Econometrica Supplementary Material

SUPPLEMENT TO BAYESIAN ESTIMATION OF DYNAMIC DISCRETE CHOICE MODELS

(*Econometrica*, Vol. 77, No. 6, November 2009, 1865–1899)

BY SUSUMU IMAI, NEELAM JAIN, AND ANDREW CHING

This material contains three appendices: Appendix A includes results of three Monte Carlo experiments; Appendix B contains all proofs of the theoretical results reported in the main body of the paper; Appendix C contains all the diagrams that illustrate the estimation results of Experiment A1 in the main body of the paper.

APPENDIX A: EXPERIMENTS

Experiment A1: Basic Model

ALL PRIORS ARE SET TO BE DIFFUSE. We set the initial guesses of the parameters to be the true parameter values given by θ^0, and set the initial guess of the expected value function to be 0. We used the same 200 grid points in each iteration as used to generate the data. The pseudo-MCMC sampler was generated 10,000 times. The posterior mean and standard errors from the 5,001st iteration up to 10,000th iteration are shown in the second, third, and fourth columns of Table AI.[1] As we can see, the sample averages of the Bayesian DP posterior means are very close to the true parameter values. We also conducted the same simulation-estimation exercises for full-solution-based Bayesian estimation and ML estimation. The CPU time required for the full-solution-based Bayesian estimation was 15 minutes 44 seconds for sample size of 2,000, 21 minutes 31 seconds for sample size of 5,000, and 29 minutes 44 seconds for sample size of 10,000. Note that for the Bayesian DP estimation, as the sample size decreases from 10,000 to 2,000, the CPU time decreases from 18 minutes 17 seconds to 4 minutes 53 seconds—a factor of 3.5 to 1. On the other hand, for the full-solution-based Bayesian estimation, there is only a 50% decrease. That is, as the sample size decreases, relatively more CPU time is spent on the solution of the model than on computing the likelihood. Hence, the computational advantage of the Bayesian DP algorithm becomes more apparent. Similar estimation exercises using the full-solution-based ML required only 17 seconds for sample size of 10,000, with parameter estimates that are similar to the posterior means of the Bayesian estimates. Thus, it is clear that for a simple DP model, ML estimation outperforms Bayesian DP. The advantages of the Bayesian DP algorithm become more apparent as the model becomes more complex, as we show in the next experiment.

To check robustness of the Bayesian DP algorithm, we also ran a simulation-estimation exercise where the starting parameter value was set to be half of

[1]The posterior means and posterior standard deviations presented are the sample averages of the posterior means and posterior standard deviations across the 50 replications.

 DOI: 10.3982/ECTA5658

TABLE AI

POSTERIOR MEANS AND STANDARD DEVIATIONS[a]

Parameter	$\overline{\text{PM}}$	$\overline{\text{PM}}$	$\overline{\text{PM}}$	Starting Value: $0.5\theta^0$	True
δ	0.4046 (0.0295)	0.4043 (0.0195)	0.4021 (0.0135)	0.3991 (0.0134)	0.4
α	0.1001 (0.00620)	0.1000 (0.00397)	0.1003 (0.00288)	0.1000 (0.00288)	0.1
σ_{ϵ_1}	0.3005 (0.00734)	0.3003 (0.00481)	0.3001 (0.00354)	0.3004 (0.00350)	0.3
σ_{ϵ_2}	0.3101 (0.0214)	0.3035 (0.0159)	0.3027 (0.0122)	0.2986 (0.0125)	0.3
b_0	0.00310 (0.0121)	0.000739 (0.00770)	0.000218 (0.00547)	0.000265 (0.00547)	0.0
b_2	0.3920 (0.0236)	0.3973 (0.0152)	0.3993 (0.0106)	0.3994 (0.0107)	0.4
b_e	0.4966 (0.0242)	0.5043 (0.0157)	0.5015 (0.0111)	0.5015 (0.0112)	0.5
σ_u	0.3981 (0.00729)	0.3994 (0.00465)	0.3997 (0.00328)	0.3997 (0.00330)	0.4
Sample	2,000	5,000	10,000	10,000	
CPU time	4 min 53 s	9 min 53 s	18 min 17 s	18 min 11 s	

[a]Standard deviations are in parentheses.

the true values. As we can see from the results reported in the fifth column, the posterior means and the standard deviations are almost the same as those of the fourth column, where the initial parameter values were set to be the true ones. These results confirm the theorems on convergence in Section 3, that stated that the estimation algorithm is not sensitive to the initial values.

Experiment A2: Estimation of Discount Factor

In Table AII we present the results of the Bayesian DP estimation where we also estimate the discount factor. As we can see, the posterior means and standard errors are very close to the true values.

TABLE AII

POSTERIOR MEANS AND STANDARD DEVIATIONS[a]

Parameter	Basic Model	True Value
δ	0.3938 (0.0130)	0.4
α	0.09889 (0.00287)	0.1
σ_{ϵ_1}	0.3023 (0.00348)	0.3
σ_{ϵ_2}	0.2877 (0.0112)	0.3
b_0	−0.001686 (0.00554)	0.0
b_2	0.4038 (0.0108)	0.4
b_e	0.5069 (0.0112)	0.5
σ_u	0.3984 (0.00331)	0.4
β	0.9727 (0.0104)	0.98
CPU time	18 min 12 s	

[a]Standard deviations are in parentheses.

TABLE AIII

FULL-SOLUTION-BASED ESTIMATION[a]

Parameter	Full-Solution Bayes	Full-Solution ML	True Value
δ	0.3981 (0.0182)	0.3795 (0.0171)	0.4
α	0.1977 (0.0105)	0.1701(0.0135)	0.2
σ_α	0.1008 (0.00729)	0.09326 (0.0140)	0.1
σ_{ϵ_1}	0.3017 (0.00302)	0.3025 (0.00317)	0.3
σ_{ϵ_2}	0.3022 (0.0149)	0.2805 (0.0176)	0.3
b_1	0.1000 (0.00487)	0.1004 (0.00530)	0.1
b_2	0.3971 (0.00978)	0.4003 (0.0101)	0.4
b_e	0.4965 (0.0137)	0.5054 (0.0145)	0.5
σ_u	0.4003 (0.00321)	0.3990 (0.00317)	0.4
Sample size	100×100	100×100	
CPU time	30 h 59 min	20 h 47 min	

[a]Standard deviations are in parentheses.

Experiment A3: Full-Solution-Based Bayesian and ML Estimation for the Random Effects Model

In Table AIII we report the results of the full-solution-based Bayesian estimation (second column) and ML estimation (third column) with 100 ϵ_i draws (i.e., $M_\epsilon = 100$) used to integrate over ϵ_i to evaluate the expected value function, 100 α_i draws per firm to simulate the likelihood (i.e., $M_\alpha = 100$). For the ML algorithm, we used the Newton–Raphson routine. Since we took numerical derivatives, in addition to the likelihood evaluation under the original parameter θ, we calculated the likelihood for the 9 parameter perturbations $\theta + \Delta\theta_i$, $i = 1, \dots, 9$. We stopped running the program when either the absolute values of all the gradients were less than 0.01 or the step size became less than $1.0D - 5$. In all the simulation-estimation results in Tables AIII and AIV, we report the average across 10 simulation-estimation exercises.

As was discussed in the main part of the paper, the posterior means of the full-solution-based Bayesian estimates are close to the true values. On the other hand, some parameter estimates of the full-solution-based simulated ML estimates have biases that make them more than 1 standard deviation away from the true values. They are δ and α. In the second column of Table AIV we report the estimation results of the simulated ML when we reduce the number of α_i draws, M_α, to 20. Then, even though the bias of δ and σ_{ϵ_2} is similar to that with $M_\alpha = 100$, the bias of α increases substantially, with the posterior mean being 0.1450. In third column of Table AIV we report the estimation results of the simulated ML when we reduce the number of ϵ_i draws, M_ϵ, to 20. Then, as we can see, the magnitude of bias is similar to that with $M_\epsilon = 100$, but there is not much reduction in computational time either, dropping only from 20 hours 47 minutes to 18 hours 15 minutes.

TABLE AIV

SIMULATED ML ESTIMATION[a]

Parameter	$20\,\alpha_i$ Draws	$20\,\epsilon$ Draws	True Value
δ	0.3795 (0.0173)	0.3895 (0.0192)	0.4
α	0.1450 (0.0123)	0.1764 (0.0157)	0.2
σ_α	0.1076 (0.0203)	0.09527 (0.0126)	0.1
σ_{ϵ_1}	0.3030 (0.00315)	0.3028 (0.00315)	0.3
σ_{ϵ_2}	0.2790 (0.0177)	0.2810 (0.0181)	0.3
b_1	0.1003 (0.00526)	0.09977 (0.00524)	0.1
b_2	0.3999 (0.0100)	0.4000 (0.00996)	0.4
b_e	0.5030 (0.0146)	0.5048 (0.0145)	0.5
σ_u	0.3988 (0.00318)	0.3988 (0.00317)	0.4
CPU time	8 h 43 min	18 h 15 min	

[a]Standard deviations are in parentheses.

APPENDIX B: PROOFS

The following two lemmas establish some properties that are used in the later proofs.

LEMMA 1: *Let $h(\cdot)$, ε_0, and $g(\cdot)$ be defined as*

$$h(\theta^*) \equiv \inf_{\theta\in\Theta} q(\theta, \theta^*), \quad \varepsilon_0 = \int h(\tilde{\theta})\, d\tilde{\theta}, \quad g(\theta^*) \equiv \frac{h(\theta^*)}{\int h(\tilde{\theta})\, d\tilde{\theta}}.$$

Then $0 < \varepsilon_0 \le 1$ and for any $\theta, \theta^* \in \Theta$,

$$\varepsilon_0 g(\theta^*) \le q(\theta, \theta^*).$$

PROOF: By Assumption 1 (compactness of parameter space), for any $\theta^* \in \Theta$, $h(\theta^*) \equiv \inf_{\theta\in\Theta} q(\theta, \theta^*)$ exists, is strictly positive, and is uniformly bounded below. Notice that $h(\cdot)$ is Lebesgue integrable. Furthermore, for any $\theta \in \Theta$, $\varepsilon_0 g(\theta^*) = h(\theta^*) \le q(\theta, \theta^*)$. Next, since q satisfies Assumption 1, $g(\cdot)$ is strictly positive and bounded, and $\int g(\theta)\, d\theta = 1$. Hence, $g(\cdot)$ is a density function. Also, by construction, ε_0 is a strictly positive constant. Finally, since both $g(\cdot)$ and $q(\theta, \cdot)$ are densities and integrate to 1, $0 < \varepsilon_0 \le 1$. *Q.E.D.*

Lemma 1 implies that the proposal density of the modified Metropolis–Hastings algorithm has an important property: regardless of the current parameter values or the number of iterations, every neighborhood in the compact parameter space is visited with a strictly positive probability.

LEMMA 2: *Let $\widetilde{h}(\cdot)$ be a continuously differentiable and uniformly bounded function which satisfies the inequality $\widetilde{h}(\theta^*) \geq \sup_{\theta\in\Theta} q(\theta, \theta^*)$. Let ε_1 and $\widetilde{g}(\cdot)$ be defined as*

$$\varepsilon_1 \equiv \int \widetilde{h}(\tilde{\theta})\, d\tilde{\theta}, \quad \widetilde{g}(\theta^*) \equiv \frac{\widetilde{h}(\theta^*)}{\int \widetilde{h}(\tilde{\theta})\, d\tilde{\theta}}.$$

Then $1 \leq \varepsilon_1 < \infty$ and for any $\theta, \theta^* \in \Theta$,

$$q(\theta, \theta^*) \leq \varepsilon_1 \widetilde{g}(\theta^*).$$

PROOF: Using similar logic as in Lemma 1, one can show that for any $\theta^* \in \Theta$, $\sup_{\theta\in\Theta} q(\theta, \theta^*)$ exists and is bounded. Then $\widetilde{g}(\theta)$ and ε_1 satisfy the conditions of the lemma. *Q.E.D.*

Lemma 2 implies that the proposal density is bounded above, the bound being independent of the current parameter value or the number of iterations.

