1$. Proposition 9 shows that $\sum_{t=0}^{\infty} z_t < \infty$. The lemmas are stated and proved below. *Q.E.D.*

LEMMA 1: *Given $\tilde{\epsilon} > 0$, there exist $\delta > 0$ and T such that for any $\theta \in \Theta$, $s \in S$, and $t > T$,*

$$P\big(|A_1^t(\theta, s)| > \tilde{\epsilon}\big) \le e^{-\delta \tilde{N}(t)\hat{N}(t-N(t))} \le e^{-0.5\delta t^{\gamma_1}}. \tag{17}$$

PROOF: Fix a combination $m = \{m_1, \dots, m_{\tilde{N}(t)}\}$ from $\{t - N(t), \dots, t-1\}$. Let

$$\begin{aligned} &X(\omega^{t-1}, \theta, s, d, m) \qquad (18)\\ &\quad = \Bigg|\sum_{i=1}^{\tilde{N}(t)} \sum_{j=1}^{\hat{N}(m_i)} \Bigg(\big(V(s^{m_i,j}; \theta) - E[V(s'; \theta) \mid s, d; \theta]\big) \end{aligned}$$

$$\times f(s^{m_i,j} \mid s, d; \theta)/g(s^{m_i,j}) \Bigg/ \left(\sum_{r=1}^{\tilde{N}(t)} \sum_{q=1}^{\hat{N}(m_r)} f(s^{m_r,q} \mid s, d; \theta)/g(s^{m_r,q})\right)\Bigg)\Bigg|.$$

Since the importance sampling weights are bounded away from zero by $\underline{f} > 0$ (see Assumption 4),

$$\begin{aligned}
(19)\qquad & [X(\omega^{t-1}, \theta, s, d, m) > \tilde{\epsilon}] \\
& \subset \Bigg[\Bigg|\sum_{i=1}^{\tilde{N}(t)} \sum_{j=1}^{\hat{N}(m_i)} \Bigg(\big(V(s^{m_i,j}; \theta) - E[V(s'; \theta) \mid s, d; \theta]\big) \\
& \quad \times f(s^{m_i,j} \mid s, d; \theta)/g(s^{m_i,j}) \\
& \quad \Bigg/ \left(\sum_{r=1}^{\tilde{N}(t)} \hat{N}(m_r) \inf_{\theta,s,s',d} f(s' \mid s, d; \theta)/g(s')\right)\Bigg)\Bigg| > \tilde{\epsilon}\Bigg] \\
& = \Bigg[\Bigg|\sum_{i=1}^{\tilde{N}(t)} \sum_{j=1}^{\hat{N}(m_i)} \big(V(s^{m_i,j}; \theta) - E[V(s'; \theta) \mid s, d; \theta]\big) \frac{f(s^{m_i,j} \mid s, d; \theta)}{g(s^{m_i,j})}\Bigg| \\
& \quad > \tilde{\epsilon}\underline{f} \sum_{i=1}^{\tilde{N}(t)} \hat{N}(m_i)\Bigg].
\end{aligned}$$

Using (19) and then applying Hoeffding (1963)'s inequality, we get

$$\begin{aligned}
(20)\qquad & P(X(\omega^{t-1}, \theta, s, d, m) > \tilde{\epsilon}) \\
& \leq P\Bigg[\Bigg|\sum_{i=1}^{\tilde{N}(t)} \sum_{j=1}^{\hat{N}(m_i)} \big(V(s^{m_i,j}; \theta) - E[V(s'; \theta) \mid s, d; \theta]\big) \frac{f(s^{m_i,j} \mid s, d; \theta)}{g(s^{m_i,j})}\Bigg| \\
& \qquad > \tilde{\epsilon}\underline{f} \sum_{i=1}^{\tilde{N}(t)} \hat{N}(m_i)\Bigg] \\
& \leq 2\exp\left\{\frac{-2\underline{f}^2\tilde{\epsilon}^2}{(b-a)^2} \sum_{r=1}^{\tilde{N}(t)} \hat{N}(m_r)\right\},
\end{aligned}$$

where a and b are correspondingly the lower and upper bounds on $(V(s^{m_i,j};\theta) - E[V(s';\theta) \mid s;\theta])f(s^{m_i,j} \mid s,d;\theta)/g(s^{m_i,j})$. Hoeffding's inequality applies since $s^{m_i,j}$ are independent, the summands have expectations equal to zero,

$$\int \frac{(V(s^{m_i,j};\theta) - E[V(s';\theta) \mid s;\theta])f(s^{m_i,j} \mid s,d;\theta)}{g(s^{m_i,j})} g(s^{m_i,j})\, ds^{m_i,j} = 0, \tag{21}$$

and a and b are finite by Assumptions 1 and 2.

Since $\hat{N}(\cdot)$ is nondecreasing, (20) implies

$$\begin{aligned} P(X(\omega^{t-1},\theta,s,d,m) > \tilde{\epsilon}) &\le 2\exp\left\{\frac{-2\underline{f}^2\tilde{\epsilon}^2}{(b-a)^2}\tilde{N}(t)\hat{N}(t-N(t))\right\} \\ &= 2\exp\{-4\delta\tilde{N}(t)\hat{N}(t-N(t))\}, \end{aligned} \tag{22}$$

where the last equality defines $\delta > 0$.

Since $|A_1^t(\theta,s,d)| < \max_m X(\omega^{t-1},\theta,s,d,m)$,

$$\begin{aligned} &P(|A_1^t(\theta,s,d)| > \tilde{\epsilon}) \\ &\quad \le P\Big[\max_m X(\omega^{t-1},\theta,s,d,m) > \tilde{\epsilon}\Big] \\ &\quad = P\Big(\bigcup_m [X(\omega^{t-1},\theta,s,d,m) > \tilde{\epsilon}]\Big) \\ &\quad \le \sum_m P[X(\omega^{t-1},\theta,s,d,m) > \tilde{\epsilon}] \\ &\quad \le 2\exp\{-4\delta\tilde{N}(t)\hat{N}(t-N(t))\}\frac{N(t)!}{(N(t)-\tilde{N}(t))!\tilde{N}(t)!}, \end{aligned} \tag{23}$$

where the summation, the maximization, and the union are taken over all possible combinations m, and $N(t)!/((N(t)-\tilde{N}(t))!\tilde{N}(t)!)$ is the number of the possible combinations.

Assumption 6 and Proposition 7 show that $\exists T_1$ such that $\forall t > T_1$,

$$\begin{aligned} &\exp\{-4\delta\tilde{N}(t)\hat{N}(t-N(t))\}\frac{N(t)!}{(N(t)-\tilde{N}(t))!\tilde{N}(t)!} \\ &\quad \le \exp\{-2\delta\tilde{N}(t)\hat{N}(t-N(t))\}. \end{aligned} \tag{24}$$

Finally,

$$\begin{aligned} &P(|A_1^t(\theta,s)| > \tilde{\epsilon}) \\ &\quad = P\Big(\max_{d\in D}|A_1^t(\theta,s,d)| > \tilde{\epsilon}\Big) \end{aligned} \tag{25}$$

$$= P\Big(\bigcup_{d\in D}\big[|A_1^t(\theta, s, d)| > \tilde{\epsilon}\big]\Big)$$
$$\le \text{card}(D)2\exp\big\{-2\delta\tilde{N}(t)\hat{N}(t-N(t))\big\} \quad (\forall t > T_1)$$
$$\le \exp\big\{-\delta\tilde{N}(t)\hat{N}(t-N(t))\big\} \quad (\forall t > T_2 \ge T_1),$$

where such T_2 exists since $\text{card}(D)2\exp\{-\delta\tilde{N}(t)\hat{N}(t-N(t))\} \to 0$. The last inequality in (17) follows since $\tilde{N}(t)\hat{N}(t-N(t)) \ge t^{\gamma_1} - t^{\gamma_1-\gamma_2} \ge 0.5t^{\gamma_1}$ for any t larger than some $T \ge T_2$. *Q.E.D.*

LEMMA 2: *Given $\tilde{\epsilon} > 0$, there exist $\delta > 0$ and T such that for any $\theta \in \Theta$, $s \in S$, and $t > T$,*

$$P\big(|A_2^t(\theta, s)| > \tilde{\epsilon}\big) \le e^{-\delta(N(t)-\tilde{N}(t))} \le e^{-0.5\delta t^{\gamma_1}}. \tag{26}$$

PROOF: Let us find an event encompassing $[|A_2^t(\theta, s)| > \tilde{\epsilon}]$, for which the probability can be easily bounded

$$\big[|A_2^t(\theta, s, d)| > \tilde{\epsilon}\big] \tag{27}$$
$$= \left[\left|\sum_{i=1}^{\tilde{N}(t)}\sum_{j=1}^{\hat{N}(t)}(V(s^{k_i,j};\theta^{k_i}) - V(s^{k_i,j};\theta))W_{k_i,j,t}(s,d,\theta)\right| > \tilde{\epsilon}\right]$$
$$\subset \left[\sum_{i=1}^{\tilde{N}(t)}\sum_{j=1}^{\hat{N}(t)}|V(s^{k_i,j};\theta^{k_i}) - V(s^{k_i,j};\theta)|W_{k_i,j,t}(s,d,\theta) > \tilde{\epsilon}\right]$$
$$\subset \big[\exists k_i, j: |V(s^{k_i,j};\theta^{k_i}) - V(s^{k_i,j};\theta)| > \tilde{\epsilon}\big].$$

Since $V(s;\theta)$ is continuous and $\Theta \times S$ is a compact, $\exists\tilde{\delta}_{\tilde{\epsilon}} > 0$ such that $\|(s_1,\theta_1) - (s_2,\theta_2)\| \le \tilde{\delta}_{\tilde{\epsilon}}$ implies $|V(s_1;\theta_1) - V(s_2;\theta_2)| \le \tilde{\epsilon}$. Therefore,

$$\big[\exists k_i, j: |V(s^{k_i,j};\theta^{k_i}) - V(s^{k_i,j};\theta)| > \tilde{\epsilon}\big] \tag{28}$$
$$\subset \big[\exists k_i, j: \|(s^{k_i,j},\theta^{k_i}) - (s^{k_i,j},\theta)\| > \tilde{\delta}_{\tilde{\epsilon}}\big] = [\exists k_i: \|\theta^{k_i} - \theta\| > \tilde{\delta}_{\tilde{\epsilon}}].$$

Because k_i are the indices of the parameters from the previous iterations that are the closest to θ,

$$[\exists k_i: \|\theta^{k_i} - \theta\| > \tilde{\delta}_{\tilde{\epsilon}}] \tag{29}$$
$$\subset \big[\forall j \in \{t-N(t),\dots,t-1\} \setminus \big\{k_1,\dots,k_{\tilde{N}(t)}\big\}: \|\theta^j - \theta\| > \tilde{\delta}_{\tilde{\epsilon}}\big]$$

$$\subset \bigcup_{\substack{(j_1,\ldots,j_{N(t)-\tilde{N}(t)})\\ j_m\in\{t-N(t),\ldots,t-1\},\\ m\neq l\Rightarrow j_m\neq j_l}} \bigcap_{m=1}^{N(t)-\tilde{N}(t)} [\|\theta^{j_m}-\theta\| > \tilde{\delta}_{\tilde{\epsilon}}].$$

Fix some $(j_1, \ldots, j_{N(t)-\tilde{N}(t)})$. Then by Assumption 5,

$$\begin{aligned} P\big([\|\theta^{j_m}-\theta\| > \tilde{\delta}_{\tilde{\epsilon}}] \mid \omega^{j_m-1}\big) &= 1 - P\big([\|\theta^{j_m}-\theta\| < \tilde{\delta}_{\tilde{\epsilon}}] \mid \omega^{j_m-1}\big) \\ &\le 1 - \hat{\delta}\lambda\big[B_{\tilde{\delta}_{\tilde{\epsilon}}}(\theta)\cap\Theta\big] \\ &\le 1 - \hat{\delta}[\tilde{\delta}_{\tilde{\epsilon}}/J_{\Theta}^{0.5}]^J \\ &= \exp\big\{-4\big(-0.25\log(1-\hat{\delta}[\tilde{\delta}_{\tilde{\epsilon}}/J_{\Theta}^{0.5}]^J)\big)\big\} \\ &= e^{-4\delta}, \end{aligned} \tag{30}$$

where the last equality defines $\delta > 0$, J_Θ is the dimensionality of rectangle Θ, and $B_{\cdot}(\cdot)$ is a ball in R^{J_Θ}. It holds for any history ω^{j_m-1}. Thus for fixed $(j_1, \ldots, j_{N(t)-\tilde{N}(t)})$,

$$P\left(\bigcap_{m=1}^{N(t)-\tilde{N}(t)} [\|\theta^{j_m}-\theta\| > \tilde{\delta}_{\tilde{\epsilon}}]\right) \le e^{-4\delta(N(t)-\tilde{N}(t))}. \tag{31}$$