PROOF OF THEOREM 1: For notational convenience, in the subsequent proofs we omit $\mathcal{H}$. We need to show that for any $s \in S$, ϵ, and $\theta \in \Theta$,

$$V^{(t)}(s, \epsilon, \theta) \overset{p}{\to} V(s, \epsilon, \theta) \quad \text{uniformly, as } t \to \infty.$$

But since

$$V^{(t)}(s, \epsilon, \theta) = \max_{a\in A} \mathcal{V}^{(t)}(s, a, \epsilon, \theta),$$
$$V(s, \epsilon, \theta) = \max_{a\in A} \mathcal{V}(s, a, \epsilon, \theta),$$

it suffices to show that for any $s \in S$, $a \in A$, ϵ, and $\theta \in \Theta$,

$$\mathcal{V}^{(t)}(s, a, \epsilon, \theta) \overset{p}{\to} \mathcal{V}(s, a, \epsilon, \theta) \quad \text{as } t \to \infty.$$

Define

$$W_{N(t),h}\big(\theta, \theta^{(t-n)}\big) \equiv \frac{K_h(\theta - \theta^{(t-n)})}{\sum_{k=1}^{N(t)} K_h(\theta - \theta^{(t-k)})}.$$

Then the difference between the true value function of action a and that obtained by the Bayesian dynamic programming iteration can be decomposed into two parts as follows.

$$\begin{aligned}
&\mathcal{V}(s,a,\epsilon,\theta)-\mathcal{V}^{(t)}(s,a,\epsilon,\theta)\\
&=\beta\Bigg[\int\sum_{s'}V(s',\epsilon',\theta)f(s'|s,a,\theta)\,dF_{\epsilon'}(\epsilon',\theta)\\
&\quad-\sum_{n=1}^{N(t)}\sum_{s'}V^{(t-n)}\big(s',\epsilon^{(t-n)},\theta^{*(t-n)}\big)f\big(s'|s,a,\theta^{*(t-n)}\big)\\
&\quad\times W_{N(t),h}\big(\theta,\theta^{*(t-n)}\big)\Bigg]\\
&=\beta\Bigg[\int\sum_{s'}V(s',\epsilon',\theta)f(s'|s,a,\theta)\,dF_{\epsilon'}(\epsilon',\theta)\\
&\quad-\sum_{n=1}^{N(t)}\sum_{s'}V\big(s',\epsilon^{(t-n)},\theta^{*(t-n)}\big)f\big(s'|s,a,\theta^{*(t-n)}\big)W_{N(t),h}\big(\theta,\theta^{*(t-n)}\big)\Bigg]\\
&\quad+\beta\Bigg[\sum_{n=1}^{N(t)}\sum_{s'}\big[V\big(s',\epsilon^{(t-n)},\theta^{*(t-n)}\big)-V^{(t-n)}\big(s',\epsilon^{(t-n)},\theta^{*(t-n)}\big)\big]\\
&\quad\times f\big(s'|s,a,\theta^{*(t-n)}\big)W_{N(t),h}\big(\theta,\theta^{*(t-n)}\big)\Bigg]\\
&\equiv A_1^{(t)}(\theta)+A_2^{(t)}(\theta).
\end{aligned}$$

The kernel smoothing part is difficult to handle because the underlying distribution of $\theta^{*(s)}$ has a density function conditional on $\theta^{(s-1)}$. Therefore, instead of deriving the asymptotic value of $\frac{1}{N(t)}\sum_{k=1}^{N(t)}K_h(\theta-\theta^{*(t-k)})$, as is done in standard nonparametric kernel asymptotics, we sometimes derive and use its asymptotic lower bound and upper bound. Lemma 1 is used for the derivation of the asymptotic lower bound. Lemma 2 is used for the derivation of the asymptotic upper bound. Using the results of Lemmas 1 and 2, we prove the following lemma.

LEMMA 3: $|A_1^{(t)}(\theta)|\xrightarrow{P}0$ *uniformly in* Θ *as* $t\to\infty$. *Furthermore*,

$$\Pr\Big[\sup_{\theta\in\Theta}\big|A_1^{(t)}(\theta)\big|<C_1h^{(3/2)J}\Big]>1-\frac{C_2}{\sqrt{N(t)h^{5J}}},$$

where C_1 *and* C_2 *are positive constants*.

PROOF: Denote

$$E_{s'}V(s,\epsilon',a,\theta)\equiv\sum_{s'}V(s',\epsilon',\theta)f(s'|s,a,\theta).$$

Recall that

$$\left|\frac{A_1^{(t)}(\theta)}{\beta}\right|=\left|\int E_{s'}V(s,\epsilon',a,\theta)\,dF_{\epsilon'}(\epsilon',\theta)-\sum_{n=1}^{N(t)}E_{s'}V\big(s,\epsilon^{(t-n)},a,\theta^{*(t-n)}\big)W_{N(t),h}\big(\theta,\theta^{*(t-n)}\big)\right|.$$

Rewrite the preceding equation as

$$\left|\frac{A_1^{(t)}(\theta)}{\beta}\right|=\left|\frac{1}{N(t)}\sum_{n=1}^{N(t)}\left(\int E_{s'}V(s,\epsilon',a,\theta)\,dF_{\epsilon'}(\epsilon',\theta)-E_{s'}V\big(s,\epsilon^{(t-n)},a,\theta^{*(t-n)}\big)\right)K_h\big(\theta-\theta^{*(t-n)}\big)\Big/\left(\frac{1}{N(t)}\sum_{k=1}^{N(t)}K_h\big(\theta-\theta^{*(t-k)}\big)\right)\right|.$$

We show that the numerator goes to 0 in probability uniformly in Θ and the denominator is bounded below by a positive number uniformly in Θ with probability arbitrarily close to 1 as $t\to\infty$.

Let

$$X_{N(t),t-n}(\theta)\equiv\left[\int E_{s'}V(s,\epsilon',a,\theta)\,dF_{\epsilon'}(\epsilon',\theta)-E_{s'}V\big(s,\epsilon^{(t-n)},a,\theta^{*(t-n)}\big)\right]\times K_h\big(\theta-\theta^{*(t-n)}\big),$$

$$X_{N(t)}(\theta)\equiv\frac{1}{N(t)}\sum_{n=1}^{N(t)}X_{N(t),t-n}(\theta).$$

First, we show that $E[X_{N(t)}(\theta)]\to 0$ uniformly in Θ as $N(t)\to\infty$ $(h\to 0)$. Since $|\int E_{s'}V(s,\epsilon',a,\theta^{*(t-n)})\,dF_{\epsilon'}(\epsilon',\theta^{*(t-n)})|$ is uniformly bounded, by change of variable and by the bounded convergence theorem, we get

$$E\Big[\int E_{s'}V\big(s,\epsilon',a,\theta^{*(t-n)}\big)\,dF_{\epsilon'}\big(\epsilon',\theta^{*(t-n)}\big)\,K_h\big(\theta-\theta^{*(t-n)}\big)\Big|\theta^{(t-n)}\Big]$$
$$=\int\Big[\int E_{s'}V(s,\epsilon',a,\theta-hz)\,dF_{\epsilon'}(\epsilon',\theta-hz)\Big]$$
$$\times K(z)q\big(\theta^{(t-n)},\theta-hz\big)\,dz$$

and

$$\Big[\int E_{s'}V(s,\epsilon',a,\theta-hz)\,dF_{\epsilon'}(\epsilon',\theta-hz)\Big]K(z)q\big(\theta^{(t-n)},\theta-hz\big)$$
$$\to\Big[\int E_{s'}V(s,\epsilon',a,\theta)\,dF_{\epsilon'}(\epsilon',\theta)\Big]K(z)q\big(\theta^{(t-n)},\theta\big)\quad\text{as } h\to 0.$$

Because $|\int E_{s'}V(s,\epsilon',a,\theta)\,dF_{\epsilon'}q|$ is uniformly bounded, by the bounded convergence theorem, we obtain

$$E\Big[\int E_{s'}V\big(s,\epsilon',a,\theta^{*(t-n)}\big)\,dF_{\epsilon'}\big(\epsilon',\theta^{*(t-n)}\big)\,K_h\big(\theta-\theta^{*(t-n)}\big)\Big|\theta^{(t-n)}\Big]$$
$$\to\int E_{s'}V(s,\epsilon',a,\theta)\,dF_{\epsilon'}(\epsilon',\theta)\,q\big(\theta^{(t-n)},\theta\big)\quad\text{as } h\to 0.$$

Thus, we have shown that

$$E\big[X_{N(t),t-n}(\theta)\big|\theta^{(t-n)}\big]\to 0\quad\text{as } h\to 0.$$

Because $|E[X_{N(t),t-n}(\theta)|\theta^{(t-n)}]|$ is uniformly bounded, by the bounded convergence theorem,

$$\text{(A1)}\qquad E\big[X_{N(t)}(\theta)\big]=E\Bigg[\frac{1}{N(t)}\sum_{n=1}^{N(t)}E\big[X_{N(t),t-n}(\theta)\big|\theta^{(t-n)}\big]\Bigg]\to 0,$$
$$\text{as } N(t)\to\infty.$$

We can also show that the above convergence is uniform. For some $h>0$,

$$\text{(A2)}\qquad \sup_{\theta\in\Theta}\big|E\big[X_{N(t),t-n}(\theta)\big]\big|\le\sup_{\theta\in\Theta}\bigg|E\bigg[\Big[\int E_{s'}V(s,\epsilon',a,\theta)\,dF_{\epsilon'}(\epsilon',\theta)$$
$$-\int E_{s'}V\big(s,\epsilon',a,\theta^{*(t-n)}\big)\,dF_{\epsilon'}\big(\epsilon',\theta^{*(t-n)}\big)\Big]$$
$$\times K_h\big(\theta-\theta^{*(t-n)}\big)I\big(\big|\theta-\theta^{*(t-n)}\big|\le\sqrt{h}\big)\bigg]\bigg|$$

$$
\begin{aligned}
&+\sup_{\theta\in\Theta}\left|E\left[\left[\int E_{s'}V(s,\epsilon',a,\theta)\,dF_{\epsilon'}(\epsilon',\theta)\right.\right.\right.\\
&-\int E_{s'}V\big(s,\epsilon',a,\theta^{*(t-n)}\big)\,dF_{\epsilon'}\big(\epsilon',\theta^{*(t-n)}\big)\Big]\\
&\left.\left.\times K_h\big(\theta-\theta^{*(t-n)}\big)I\big(\big|\theta-\theta^{*(t-n)}\big|>\sqrt{h}\big)\right]\right|.
\end{aligned}
$$

By Lemma 2, $q(\theta,\theta')\le\varepsilon_1 g(\theta')$ for any $\theta,\theta'\in\Theta$. Furthermore, because both $E_{s'}V(s,\epsilon',a,\theta)$ and $dF_{\epsilon'}(\epsilon',\theta)$ are assumed to satisfy the Lipschitz conditions, there exists a constant B such that

$$
\begin{aligned}
&\text{right hand side (RHS) of (A2)}\\
&\le E\Big[\sup_{\theta,\theta'\in\Theta,|\theta-\theta^*|\le\sqrt{h}} B|\theta-\theta^*|K_h(\theta-\theta^*)\Big]\\
&\quad+2\sup_{\theta\in\Theta}\left|\int E_{s'}V(s,\epsilon',a,\theta)\,dF_{\epsilon'}(\epsilon',\theta)\right|\\
&\quad\times\varepsilon_1\int_{|z|>1/\sqrt{h}}K(z)\widetilde{g}(\theta-hz)\,dz\\
&\le\varepsilon_1\sup_{\theta,\theta'\in\Theta,|\theta-\theta'|\le\sqrt{h}}\left|\int B|\theta-\theta^*|K_h(\theta-\theta^*)\,d\widetilde{g}(\theta^*)\right|\\
&\quad+2\sup_{\theta\in\Theta}\left|\int E_{s'}V(s,\epsilon',a,\theta)\,dF_{\epsilon'}(\epsilon',\theta)\right|\\
&\quad\times\varepsilon_1\int_{|z|>1/\sqrt{h}}K(z)\widetilde{g}(\theta-hz)\,dz.
\end{aligned}
$$