Since the union in (29) is taken over $N(t)!/(\tilde{N}(t)!(N(t)-\tilde{N}(t))!)$ events,

$$\begin{aligned} P\big[|A_2^t(x,\theta,\epsilon)| > \tilde{\epsilon}\big] &\le e^{-4\delta(N(t)-\tilde{N}(t))}\frac{N(t)!}{\tilde{N}(t)!(N(t)-\tilde{N}(t))!} \\ &\le e^{-2\delta(N(t)-\tilde{N}(t))} \quad \forall t > T_2, \end{aligned} \tag{32}$$

where the second inequality and existence of T_2 follows from Assumption 6 and Proposition 7. Finally,

$$\begin{aligned} P\big(|A_2^t(\theta,s)| > \tilde{\epsilon}\big) &= P\Big(\max_{d\in D}|A_2^t(\theta,s,d)| > \tilde{\epsilon}\Big) \\ &= P\left(\bigcup_{d\in D}\big[|A_2^t(\theta,s,d)| > \tilde{\epsilon}\big]\right) \\ &\le \mathrm{card}(D)e^{-2\delta(N(t)-\tilde{N}(t))} \quad (\forall t > T_2) \\ &\le e^{-\delta(N(t)-\tilde{N}(t))} \quad (\forall t > T_3 \ge T_2), \end{aligned} \tag{33}$$

where such T_3 exists since $\mathrm{card}(D)e^{-\delta(N(t)-\tilde{N}(t))} \to 0$. The last inequality in (26) follows since $N(t)-\tilde{N}(t) \ge [t^{\gamma_1}] - [t^{\gamma_2}] \ge t^{\gamma_1} - 1 - t^{\gamma_2} \ge 0.5t^{\gamma_1}$ for any t larger than some $T \ge T_3$. *Q.E.D.*

LEMMA 3: *Given* $\tilde{\epsilon} > 0$, *there exist* $\delta > 0$ *and* T *such that* $\forall \theta \in \Theta$, $\forall s \in S$, *and* $\forall t > T$,

$$P\big(|A_3^t(\theta, s)| > \tilde{\epsilon}\big) \le e^{-\delta t^{\gamma_0 \gamma_1}}. \tag{34}$$

PROOF: First let us show that for any positive integer m, $\forall \theta \in \Theta$, and $\forall s \in S$,

$$\begin{aligned} A_3^t(\theta, s) \le \frac{\beta}{1-\beta}\Big[&\max_{i=t-mN(t),t-1}\Big(\max_{j=1,\hat{N}(i)} A_1^i(\theta^i, s^{i,j})\Big) \\ &+ \max_{i=t-mN(t),t-1}\Big(\max_{j=1,\hat{N}(i)} A_2^i(\theta^i, s^{i,j})\Big)\Big] \\ &+ \beta^m \max_{i=t-mN(t),t-1}\Big(\max_{j=1,\hat{N}(i)} A_3^i(\theta^i, s^{i,j})\Big). \end{aligned} \tag{35}$$

By definition,

$$A_3^t(\theta, s, d) = \left| \sum_{i=1}^{\tilde{N}(t)} \sum_{j=1}^{\hat{N}(k_i)} \big(V^{k_i}(s^{k_i,j}; \theta^{k_i}) - V(s^{k_i,j}; \theta^{k_i})\big) W_{k_i,j,t}(s, d, \theta) \right|. \tag{36}$$

Since $\max_d a(d) - \max_d b(d) \le \max_d\{a(d) - b(d)\}$,

$$\begin{aligned} &|V^{k_i}(s^{k_i,j}; \theta^{k_i}) - V(s^{k_i,j}; \theta^{k_i})| \\ &\quad = \Big| \max_{d \in D}\big\{u(s^{k_i,j}, d) + \beta \hat{E}^{(k_i)}[V(s'; \theta^{k_i}) \mid s^{k_i,j}, d; \theta^{k_i}]\big\} \\ &\qquad - \max_{d \in D}\big\{u(s^{k_i,j}, d) + \beta E[V(s'; \theta^{k_i}) \mid s^{k_i,j}, d; \theta^{k_i}]\big\} \Big| \\ &\quad \le \Big| \max_{d \in D}\big\{\beta \hat{E}^{(k_i)}[V(s'; \theta^{k_i}) \mid s^{k_i,j}, d; \theta^{k_i}] \\ &\qquad - \beta E[V(s'; \theta^{k_i}) \mid s^{k_i,j}, d; \theta^{k_i}]\big\} \Big|. \end{aligned} \tag{37}$$

From (37) and definition of $A_l^t(\cdot)$ given in Theorem 4,

$$\begin{aligned} &|V^{k_i}(s^{k_i,j}; \theta^{k_i}) - V(s^{k_i,j}; \theta^{k_i})| \\ &\quad \le \beta \max_{d \in D}\big(A_1^{k_i}(\theta^{k_i}, s^{k_i,j}, d) + A_2^{k_i}(\theta^{k_i}, s^{k_i,j}, d) + A_3^{k_i}(\theta^{k_i}, s^{k_i,j}, d)\big) \\ &\quad \le \beta\big(A_1^{k_i}(\theta^{k_i}, s^{k_i,j}) + A_2^{k_i}(\theta^{k_i}, s^{k_i,j}) + A_3^{k_i}(\theta^{k_i}, s^{k_i,j})\big). \end{aligned} \tag{38}$$

Combining (36) and (38) gives

$$A_3^t(\theta, s, d) \le \beta \sum_{i=1}^{\tilde{N}(t)} \sum_{j=1}^{\hat{N}(k_i)} \big(A_1^{k_i}(\theta^{k_i}, s^{k_i,j}) + A_2^{k_i}(\theta^{k_i}, s^{k_i,j}) \tag{39}$$

$$
\begin{aligned}
&+ A_3^{k_i}(\theta^{k_i}, s^{k_i,j})\big)W_{k_i,j,t}(s,d,\theta)\\
&\le \beta \max_{i=t-N(t),t-1}\Big(\max_{j=1,\hat N(i)} A_1^i(\theta^i, s^{i,j})\Big)\\
&\quad + \beta \max_{i=t-N(t),t-1}\Big(\max_{j=1,\hat N(i)} A_2^i(\theta^i, s^{i,j})\Big)\\
&\quad + \beta \max_{i=t-N(t),t-1}\Big(\max_{j=1,\hat N(i)} A_3^i(\theta^i, s^{i,j})\Big),
\end{aligned}
$$

where the second inequality follows from the fact that $\forall i \in \{1,\dots,\tilde N(t)\}$, $k_i \in \{t-N(t),\dots,t-1\}$ and the weights sum to 1. Since the right-hand side (r.h.s.) of (39) does not depend on d,

$$
\begin{aligned}
A_3^t(\theta,s) &\le \beta \max_{i=t-N(t),t-1}\Big(\max_{j=1,\hat N(i)} A_1^i(\theta^i, s^{i,j})\Big)\\
&\quad + \beta \max_{i=t-N(t),t-1}\Big(\max_{j=1,\hat N(i)} A_2^i(\theta^i, s^{i,j})\Big)\\
&\quad + \beta \max_{i=t-N(t),t-1}\Big(\max_{j=1,\hat N(i)} A_3^i(\theta^i, s^{i,j})\Big).
\end{aligned}
\tag{40}
$$

To facilitate the description of the iterative process on (40) that will lead to (35), let $M(t,0)=t$ and $M(t,i)=M(t,i-1)-N(M(t,i-1))$. Then

$$
\begin{aligned}
A_3^t(\theta,s) &\le \beta \max_{i=t-N(t),t-1}\Big(\max_{j=1,\hat N(i)} A_1^i(\theta^i, s^{i,j})\Big)\\
&\quad + \beta \max_{i=t-N(t),t-1}\Big(\max_{j=1,\hat N(i)} A_2^i(\theta^i, s^{i,j})\Big)\\
&\quad + \beta^2 \max_{i=t-N(t)-N[t-N(t)],t-2}\Big(\max_{j=1,\hat N(i)} A_1^i(\theta^i, s^{i,j})\Big)\\
&\quad + \beta^2 \max_{i=t-N(t)-N[t-N(t)],t-2}\Big(\max_{j=1,\hat N(i)} A_2^i(\theta^i, s^{i,j})\Big)\\
&\quad + \beta^2 \max_{i=t-N(t)-N[t-N(t)],t-2}\Big(\max_{j=1,\hat N(i)} A_3^i(\theta^i, s^{i,j})\Big)\\
&\le \sum_{k=1}^{m} \beta^k \Big[\max_{i=M(t,k),t-k}\Big(\max_{j=1,\hat N(i)} A_1^i(\theta^i, s^{i,j})\Big)\\
&\quad + \max_{i=M(t,k),t-k}\Big(\max_{j=1,\hat N(i)} A_2^i(\theta^i, s^{i,j})\Big)\Big]\\
&\quad + \beta^m \max_{i=M(t,m),t-m}\Big(\max_{j=1,\hat N(i)} A_3^i(\theta^i, s^{i,j})\Big),
\end{aligned}
\tag{41}
$$

from which (35) follows since $\sum_{k=1}^{m}\beta^k < \beta/(1-\beta)$ and $M(t,m) \geq t - mN(t)\ \forall m$.

Inequality in (35) is shown to hold for any m. Let $m(t) = [(t - t^{\gamma_0})/N(t)]$ ($[x]$ is the integer part of x) and notice that $M(t, m(t)) \geq t - m(t)N(t) \geq t^{\gamma_0}$. Since $A_3^i(\theta^i, s^{i,j})$ is bounded above by some $\bar{A}_3 < \infty$ (utility function, and state and parameter spaces are bounded),

$$
\begin{aligned}
&P\big[|A_3^t(\theta, s)| > \tilde{\epsilon}\big] \\
&\quad \leq P\bigg[\frac{\beta}{1-\beta}\bigg\{\max_{i=t-m(t)N(t),t-1}\Big(\max_{j=1,\hat{N}(i)} A_1^i(\theta^i, s^{i,j})\Big) \\
&\qquad + \max_{i=t-m(t)N(t),t-1}\Big(\max_{j=1,\hat{N}(i)} A_2^i(\theta^i, s^{i,j})\Big)\bigg\} \\
&\qquad + \beta^{m(t)} \max_{i=t-m(t)N(t),t-1}\Big(\max_{j=1,\hat{N}(i)} A_3^i(\theta^i, s^{i,j})\Big) > \tilde{\epsilon}\bigg] \\
&\quad \leq P\bigg[\max_{i=t-m(t)N(t),t-1}\Big(\max_{j=1,\hat{N}(i)} A_1^i(\theta^i, s^{i,j})\Big) > \frac{\tilde{\epsilon}(1-\beta)}{3\beta}\bigg] \\
&\qquad + P\bigg[\max_{i=t-m(t)N(t),t-1}\Big(\max_{j=1,\hat{N}(i)} A_2^i(\theta^i, s^{i,j})\Big) > \frac{\tilde{\epsilon}(1-\beta)}{3\beta}\bigg] \\
&\qquad + P\bigg[\beta^{m(t)}\bar{A}_3 > \frac{\tilde{\epsilon}}{3}\bigg] \\
&\quad \leq \sum_{i=t-m(t)N(t)}^{t-1}\sum_{j=1}^{\hat{N}(i)}\bigg\{P\bigg[A_1^i(\theta^i, s^{i,j}) > \frac{\tilde{\epsilon}(1-\beta)}{3\beta}\bigg] \\
&\qquad + P\bigg[A_2^i(\theta^i, s^{i,j}) > \frac{\tilde{\epsilon}(1-\beta)}{3\beta}\bigg]\bigg\}.
\end{aligned}
$$

The last inequality holds for $t > T_3$, where T_3 satisfies $P(\beta^{m(t)}\bar{A}_3 > \tilde{\epsilon}/3) = 0\ \forall t > T_3$. Such T_3 exists since $m(t) \to \infty$.

Since $t - m(t)N(t) \to \infty$, by Lemma 1 and Lemma 2, there exist $\delta_1 > 0$, $\delta_2 > 0$, T_1, and T_2 such that $\forall t > \max(T_1, T_2, T_3)$,

$$
\begin{aligned}
(42)\qquad &P(|A_3^t(\theta, s)| > \tilde{\epsilon}) \\
&\quad \leq \sum_{i=t-m(t)N(t)}^{t-1} \hat{N}(i)\big[e^{-\delta_1 \tilde{N}(i)\hat{N}(i-N(i))} + e^{-\delta_2(N(i)-\tilde{N}(i))}\big].
\end{aligned}
$$

Proposition 8 shows that there exist $\delta > 0$ and T_4 such that the r.h.s. of (42) is no larger than $\exp(-\delta t^{\gamma_0\gamma_1})\ \forall t > T_4$. Thus, setting $T = \max(T_1, T_2, T_3, T_4)$ completes the proof. *Q.E.D.*

S2. EXTENSION TO THE UNIFORM CONVERGENCE

First, note that the approximation error is not a continuous function of (θ, s). Thus, we cannot apply the standard results to show the measurability of the supremum of the approximation error over the state and parameter spaces. Proposition 1 below and Proposition 3 establish the measurability in this case. Next, Lemma 4 shows that a uniform version of Lemma 1 holds; Lemma 5 shows that a uniform version of Lemma 2 also holds; a uniform version of Lemma 3 holds trivially since the right-hand side of the key inequality (35) does not depend on (θ, s). Theorem 1 follows from the uniform versions of the lemmas in the same way as Theorem 4 follows from Lemmas 1–3.

PROPOSITION 1: *Let $f(\omega, \theta)$ be a measurable function on $(\Omega \times \Theta, \sigma(\mathcal{A} \times \mathcal{B}))$ with values in R. Assume that Θ has a countable subset $\tilde{\Theta}$ and that for any $\omega \in \Omega$ and any $\theta \in \Theta$ there exists a sequence in $\tilde{\Theta}$, $\{\tilde{\theta}_n\}$ such that $f(\omega, \tilde{\theta}_n) \to f(\omega, \theta)$. Then $\sup_{\theta\in\Theta} f(\omega, \theta)$ is measurable with respect to (w.r.t.) $(\Omega, \mathcal{A})$ (the proposition can be used to show that the supremum of a random function with some simple discontinuities, e.g., jumps, on a separable space is measurable).*

PROOF: First, let us show that for an arbitrary t,

$$\bigcup_{\theta\in\Theta}[f(\omega, \theta) > t] = \bigcup_{\theta\in\tilde{\Theta}}[f(\omega, \theta) > t]. \tag{43}$$

Assume $\omega_1 \in \bigcup_{\theta\in\Theta}[f(\omega, \theta) > t]$. This means there exists $\theta_1 \in \Theta$ such that $f(\omega_1, \theta_1) > t$. By the theorem's assumption, $\exists\{\tilde{\theta}_n\}$ is such that $f(\omega_1, \tilde{\theta}_n) \to f(\omega_1, \theta_1)$. Then $\exists n$, $f(\omega_1, \tilde{\theta}_n) > t$. Thus, $\omega_1 \in \bigcup_{\theta\in\tilde{\Theta}}[f(\omega, \theta) > t]$ and (43) is proved.

Note that $[\sup_{\theta\in\Theta} f(\omega, \theta) > t] = \bigcup_{\theta\in\Theta}[f(\omega, \theta) > t] = \bigcup_{\theta\in\tilde{\Theta}}[f(\omega, \theta) > t]$ is a countable union of sets from $\mathcal{A}$ and thus also belongs to $\mathcal{A}$. *Q.E.D.*

To apply the proposition for establishing the measurability of the supremum of the approximation errors, let the set of rational numbers contained in $\Theta \times S$ play the role of the countable subset $\tilde{\Theta}$. Proposition 3 shows that for any given history ω^{t-1} and any (θ, s), it is always possible to find a sequence with rational coordinates $(\tilde{\theta}_n) \to \theta$ such that for all n, $(\tilde{\theta}_n)$ and θ have the same iteration indices for the nearest neighbors. For a given history ω^{t-1}, the approximation error is continuous in (θ, s) on the subsets of $\Theta \times S$ that give the same iteration indices of the nearest neighbors. Using any rational sequence $s^n \to s$ gives $f(\omega, (\tilde{\theta, s})_n) \to f(\omega, (\theta, s))$ as required in the proposition. Thus, the supremum of the approximation error is measurable.

LEMMA 4: *Given $\tilde{\epsilon} > 0$, there exist $\delta > 0$ and T such that $\forall t > T$,*

$$P\Big(\sup_{\theta\in\Theta, s\in S} |A_1^t(\theta, s)| > \tilde{\epsilon}\Big) \le e^{-\delta\tilde{N}(t)\hat{N}(t-N(t))} \le e^{-0.5\delta t^{\gamma_1}}. \tag{44}$$

PROOF: Fix a combination $m = \{m_1, \dots, m_{\tilde{N}(t)}\}$ from $\{t - N(t), \dots, t - 1\}$. Lemma 1 defines $X(\omega^{t-1}, \theta, s, d, m)$ in (18). By Proposition 5, $\{X(\omega^{t-1}, \theta, s, d, m)\}_{\omega^{t-1}}$ are equicontinuous on $\Theta \times S$: there exists $\tilde{\delta}(\tilde{\epsilon}) > 0$ such that $\|(\theta_1, s_1) - ((\theta_2, s_2))\| < \tilde{\delta}(\tilde{\epsilon})$ implies $|X(\omega^{t-1}, \theta_1, s_1, d, m) - X(\omega^{t-1}, \theta_2, s_2, d, m)| < \tilde{\epsilon}/2$. Since $\Theta \times S$ is a compact set, it can be covered by M balls: $\Theta \times S \subset \bigcup_{i=1}^{M} B_i$ with radius $\tilde{\delta}(\tilde{\epsilon})$ and centers at (θ_i, s_i), where $M < \infty$ depends only on $\tilde{\epsilon}$. It follows that

$$\Big[\sup_{\theta,s} X(\omega^{t-1}, \theta, s, d, m) > \tilde{\epsilon}\Big] = \bigcup_{\theta,s} [X(\omega^{t-1}, \theta, s, d, m) > \tilde{\epsilon}] \tag{45}$$

$$= \bigcup_{i=1}^{M} \bigcup_{(\theta,s)\in B_i} [X(\omega^{t-1}, \theta, s, d, m) > \tilde{\epsilon}].$$

Let us show that

$$\bigcup_{(\theta,s)\in B_i} [X(\omega^{t-1}, \theta, s, d, m) > \tilde{\epsilon}] \subset \Big[X(\omega^{t-1}, \theta_i, s_i, d, m) > \frac{\tilde{\epsilon}}{2}\Big]. \tag{46}$$

If $\omega_*^{t-1} \in \bigcup_{(\theta,s)\in B_i} [X(\omega^{t-1}, \theta, s, d, m) > \tilde{\epsilon}]$, then $\exists(\theta^*, s^*) \in B_i(\theta_i, s_i)$ such that $X(\omega_*^{t-1}, \theta^*, s^*, d, m) > \tilde{\epsilon}$. Since $\|(\theta^*, s^*) - (\theta_i, s_i)\| \le \tilde{\delta}(\tilde{\epsilon})$, $X(\omega_*^{t-1}, \theta_i, s_i, d, m) \ge X(\omega_*^{t-1}, \theta^*, s^*, d, m) - \tilde{\epsilon}/2$. This implies $\omega_*^{t-1} \in [X(\omega^{t-1}, \theta_i, s_i, d, m) > \frac{\tilde{\epsilon}}{2}]$.

Since $\sup_{\theta,s} |A_1^t(\theta, s, d)| < \max_m \sup_{\theta,s} X(\omega^{t-1}, \theta, s, d, m)$,

$$\begin{aligned} & P\Big(\sup_{\theta,s} |A_1^t(\theta, s, d)| > \tilde{\epsilon}\Big) \\ & \le P\Big[\max_m \sup_{\theta,s} X(\omega^{t-1}, \theta, s, d, m) > \tilde{\epsilon}\Big] \\ & \quad \text{(max is over all possible combinations } m) \\ & \le P\Big(\bigcup_m \Big[\sup_{\theta,s} X(\omega^{t-1}, \theta, s, d, m) > \tilde{\epsilon}\Big]\Big) \\ & \le \sum_m P\Big[\sup_{\theta,s} X(\omega^{t-1}, \theta, s, d, m) > \tilde{\epsilon}\Big] \\ & \le \sum_m P\Big(\bigcup_{i=1}^{M} \Big[X(\omega^{t-1}, \theta_i, s_i, d, m) > \frac{\tilde{\epsilon}}{2}\Big]\Big) \quad \text{(by (45) and (46))} \\ & \le M \frac{N(t)!}{(N(t) - \tilde{N}(t))!\tilde{N}(t)!} 2 \exp\big\{-4\delta \tilde{N}(t) \hat{N}(t - N(t))\big\}, \end{aligned} \tag{47}$$

where $N(t)!/((N(t)-\tilde{N}(t))!\tilde{N}(t)!)$ is the number of different combinations m and $2\exp\{-4\delta\tilde{N}(t)\hat{N}(t-N(t))\}$ is the bound from (22) in Lemma 1. From the last inequality, the proof follows steps of the argument starting after (23) in the proof of Lemma 1. *Q.E.D.*

LEMMA 5: *Given* $\tilde{\epsilon}>0$, *there exist* $\delta>0$ *and* T *such that* $\forall t>T$,

$$P\Big(\sup_{\theta,s}|A_2^t(\theta,s)|>\tilde{\epsilon}\Big)\le e^{-\delta(N(t)-\tilde{N}(t))}\le e^{-0.5\delta t^{\gamma_1}}. \tag{48}$$

PROOF: From Lemma 2,

$$\begin{aligned}&\big[|A_2^t(\theta,s,d)|>\tilde{\epsilon}\big]\\ &\quad\subset\bigcup_{\substack{(j_1,\dots,j_{N(t)-\tilde{N}(t)})\\ j_m\in\{t-N(t),\dots,t-1\},\\ m\neq l\Rightarrow j_m\neq j_l}}\ \bigcap_{m=1}^{N(t)-\tilde{N}(t)}[\|\theta^{j_m}-\theta\|>\tilde{\delta}_{\tilde{\epsilon}}].\end{aligned} \tag{49}$$

This implies that

$$\begin{aligned}&\Big[\sup_{\theta,s}|A_2^t(\theta,s,d)|>\tilde{\epsilon}\Big]\\ &\quad=\bigcup_{\theta,s}\big[|A_2^t(\theta,s,d)|>\tilde{\epsilon}\big]\\ &\quad\subset\bigcup_{\theta\in\Theta}\left\{\bigcup_{\substack{(j_1,\dots,j_{N(t)-\tilde{N}(t)})\\ j_m\in\{t-N(t),\dots,t-1\},\\ m\neq l\Rightarrow j_m\neq j_l}}\ \bigcap_{m=1}^{N(t)-\tilde{N}(t)}[\|\theta^{j_m}-\theta\|>\tilde{\delta}_{\tilde{\epsilon}}]\right\}.\end{aligned} \tag{50}$$

Since Θ is a rectangle in R^{J_Θ}, it can be covered by a finite number of balls with radius $\tilde{\delta}_{\tilde{\epsilon}}/2$:

$$\Theta\subset\bigcup_{i=1}^{M}B(\theta_i),\qquad M=\text{const}\cdot(\tilde{\delta}_{\tilde{\epsilon}}/2)^{-J_\Theta}. \tag{51}$$

Let us prove the fact

$$\bigcup_{\theta\in B(\theta_i)}\ \bigcap_{m=1}^{N(t)-\tilde{N}(t)}[\|\theta^{j_m}-\theta\|>\tilde{\delta}_{\tilde{\epsilon}}]\subset\bigcap_{m=1}^{N(t)-\tilde{N}(t)}[\theta^{j_m}\notin B(\theta_i)]. \tag{52}$$

Assume $\omega^{t-1} \in (\bigcap_{m=1}^{N(t)-\tilde{N}(t)}[\theta^{j_m} \notin B(\theta_i)])^c$. There exists m such that $\theta^{j_m} \in B(\theta_i)$. It follows that $\forall \theta \in B(\theta_i), \exists \theta^{j_m}, \|\theta^{j_m} - \theta\| \leq \tilde{\delta}_{\tilde{\epsilon}}$. Thus, ω^{t-1} belongs to the set

$$\bigcap_{\theta \in B(\theta_i)} \bigcup_{m=1}^{N(t)-\tilde{N}(t)} [\|\theta^{j_m} - \theta\| \leq \tilde{\delta}_{\tilde{\epsilon}}] = \left(\bigcup_{\theta \in B(\theta_i)} \bigcap_{m=1}^{N(t)-\tilde{N}(t)} [\|\theta^{j_m} - \theta\| > \tilde{\delta}_{\tilde{\epsilon}}] \right)^c. \tag{53}$$

Therefore, the claim in (52) is proved.