From Assumption 8, there exists a constant $A>0$ such that

$$\int_{|z|>1/\sqrt{h}}K(z)\,dz\le Ah^{2J}.$$

Therefore,

$$
\begin{aligned}
\text{(A3)}\qquad&\text{RHS of (A2)}\\
&\le\varepsilon_1\sup|K|\sup|\widetilde{g}|\int_{|z|\le\sqrt{h}}Bh|z|\,dz\\
&\quad+2\sup_{\theta\in\Theta}\left|\int E_{s'}V(s,\epsilon',a,\theta)\,dF_{\epsilon'}(\epsilon',\theta)\right|\sup_{\theta}\widetilde{g}(\theta)\varepsilon_1 Ah^{2J}
\end{aligned}
$$

$$
\begin{aligned}
&\le \varepsilon_1 \sup|K| \sup|\widetilde{g}| B h^{2J}/2 \\
&\quad + 2 \sup_{\theta\in\Theta} \left| \int E_{s'} V(s, \epsilon', a, \theta)\, dF_{\epsilon'}(\epsilon', \theta) \right| \sup_{\theta} \widetilde{g}(\theta) \varepsilon_1 A h^{2J} \\
&= B_1 h^{2J}
\end{aligned}
$$

for some positive constant B_1. Therefore, we have shown that the convergence in (A1) is uniform in Θ.

Next, we show that $X_{N(t)}(\theta)$ converges to zero uniformly in Θ. Here we follow Section 10.3 of Bierens (1994) closely. Denote

$$
R_{N(t)}(\theta) \equiv \frac{1}{N(t)} \sum_{n=1}^{N(t)} E_{s'} V\big(s, \epsilon^{(t-n)}, a, \theta^{*(t-n)}\big) K_h\big(\theta - \theta^{*(t-n)}\big).
$$

By using the Fourier transform, we can express the kernel as

$$
K(x) = \left(\frac{1}{2\pi}\right)^J \int \exp(-iz'x)\psi(z)\, dz,
$$

where $\psi(z) = \int \exp(iz'x) K(x)\, dx$. By Assumption 6, $\int |\psi(z)|\, dz < \infty$. Then, by Fourier inversion,

$$
\begin{aligned}
\text{(A4)}\qquad & R_{N(t)}(\theta) \\
&= \left[\frac{1}{2\pi}\right]^J \frac{1}{N(t) h^J} \sum_{n=1}^{N(t)} EV_{s'}\big(s, \epsilon^{(t-n)}, a, \theta^{*(t-n)}\big) \\
&\quad \times \int \exp\left(\frac{-iz'(\theta - \theta^{*(t-n)})}{h}\right) \psi(z)\, dz \\
&= \left[\frac{1}{2\pi}\right]^J \frac{1}{N(t)} \int \left[\sum_{n=1}^{N(t)} EV_{s'}\big(s, \epsilon^{(t-n)}, a, \theta^{*(t-n)}\big) \exp\big(iz'\theta^{*(t-n)}\big)\right] \\
&\quad \times \exp(-iz'\theta)\psi(hz)\, dz.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\text{(A5)}\qquad & E\Big[\sup_{\theta\in\Theta} \big| R_{N(t)}(\theta) - E\big[R_{N(t)}(\theta)\big]\big|\Big] \\
&\le \left[\frac{1}{2\pi}\right]^J \int E \left| \frac{1}{N(t)} \sum_{n=1}^{N(t)} \Big\{ EV_{s'}\big(s, \epsilon^{(t-n)}, a, \theta^{*(t-n)}\big) \exp\big(iz'\theta^{*(t-n)}\big) \right. \\
&\quad \left. - E\big[EV_{s'}\big(s, \epsilon^{(t-n)}, a, \theta^{*(t-n)}\big) \exp\big(iz'\theta^{*(t-n)}\big)\big]\Big\} \right| |\psi(hz)|\, dz.
\end{aligned}
$$

Using the Liapunov inequality and $\exp(ia) = \cos(a) + i\sin(a)$, we get

$$\begin{aligned}
&E\Bigg|\frac{1}{N(t)}\sum_{n=1}^{N(t)}\big\{EV_{s'}\big(s,\epsilon^{(t-n)},a,\theta^{*(t-n)}\big)\exp\big(iz'\theta^{*(t-n)}\big)\\
&\qquad - E\big[EV_{s'}\big(s,\epsilon^{(t-n)},a,\theta^{*(t-n)}\big)\exp\big(iz'\theta^{*(t-n)}\big)\big]\big\}\Bigg|\\
&\quad\le\Bigg\{\mathrm{Var}\Bigg[\frac{1}{N(t)}\sum_{n=1}^{N(t)}EV_{s'}\big(s,\epsilon^{(t-n)},a,\theta^{*(t-n)}\big)\cos\big(z'\theta^{*(t-n)}\big)\Bigg]\\
&\qquad+\mathrm{Var}\Bigg[\frac{1}{N(t)}\sum_{n=1}^{N(t)}EV_{s'}\big(s,\epsilon^{(t-n)},a,\theta^{*(t-n)}\big)\sin\big(z'\theta^{*(t-n)}\big)\Bigg]\Bigg\}^{1/2}.
\end{aligned}$$

Now, because $\epsilon^{(t-n)}, \epsilon^{(t-m)}$ $n \neq m$ are i.i.d.,

$$\begin{aligned}
&\mathrm{Cov}\big[EV_{s'}\big(s,\epsilon^{(t-n)},a,\theta^{*(t-n)}\big)\cos\big(z'\theta^{*(t-n)}\big),\\
&\quad EV_{s'}\big(s,\epsilon^{(t-m)},a,\theta^{*(t-m)}\big)\cos\big(z'\theta^{*(t-m)}\big)\big]=0.
\end{aligned}$$

Hence,

$$\begin{aligned}
&\mathrm{Var}\Bigg[\frac{1}{N(t)}\sum_{n=1}^{N(t)}EV_{s'}\big(s,\epsilon^{(t-n)},a,\theta^{*(t-n)}\big)\cos\big(z'\theta^{*(t-n)}\big)\Bigg]\\
&\quad=\frac{1}{N(t)^2}\sum_{n=1}^{N(t)}\sum_{m=1}^{N(t)}\mathrm{Cov}\big[EV_{s'}\big(s,\epsilon^{(t-n)},a,\theta^{*(t-n)}\big)\cos\big(z'\theta^{*(t-n)}\big),\\
&\qquad EV_{s'}\big(s,\epsilon^{(t-m)},a,\theta^{*(t-m)}\big)\cos\big(z'\theta^{*(t-m)}\big)\big]\\
&\quad=\frac{1}{N(t)^2}\sum_{n=1}^{N(t)}\mathrm{Var}\big[EV_{s'}\big(s,\epsilon^{(t-n)},a,\theta^{*(t-n)}\big)\cos\big(z'\theta^{*(t-n)}\big)\big].
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\mathrm{Var}\Bigg[\frac{1}{N(t)}\sum_{n=1}^{N(t)}EV_{s'}\big(s,\epsilon^{(t-n)},a,\theta^{*(t-n)}\big)\sin\big(z'\theta^{*(t-n)}\big)\Bigg]\\
&\quad=\frac{1}{N(t)^2}\sum_{n=1}^{N(t)}\mathrm{Var}\big[EV_{s'}\big(s,\epsilon^{(t-n)},a,\theta^{*(t-n)}\big)\sin\big(z'\theta^{*(t-n)}\big)\big].
\end{aligned}$$

Together, we derive that

$$\text{(A6)}\qquad E\Big[\sup_{\theta\in\Theta}\big|R_{N(t)}(\theta)-E\big[R_{N(t)}(\theta)\big]\big|\Big]$$
$$\leq\left[\frac{1}{2\pi}\right]^J\int\left\{\frac{1}{N(t)^2}\sum_{n=1}^{N(t)}\Big\{\mathrm{Var}\big[EV_{s'}\big(s,\epsilon^{(t-n)},a,\theta^{*(t-n)}\big)\cos\big(z'\theta^{*(t-n)}\big)\big]\right.$$
$$\left.+\mathrm{Var}\big[EV_{s'}\big(s,\epsilon^{(t-n)},a,\theta^{*(t-n)}\big)\sin\big(z'\theta^{*(t-n)}\big)\big]\Big\}\right\}^{1/2}|\psi(hz)|\,dz$$
$$\leq\left[\frac{1}{2\pi}\right]^J\left\{\frac{1}{N(t)^2}\sum_{n=1}^{N(t)}\sup_{\theta\in\Theta,\epsilon}|EV_{s'}(s,\epsilon,a,\theta)|^2\right\}^{1/2}\int|\psi(hz)|\,dz$$
$$=\left[\frac{1}{2\pi}\right]^J\left\{\frac{1}{N(t)h^{2J}}\sup_{\theta\in\Theta,\epsilon}|EV_{s'}(s,\epsilon,a,\theta)|^2\right\}^{1/2}\int|\psi(z)|\,dz.$$

Therefore, from Chebychev inequality, for $\alpha=3/2$,

$$\text{(A7)}\qquad \Pr\Big[\sup_{\theta\in\Theta}\big|R_{N(t)}(\theta)-E\big[R_{N(t)}(\theta)\big]\big|<h^{\alpha J}\Big]>1-\frac{B_2}{\sqrt{N(t)h^{2J(1+\alpha)}}},$$

where $B_2=[\frac{1}{2\pi}]^J\sup_{\theta\in\Theta,\epsilon}|V(s',\epsilon,\theta)|\int|\psi(z)|\,dz$. Thus, we have shown that the numerator of $A_1^{(t)}(\theta)\to 0$ in probability uniformly in $\theta\in\Theta$.

We next show that the denominator of $A_1^{(t)}(\theta)$ is bounded below by a positive number uniformly in Θ with probability arbitrarily close to 1 as $t\to\infty$. Let

$$R^{(t-n)}\equiv\varepsilon_0\frac{g(\theta^{*(t-n)})}{q(\theta^{(t-n)},\theta^{*(t-n)})}.$$

Then, from Lemma 1, $0\leq R^{(t-n)}\leq 1$ and $0<\varepsilon_0\leq 1$. Also, define a random variable

$$Y^{(t-n)}(\theta)=\begin{cases}K_h\big(\theta-\theta^{*(t-n)}(q)\big) & \text{with probability } R^{(t-n)},\\ 0 & \text{with probability } 1-R^{(t-n)}.\end{cases}$$

We denote $\theta^{*(t-n)}(q)$ to mean that $\theta^{*(t-n)}$ is generated from a distribution function q. Then $Y^{(t-n)}$ is a mixture of 0 and $K_h(\theta-\theta^{*(t-n)}(g))$, with the mixing probability being $1-\varepsilon_0$ and ε_0. That is,

$$Y^{(t-n)}(\theta)=\begin{cases}K_h\big(\theta-\theta^{*(t-n)}(g)\big) & \text{with probability } \varepsilon_0,\\ 0 & \text{with probability } 1-\varepsilon_0;\end{cases}$$

equivalently,

$$Y^{(t-n)}(\theta) = K_h\big(\theta - \theta^{*(t-n)}(g)\big)I^{(t-n)},$$

where

$$I^{(t-n)} = \begin{cases} 1 & \text{with probability } \varepsilon_0, \\ 0 & \text{with probability } 1 - \varepsilon_0. \end{cases}$$