By the same argument as for (31) from Lemma 2, we can establish that

$$P\left(\bigcap_{m=1}^{N(t)-\tilde{N}(t)} [\theta^{j_m} \notin B(\theta_i)] \right) \leq e^{-4\delta(N(t)-\tilde{N}(t))} \tag{54}$$

for some positive δ.

From (50), (51), and (52)

$$\begin{aligned} &\left[\sup_{\theta,s} |A_2^t(\theta, s, d)| > \tilde{\epsilon} \right] \\ &\quad \subset \bigcup_{\substack{(j_1,\ldots,j_{N(t)-\tilde{N}(t)}) \\ j_m \in \{t-N(t),\ldots,t-1\}, \\ m \neq l \Rightarrow j_m \neq j_l}} \bigcup_{i=1}^{M} \left(\bigcap_{m=1}^{N(t)-\tilde{N}(t)} [\theta^{j_m} \notin B(\theta_i)] \right). \end{aligned} \tag{55}$$

Using (54) and (55) gives

$$P\left[\sup_{\theta,s} |A_2^t(\theta, s, d)| > \tilde{\epsilon} \right] \leq \frac{N(t)!}{\tilde{N}(t)!(N(t) - \tilde{N}(t))!} M e^{-4\delta(N(t)-\tilde{N}(t))}. \tag{56}$$

The rest of the proof follows the corresponding steps in Lemma 2. *Q.E.D.*

S3. PROOF OF CONVERGENCE OF POSTERIOR EXPECTATIONS

PROOF OF THEOREM 2: First, let us introduce some notation shortcuts:

$$\begin{aligned} r &= r(\theta, \mathcal{V}, \epsilon; F(\theta, \epsilon)), \\ \hat{r} &= r(\theta, \mathcal{V}, \epsilon; \hat{F}^n(\theta, \epsilon)), \\ 1_{\{\cdot\}} &= 1_{\Theta}(\theta) \cdot \left(\prod_{i,t} 1_E(\epsilon_{t,i}) p(d_{t,i}|\mathcal{V}_{t,i}) \right) \\ &\quad \cdot \left(\prod_{i,t,k} 1_{[-\overline{\nu},\overline{\nu}]} \big(q(\theta, \mathcal{V}_{t,i}, \epsilon_{t,i}, F_{t,i}(\theta, \epsilon_{t,i})) \big) \right), \end{aligned}$$

$$\hat{1}_{\{\cdot\}} = 1_{\Theta}(\theta) \cdot \left(\prod_{i,t} 1_E(\epsilon_{t,i}) p(d_{t,i}|\mathcal{V}_{t,i}) \right)$$
$$\cdot \left(\prod_{i,t,k} 1_{[-\overline{\nu},\overline{\nu}]}\big(q(\theta, \mathcal{V}_{t,i}, \epsilon_{t,i}, \hat{F}^n_{t,i}(\theta, \epsilon_{t,i}))\big) \right),$$
$$\int h(\theta, \mathcal{V}, \epsilon)\, d(\theta, \mathcal{V}, \epsilon) = \int h,$$
$$p = p(\theta, \mathcal{V}, \epsilon; F|d, x) = \frac{r \cdot 1_{\{\cdot\}}}{\int r \cdot 1_{\{\cdot\}}},$$
$$\hat{p} = p(\theta, \mathcal{V}, \epsilon; \hat{F}^n|d, x) = \frac{\hat{r} \cdot \hat{1}_{\{\cdot\}}}{\int \hat{r} \cdot \hat{1}_{\{\cdot\}}}.$$

The probability that the approximation error exceeds $\varepsilon > 0$ can be bounded by the sum of two terms

$$P\left[\left| \int h \cdot p - \int h \cdot \hat{p} \right| > \varepsilon \right] \tag{57}$$
$$\leq P(\|F - \hat{F}\| > \delta_F)$$
$$+ P\left(\left[\left| \int h \cdot p - \int h \cdot \hat{p} \right| > \varepsilon \right] \cap [\|F - \hat{F}\| \leq \delta_F] \right), \tag{58}$$

where $\|F - \hat{F}\| = \sup_{s,\theta,d} |F(s, \theta, d) - \hat{F}(s, \theta, d)|$, $F(s, \theta, d)$ is the expected value function (or the difference of expected value functions, depending on the parameterization of the Gibbs sampler), and $\hat{F}$ is the approximation to F from the DP solving algorithm on its iteration n (fixed in this proof). I will show that for a sufficiently small $\delta_F > 0$, the set in (58) is empty. Then, by Theorem 1, the term in (57) can be bounded by z_n corresponding to δ_F:

$$\left[\left| \int h \cdot p - \int h \cdot \hat{p} \right| > \varepsilon \right] \cap [\|F - \hat{F}\| \leq \delta_F]$$
$$\subset \left[\int |p - \hat{p}| > \varepsilon / \|h\| \right] \cap [\|F - \hat{F}\| \leq \delta_F]$$
$$\subset \left(\left[\int_{\hat{1}_{\{\cdot\}} = 1_{\{\cdot\}}} |p - \hat{p}| > \varepsilon / (2\|h\|) \right] \cap [\|F - \hat{F}\| \leq \delta_F] \right) \tag{59}$$
$$\cup \left(\left[\int_{\hat{1}_{\{\cdot\}} \neq 1_{\{\cdot\}}} |p - \hat{p}| > \varepsilon / (2\|h\|) \right] \cap [\|F - \hat{F}\| \leq \delta_F] \right). \tag{60}$$

Let us start with (59):

$$\left(\left[\int_{\hat{1}_{\{\cdot\}}=1_{\{\cdot\}}} |p-\hat{p}| > \frac{\varepsilon}{(2\|h\|)}\right] \cap [\|F-\hat{F}\| \leq \delta_F]\right) \tag{61}$$

$$= \left[\int_{\hat{1}_{\{\cdot\}}=1_{\{\cdot\}}} \left|\frac{r}{\int r \cdot 1_{\{\cdot\}}} - \frac{\hat{r}}{\int \hat{r} \cdot \hat{1}_{\{\cdot\}}}\right| > \frac{\varepsilon}{2\|h\|}\right] \cap [\|F-\hat{F}\| \leq \delta_F]$$

$$\subset \left[\left\|\frac{r}{\int r \cdot 1_{\{\cdot\}}} - \frac{\hat{r}}{\int \hat{r} \cdot \hat{1}_{\{\cdot\}}}\right\| > \frac{\varepsilon}{2\|h\|\overline{\lambda}}\right] \cap [\|F-\hat{F}\| \leq \delta_F],$$

where $\overline{\lambda} < \infty$ is the Lebesgue measure of the space for the parameters and the latent variables. For $\delta_{\mathrm{Sp}} \in (0, \int r \cdot 1_{\{\cdot\}})$:

$$\left[\left\|\frac{r}{\int r \cdot 1_{\{\cdot\}}} - \frac{\hat{r}}{\int \hat{r} \cdot \hat{1}_{\{\cdot\}}}\right\| > \frac{\varepsilon}{2\|h\|\overline{\lambda}}\right] \cap [\|F-\hat{F}\| \leq \delta_F]$$

$$= \left(\left[\left\|\frac{r}{\int r \cdot 1_{\{\cdot\}}} - \frac{\hat{r}}{\int \hat{r} \cdot \hat{1}_{\{\cdot\}}}\right\| > \frac{\varepsilon}{2\|h\|\overline{\lambda}}\right] \cap [\|F-\hat{F}\| \leq \delta_F]\right.$$

$$\left.\cap \left[\left|\int r \cdot 1_{\{\cdot\}} - \int \hat{r} \cdot \hat{1}_{\{\cdot\}}\right| > \delta_{\mathrm{Sp}}\right]\right) \tag{62}$$

$$\cup \left(\left[\left\|\frac{r}{\int r \cdot 1_{\{\cdot\}}} - \frac{\hat{r}}{\int \hat{r} \cdot \hat{1}_{\{\cdot\}}}\right\| > \frac{\varepsilon}{2\|h\|\overline{\lambda}}\right] \cap [\|F-\hat{F}\| \leq \delta_F]\right.$$

$$\left.\cap \left[\left|\int r \cdot 1_{\{\cdot\}} - \int \hat{r} \cdot \hat{1}_{\{\cdot\}}\right| \leq \delta_{\mathrm{Sp}}\right]\right). \tag{63}$$

By Proposition 2 for δ_{Sp} there exists $\delta_F^1 > 0$ such that $[|\int r \cdot 1_{\{\cdot\}} - \int \hat{r} \cdot \hat{1}_{\{\cdot\}}| > \delta_{\mathrm{Sp}}] = \emptyset$. Thus, (62) (the whole two-line expression in parentheses) is the empty set for any $\delta_F < \delta_F^1$. Now, let us work with (63) (again, both lines in parentheses):

$$\left\|\frac{r}{\int r \cdot 1_{\{\cdot\}}} - \frac{\hat{r}}{\int \hat{r} \cdot \hat{1}_{\{\cdot\}}}\right\| \leq \frac{\|r\| \cdot \left|\int r \cdot 1_{\{\cdot\}} - \int \hat{r} \cdot \hat{1}_{\{\cdot\}}\right|}{\int r \cdot 1_{\{\cdot\}} \cdot \int \hat{r} \cdot \hat{1}_{\{\cdot\}}} + \frac{\|\hat{r} - r\|}{\int \hat{r} \cdot \hat{1}_{\{\cdot\}}} \tag{64}$$

$$\leq \frac{\|r\| \cdot \left| \int r \cdot 1_{\{\cdot\}} - \int \hat{r} \cdot \hat{1}_{\{\cdot\}} \right|}{\int r \cdot 1_{\{\cdot\}} \cdot \left(\int r \cdot 1_{\{\cdot\}} - \delta_{\mathrm{Sp}} \right)} + \frac{\|\hat{r} - r\|}{\int r \cdot 1_{\{\cdot\}} - \delta_{\mathrm{Sp}}}.$$

This inequality shows that (63) is a subset of the union of the two sets

$$\left[\frac{\|r\| \cdot \left| \int r \cdot 1_{\{\cdot\}} - \int \hat{r} \cdot \hat{1}_{\{\cdot\}} \right|}{\int r \cdot 1_{\{\cdot\}} \cdot \left(\int \int r \cdot 1_{\{\cdot\}} - \delta_{\mathrm{Sp}} \right)} > \frac{\varepsilon}{4\|h\|\overline{\lambda}} \right] \cap [\|F - \hat{F}\| \leq \delta_F] \cap \left[\left| \int r \cdot 1_{\{\cdot\}} - \int \hat{r} \cdot \hat{1}_{\{\cdot\}} \right| \leq \delta_{\mathrm{Sp}} \right] \tag{65}$$

and

$$\left[\frac{\|\hat{r} - r\|}{\int r \cdot 1_{\{\cdot\}} - \delta_{\mathrm{Sp}}} > \frac{\varepsilon}{4\|h\|\overline{\lambda}} \right] \cap [\|F - \hat{F}\| \leq \delta_F] \cap \left[\left| \int r \cdot 1_{\{\cdot\}} - \int \hat{r} \cdot \hat{1}_{\{\cdot\}} \right| \leq \delta_{\mathrm{Sp}} \right]. \tag{66}$$

I will show that both of them are empty for sufficiently small δ_F. By Proposition 2, there exists $\delta_F^2 > 0$ such that

$$\left[\left| \int r \cdot 1_{\{\cdot\}} - \int \hat{r} \cdot \hat{1}_{\{\cdot\}} \right| > \frac{\varepsilon \cdot \int r \cdot 1_{\{\cdot\}} \cdot \left(\int r \cdot 1_{\{\cdot\}} - \delta_{\mathrm{Sp}} \right)}{4\|h\|\overline{\lambda}\|r\|} \right] = \emptyset$$

whenever $\|F - \hat{F}\| \leq \delta_F^2$. Therefore, (65) is equal to the empty set for $\delta_F \leq \delta_F^2$. Since r is continuous in components of F, there exists $\delta_F^3 > 0$ such that

$$\|\hat{r} - r\| < \frac{\varepsilon \cdot \left(\int r \cdot 1_{\{\cdot\}} - \delta_{\mathrm{Sp}} \right)}{4\|h\|\overline{\lambda}\|r\|}$$

whenever $\|F - \hat{F}\| \leq \delta_F^3$. Therefore, for $\delta_F \leq \delta_F^3$, (66) is equal to the empty set and so is (63). Thus far we showed that (59) is equal to the empty set for $\delta_F \leq \min_{i=1,2,3}(\delta_F^i)$.

Now, let us work with (60). Note that

$$\int_{\hat{1}_{\{\cdot\}}\neq 1_{\{\cdot\}}} |p-\hat{p}| \le \left(\frac{\|r\|}{\int r\cdot 1_{\{\cdot\}}} + \frac{\|\hat{r}\|}{\int \hat{r}\cdot \hat{1}_{\{\cdot\}}}\right)\int_{\hat{1}_{\{\cdot\}}\neq 1_{\{\cdot\}}} 1$$

$$\le \left(\frac{\|r\|}{\int r\cdot 1_{\{\cdot\}}} + \frac{\|\hat{r}\|}{\int r\cdot 1_{\{\cdot\}} - \delta_{\mathrm{Sp}}}\right)\int_{\hat{1}_{\{\cdot\}}\neq 1_{\{\cdot\}}} 1.$$

Thus, (60) is a subset of the set

$$\left(\left[\int_{\hat{1}_{\{\cdot\}}\neq 1_{\{\cdot\}}} |p-\hat{p}| > \frac{\varepsilon}{(2\|h\|)}\right]\cap[\|F-\hat{F}\|\le\delta_F]\right) \tag{67}$$

$$\subset \left(\left[\int_{\hat{1}_{\{\cdot\}}\neq 1_{\{\cdot\}}} 1 > \varepsilon\Big/\left(2\|h\|\left(\frac{\|r\|}{\int r\cdot 1_{\{\cdot\}}} + \frac{\|\hat{r}\|}{\int r\cdot 1_{\{\cdot\}} - \delta_{\mathrm{Sp}}}\right)\right)\right]\right.$$

$$\left.\cap[\|F-\hat{F}\|\le\delta_F]\right).$$

Using the same argument as the one starting from (71) in Proposition 2, I can show that there exists $\delta_F^4 > 0$ such that $\forall \delta_F < \delta_F^4$, (60) will be the empty set. Setting $\delta_F = \min_{i=1,2,3,4}\{\delta_F^i\}$ completes the proof of the theorem. *Q.E.D.*

PROPOSITION 2: *For any* $\varepsilon > 0$, *there exists* $\delta_F > 0$ *such that*

$$[\|F-\hat{F}\| < \delta_F]\cap\left[\left|\int \hat{r}\cdot\hat{1}_{\{\cdot\}} - \int r\cdot 1_{\{\cdot\}}\right| > \varepsilon\right] = \emptyset. \tag{68}$$

PROOF:

$$\left[\left|\int \hat{r}\cdot\hat{1}_{\{\cdot\}} - \int r\cdot 1_{\{\cdot\}}\right| > \varepsilon\right] \subset \left[\int \left|\hat{r}\cdot\hat{1}_{\{\cdot\}} - r\cdot 1_{\{\cdot\}}\right| > \varepsilon\right]$$

$$\subset \left[\int_{\hat{1}_{\{\cdot\}}=1_{\{\cdot\}}} \left|\hat{r}\cdot\hat{1}_{\{\cdot\}} - r\cdot 1_{\{\cdot\}}\right| > \varepsilon/2\right] \tag{69}$$

$$\cup \left[\int_{\hat{1}_{\{\cdot\}}\neq 1_{\{\cdot\}}} \left|\hat{r}\cdot\hat{1}_{\{\cdot\}} - r\cdot 1_{\{\cdot\}}\right| > \varepsilon/2\right]. \tag{70}$$

Let us show that the intersection of (69) and $[\|F-\hat{F}\| < \delta_F]$ is the empty set for a sufficiently small δ_F:

$$\left[\int_{\hat{1}_{\{\cdot\}}=1_{\{\cdot\}}} \left|\hat{r}\cdot\hat{1}_{\{\cdot\}} - r\cdot 1_{\{\cdot\}}\right| > \varepsilon/2\right] \subset \left[\int_{\hat{1}_{\{\cdot\}}=1_{\{\cdot\}}} \left|\hat{r} - r\right| > \varepsilon/2\right]$$

$$\subset [\|\hat{r} - r\| > \varepsilon/(2\overline{\lambda})],$$

where $\overline{\lambda} < \infty$ is the Lebesgue measure of the bounded space for the parameters and the latent variables on which the integration is performed: $\Theta \times E \times \cdots \times E \times \mathbf{V} \times \cdots \times \mathbf{V}$, where $\mathbf{V} \subset R$ is the space for the alternative specific value functions $\mathcal{V}_{t,d,i}$. By Assumption 7, r is continuous in components of F. Thus, $\exists \delta_F^1 > 0$ such that $\|F - \hat{F}\| < \delta_F^1$ implies $\|\hat{r} - r\| < \varepsilon/(2\overline{\lambda})$, which means that the intersection of (69) and $[\|F - \hat{F}\| < \delta_F]$ is the empty set for $\forall \delta_F < \delta_F^1$.

Let us show that the intersection of (70) and $[\|F - \hat{F}\| < \delta_F]$ is the empty set for a sufficiently small δ_F. First, note that

$$\int_{\hat{1}_{\{\cdot\}} \neq 1_{\{\cdot\}}} \left|\hat{r} \cdot \hat{1}_{\{\cdot\}} - r \cdot 1_{\{\cdot\}}\right| \leq (\|r\| + \|\hat{r}\|) \int_{\hat{1}_{\{\cdot\}} \neq \cdot 1_{\{\cdot\}}} 1, \tag{71}$$

where $\|r\| < \infty$ and $\|\hat{r}\| < \overline{r} < \infty$ for any $\hat{F}$ (everything is bounded in the model). Thus,

$$\begin{aligned} &[\|F - \hat{F}\| < \delta_F] \cap \left[\int_{\hat{1}_{\{\cdot\}} \neq 1_{\{\cdot\}}} \left|\hat{r} \cdot \hat{1}_{\{\cdot\}} - r \cdot 1_{\{\cdot\}}\right| > \varepsilon/2\right] \\ &\quad \subset [\|F - \hat{F}\| < \delta_F] \cap \left[\int_{\hat{1}_{\{\cdot\}} \neq 1_{\{\cdot\}}} 1 > \varepsilon/\big(2(\|r\| + \|\hat{r}\|)\big)\right] \\ &\quad = [\|F - \hat{F}\| < \delta_F] \cap \big[\lambda\big[\hat{1}_{\{\cdot\}} \neq 1_{\{\cdot\}}\big] > \varepsilon/\big(2(\|r\| + \|\hat{r}\|)\big)\big], \end{aligned} \tag{72}$$

where $\lambda(\cdot)$ is the Lebesgue measure on the space of the parameters and the latent variables.

By Assumption 7, q_k is continuous in components of F. Thus, for any $\delta_q > 0$, there exists $\delta_F(\delta_q) > 0$ such that $\|F - \hat{F}\| < \delta_F(\delta_q)$ implies $\max_k \|\hat{q}_k - q_k\| < \delta_q$. On the space of the parameters and the latent variables (these are not subsets of the underlying probability space),

$$\begin{aligned} &\big[(\theta, \mathcal{V}, \epsilon) : \hat{1}_{\{\cdot\}} \neq 1_{\{\cdot\}}\big] \\ &\quad \subset \bigcup_{i,t,k} \big[(\theta, \mathcal{V}, \epsilon) : q_k(\theta, \mathcal{V}_{t,i}, \epsilon_{t,i}, F_{t,i}) \in B_{\delta_q}(\overline{\nu}) \cup B_{\delta_q}(-\overline{\nu})\big] \end{aligned} \tag{73}$$

if $\|F - \hat{F}\| < \delta_F(\delta_q)$. To prove this claim, assume $\forall i, t, k$ $q_k(\theta, \mathcal{V}_{t,i}, \epsilon_{t,i}, F_{t,i}) \notin B_{\delta_q}(\overline{\nu}) \cup B_{\delta_q}(-\overline{\nu})$. So the distance between q_k and the truncation region edges $-\overline{\nu}$ and $\overline{\nu}$ is larger than δ_q for all i, t, k. But then, since $\|\hat{q}_k - q_k\| < \delta_q$, $\hat{1}_{\{\cdot\}} = 1_{\{\cdot\}}$ and the claim (73) is proved.

Note that

$$\lim_{\delta_q \to 0} \lambda\left(\bigcup_{i,t,k}\left[(\theta, \mathcal{V}, \epsilon) : q_k(\theta, \mathcal{V}_{t,i}, \epsilon_{t,i}, F_{t,i}) \in B_{\delta_q}(\overline{\nu}) \cup B_{\delta_q}(-\overline{\nu})\right]\right) \tag{74}$$
$$\leq \sum_{i,t,k} \lim_{\delta_q \to 0} \lambda\left[(\theta, \mathcal{V}, \epsilon) : q_k(\theta, \mathcal{V}_{t,i}, \epsilon_{t,i}, F_{t,i}) \in B_{\delta_q}(\overline{\nu}) \cup B_{\delta_q}(-\overline{\nu})\right]$$
$$= \sum_{i,t,k} \lambda\left[(\theta, \mathcal{V}, \epsilon) : q_k(\theta, \mathcal{V}_{t,i}, \epsilon_{t,i}, F_{t,i}) \in \{\overline{\nu}, -\overline{\nu}\}\right],$$

where the last equality holds by the monotone property of measures (the Lebesgue measure in this case) and by the fact that $\bigcap_{\delta_q > 0}[q_k \in B_{\delta_q}(\overline{\nu})] = [q_k = \overline{\nu}]$.