From the construction of $Y^{(t-n)}$,

$$Y^{(t-n)}(\theta) \leq K_h\big(\theta - \theta^{*(t-n)}(q)\big).$$

Now, because $\theta^{*(t-n)}(g)$, $n = 1, \dots, N(t)$, are i.i.d., we can follow Bierens (1994, Sections 10.1 and 10.3) to work with the Fourier transform of the kernel:

$$\begin{aligned} &\frac{1}{N(t)} \sum_{n=1}^{N(t)} Y^{(t-n)}(\theta) \\ &\quad = \frac{1}{N(t)} \sum_{n=1}^{N(t)} I^{(t-n)} \left(\frac{1}{2\pi h}\right)^J \int \exp\left(\frac{-iz'(\theta - \theta^{*(t-n)}(g))}{h}\right) \psi(z)\, dz \\ &\quad = \left(\frac{1}{2\pi}\right)^J \int \left[\frac{1}{N(t)} \sum_{n=1}^{N(t)} I^{(t-n)} \exp\big(iz'\theta^{*(t-n)}(g)\big)\right] \\ &\qquad \times \exp(-iz'\theta)\psi(hz)\, dz. \end{aligned}$$

Hence, following (10.3.5) and (10.3.6) of Bierens (1994), we get

$$\begin{aligned} &E\left[\sup_\theta \left|\frac{1}{N(t)} \sum_{n=1}^{N(t)} Y^{(t-n)}(\theta) - E\left[\frac{1}{N(t)} \sum_{n=1}^{N(t)} Y^{(t-n)}(\theta)\right]\right|\right] \\ &\quad \leq \left(\frac{1}{2\pi}\right)^J \int E\left|\frac{1}{N(t)} \sum_{n=1}^{N(t)} \big\{I^{(t-n)} \exp\big(iz'\theta^{*(t-n)}(g)\big)\right. \\ &\qquad \left. - E\big[I^{(t-n)} \exp\big(iz'\theta^{*(t-n)}(g)\big)\big]\big\}\right| |\psi(hz)|\, dz \\ &\quad \leq \sqrt{\frac{E[I^2]}{N(t)}} \left(\frac{1}{2\pi}\right)^J \int |\psi(hz)|\, dz \\ &\quad = \sqrt{\frac{E[I^2]}{N(t)h^{2J}}} \left(\frac{1}{2\pi}\right)^J \int |\psi(z)|\, dz. \end{aligned}$$

Furthermore,

$$E\left[\frac{1}{N(t)}\sum_{n=1}^{N(t)} Y^{(t-n)}(\theta)\right] = \varepsilon_0 g(\theta).$$

From Chebychev inequality, for any $\kappa > 0$,

$$\Pr\left[\sup_{\theta\in\Theta}\left|\frac{1}{N(t)}\sum_{n=1}^{N(t)} Y^{(t-n)}(\theta) - \varepsilon_0 g(\theta)\right| < \kappa\right] > 1 - \frac{B_3}{\kappa\sqrt{N(t)}h^J}, \tag{A8}$$

where $B_3 = (\frac{1}{2\pi})^J\sqrt{EI^2}\int|\psi(z)|\,dz$. That is,

$$\Pr\left[\inf_{\theta\in\Theta}\frac{1}{N(t)}\sum_{n=1}^{N(t)} Y^{(t-n)}(\theta) + \kappa > \varepsilon_0 \inf_{\theta\in\Theta} g(\theta)\right] > 1 - \frac{B_3}{\kappa\sqrt{N(t)}h^J}.$$

Now, choose $\kappa = \frac{1}{2}\inf_{\theta\in\Theta}\varepsilon_0 g(\theta)$. Then

$$\Pr\left[\inf_{\theta\in\Theta}\frac{1}{N(t)}\sum_{n=1}^{N(t)} Y^{(t-n)}(\theta) > \frac{1}{2}\varepsilon_0 \inf_{\theta\in\Theta} g(\theta)\right] > 1 - \frac{B_3}{\sqrt{N(t)}h^J},$$

where

$$B_3 = \frac{2\sqrt{EI^2}\displaystyle\int|\psi(z)|\,dz}{(2\pi)^J\inf_{\theta\in\Theta}\varepsilon_0 g(\theta)}.$$

Since

$$\sum_{n=1}^{N(t)} K_h\big(\theta - \theta^{*(t-n)}(q)\big) \geq \sum_{n=1}^{N(t)} Y^{(t-n)},$$

we conclude that

$$\Pr\left[\inf_{\theta\in\Theta}\frac{1}{N(t)}\sum_{n=1}^{N(t)} K_h\big(\theta - \theta^{*(t-n)}(q)\big) > \frac{1}{2}\varepsilon_0 \inf_{\theta\in\Theta} g(\theta)\right] > 1 - \frac{B_3}{\sqrt{N(t)}h^J}. \tag{A9}$$

Now, we use (A3), (A7), and (A9) to derive the following inequality for $h < 1$:

$$\text{(A10)} \quad \Pr\left[\sup_{\theta\in\Theta}\left|\frac{A_1^{(t)}(\theta)}{\beta}\right| \leq \frac{(1+B_1)h^{\alpha J}}{\frac{1}{2}\varepsilon_0 \inf_{\theta\in\Theta} g(\theta)}\right]$$
$$\geq \Pr\left[\frac{\sup_{\theta\in\Theta}||E[X_{N(t)}(\theta)]| + |R_{N(t)}(\theta) - ER_{N(t)}(\theta)||}{\inf_{\theta\in\Theta}\left|\frac{1}{N(t)}\sum_{n=1}^{N(t)} K_h(\theta - \theta^{*(t-n)})\right|}\right.$$
$$\left.\leq \frac{h^{\alpha J} + B_1 h^{2J}}{\frac{1}{2}\varepsilon_0 \inf_{\theta\in\Theta} g(\theta)}\right]$$
$$\geq \Pr\left\{\left[\sup_{\theta\in\Theta}\left|E[X_{N(t)}(\theta)]\right| \leq B_1 h^{2J}\right]\right.$$
$$\cap \left[\sup_{\theta\in\Theta}\left|R_{N(t)}(\theta) - ER_{N(t)}(\theta)\right| \leq h^{\alpha J}\right]$$
$$\left.\cap \left[\inf_{\theta\in\Theta}\left|\frac{1}{N(t)}\sum_{n=1}^{N(t)} K_h\big(\theta - \theta^{*(t-n)}\big)\right| > \frac{1}{2}\varepsilon_0 \inf_{\theta\in\Theta} g(\theta)\right]\right\}$$
$$\geq 1 - \Pr\left[\sup_{\theta\in\Theta}\left|E\big[X_{N(t)}(\theta)\big]\right| > B_1 h^{2J}\right]$$
$$- \Pr\left[\sup_{\theta\in\Theta}\left|R_{N(t)}(\theta) - ER_{N(t)}(\theta)\right| > h^{\alpha J}\right]$$
$$- \Pr\left[\inf_{\theta\in\Theta}\left|\frac{1}{N(t)}\sum_{n=1}^{N(t)} K_h\big(\theta - \theta^{*(t-n)}\big)\right| \leq \frac{1}{2}\varepsilon_0 \inf_{\theta\in\Theta} g(\theta)\right]$$
$$\geq 1 - \frac{B_2}{\sqrt{N(t) h^{2J(1+\alpha)}}} - \frac{B_3}{\sqrt{N(t)} h^J}$$
$$\geq 1 - \frac{(B_2 + B_3)}{\sqrt{N(t) h^{2J(1+\alpha)}}}.$$

Letting $C_1 = 1 + B_1$ and $C_2 = B_2 + B_3$, Lemma 3 follows. *Q.E.D.*

Next, we proceed to prove $A_2^{(t)}(\theta) \xrightarrow{P} 0$ as $t \to \infty$ uniformly in Θ. Recall that

$$\text{(A11)} \quad \mathcal{V}(s,a,\epsilon,\theta) - \mathcal{V}^{(t)}(s,a,\epsilon,\theta) = A_1^{(t)}(\theta) + \beta\Bigg[\sum_{n=1}^{N(t)}\big[EV_{s'}\big(s,\epsilon^{(t-n)},a,\theta^{*(t-n)}\big) - EV_{s'}^{(t-n)}\big(s,\epsilon^{(t-n)},a,\theta^{*(t-n)}\big)\big]W_{N(t),h}\big(\theta,\theta^{*(t-n)}\big)\Bigg].$$

Notice that if $V(s,\epsilon,\theta) \ge V^{(t)}(s,\epsilon,\theta)$, then

$$\begin{aligned} 0 &\le V(s,\epsilon,\theta) - V^{(t)}(s,\epsilon,\theta) \\ &= \max_{a\in A}\mathcal{V}(s,a,\epsilon,\theta) - \max_{a\in A}\mathcal{V}^{(t)}(s,a,\epsilon,\theta) \\ &\le \max_{a\in A}\big[\mathcal{V}(s,a,\epsilon,\theta) - \mathcal{V}^{(t)}(s,a,\epsilon,\theta)\big] \\ &\le \max_{a\in A}\big|\mathcal{V}(s,a,\epsilon,\theta) - \mathcal{V}^{(t)}(s,a,\epsilon,\theta)\big|. \end{aligned}$$

Similarly, if $V(s,\epsilon,\theta) \le V^{(t)}(s,\epsilon,\theta)$, then

$$\begin{aligned} 0 &\le V^{(t)}(s,\epsilon,\theta) - V(s,\epsilon,\theta) \\ &= \max_{a\in A}\mathcal{V}^{(t)}(s,a,\epsilon,\theta) - \max_{a\in A}\mathcal{V}(s,a,\epsilon,\theta) \\ &\le \max_{a\in A}\big[\mathcal{V}^{(t)}(s,a,\epsilon,\theta) - \mathcal{V}(s,a,\epsilon,\theta)\big] \\ &\le \max_{a\in A}\big|\mathcal{V}(s,a,\epsilon,\theta) - \mathcal{V}^{(t)}(s,a,\epsilon,\theta)\big|. \end{aligned}$$

Hence, taking the supremum over s on the right-hand side of (A11) and then taking absolute values on both sides, we obtain

$$\text{(A11}'\text{)} \quad \begin{aligned} &\big|V(s,\epsilon,\theta) - V^{(t)}(s,\epsilon,\theta)\big| \\ &\le \max_{a\in A}\big|\mathcal{V}(s,a,\epsilon,\theta) - \mathcal{V}^{(t)}(s,a,\epsilon,\theta)\big| \\ &\le \sup_{s'\in S}\big|A_1^{(t)}(\theta)\big| + \beta\Bigg[\sum_{n=1}^{N(t)}\sup_{\widehat{s}\in S,a}\big|EV_{s'}\big(\widehat{s},\epsilon^{(t-n)},a,\theta^{*(t-n)}\big) - EV_{s'}^{(t-n)}\big(\widehat{s},\epsilon^{(t-n)},a,\theta^{*(t-n)}\big)\big|W_{N(t),h}\big(\theta,\theta^{*(t-n)}\big)\Bigg] \end{aligned}$$

$$\leq \sup_{s'\in S}\left|A_1^{(t)}(\theta)\right| + \beta\left[\sum_{n=1}^{N(t)} \sup_{\widehat{s}\in S}\left|V\left(\widehat{s}, \epsilon^{(t-n)}, \theta^{*(t-n)}\right)\right.\right.$$
$$\left.\left. - V^{(t-n)}\left(\widehat{s}, \epsilon^{(t-n)}, \theta^{*(t-n)}\right)\right| W_{N(t),h}\left(\theta, \theta^{*(t-n)}\right)\right].$$

Denote

$$\Delta V^{(t)}\left(\epsilon^{(t)}, \theta^{*(t)}\right) \equiv \sup_{\widehat{s}\in S}\left|V\left(\widehat{s}, \epsilon^{(t)}, \theta^{*(t)}\right) - V^{(t)}\left(\widehat{s}, \epsilon^{(t)}, \theta^{*(t)}\right)\right|.$$