By Assumption 7, $\lambda[(\theta, \mathcal{V}, \epsilon) : q_k = \overline{\nu}] = \lambda[(\theta, \mathcal{V}, \epsilon) : q_k = -\overline{\nu}] = 0$. Therefore, the limit in (74) is equal to zero and there exists $\delta_q^* > 0$ such that if $\|F - \hat{F}\| < \delta_F(\delta_q^*)$, then

$$\lambda\left[\hat{1}_{\{\cdot\}} \neq 1_{\{\cdot\}}\right] < \varepsilon / \left(2(\|r\| + \|\hat{r}\|)\right).$$

So $\forall \delta_F \in (0, \delta_F(\delta_q^*)]$, the intersection of (70) and $[\|F - \hat{F}\| < \delta_F]$ is the empty set. Setting $\delta_F = \min\{\delta_F(\delta_q^*), \delta_F^1\}$ completes the proof of the proposition. *Q.E.D.*

PROOF OF THEOREM 3: The proof is the same as the proofs of analogous results in Roberts and Rosenthal (2006). Conditions (a) and (b) here can be used instead of their assumptions (a) and (b) on page 8. Result (i) follows from their proof of Theorem 5 and result (ii) follows from their proof of Theorem 23 (the only required change in the proofs is to use the exact transition kernel from my setup instead of $P_{\Gamma_{K-N}}$ in their formula (3)). *Q.E.D.*

S4. AUXILIARY RESULTS

PROPOSITION 3: *For any* $\{\theta^1, \ldots, \theta^N\}$ *and* θ *in* R^n, *and any* $\tilde{N} \leq N$, *there exists a sequence of rational numbers* $q_m \to \theta$ *such that for any* m, q_m *and* θ *have the same set of indices for the nearest neighbors*: $\{k_1, \ldots, k_{\tilde{N}}\}$ *defined by* (11).

PROOF: The outcomes of selecting the nearest neighbors can be classified into two cases. The trivial case occurs when there exists a ball around θ with radius r such that $\|\theta^{k_i} - \theta\| < r$ and $\|\theta^j - \theta\| > r + d$ for $d > 0$ and $j \neq k_i$. Then, applying the triangle inequality twice, we get $\forall q \in B_{d/4}(\theta)$, $\|\theta^{k_i} - q\| < r + d/2 < \|\theta^j - q\|$ $\forall j \neq k_i$. For this case the proposition holds trivially.

The other case occurs when there exists a ball at θ with radius r_1 such that the closure of the ball includes all the nearest neighbors and the boundary of the ball includes one or more θ^j that are not included in the set of the nearest neighbors. For this case, I will construct a ball in the vicinity of θ such that it can be made as close to θ as needed and such that for any point inside this ball, the set of the nearest neighbors is the same as for θ.

As described in the main paper (see (11)), the selection of the nearest neighbors on the boundary of $B_{r_1}(\theta)$ is conducted by the lexicographic comparison of $(\theta^j - \theta)$. Let us denote vectors $(\theta^j - \theta)$ such that θ^j is on the boundary of $B_{r_1}(\theta)$: $\|\theta^j - \theta\| = r_1$ by $x^{0,i}$, $i = 1, \dots, M_x^0$. The results of the lexicographic selection process can be represented as

$$
\begin{aligned}
z^{k,i} &= (r_1 - a_1, \dots, r_{k-1} - a_{k-1}, z_k^{k,i}, \dots, z_n^{k,i}), \\
x^{k,i} &= (r_1 - a_1, \dots, r_{k-1} - a_{k-1}, r_k - a_k, x_{k+1}^{k,i}, \dots, x_n^{k,i}), \\
y^{k,i} &= (r_1 - a_1, \dots, r_{k-1} - a_{k-1}, y_k^{k,i}, \dots, y_n^{k,i}),
\end{aligned}
\tag{75}
$$

where a geometric interpretation of variables r_k and a_k is given in Figure 1,

$$
z_k^{k,i} > r_k - a_k > y_k^{k,i},
\tag{76}
$$

and $k = 1, \dots, K$ for some $K \leq n$. Vectors $z^{k,i}$, $i = 1, \dots, M_z^k$, are those vectors included in the set of nearest neighbors for which the decision of inclusion was obtained from the lexicographic comparison for the coordinate k. Vectors $x^{k,i}$, $i = 1, \dots, M_x^k$, are the vectors for which the decision has not yet been made after comparing coordinates k. Vectors $y^{k,i}$, $i = 1, \dots, M_y^k$, are the vectors for which the decision of not including them in the set of the nearest neighbors was obtained from comparing coordinate k. Vectors $x^{k+1,i}$, $y^{k+1,i}$, and $z^{k+1,i}$ are all selected from $x^{k,i}$. The lexicographic selection will end at some coordinate K with unique x^K. This vector is denoted by x, not by z, to emphasize the fact that if there are multiple repetitions of $\theta + x^K = \theta^i = \theta^j$, $i \neq j$, in the history, then not all the repetitions have to be selected for the set of nearest neighbors

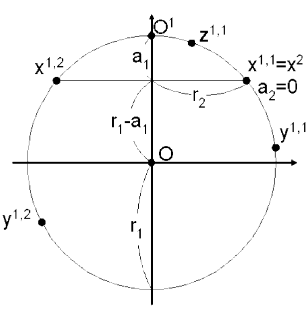

FIGURE 1.—Nearest neighbors.

(the ones with larger iteration number will be selected first). Of course, this is true only for the last selected nearest neighbor; for all the previous ones, all the repetitions are included. Note that vectors $z^{k,i}$, $x^{k,i}$, and $y^{k,i}$ are constructed in the system of coordinates with the origin at θ, so we should add θ to all of them to get back to the original coordinate system.

A graphical illustration might be helpful for understanding the idea of the proof (the proof was actually constructed from similar graphical examples in R^2 and R^3). The figure shows an example in which two nearest neighbors have to be chosen for point O. Since the required number of nearest neighbors is smaller than the number of points on the circle, we can always find a_1 such that all the points with the first coordinate strictly above $r_1 - a_1$ will be included in the set of the nearest neighbors and all the points with the first coordinate strictly below $r_1 - a_1$ will not be. For the points with the coordinate equal to $r_1 - a_1$, the selection process continues to the next dimension.

If we did not use the lexicographic comparison and just resolved the multivaluedness of arg min by choosing vectors with larger iteration numbers first, then the proposition would not hold (a counterexample could be easily found in R^2).

If the following conditions hold, then the same nearest neighbors from the surface of $B_{r_1}(\theta)$ will be chosen for $(\theta + b)$ and θ:

$$\|b - y^{k,i}\| > \|b - x^K\| > \|b - z^{k,i}\| \quad \forall k, i. \tag{77}$$

The condition says that $(x^K + \theta)$, which is the last nearest neighbor selected for θ, also has to be selected last for $(\theta + b)$ and that vectors on the boundary of $B_{r_1}(\theta)$ that are not the selected nearest neighbors for θ ($y^{k,i}, \forall k, i$) should not be the selected nearest neighbors for $(\theta + b)$. Since $\|y^{k,i}\| = \|x^K\| = \|z^{k,i}\| = r_1$, these conditions are equivalent to

$$b^T(x^K - y^{k,i}) > 0 \quad \text{and} \quad b^T(z^{k,i} - x^K) > 0. \tag{78}$$

Define

$$d = \min_{k=1,K} \min\Big\{\min_i [z_k^{k,i} - (r_k - a_k)],$$
$$\min_i [(r_k - a_k) - y_k^{k,i}]\Big\}, \qquad d > 0 \text{ by construction.}$$

For given $\epsilon_1 > 0$, let

$$\begin{aligned} &\epsilon_{k+1} = \min\{\epsilon_k, \epsilon_k d/(4nr_1)\}, \\ &\epsilon(\epsilon_1) = (\epsilon_1, \ldots, \epsilon_n), \\ &\delta(\epsilon_1) = \epsilon_n d/(8nr_1). \end{aligned} \tag{79}$$

Let $b \in B_{\delta(\epsilon_1)}(\epsilon(\epsilon_1))$ and $l = b - \epsilon(\epsilon_1)$. Let us show that (78) holds for any such b:

$$b^T(x^K - y^{k,i}) = (r_k - a_k - y_k^{k,i})\epsilon_k + \sum_{m=k+1}^{n}(x_m^K - y_m^{k,i})\epsilon_m + \sum_{m=k}^{n}(x_m^K - y_m^{k,i})l_m. \tag{80}$$

Note that $|l_k| \le \delta(\epsilon_1)$ and $|x_m^K - y_m^{k,i}| \le 2r_1$:

$$\begin{aligned} b^T(x^K - y^{k,i}) &\ge (r_k - a_k - y_k^{k,i})\epsilon_k - n2r_1 \max_{m=k+1,n}\epsilon_m - n2r_1\delta(\epsilon_1) \\ &\ge d\epsilon_k - n2r_1\frac{\epsilon_k d}{4nr_1} - n2r_1\frac{\epsilon_k d}{8nr_1} \\ &= \frac{d\epsilon_k}{4} > 0. \end{aligned} \tag{81}$$

Analogously,

$$\begin{aligned} b^T(z^{k,i} - x^K) &\ge [z_k^{k,i} - (r_k - a_k)]\epsilon_k + \sum_{m=k+1}^{n}(z_m^{k,i} - x_m^K)\epsilon_m \\ &\quad + \sum_{m=k}^{n}(z_m^{k,i} - x_m^K)l_m \\ &\ge d\epsilon_k - n2r_1 \max_{m=k+1,n}\epsilon_m - n2r_1\delta(\epsilon_1) \\ &\ge \frac{d\epsilon_k}{4} > 0. \end{aligned} \tag{82}$$

Thus, the order of selecting the nearest neighbors on the surface of $B_{r_1}(\theta)$ is the same for θ and any $\theta + b$ if $b \in B_{\delta(\epsilon_1)}(\epsilon(\epsilon_1))$ for any $\epsilon_1 > 0$. Making ϵ_1 sufficiently small, we can guarantee that all θ^j satisfying $\|\theta^j - \theta\| < r_1$ will be chosen as the nearest neighbors for $\theta + b$ before the vectors on the surface of $B_{r_1}(\theta)$ and that θ^j satisfying $\|\theta^j - \theta\| > r_1$ will not be chosen at all. For any $\epsilon_1 > 0$, $B_{\delta(\epsilon_1)}(\theta + \epsilon(\epsilon_1))$ will contain rational numbers. Letting a positive sequence $\{\epsilon_1^m\}$ go to zero and choosing $q_m \in B_\delta(\theta + \epsilon^m) \cap Q$ will give the sought sequence $\{q_m\}$. *Q.E.D.*

PROPOSITION 4: *If Θ and S are compact, $u(s, d; \theta)$ is continuous in (s, θ), and $f(s' \mid s, d; \theta)$ is continuous in (θ, s, s'), then $V(s; \theta)$ and $E\{V(s'; \theta) \mid s, d; \theta\}$ are continuous in (θ, s).*

PROOF: The proof of the proposition follows closely the standard proof of the continuity of value functions with respect to the state variables (see, for example, Chapters 3 and 4 of Stokey and Lucas (1989)). Let us consider the Bellman operator Γ on the Banach space of bounded functions B with sup norm: $V : \Theta \times S \to X$, where X is a bounded subset of R:

$$\Gamma(V)(s;\theta) = \max_d \left\{ u(s, d; \theta) + \beta \int V(s'; \theta) f(s' \mid s, d; \theta)\, ds' \right\}.$$

Blackwell's sufficient conditions for contraction are satisfied for this operator, so Γ is a contraction mapping on B. The set of continuous functions C is a closed subset in B. Thus, it suffices show that $\Gamma(C) \subset C$ (this trivially implies that the fixed point of Γ is a continuous function).