Now, $|V(s,\epsilon,\theta) - V^{(t)}(s,\epsilon,\theta)|$ appears on the left-hand side (LHS) and $|V(\widehat{s}, \epsilon^{(t-n)}, \theta^{*(t-n)}) - V^{(t-n)}(\widehat{s}, \epsilon^{(t-n)}, \theta^{*(t-n)})|$ appears on the RHS of equation (A11′). Using this, we can recursively substitute away $|V(\widehat{s}, \epsilon^{(t-n)}, \theta^{*(t-n)}) - V^{(t-n)}(\widehat{s}, \epsilon^{(t-n)}, \theta^{*(t-n)})|$. This logic is used in the following lemma. Before we proceed with the lemma and its proof, we introduce some additional notation. For $\tau < t$, let

$$\widetilde{W}(t,\tau) \equiv \beta W_{N(t),h}(\theta, \theta'),$$

where θ is the parameter vector at iteration t and θ' the parameter vector at iteration τ. Now, for $\underline{N} \geq 1$ and for m such that $0 < m \leq \underline{N} + 1$, define

$$\Psi_m(t+\underline{N}, t, \tau) \equiv \{J_m = (t_m, t_{m-1}, \dots, t_1, t_0) :$$
$$t_m = t + \underline{N} > t_{m-1} > \cdots > t_2 > t_1 \geq t, t_0 = \tau\}.$$

That is, $\Psi_m(t+\underline{N}, t, \tau)$ is the resulting set of iterations where the largest is $t + N$ and the smallest is τ, and the other $m-1$ iterations are greater than or equal to t. Furthermore, let

$$\widehat{W}(t+\underline{N}, t, \tau) \equiv \sum_{m=1}^{\underline{N}+1}\left\{\sum_{\Psi_m(t+\underline{N},t,\tau)} \prod_{k=1}^{m} \widetilde{W}(t_k, t_{k-1})\right\}.$$

Notice that $\widehat{W}(t,t,\tau) \equiv \widetilde{W}(t,\tau)$.

LEMMA 4: *For any* $\underline{N} \geq 1$, $t > 0$,

$$\text{(A12)} \quad \Delta V^{(t+\underline{N})}(\epsilon, \theta) \leq \sup_{s'\in S}\left|A_1^{(t+\underline{N})}(\theta)\right|$$
$$+ \sum_{m=0}^{\underline{N}-1} \widehat{W}(t+\underline{N}, t+\underline{N}-m, t+\underline{N}-m-1)$$

$$\times \sup_{s' \in S} \left| A_1^{(t+\underline{N}-m-1)}\left(\theta^{*(t+\underline{N}-m-1)}\right)\right|$$

$$+ \sum_{n=1}^{N(t)} \Delta V^{(t-n)}\left(\epsilon^{(t-n)}, \theta^{*(t-n)}\right)\widehat{W}(t+\underline{N}, t, t-n).$$

Furthermore,

$$\text{(A13)} \qquad \sum_{n=1}^{N(t)} \widehat{W}(t+\underline{N}, t, t-n) \leq \beta.$$

PROOF: First, we show that inequalities (A12) and (A13) hold for $\underline{N} = 1$. For iteration $t+1$, we get

$$\begin{aligned}
\Delta V^{(t+1)}(\epsilon, \theta) &\leq \sup_{s' \in S} \left| A_1^{(t+1)}(\theta)\right| \\
&\quad + \sum_{n=1}^{N(t+1)} \Delta V^{(t+1-n)}\left(\epsilon^{(t+1-n)}, \theta^{*(t+1-n)}\right) \\
&\quad \times \widetilde{W}(t+1, t+1-n) \\
&\leq \sup_{s' \in S} \left| A_1^{(t+1)}(\theta)\right| + \Delta V^{(t)}\left(\epsilon^{(t)}, \theta^{*(t)}\right)\widetilde{W}(t+1, t) \\
&\quad + \sum_{n=1}^{N(t+1)-1} \Delta V^{(t-n)}\left(\epsilon^{(t-n)}, \theta^{*(t-n)}\right)\widetilde{W}(t+1, t-n).
\end{aligned}$$

Now, we substitute away $\Delta V(\epsilon^{(t)}, \theta^{*(t)})$ by using (A11′) and the fact that $N(t) \geq N(t+1) - 1$:

$$\begin{aligned}
\Delta V^{(t+1)}(\epsilon, \theta) &\leq \sup_{s' \in S} \left| A_1^{(t+1)}(\theta)\right| \\
&\quad + \sup_{s' \in S} \left| A_1^{(t)}\left(\theta^{*(t)}\right)\right| \widetilde{W}(t+1, t) \\
&\quad + \sum_{n=1}^{N(t)} \Delta V^{(t-n)}\left(\epsilon^{(t-n)}, \theta^{*(t-n)}\right) \\
&\quad \times \{\widetilde{W}(t+1, t)\widetilde{W}(t, t-n) + \widetilde{W}(t+1, t-n)\} \\
&= \sup_{s' \in S} \left| A_1^{(t+1)}(\theta)\right| + \sup_{s' \in S} \left| A_1^{(t)}\left(\theta^{*(t)}\right)\right| \widehat{W}(t+1, t+1, t) \\
&\quad + \sum_{n=1}^{N(t)} \Delta V^{(t-n)}\left(\epsilon^{(t-n)}, \theta^{*(t-n)}\right)\widehat{W}(t+1, t, t-n).
\end{aligned}$$

Hence, inequality (A12) holds for $\underline{N} = 1$.

We now prove (A13) for $\underline{N}=1$. Since

$$\sum_{n=1}^{N(t)} \widetilde{W}(t,t-n)/\beta = \sum_{n=1}^{N(t)} W_{N(t),h}\big(\theta^{*(t)}, \theta^{*(t-n)}\big) = 1,$$

then

$$\begin{aligned}
&\sum_{n=1}^{N(t)} \widehat{W}(t+1,t,t-n) \\
&\quad = \sum_{n=1}^{N(t)} \widetilde{W}(t+1,t)\widetilde{W}(t,t-n) + \sum_{n=1}^{N(t)} \widetilde{W}(t+1,t-n) \\
&\quad = \widetilde{W}(t+1,t)\sum_{n=1}^{N(t)} \widetilde{W}(t,t-n) + \sum_{n=1}^{N(t)} \widetilde{W}(t+1,t-n) \\
&\quad = \beta\widetilde{W}(t+1,t) + \sum_{n=1}^{N(t)} \widetilde{W}(t+1,t-n) \\
&\quad \le \sum_{n=1}^{N(t)+1} \widetilde{W}(t+1,t+1-n).
\end{aligned}$$

Since $\widetilde{W}(t+1,t+1-n)=0$ for any $n > N(t+1)$,

$$\begin{aligned}
\sum_{n=1}^{N(t)+1} \widetilde{W}(t+1,t+1-n) &= \sum_{n=1}^{N(t+1)} \widetilde{W}(t+1,t+1-n) \\
&= \beta \sum_{n=1}^{N(t+1)} W_{N(t+1),h}\big(\theta^{*(t+1)}, \theta^{*(t+1-n)}\big) = \beta.
\end{aligned}$$

Thus,

$$\sum_{n=1}^{N(t)} \widehat{W}(t+1,t,t-n) \le \beta.$$

Hence, inequality (A13) holds for $\underline{N}=1$.

Next, suppose that inequality (A12) holds for $\underline{N}=M$. Then, using $t+1$ instead of t in inequality (A12), we get

$$\begin{aligned}
\Delta V^{(t+1+M)}(\epsilon,\theta) \le \sup_{s'\in S}\big|A_1^{(t+1+M)}(\theta)\big| \\
+ \sum_{m=0}^{M-1} \widehat{W}(t+1+M, t+1+M-m, t+M-m)
\end{aligned}$$

$$\times \sup_{s' \in S} \left| A_1^{(t+M-m)}\left(\theta^{*(t+M-m)}\right)\right|$$
$$+ \Delta V^{(t)}\left(\epsilon^{(t)}, \theta^{*(t)}\right) \widehat{W}(t+1+M, t+1, t)$$
$$+ \sum_{n=2}^{N(t+1)} \Delta V^{(t+1-n)}\left(\epsilon^{(t+1-n)}, \theta^{*(t+1-n)}\right)$$
$$\times \widehat{W}(t+1+M, t+1, t+1-n).$$

Now, using (A11′) to substitute away $\Delta V^{(t)}(\epsilon^{(t)}, \theta^{*(t)})$, we get

$$\Delta V^{(t+1+M)}(\epsilon, \theta) \leq \sup_{s' \in S} \left| A_1^{(t+M+1)}(\theta)\right| \tag{A14}$$
$$+ \sum_{m=0}^{M} \widehat{W}(t+M+1, t+M+1-m, t+M-m)$$
$$\times \sup_{s' \in S} \left| A_1^{(t+M-m)}\left(\theta^{*(t+M-m)}\right)\right|$$
$$+ \sum_{n=1}^{N(t)} \Delta V^{(t-n)}\left(\epsilon^{(t-n)}, \theta^{*(t-n)}\right)$$
$$\times [\widehat{W}(t+M+1, t+1, t)\widetilde{W}(t, t-n)$$
$$+ \widehat{W}(t+M+1, t+1, t-n)].$$

CLAIM: *Now, we claim that, for any* $M \geq 1$,

$$\widehat{W}(t+M, t+1, t)\widetilde{W}(t, t-n) + \widehat{W}(t+M, t+1, t-n) = \widehat{W}(t+M, t, t-n). \tag{A15}$$

PROOF: Let

$$\Psi_{m,1}(t+M, t, \tau) \equiv \{J_m = (t_m, t_{m-1}, \dots, t_1, t_0): t_m = t+M > t_{m-1} > \cdots > t_2 \geq t+1, t_1 = t, t_0 = \tau\}.$$

Notice that

$$\Psi_m(t+M, t+1, \tau) \equiv \{J_m = (t_m, t_{m-1}, \dots, t_1, t_0): t_m = t+M > t_{m-1} > \cdots > t_2 > t_1 \geq t+1, t_0 = \tau\}.$$

Then

$$\Psi_m(t+M,t,\tau)=\Psi_{m,1}(t+M,t,\tau)\cup\Psi_m(t+M,t+1,\tau)$$

and

$$\Psi_{m,1}(t+M,t,\tau)\cap\Psi_m(t+M,t+1,\tau)=\emptyset.$$

Also,

$$\Psi_{M+1}(t+M,t+1,\tau)=\emptyset.$$

Therefore,

$$\begin{aligned}
&\widehat{W}(t+M,t,\tau)\\
&\quad\equiv\sum_{m=1}^{M+1}\left\{\sum_{\Psi_m(t+M,t,\tau)}\prod_{k=1}^{m}\widetilde{W}(t_k,t_{k-1})\right\}\\
&\quad=\sum_{m=1}^{M+1}\left\{\sum_{\Psi_{m,1}(t+M,t,\tau)}\prod_{k=1}^{m}\widetilde{W}(t_k,t_{k-1})\right\}\\
&\qquad+\sum_{m=1}^{M+1}\left\{\sum_{\Psi_m(t+M,t+1,\tau)}\prod_{k=1}^{m}\widetilde{W}(t_k,t_{k-1})\right\}\\
&\quad=\sum_{m=2}^{M+1}\left\{\sum_{\Psi_{m-1}(t+M,t+1,t)}\prod_{k=1}^{m-1}\widetilde{W}(t_k,t_{k-1})\right\}\widetilde{W}(t,\tau)\\
&\qquad+\sum_{m=1}^{M}\left\{\sum_{\Psi_m(t+M,t+1,\tau)}\prod_{k=1}^{m}\widetilde{W}(t_k,t_{k-1})\right\}\\
&\quad=\sum_{m=1}^{M}\left\{\sum_{\Psi_m(t+M,t+1,t)}\prod_{k=1}^{m}\widetilde{W}(t_k,t_{k-1})\right\}\widetilde{W}(t,\tau)\\
&\qquad+\sum_{m=1}^{M}\left\{\sum_{\Psi_m(t+M,t+1,\tau)}\prod_{k=1}^{m}\widetilde{W}(t_k,t_{k-1})\right\}\\
&\quad=\widehat{W}(t+M,t+1,t)\widetilde{W}(t,\tau)+\widehat{W}(t+M,t+1,\tau).
\end{aligned}$$

Hence, (A15) holds and the claim is proved. *Q.E.D.*

Substituting this into equation (A14) with $M+1$ instead of M yields the first part of Lemma 4 (that is, (A12)) by induction.

Next, suppose that (A13) holds for $\underline{N} = M$. That is,

$$\sum_{n=1}^{N(t)} \widehat{W}(t+M, t, t-n) \leq \beta.$$

Then, denoting $t' = t+1$ and using (A15) with $M+1$ instead of M, we get

$$\begin{aligned}
&\sum_{n=1}^{N(t)} \widehat{W}(t+M+1, t, t-n) \\
&\quad = \sum_{n=1}^{N(t)} \widehat{W}(t+M+1, t+1, t)\widetilde{W}(t, t-n) \\
&\qquad + \sum_{n=1}^{N(t)} \widehat{W}(t+M+1, t+1, t-n) \\
&\quad \leq \widehat{W}(t'+M, t', t) + \sum_{n=1}^{N(t)} \widehat{W}(t'+M, t', t-n) \\
&\quad = \sum_{n=1}^{N(t')} \widehat{W}(t'+M, t', t'-n) \leq \beta.
\end{aligned}$$

Hence, induction holds and for any $\underline{N} > 0$,

$$\sum_{n=1}^{N(t)} \widehat{W}(t+\underline{N}, t, t-n) \leq \beta,$$

which proves the second part of the lemma (that is, (A13)). Thus, Lemma 4 is proved. *Q.E.D.*

We sometimes suppress the arguments of A_1, for convenience. Now, for any $m = 1, \ldots, \widetilde{N}(l)$, if we substitute $t(l) - m$ for $t + \underline{N}$ and $t(l-1)$ for t, then equation (A12) becomes

$$\begin{aligned}
\Delta V^{(t(l)-m)}(\epsilon, \theta) \leq{}& \sup_{s' \in S} \left| A_1^{(t(l)-m)}(\theta) \right| \\
&+ \sum_{i=0}^{\widetilde{N}(l)-m} \widehat{W}(t(l)-m, t(l)-m-i, t(l)-m-i-1) \\
&\times \sup_{s' \in S} \left| A_1^{(t(l)-m-i-1)} \right|
\end{aligned}$$

$$+\sum_{n=1}^{\widetilde{N}(l-1)} \Delta V^{(t(l-1)-n)}\big(\epsilon^{(t(l-1)-n)}, \theta^{*(t(l-1)-n)}\big) \times \widehat{W}(t(l)-m, t(l-1), t(l-1)-n).$$

Now, we take a weighted sum of $\Delta V^{(t(l)-m)}(\epsilon, \theta^{*(t(l)-m)})$, $m = 1, \ldots, \widetilde{N}(l)$, where the weights are $W_{\widetilde{N}(l),h}(\theta, \theta^{*(t(l)-m)})$. Then

$$\begin{aligned}
&\sum_{m=1}^{\widetilde{N}(l)} \Delta V^{(t(l)-m)}\big(\epsilon^{(t(l)-m)}, \theta^{*(t(l)-m)}\big) W_{\widetilde{N}(l),h}\big(\theta, \theta^{*(t(l)-m)}\big) \\
&\quad \le \sum_{m=1}^{\widetilde{N}(l)} \Bigg\{ \sup_{s' \in S} \big| A_1^{(t(l)-m)} \big| \\
&\qquad + \sum_{i=0}^{\widetilde{N}(l)-m} \widehat{W}(t(l)-m, t(l)-m-i, t(l)-m-i-1) \\
&\qquad \times \sup_{s' \in S} \big| A_1^{(t(l)-m-i-1)} \big| \Bigg\} W_{\widetilde{N}(l),h}\big(\theta, \theta^{*(t(l)-m)}\big) \\
&\qquad + \sum_{m=1}^{\widetilde{N}(l)} \sum_{n=1}^{\widetilde{N}(l-1)} \Delta V^{(t(l-1)-n)}\big(\epsilon^{(t(l-1)-n)}, \theta^{*(t(l-1)-n)}\big) \\
&\qquad \times \widehat{W}(t(l)-m, t(l-1), t(l-1)-n) W_{\widetilde{N}(l),h}\big(\theta, \theta^{*(t(l)-m)}\big).
\end{aligned} \tag{A16}$$

Now, let $A(l, l) \equiv B_1(l, l) + B_2(l, l)$, where

$$B_1(l, l) = \sum_{m=1}^{\widetilde{N}(l)} \sup_{s' \in S} \big| A_1^{(t(l)-m)} \big| W_{\widetilde{N}(l),h}\big(\theta, \theta^{*(t(l)-m)}\big),$$

$$\begin{aligned}
B_2(l, l) \equiv &\sum_{m=1}^{\widetilde{N}(l)} W_{\widetilde{N}(l),h}\big(\theta, \theta^{*(t(l)-m)}\big) \\
&\times \sum_{i=0}^{\widetilde{N}(l)-m} \Big\{ \widehat{W}(t(l)-m, t(l)-m-i, t(l)-m-i-1) \\
&\times \sup_{s' \in S} \big| A_1^{(t(l)-m-i-1)} \big| \Big\},
\end{aligned}$$

and denote $h_l = h(t(l))$ as the bandwidth at iteration $t(l)$.

LEMMA 5: *$A(l,l) \xrightarrow{P} 0$ as $l \to \infty$ uniformly in Θ.*

PROOF: We first show that $B_1(l,l) \xrightarrow{P} 0$ uniformly in Θ. Now

$$B_1(l,l) = \frac{\sum_{m=1}^{\widetilde{N}(l)} \sup_{s' \in S} |A_1^{(t(l)-m)}| \frac{1}{\widetilde{N}(l)} K_{h(\widetilde{N}(l))}(\theta, \theta^{*(t(l)-m)})}{\frac{1}{\widetilde{N}(l)} \sum_{k=1}^{\widetilde{N}(l)} K_{h(\widetilde{N}(l))}(\theta, \theta^{*(t(l)-k)})}. \tag{A17}$$

Because $\int EV_{s'}(s,\epsilon',a,\theta)\,dF_{\epsilon'}(\epsilon',\theta)$ and $EV_{s'}(s,\epsilon^{(t-n)},a,\theta^{*(t-n)})$ are uniformly bounded in $a, \epsilon, \epsilon^{(t-n)}$, and $\theta,\ \theta^{*(t-n)} \in \Theta$, $A_1^{(t)}$ is uniformly bounded. Hence, there exists $\overline{A} > 0$ such that $A_1^{(t)} \le \overline{A}$ for any t. By Lemma 3,

$$\begin{aligned} &E\Big[\sup_{a,s' \in S,\theta' \in \Theta} \big|A_1^{(t)}(\theta')\big|\Big] \\ &\quad \le C_1 h^{(3/2)J} \Pr\Big[\sup_{a,s' \in S,\theta' \in \Theta} \big|A_1^{(t)}(\theta')\big| < C_1 h^{(3/2)J}\Big] \\ &\qquad + \overline{A} \Pr\Big[\sup_{a,s' \in S,\theta' \in \Theta} \big|A_1^{(t)}(\theta')\big| \ge C_1 h^{(3/2)J}\Big] \\ &\quad \le C_1 h^{(3/2)J} + \frac{\overline{A} C_2}{\sqrt{N(t) h^{5J}}} \end{aligned} \tag{A18}$$

and

$$\begin{aligned} &E\Big[\sup_{s' \in S,\theta' \in \Theta} \big|A_1^{(t)}(\theta')\big|^2\Big] \\ &\quad \le C_1^2 h^{3J} \Pr\Big[\sup_{s' \in S,\theta' \in \Theta} \big|A_1^{(t)}(\theta')\big| < C_1 h^{(3/2)J}\Big] \\ &\qquad + \overline{A}^2 \Pr\Big[\sup_{s' \in S,\theta' \in \Theta} \big|A_1^{(t)}(\theta')\big| \ge C_1 h^{(3/2)J}\Big] \\ &\quad \le C_1^2 h^{3J} + \frac{\overline{A}^2 C_2}{\sqrt{N(t) h^{5J}}}. \end{aligned} \tag{A19}$$

Now, using the Fourier transform method of Bierens (1994), we let

$$Y_m \equiv \sup_{s' \in S,\theta' \in \Theta} \big|A_1^{(t(l)-m)}(\theta')\big|.$$

Then

$$
\text{(A20)}\quad E\left[\sup_{\theta\in\Theta}\frac{1}{\widetilde{N}(l)}\sum_{m=1}^{\widetilde{N}(l)}\sup_{s'\in S,\theta'\in\Theta}\left|A_1^{(t(l)-m)}(\theta')\right|K_{h(\widetilde{N}(l))}\big(\theta,\theta^{*(t-m)}(q)\big)\right]
$$

$$
\begin{aligned}
&\le \varepsilon_1 E\left[\sup_{\theta\in\Theta}\frac{1}{\widetilde{N}(l)}\sum_{m=1}^{\widetilde{N}(l)}Y_m K_{h(\widetilde{N}(t))}\big(\theta,\theta^{*(t-m)}(\widetilde{g})\big)\right]\\
&\le \varepsilon_1\left(\frac{1}{2\pi}\right)^J\left\{E\left[\frac{1}{\widetilde{N}(l)}\sum_{m=1}^{\widetilde{N}(l)}Y_m\cos\big(\theta^{*(t-m)}(\widetilde{g})\big)\right]^2\right.\\
&\qquad\left.+E\left[\frac{1}{\widetilde{N}(l)}\sum_{m=1}^{\widetilde{N}(l)}Y_m\sin\big(\theta^{*(t-m)}(\widetilde{g})\big)\right]^2\right\}^{1/2}\int|\psi(hz)|\,dz\\
&\le \varepsilon_1\left(\frac{1}{2\pi}\right)^J\left\{2E\left[\frac{1}{\widetilde{N}(l)}\sum_{m=1}^{\widetilde{N}(l)}Y_m\right]^2\right\}^{1/2}\int|\psi(hz)|\,dz.
\end{aligned}
$$

The last inequality holds because $Y_m > 0$. Now

$$
\begin{aligned}
\left[\sum_{m=1}^{\widetilde{N}(l)}Y_m\right]^2 &= \sum_{m=1}^{\widetilde{N}(l)}Y_m^2+2\sum_{m>n}Y_mY_n\le\sum_{m=1}^{\widetilde{N}(l)}Y_m^2+\sum_{m>n}[Y_m^2+Y_n^2]\\
&\le 2\widetilde{N}(l)\sum_{m=1}^{\widetilde{N}(l)}Y_m^2.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\text{(A21)}\quad &\text{RHS of (A20)}\\
&\le \varepsilon_1\left(\frac{1}{2\pi}\right)^J\left\{2E\left[\frac{2\widetilde{N}(l)}{\widetilde{N}(l)^2}\sum_{m=1}^{\widetilde{N}(l)}Y_m^2\right]\right\}^{1/2}\int|\psi(h_lz)|\,dz\\
&\le 2\varepsilon_1\frac{1}{h_l^J}\sqrt{C_1^2h_{l-1}^{3J}+\frac{\overline{A}^2C_2}{\sqrt{N(l-1)h_l^{5J}}}}\left(\frac{1}{2\pi}\right)^J\int|\psi(z)|\,dz\\
&= 2\varepsilon_1\sqrt{C_1^2h_{l-1}^{J}+\frac{\overline{A}^2C_2}{\sqrt{N(l-1)h_l^{9J}}}}\left(\frac{1}{2\pi}\right)^J\int|\psi(z)|\,dz\\
&\to 0\quad\text{as } l\to\infty.
\end{aligned}
$$

Assumption 7 ($h(\widetilde{N}(k+1))^{9J}\widetilde{N}(k+1) \to \infty$ as $k \to \infty$) and

$$\frac{\widetilde{N}(k)}{\widetilde{N}(k)+1} \to 1$$

as $k \to \infty$ together imply $\widetilde{N}(l-1)h_l^{9J} \to \infty$ as $l \to \infty$ Therefore, the numerator of $B_1(l,l)$ in (A17) converges to zero in probability uniformly in Θ. Furthermore, from (A9), we know that the probability of the denominator of (A17) being larger than $\frac{1}{2}\varepsilon_0 \inf_{\theta\in\Theta} g(\theta)$ uniformly in Θ can be made arbitrarily close to 1 by making t large enough. Therefore, $B_1(l,l) \overset{P}{\to} 0$ as $l \to \infty$ uniformly in Θ.