Let $V(s;\theta)$ be a continuous function in B ($V \in C$). Let us show that $\Gamma(V)$ is also continuous:

$$\begin{aligned}
(83)\qquad & |\Gamma(V)(s_1;\theta_1) - \Gamma(V)(s_2;\theta_2)| \\
&\le \max_d \bigg| u(s_1, d; \theta_1) - u(s_2, d; \theta_2) \\
&\quad + \beta \int V(s'; \theta_1) f(s' \mid s_1, d; \theta_1)\, ds' \\
&\quad - \beta \int V(s'; \theta_2) f(s' \mid s_2, d; \theta_2)\, ds' \bigg| \\
&\le \max_d |u(s_1, d; \theta_1) - u(s_2, d; \theta_2)| \\
&\quad + \beta \max_d \bigg| \int [V(s'; \theta_1) f(s' \mid s_1, d; \theta_1) \\
&\quad - V(s'; \theta_2) f(s' \mid s_2, d; \theta_2)]\, ds' \bigg|.
\end{aligned}$$

Given $\epsilon > 0$, there exists $\delta_1 > 0$ such that $\|(s_1;\theta_1) - (s_2;\theta_2)\| < \delta_1$ implies $\max_d |u(s_1, d; \theta_1) - u(s_2, d; \theta_2)| < \epsilon/2$:

$$\begin{aligned}
(84)\qquad & \left| \int [V(s'; \theta_1) f(s' \mid s_1, d; \theta_1) - V(s'; \theta_2) f(s' \mid s_2, d; \theta_2)]\, ds' \right| \\
&\le \max_d \sup_{s'} \big| V(s'; \theta_1) f(s' \mid s_1, d; \theta_1) - V(s'; \theta_2) f(s' \mid s_2, d; \theta_2) \big| \cdot \lambda(S).
\end{aligned}$$

Since $V(s';\theta)f(s' \mid s,d;\theta)$ is continuous on compact $\Theta \times S \times S$, for $\epsilon > 0$ there exists $\delta_2^d > 0$ such that $\|(s_1, s'; \theta_1) - (s_2, s'; \theta_2)\| = \|(s_1;\theta_1) - (s_2;\theta_2)\| < \delta_2^d$ implies

$$\sup_{s'} |V(s';\theta_1)f(s' \mid s_1,d;\theta_1) - V(s';\theta_2)f(s' \mid s_2,d;\theta_2)| < \frac{\epsilon}{2\lambda(S)}.$$

Thus, for $\delta = \min\{\delta_1, \min_d \delta_2^d\}$, $\|(s_1;\theta_1)-(s_2;\theta_2)\| < \delta$ implies $|\Gamma(V)(s_1;\theta_1) - \Gamma(V)(s_2;\theta_2)| < \epsilon$. So $\Gamma(V)$ is a continuous function. The continuity of $E\{V(s';\theta) \mid s,d;\theta\}$ follows from the continuity of $V(s';\theta)$ by an analogous argument. *Q.E.D.*

PROPOSITION 5: *A family of functions $X(\omega^{t-1},\theta,s,d,m)$ defined in* (18) *is equicontinuous on $\Theta \times S$ if Θ and S are compacts, $V(s;\theta)$ and $E[V(s';\theta) \mid s,d;\theta]$ are continuous in (θ,s), and $f(s' \mid s,d;\theta)/g(s')$ is continuous in (θ,s,s') and satisfies Assumption* 4.

PROOF: Let us introduce the following notation shortcuts: T will denote the number of terms in the sum defining $X(\omega^{t-1},\theta,s,d,m)$. Consider two arbitrary points (θ_1,s_1) and (θ_2,s_2), let $V_j^i = V(s^j;\theta_i) - EV(\theta_i,s_i)$ and

$$W_j^i = \frac{f_j^i/g_j^i}{\sum f_k^i/g_k^i} = \frac{f(s^j \mid s_i,d;\theta_i)/g(s^j)}{\sum f(s^k \mid s_i,d;\theta_i)/g(s^k)}.$$

Then

$$\begin{aligned}
&|X(\omega^{t-1},\theta_1,s_1,d,m) - X(\omega^{t-1},\theta_2,s_2,d,m)| \\
&\quad = \left| \sum_{j=1}^T V_j^1 W_j^1 - \sum_{j=1}^T V_j^2 W_j^2 \pm \sum_{j=1}^T V_j^2 W_j^1 \right| \\
&\quad \le \left| \sum_{j=1}^T (V_j^1 - V_j^2) W_j^1 \right| \qquad (85)\\
&\qquad + \left| \sum_{j=1}^T V_j^2 (W_j^1 - W_j^2) \right|. \qquad (86)
\end{aligned}$$

By the proposition's hypothesis, $V(s;\theta)$ and $E[V(s';\theta) \mid s,d;\theta]$ are continuous in (θ,s) on a compact set. Thus, given $\epsilon > 0$ $\exists \delta_1 > 0$ such that

$$\|(\theta_1,s_1,s^j) - (\theta_2,s_2,s^j)\| = \|(\theta_1,s_1) - (\theta_2,s_2)\| < \delta_1$$

implies $|V(s^j;\theta_1) - EV(\theta_1, s_1) - (V(s^j;\theta_2) - EV(\theta_2, s_2))| < \epsilon/2$. Since the weights sum to 1, (85) is bounded above by $\epsilon/2$. Let us similarly bound (86):

$$
\begin{aligned}
(87)\qquad & \left|\sum_{j=1}^{T} V_j^2 (W_j^1 - W_j^2)\right| \\
&= \left|\sum_{j=1}^{T} V_j^2 \left(\frac{f_j^1/g_j^1}{\sum f_k^1/g_k^1} - \frac{f_j^2/g_j^2}{\sum f_k^2/g_k^2}\right)\right| \\
&= \Bigg|\Big(\Big(\sum f_k^2/g_k^2\Big)\Big[\sum V_j^2 (f_j^1/g_j^1 - f_j^2/g_j^2)\Big] \\
&\quad + \Big[\sum f_k^2/g_k^2 - \sum f_k^1/g_k^1\Big]\Big(\sum V_j^2 f_j^2/g_j^2\Big)\Big) \\
&\quad \Big/\Big(\sum f_k^1/g_k^1 \cdot \sum f_k^2/g_k^2\Big)\Bigg| \\
&\le \frac{\overline{V} \cdot \max_j |f_j^1/g_j^1 - f_j^2/g_j^2| \cdot T}{\underline{f} T} + \frac{T \cdot \max_j |f_j^1/g_j^1 - f_j^2/g_j^2| \cdot \overline{V} \cdot \overline{f} \cdot T}{\underline{f}^2 T^2} \\
&\le \max_j \left|\frac{f_j^1}{g_j^1} - \frac{f_j^2}{g_j^2}\right| \cdot \overline{V}\left(\frac{1}{\underline{f}} + \frac{\overline{f}}{\underline{f}^2}\right),
\end{aligned}
$$

where $\overline{f}$ and $\underline{f}$ are the upper and lower bounds on f/g introduced in Assumption 4, and $\overline{V} < \infty$ is an upper bound on V_j^i. Since $f(s' \mid s, d; \theta)/g(s')$ is continuous in (θ, s, s') on compact $\Theta \times S \times S$, for any $\epsilon > 0$, there exists $\delta_2 > 0$ such that $\|(\theta_1, s_1, s^j) - (\theta_2, s_2, s^j)\| = \|(\theta_1, s_1) - (\theta_2, s_2)\| < \delta_2$ implies

$$
\left|\frac{f(s^j \mid s_1, d; \theta_1)}{g(s^j)} - \frac{f(s^j \mid s_2, d; \theta_2)}{g(s^j)}\right| < \frac{\epsilon/2}{\|V\|\left(\frac{1}{\underline{f}} + \frac{\overline{f}}{\underline{f}^2}\right)} \quad \forall j.
$$

Thus, (86) is also bounded above by $\epsilon/2$. For given $\epsilon > 0$, let $\delta = \min\{\delta_1, \delta_2\}$. Then $\|(\theta_1, s_1) - (\theta_2, s_2)\| < \delta$ implies $|X(\omega^{t-1}, \theta_1, s_1, d, m) - X(\omega^{t-1}, \theta_1, s_1, d, m)| < \epsilon/2 + \epsilon/2 = \epsilon$. *Q.E.D.*

PROPOSITION 6: *Assume that in the DP solving algorithm, the same random grid over the state space is used at each iteration:* $s^{m_1,j} = s^{m_2,j} = s^j$ *for any* m_1, m_2, *and* j, *where* $s^j \overset{\text{i.i.d.}}{\sim} g(\cdot)$. *If the number of the nearest neighbors is constant,* γ_2 *in Assumption 6 is equal to zero and* $\tilde{N}(t) = \tilde{N}$, *then all the theoretical results proved in the paper will hold.*

PROOF: Only the proof of Lemma 1 is affected by the change, since in the other parts, I use only one fact about the weights in the importance sampling: the weights are in [0, 1]. Thus let us show that Lemma 1 holds.

In Lemma 1, the terms in the sum (19) corresponding to the same s^j should be grouped into one term multiplied by the number of such terms:

$$
\begin{aligned}
(88)\qquad & P(X(\omega^{t-1}, \theta, s, d, m) > \tilde{\epsilon}) \\
&= P\Bigg[\Bigg|\sum_{j=1}^{\hat{N}(\max\{m_i\})} \big(M_j(t,m)(V(s^j;\theta) - E[V(s';\theta) \mid s, d; \theta]) \\
&\qquad \times f(s^j \mid s, d; \theta)/g(s^j)\big) \\
&\qquad \Big/ \Bigg(\sum_{r=1}^{\hat{N}(\max\{m_r\})} M_r(t,m) f(s^r \mid s, d; \theta)/g(s^r)\Bigg)\Bigg| > \tilde{\epsilon}\Bigg] \\
&\le P\Bigg[\Bigg|\sum_{j=1}^{\hat{N}(\max\{m_i\})} M_j(t,m)(V(s^j;\theta) - E[V(s';\theta) \mid s, d; \theta]) \\
&\qquad \times \frac{f(s^j \mid s, d; \theta)}{g(s^j)}\Bigg| > \tilde{\epsilon}\underline{f}\hat{N}(\max\{m_i\})\Bigg],
\end{aligned}
$$

where $M_j(t,m) \in \{1, \ldots, \tilde{N}(t)\}$ denotes the number of the terms corresponding to s^j and $\hat{N}(\max\{m_r\})$ is the largest grid size. The inequality above follows since

$$
\sum_{j=1}^{\hat{N}(\max\{m_i\})} M_j(t,m) f(s^j \mid s, d; \theta)/g(s^j) \ge \underline{f}\hat{N}(\max\{m_i\}).
$$

The summands in (88) are bounded by $(\tilde{N}a, \tilde{N}b)$, where a and b were defined in Lemma 1. Application of Hoeffding's inequality to (88) gives

$$
\begin{aligned}
(89)\qquad P(X(\omega^{t-1}, \theta, s, d, m) > \tilde{\epsilon}) &\le 2\exp\big\{-4\delta\tilde{N}\hat{N}(\max\{m_i\})\big\} \\
&\le 2\exp\big\{-4\delta\tilde{N}\hat{N}(t - N(t))\big\},
\end{aligned}
$$

where $0 < \delta = \tilde{\epsilon}^2\underline{f}^2/(2(b-a)^2\tilde{N}^3)$. The rest of the argument follows the steps in Lemma 1 starting after (22). *Q.E.D.*

PROOF OF COROLLARY 1: Given the assumptions made in the first part of this proposition, the proofs of Lemma 1 and its uniform extension Lemma 4 apply without any changes. The rest of the results are not affected at all. *Q.E.D.*

PROOF OF COROLLARY 2: If (14) is used for approximating the expectations, then in the proofs of Lemmas 1 and 4 let us separate the expression for $X(\cdot)$ into $K = \text{card}(S_f(s_f, d))$ terms corresponding to each possible future discrete state:

$$
\begin{aligned}
(90)\qquad & X(\omega^{t-1}, \theta, s, d, m) \\
&= f(s'_{f,1} \mid s_f, d; \theta) \\
&\quad \times \left\{ \sum_{i=1}^{\tilde{N}(m)} \sum_{j=1}^{\hat{N}(k_i)} \frac{V^{k_i}(s'_{f,1}, s^{k_i,j}; \theta^{k_i}) f(s^{k_i,j} \mid s_c, d; \theta)/g(s^{k_i,j})}{\sum_{r=1}^{\tilde{N}(m)} \sum_{q=1}^{\hat{N}(k_r)} f(s^{k_r,q} \mid s_c; \theta)/g(s^{k_r,q})} \right. \\
&\quad \left. - E[V(s'; \theta) \mid s'_f = s'_{f,1}, s_c, d; \theta] \right\} + \cdots \\
&\quad + f(s'_{f,K} \mid s_f, d; \theta) \\
&\quad \times \left\{ \sum_{i=1}^{\tilde{N}(m)} \sum_{j=1}^{\hat{N}(k_i)} \frac{V^{k_i}(s'_{f,K}, s^{k_i,j}; \theta^{k_i}) f(s^{k_i,j} \mid s_c, d; \theta)/g(s^{k_i,j})}{\sum_{r=1}^{\tilde{N}(m)} \sum_{q=1}^{\hat{N}(k_r)} f(s^{k_r,q} \mid s_c; \theta)/g(s^{k_r,q})} \right. \\
&\quad \left. - E[V(s'; \theta) \mid s'_f = s'_{f,K}, s_c, d; \theta] \right\}.
\end{aligned}
$$