We next show that $B_2(l,l) \overset{P}{\to} 0$ uniformly as $t \to \infty$.

For any $t' > t > 0$, let, $\widetilde{K}(t',t) \equiv K_h(\theta^{*(t')} - \theta^{*(t)})$. For $\tau_1 > \tau_2 > \tau$ and for l such that $t(l-1) < \tau_1 \leq t(l)$, define $W^*(\tau_1, \tau_2, \tau, j)$ recursively to be

$$W^*(\tau_1, \tau_2, \tau, 1) \equiv \widetilde{W}(\tau_1, \tau),$$

$$W^*(\tau_1, \tau_2, \tau, 2) \equiv \sum_{j=1}^{\tau_1-\tau_2} \widetilde{W}(\tau_1, \tau_1 - j)W^*(\tau_1 - j, \tau_2, \tau, 1),$$

$$\vdots$$

$$W^*(\tau_1, \tau_2, \tau, k) \equiv \sum_{j=1}^{\tau_1-\tau_2-(k-2)} \widetilde{W}(\tau_1, \tau_1 - j)W^*(\tau_1 - j, \tau_2, \tau, k-1).$$

Notice that for $\tau < \tau_2 - N(\tau_2)$, $W^*(\tau_1, \tau_2, \tau, k) = 0$ for all k. Similarly,

$$K^*(\tau_1, \tau_2, \tau, 1) \equiv \frac{1}{\widetilde{N}(l)}\widetilde{K}(\tau_1, \tau),$$

$$K^*(\tau_1, \tau_2, \tau, 2) \equiv \sum_{j=1}^{\tau_1-\tau_2} \frac{1}{\widetilde{N}(l)}\widetilde{K}(\tau_1, \tau_1 - j)K^*(\tau_1 - j, \tau_2, \tau, 1),$$

$$\vdots$$

$$K^*(\tau_1, \tau_2, \tau, k) \equiv \sum_{j=1}^{\tau_1-\tau_2-(k-2)} \frac{1}{\widetilde{N}(l)}\widetilde{K}(\tau_1, \tau_1 - j)K^*(\tau_1 - j, \tau_2, \tau, k-1),$$

and, for $\tau < \tau_2 - N(\tau_2)$,

$$K^*(\tau_1, \tau_2, \tau, k) = 0.$$

Then, for any $\tau_1 > \tau_2 > \tau$,

$$\text{(A22)}\quad \widehat{W}(\tau_1, \tau_2, \tau) \equiv \sum_{m=1}^{\widetilde{N}(l)+1} \left\{ \sum_{\Psi_m(\tau_1,\tau_2,\tau)} \prod_{k=1}^{m} \widetilde{W}(t_k, t_{k-1}) \right\} = \sum_{k=1}^{\tau_1-\tau_2+1} W^*(\tau_1, \tau_2, \tau, k).$$

Hence,

$$\begin{aligned}
&\sum_{i=0}^{\widetilde{N}(l)-m} \left\{ \widehat{W}(t(l)-m, t(l)-m-i, t(l)-m-i-1) \sup_{s'\in S} \left| A_1^{(t(l)-m-i-1)} \right| \right\} \\
&= \sum_{i=0}^{\widetilde{N}(l)-m} \left\{ \sum_{k=1}^{i+1} W^*(t(l)-m, t(l)-m-i, t(l)-m-i-1, k) \right. \\
&\qquad \left. \times \sup_{s'\in S} \left| A_1^{(t(l)-m-i-1)} \right| \right\} \\
&= \sum_{k=1}^{\widetilde{N}(l)-m} \left\{ \sum_{i=k-1}^{\widetilde{N}(l)-m} W^*(t(l)-m, t(l)-m-i, t(l)-m-i-1, k) \right. \\
&\qquad \left. \times \sup_{s'\in S} \left| A_1^{(t(l)-m-i-1)} \right| \right\}.
\end{aligned}$$

Also, notice that, for any $\widetilde{t}$ such that $t(l-1) < \widetilde{t} \leq t(l)$,

$$\begin{aligned}
\text{(A23)}\quad & W^*(\widetilde{t}, \widetilde{t}-i, \widetilde{t}-i-1, k) \\
&= \sum_{\Psi_k(\widetilde{t},\widetilde{t}-i,\widetilde{t}-i-1)} \prod_{j=1}^{k} \widetilde{W}(t_j, t_{j-1}) \\
&= \sum_{\Psi_k(\widetilde{t},\widetilde{t}-i,\widetilde{t}-i-1)} \prod_{j=1}^{k} \beta \frac{\widetilde{K}(t_j, t_{j-1})}{\sum_{i=1}^{N(t_j)} \widetilde{K}(t_j, t_{j-i})}
\end{aligned}$$

$$\le \beta^k \left[\inf_{t(l-1)<t\le t(l)} \sum_{i=1}^{N(t)} \widetilde{K}(t, t-i) \right]^{-k} \sum_{\Psi_k(\widetilde{t}, \widetilde{t}-i, \widetilde{t}-i-1)} \prod_{j=1}^{k} \widetilde{K}(t_j, t_{j-1})$$

$$= \beta^k \left[\frac{1}{\widetilde{N}(l)} \inf_{t(l-1)<t\le t(l)} \sum_{i=1}^{N(t)} \widetilde{K}(t, t-i) \right]^{-k} \sum_{\Psi_k(\widetilde{t}, \widetilde{t}-i, \widetilde{t}-i-1)} \prod_{j=1}^{k} \frac{\widetilde{K}(t_j, t_{j-1})}{\widetilde{N}(l)}$$

$$= \beta^k \left[\frac{1}{\widetilde{N}(l)} \inf_{t(l-1)<t\le t(l)} \sum_{i=1}^{N(t)} \widetilde{K}(t, t-i) \right]^{-k} K^*(\widetilde{t}, \widetilde{t}-i, \widetilde{t}-i-1, k).$$

Hence, for any $0 < \delta < 1$ we get

$$\text{(A24)} \quad \Pr\Bigg[\sum_{k=1}^{\widetilde{N}(l)} \sum_{m=1}^{\widetilde{N}(l)} W_{\widetilde{N}(l),h}\big(\theta, \theta^{*(t(l)-m)}\big) \times \sum_{i=k-1}^{\widetilde{N}(l)-m} W^*(t(l)-m, t(l)-m-i, t(l)-m-i-1, k) \times \sup_{s'\in S} \big| A_1^{(t(l)-m-i-1)} \big| \ge \frac{\delta - \delta^{\widetilde{N}(l)+1}}{1-\delta} \Bigg]$$

$$\le \Pr\Bigg[\bigcup_{k=1}^{\widetilde{N}(l)} \Bigg\{ \sum_{m=1}^{\widetilde{N}(l)} W_{\widetilde{N}(l),h}\big(\theta, \theta^{*(t(l)-m)}\big) \times \sum_{i=k-1}^{\widetilde{N}(l)-m} W^*(t(l)-m, t(l)-m-i, t(l)-m-i-1, k) \times \sup_{s'\in S} \big| A_1^{(t(l)-m-i-1)} \big| \ge \delta^k \Bigg\} \Bigg]$$

$$\le \sum_{k=1}^{\widetilde{N}(l)} \Pr\Bigg[\sum_{m=1}^{\widetilde{N}(l)} W_{\widetilde{N}(l),h}\big(\theta, \theta^{*(t(l)-m)}\big) \times \sum_{i=k-1}^{\widetilde{N}(l)-m} W^*(t(l)-m, t(l)-m-i, t(l)-m-i-1, k)$$

$$
\begin{aligned}
&\times \sup_{s' \in S} \left| A_1^{(t(l)-m-i-1)} \right| \geq \delta^k \Bigg] \\
&\leq \sum_{k=1}^{\widetilde{N}(l)} \Pr \Bigg\{ \Bigg[\sum_{m=1}^{\widetilde{N}(l)} K_{h(\widetilde{N}(l))}\big(\theta, \theta^{*(t(l)-m)}\big) \\
&\times \sum_{i=k-1}^{\widetilde{N}(l)-m} K^*(t(l)-m, t(l)-m-i, t(l)-m-i-1, k) \\
&\times \sup_{s' \in S} \left| A_1^{(t(l)-m-i-1)} \right| \geq \left[\frac{\delta}{4A\beta} \varepsilon_0 \inf_\theta g(\theta) \right]^k \Bigg] \\
&\cup \left[\inf_{t(l-1)<t\leq t(l)} \left[\frac{1}{\widetilde{N}(l)} \sum_{i=1}^{N(t)} \widetilde{K}(t, t-i) \right] < \frac{1}{4A} \varepsilon_0 \inf_\theta g(\theta) \right] \Bigg\} \\
&\leq \sum_{k=1}^{\widetilde{N}(l)} \Pr \Bigg[\sum_{m=1}^{\widetilde{N}(l)} K_{h(\widetilde{N}(l))}\big(\theta, \theta^{*(t(l)-m)}\big) \\
&\times \sum_{i=k-1}^{\widetilde{N}(l)-m} K^*(t(l)-m, t(l)-m-i, t(l)-m-i-1, k) \\
&\times \sup_{s' \in S} \left| A_1^{(t(l)-m-i-1)} \right| \geq \left[\frac{\delta}{4A\beta} \varepsilon_0 \inf_\theta g(\theta) \right]^{k+1} \Bigg] \\
&+ \Pr \left[\inf_{t(l-1)<t\leq t(l)} \left[\frac{1}{\widetilde{N}(l)} \sum_{i=1}^{N(t)} \widetilde{K}(t, t-i) \right] < \frac{1}{4A} \varepsilon_0 \inf_\theta g(\theta) \right].
\end{aligned}
$$

First, we consider the first term of the RHS of equation (A24). We prove the following claim.