Then, applying the argument from Lemmas 1 and 4, we can bound the probabilities for $k = 1, \ldots, K$,

$$
\begin{aligned}
(91)\qquad & P\left[\left| f(s'_{f,k} \mid s_f, d; \theta) \sum_{i=1}^{\tilde{N}(m)} \sum_{j=1}^{\hat{N}(k_i)} \frac{V^{k_i}(s'_{f,k}, s^{k_i,j}; \theta^{k_i}) f(s^{k_i,j} \mid s_c, d; \theta)/g(s^{k_i,j})}{\sum_{r=1}^{\tilde{N}(m)} \sum_{q=1}^{\hat{N}(k_r)} f(s^{k_r,q} \mid s_c; \theta)/g(s^{k_r,q})} \right. \right. \\
(92)\qquad & \left. \left. \quad - E[V(s'; \theta) \mid s'_f = s'_{f,k}, s_c, d; \theta] \right| > \frac{\epsilon}{K} \right],
\end{aligned}
$$

and Lemmas 1 and 4 will hold. The proofs of the other lemmas are not affected at all, since the weights on the value functions in expectation approximations are still nonnegative and sum to 1. *Q.E.D.*

PROPOSITION 7: *If* x_t, z_t, *and* y_t *are integer sequences with* $\lim_{t\to\infty} y_t/z_t = 0$, $\lim_{t\to\infty} z_t = \infty$, *and* $\limsup_{t\to\infty} z_t/x_t < \infty$, *then* $\forall\delta > 0\ \exists T$ *such that* $\forall t > T$,

$$e^{-\delta x_t}\frac{z_t!}{(z_t-y_t)!y_t!} \le e^{-0.5\delta x_t}.$$

PROOF: To prove the inequality let us work with the logarithm of the left hand side

$$\begin{aligned}
(93)\qquad & \log\left[e^{-\delta x_t}\frac{z_t!}{(z_t-y_t)!y_t!}\right] \\
&= -\delta x_t + \sum_{i=z_t-y_t+1}^{z_t}\log(i) - \sum_{i=1}^{y_t}\log(i) \\
&\le -\delta x_t + \int_{z_t-y_t+1}^{z_t+1}\log(i)\,di - \int_1^{y_t}\log(i)\,di \\
&= -\delta x_t + (z_t+1)\log(z_t+1) - (z_t-y_t+1)\log(z_t-y_t+1) \\
&\quad - [(z_t+1)-(z_t-y_t+1)] - \{y_t\log(y_t) - 1\log(1) - [y_t-1]\} \\
&= -\delta x_t + z_t[\log(z_t+1) - \log(z_t-y_t+1)] \\
&\quad + y_t[\log(z_t-y_t+1) - \log(y_t)] \\
&\quad + \log(z_t+1) - \log(z_t-y_t+1) - y_t\log(y_t) - 1 \\
&\le -\delta x_t + z_t\log\frac{z_t+1}{z_t-y_t+1} + y_t\log\frac{z_t-y_t+1}{y_t} + \log\frac{z_t+1}{z_t-y_t+1} \\
&= x_t\left[-\delta + \frac{z_t}{x_t}\log\frac{z_t+1}{z_t-y_t+1}\right. \\
&\quad \left. + \frac{(z_t-y_t+1)y_t}{x_t(z_t-y_t+1)}\log\frac{z_t-y_t+1}{y_t} + \frac{1}{x_t}\log\frac{z_t+1}{z_t-y_t+1}\right] \\
&\le -0.5\delta x_t \quad \forall t > T.
\end{aligned}$$

There exists such T that the last inequality holds, since all the terms in (93) converge to zero. Exponentiating the obtained inequality completes the proof. *Q.E.D.*

PROPOSITION 8: *For any* $\delta_1 > 0$ *and* $\delta_2 > 0$, *there exist* $\delta > 0$ *and* T *such that* $\forall t > T$,

$$(94)\qquad \sum_{i=t-m(t)N(t)}^{t-1}\hat{N}(i)\left[e^{-\delta_1\tilde{N}(i)\hat{N}(i-N(i))} + e^{-\delta_2(N(i)-\tilde{N}(i))}\right] \le \exp\{-\delta t^{\gamma_0\gamma_1}\}.$$

PROOF: The following inequalities will be used below:

$$t - m(t)N(t) \geq t - \frac{t - t^{\gamma_0}}{N(t)} N(t) = t^{\gamma_0}, \tag{95}$$

$$t - m(t)N(t) \leq t - \left(\frac{t - t^{\gamma_0}}{N(t)} - 1\right) N(t) \leq t^{\gamma_0} + t^{\gamma_1} < 2t^{\gamma_0}, \tag{96}$$

$$\begin{aligned} \tilde{N}(t - m(t)N(t)) &= \left[(t - m(t)N(t))^{\gamma_2}\right] \\ &\geq [t^{\gamma_0\gamma_2}] \geq t^{\gamma_0\gamma_2} - 1 \geq 0.5t^{\gamma_0\gamma_2} \quad \forall t > T_1 = 2^{1/(\gamma_0\gamma_2)}, \end{aligned} \tag{97}$$

$$N(t - m(t)N(t)) = \left[(t - m(t)N(t))^{\gamma_1}\right] \leq (2t^{\gamma_0})^{\gamma_1} \leq 2^{\gamma_1} t^{\gamma_0\gamma_1}, \tag{98}$$

$$\begin{aligned} &\hat{N}\big(t - m(t)N(t) - N(t - m(t)N(t))\big) \\ &\quad = \left[\big(t - m(t)N(t) - N(t - m(t)N(t))\big)^{\gamma_1 - \gamma_2}\right] \\ &\quad \geq (t^{\gamma_0} - 2^{\gamma_1} t^{\gamma_0\gamma_1})^{\gamma_1 - \gamma_2} - 1 \quad \text{(by (95) and (98))} \\ &\quad \geq \frac{t^{\gamma_0(\gamma_1 - \gamma_2)}}{2^{\gamma_1 - \gamma_2}} - 1 \quad (\forall t > 2^{(1+\gamma_1)/(\gamma_0(1-\gamma_1))}) \\ &\quad \geq \frac{t^{\gamma_0(\gamma_1 - \gamma_2)}}{2^{1+\gamma_1 - \gamma_2}} \quad (\forall t > T_2), \end{aligned} \tag{99}$$

where $T_2 = \max\{2^{(1+\gamma_1-\gamma_2)/(\gamma_0(\gamma_1-\gamma_2))}, 2^{(1+\gamma_1)/(\gamma_0(1-\gamma_1))}\}$.

Combining (97) and (99) gives

$$\begin{aligned} &\exp\big\{-\delta_1 \tilde{N}(t - m(t)N(t)) \hat{N}\big(t - m(t)N(t) - N(t - m(t)N(t))\big)\big\} \\ &\quad \leq \exp\left\{-\frac{\delta_1 t^{\gamma_0\gamma_1}}{2^{2+\gamma_1-\gamma_2}}\right\} = \exp\{-\tilde{\delta}_1 t^{\gamma_0\gamma_1}\}, \end{aligned} \tag{100}$$

where the last equality defines $\tilde{\delta}_1 > 0$, and

$$\begin{aligned} &N(t - m(t)N(t)) - \tilde{N}(t - m(t)N(t)) \\ &\quad = \left[(t - m(t)N(t))^{\gamma_1}\right] - \left[(t - m(t)N(t))^{\gamma_2}\right] \\ &\quad \geq [t^{\gamma_0\gamma_1}] - [2^{\gamma_2} t^{\gamma_0\gamma_2}] \quad \text{(by (95) and (96))} \\ &\quad \geq t^{\gamma_0\gamma_1} - 1 - 2^{\gamma_2} t^{\gamma_0\gamma_2} \\ &\quad \geq 0.5t^{\gamma_0\gamma_1} \quad \text{for } t \text{ larger than some } T_3, \end{aligned} \tag{101}$$

where such T_3 exists since $(0.5t^{\gamma_0\gamma_1} - 1 - 2^{\gamma_2} t^{\gamma_0\gamma_2}) \to \infty$.

Taking an upper bound on summands in (94) and multiplying it by the number of terms in the sum gives the upper bound on the sum:

$$\sum_{i=t-m(t)N(t)}^{t-1} \hat{N}(i)\big[e^{-\delta_1 \tilde{N}(i)\hat{N}(i-N(i))} + e^{-\delta_2 (N(i)-\tilde{N}(i))}\big] \tag{102}$$
$$\leq \big((t-1) - (t-m(t)N(t)) + 1\big) \times \hat{N}(t-1)$$
$$\times \big[e^{-\delta_1 \tilde{N}(t-m(t)N(t))\hat{N}(t-m(t)N(t)-N(t-m(t)N(t)))}$$
$$+ e^{-\delta_2 (N(t-m(t)N(t))-\tilde{N}(t-m(t)N(t)))}\big].$$

Inequalities in (100), (101), and (102) imply

$$\sum_{i=t-m(t)N(t)}^{t-1} \hat{N}(i)\big[e^{-\delta_1 \tilde{N}(i)\hat{N}(i-N(i))} + e^{-\delta_2 (N(i)-\tilde{N}(i))}\big] \tag{103}$$
$$\leq t^{1+\gamma_1-\gamma_2}(\exp\{-\tilde{\delta}_1 t^{\gamma_0\gamma_1}\} + \exp\{-0.5\delta_2 t^{\gamma_0\gamma_1}\})$$
$$\leq 2t^{1+\gamma_1-\gamma_2}\exp\{-\min(\tilde{\delta}_1, 0.5\delta_2)t^{\gamma_0\gamma_1}\},$$

where $\tilde{\delta}_1$ was defined in (100).

Note that $(2t^{1+\gamma_1-\gamma_2}\exp\{-0.5\min(\tilde{\delta}_1, 0.5\delta_2)t^{\gamma_0\gamma_1}\}) \to \infty$ and therefore $\exists T \geq \max(T_1, T_2, T_3)$ such that $\forall t > T$,

$$2t^{1+\gamma_1-\gamma_2}\exp\{-\min(\tilde{\delta}_1, 0.5\delta_2)t^{\gamma_0\gamma_1}\} \leq \exp\{-\delta t^{\gamma_0\gamma_1}\}, \tag{104}$$

where $\delta = 0.5\min(\tilde{\delta}_1, 0.5\delta_2)$. This completes the proof. *Q.E.D.*

PROPOSITION 9: *For any* $a > 0$ *and* $\delta > 0$, $\sum_{t=1}^{\infty}\exp\{-\delta t^a\} < \infty$.

PROOF—Sketch: The sum above is a lower sum for the improper integral $\int_0^\infty \exp\{-\delta t^a\}\,dt$. One way to show that it is finite is to do a transformation of variables $y = t^a$ and then bound the obtained integral by an integral of the form $\int_0^\infty y^n \exp\{-\delta y\}\,dy$, where n is an integer. It follows by induction and integration by parts that this integral is finite. *Q.E.D.*

REFERENCES

HOEFFDING, W. (1963): "Probability Inequalities for Sums of Bounded Random Variables," *Journal of the American Statistical Association*, 58, 13–30.

ROBERTS, G., AND ROSENTHAL, J. (2006): "Coupling and Ergodicity of Adaptive MCMC," available at http://www.probability.ca/jeff/research.html.

STOKEY, N., AND LUCAS, R. (1989): *Recursive Methods in Economic Dynamics*. Cambridge, MA: Harvard University Press.

Dept. of Economics, Princeton University, 313 Fisher Hall, Princeton, NJ 08544, U.S.A.; anorets@princeton.edu.

Manuscript received July, 2007; final revision received March, 2009.