CLAIM 1: *The following inequalities hold*:

$$
\begin{aligned}
\text{(A25)} \quad & E\Bigg\{ \sum_{i=k-1}^{\widetilde{N}(l)-m} K^*(t(l)-m, t(l)-m-i, t(l)-m-i-1, k) \\
&\times \sup_{s' \in S, \theta' \in \Theta} \left| A_1^{(t(l)-m-i-1)} \right| \Bigg\} \\
&\leq \frac{\varepsilon_1^{k+1}}{(k-1)!} \Big\{ \sup_{\theta' \in \Theta} E_\theta \big[K_{h_l}(\theta' - \theta(\widetilde{g})) \big] \Big\}^k E\Big[\sup_{s' \in S, \theta \in \Theta} \left| A_1^{(t(l)-m-i-1)} \right| \Big],
\end{aligned}
$$

$$E\left\{\sum_{i=k-1}^{\widetilde{N}(l)-m} K^*(t(l)-m, t(l)-m-i, t(l)-m-i-1, k) \times \sup_{s'\in S,\theta'\in\Theta}\left|A_1^{(t(l)-m-i-1)}\right|\right\}^2 \leq 2\left(\frac{\varepsilon_1^{k+1}}{(k-1)!}\right)^2 \left\{\sup_{\theta'\in\Theta} E_\theta\left[K_{h_l}(\theta'-\theta^*(\widetilde{g}))\right]\right\}^{2k} \times E\left[\sup_{s'\in S,\theta\in\Theta}\left|A_1^{(t(l)-m-i-1)}\right|^2\right]. \tag{A26}$$

PROOF: First, by definition of K^*, note that

$$\begin{aligned} &E\left\{\sum_{i=k-1}^{\widetilde{N}(l)-m} K^*(t(l)-m, t(l)-m-i, t(l)-m-i-1, k) \times \sup_{s'\in S,\theta'\in\Theta}\left|A_1^{(t(l)-m-i-1)}\right|\right\} \\ &= \frac{1}{\widetilde{N}(l)^k}\sum_{i=k-1}^{\widetilde{N}(l)-m}\sum_{j_1,\dots,j_{k-1}} I(j_0 = t(l)-m-i-1, \\ &\quad t(l)-m-i \leq j_1 < j_2 < \cdots < j_k = t(l)-m) \\ &\quad \times E\left[\left\{\prod_{s=0}^{k-1}\left[K_h\left(\theta^{*(j_{s+1})}-\theta^{*j_s}\right)\right]\right\}\sup_{s'\in S,\theta\in\Theta}\left|A_1^{(t(l)-m-i-1)}\right|\right]. \end{aligned} \tag{A27}$$

For $k \geq 1$, let $(j_0, j_1, \dots, j_k)$ satisfy $t(l)-m-i-1 = j_0 < j_1 < j_2 < \cdots < j_{k-1} < j_k = t(l)-m$. Denote the conditional transition probability from $\theta^{*(t)}$ to $\theta^{*(t+1)}$ given $\mathcal{H}^{(t)}$ as $p^*(\theta^{*(t)}, \theta^{*(t+1)}|\mathcal{H}^{(t)})$ or, in shorthand, $p^{*(t+1)}$, and denote $f(\theta^{(t+1)}|\theta^{*(t)}, \mathcal{H}^{(t)})$ to be the conditional distribution of $\theta^{(t+1)}$ given $\theta^{*(t)}, \mathcal{H}^{(t)}$. Then

$$\begin{aligned} p^*\left(\theta^{*(t)}, \theta^{*(t+1)}\middle|\mathcal{H}^{(t)}\right) &= \int f\left(\theta^{(t+1)}\middle|\theta^{*(t)}, \mathcal{H}^{(t)}\right) q\left(\theta^{(t+1)}, \theta^{*(t+1)}\right) d\theta^{(t+1)} \\ &\leq \varepsilon_1 \int f\left(\theta^{(t+1)}\middle|\theta^{*(t)}, \mathcal{H}^{(t)}\right)\widetilde{g}\left(\theta^{*(t+1)}\right) d\theta^{(t+1)} \\ &= \varepsilon_1 \widetilde{g}\left(\theta^{*(t+1)}\right). \end{aligned}$$

Therefore,

$$\text{(A28)}\qquad \left\{\prod_{s=2}^{t(l)-m} p^*\big(\theta^{*(s-1)}, \theta^{*(s)}\big|\mathcal{H}^{(s-1)}\big)\right\} \leq \left[\prod_{m=0}^{k-1} \varepsilon_1 \widetilde{g}\big(\theta^{*(j_m)}\big)\right] \times \left\{\prod_{s=2}^{t(l)-m} \big[p^*\big(\theta^{*(s-1)}, \theta^{*(s)}\big|\mathcal{H}^{(s-1)}\big)1(s \neq \{j_m\}_{m=0}^{k-1}) + 1(s = \{j_m\}_{m=0}^{k-1})\big]\right\}.$$

Because $K_h(\cdot) \geq 0$ for any $0 < t < t'$,

$$E\big[K_h\big(\theta^{*(t')} - \theta^{*(t)}\big)\big] = E\big[K_h\big(\theta^{*(t')}\big(p^{*(t')}\big) - \theta^{*(t)}\big(p^{*(t)}\big)\big)\big] \leq \varepsilon_1^2 E_{t'}\big\{E_t\big[K_h\big(\theta^{*(t')}(\widetilde{g}) - \theta^{*(t)}(\widetilde{g})\big)\big]\big\}.$$

Furthermore,

$$\text{(A29)}\qquad E_{\theta',\theta}\big[K_h(\theta'(\widetilde{g}) - \theta(\widetilde{g}))\big] = E_{\theta'}\big[E_\theta\big\{K_h(\theta'(\widetilde{g}) - \theta(\widetilde{g}))\big\}\big] \leq \sup_{\widetilde{\theta}\in\Theta} E_\theta\big[K_h(\widetilde{\theta} - \theta(\widetilde{g}))\big].$$

Using (A28) and (A29),

$$\text{(A30)}\qquad E\left[\left\{\prod_{i=0}^{k-1}\big[K_h\big(\theta^{*(j_{i+1})} - \theta^{*(j_i)}\big)\big]\right\} \sup_{s'\in S}\big|A_1^{(t(l)-m-i-1)}\big|\Big|\mathcal{H}^{(j_0)}\right]$$
$$\leq \varepsilon_1^{k+1} E\left[\left\{\prod_{i=0}^{k-1}\big[K_h\big(\theta^{*(j_{i+1})}(\widetilde{g}) - \theta^{*(j_i)}(\widetilde{g})\big)\big]\right\} \sup_{s'\in S}\big|A_1^{(t(l)-m-i-1)}\big|\right]$$
$$\leq \varepsilon_1^{k+1} E\left[\left\{\prod_{i=0}^{k-1}\sup_{\theta'\in\Theta}\big[K_{h_l}\big(\theta' - \theta^{*(j_i)}(\widetilde{g})\big)\big]\right\} \sup_{s'\in S}\big|A_1^{(t(l)-m-i-1)}\big|\right]$$
$$= \varepsilon_1^{k+1}\left\{\sup_{\theta'\in\Theta} E_\theta\big[K_{h_l}(\theta' - \theta^*(\widetilde{g}))\big]\right\}^k E\Big[\sup_{s'\in S,\theta'\in\Theta}\big|A_1^{(t(l)-m-i-1)}(\theta')\big|\Big].$$

Furthermore, for any i, m such that $0 < m + i \leq \widetilde{N}(l)$ and for any $k > 1$ such that $k \leq m + i$,

$$\text{(A31)}\quad \frac{1}{\widetilde{N}(l)^{k-1}} \sum_{j_1,\dots,j_{k-1}} I(t(l) - m - i \leq j_1 < \cdots < j_{k-1} < t(l) - m)$$
$$= \frac{1}{\widetilde{N}(l)^{k-1}} \left(\frac{[i]!}{(k-1)!(i-(k-1))!} \right)$$
$$\leq \frac{[[i]/\widetilde{N}(l)]^{k-1}}{(k-1)!} \leq \frac{1}{(k-1)!}.$$

Substituting (A30) and (A31) into (A27), (A25) follows and hence the first part of Claim 1 is proved.

Now, let $Y_i = K^*(t(l) - m, t(l) - m - i, t(l) - m - i - 1, k) \times \sup_{s' \in S, \theta \in \Theta} |A_1^{(t(l)-m-i-1)}|$. Then

$$E\left\{ \sum_{i=k-1}^{\widetilde{N}(l)-m} K^*(t(l) - m, t(l) - m - i, t(l) - m - i - 1, k) \times \sup_{s' \in S, \theta \in \Theta} \left| A_1^{(t(l)-m-i-1)} \right| \right\}^2$$
$$\leq E\left[\sum_{i=k-1}^{\widetilde{N}(l)-m} Y_i \right]^2 \leq 2\widetilde{N}(l) E\left[\sum_{i=k-1}^{\widetilde{N}(l)-m} Y_i^2 \right]$$
$$\leq \frac{2\widetilde{N}(l)}{\widetilde{N}(l)^{2k}} \times \sum_{i=k-1}^{\widetilde{N}(l)-m} E\Bigg[\sum_{j_1,\dots,j_{k-1}} I(t(l) - m - i \leq j_1 < \cdots < j_{k-1} < t(l) - m) \times \prod_{s=0}^{k-1} \left[K_h\left(\theta^{*(j_{s+1})} - \theta^{*(j_s)}\right) \right] \sup_{s' \in S} \left| A_1^{(t(l)-m-i-1)} \right| \Bigg]^2$$
$$\leq \frac{2\varepsilon_1^{2(k+1)}}{\widetilde{N}(l)^{2k-1}} \times \sum_{i=k-1}^{\widetilde{N}(l)-m} \left[\sum_{j_1,\dots,j_{k-1}} I(t(l) - m - i \leq j_1 < \cdots < j_{k-1} < t(l) - m) \right]^2$$

$$\times E\left[\left\{\prod_{s=0}^{k-1}\left[K_h\left(\theta^{*(j_{s+1})}(\widetilde{g})-\theta^{*(j_s)}(\widetilde{g})\right)\right]\sup_{s'\in S,\theta\in\Theta}\left|A_1^{(t(l)-m-i-1)}\right|\right\}^2\right]$$

$$\leq 2\left(\frac{\varepsilon_1^{k+1}}{(k-1)!}\right)^2\left\{\sup_{\theta'\in\Theta}E_\theta\left[K_{h_l}(\theta'-\theta^*(\widetilde{g}))\right]\right\}^{2k}$$

$$\times E\left[\sup_{s'\in S,\theta\in\Theta}\left|A_1^{(t(l)-m-i-1)}\right|^2\right].$$

Hence, the second part of Claim 1 is proved. *Q.E.D.*

Now, using the law of iterated expectations and Claim 1, we obtain the inequality

$$E\left[\frac{1}{\widetilde{N}(l)}\sum_{m=1}^{\widetilde{N}(l)}K_{h(\widetilde{N}(l))}\left(\theta,\theta^{*(t(l)-m)}\right)\right.$$

$$\times\sum_{i=k-1}^{\widetilde{N}(l)-m}K^*(t(l)-m,t(l)-m-i,t(l)-m-i-1,k)$$

$$\left.\times\sup_{s'\in S}\left|A_1^{(t(l)-m-i-1)}\right|\right]$$

$$\leq\frac{\varepsilon_1}{\widetilde{N}(l)}\sum_{m=1}^{\widetilde{N}(l)}E\left[K_{h(\widetilde{N}(l))}\left(\theta,\theta^{*(t(l)-m)}(\widetilde{g})\right)\right.$$

$$\left.\times\sum_{i=k-1}^{\widetilde{N}(l)-m}K^*(t(l)-m,t(l)-m-i,t(l)-m-i-1,k)\right]$$

$$\times\left[C_1h_{l-1}^{(3/2)J}+\frac{\overline{A}C_2}{\sqrt{N(l-1)h_l^{5J}}}\right]$$

$$\leq\left[\varepsilon_1^{k+2}\sup_{\theta'\in\Theta}E_\theta\left[K_{h_l}(\theta'-\theta(\widetilde{g}))\right]^{k+1}\frac{1}{(k-1)!}\right]$$

$$\times\left[C_1h_{l-1}^{(3/2)J}+\frac{\overline{A}C_2}{\sqrt{N(l-1)h_l^{5J}}}\right].$$

Chebychev inequality implies

$$\text{(A32)}\quad \Pr\Bigg[\sum_{m=1}^{\widetilde{N}(l)} \frac{1}{\widetilde{N}(l)} K_{h(\widetilde{N}(l))}\big(\theta, \theta^{*(t(l)-m)}\big) \times \sum_{i=k-1}^{\widetilde{N}(l)-m} K^*(t(l)-m, t(l)-m-i, t(l)-m-i-1, k) \times \sup_{s'\in S}\big|A_1^{(t(l)-m-i-1)}\big| > \left[\frac{\delta}{4A\beta}\varepsilon_0 \inf_\theta g(\theta)\right]^k\Bigg]$$

$$\leq \left[C_1 h_{l-1}^{(3/2)J} + \frac{\overline{A}C_2}{\sqrt{N(l-1)h_l^{5J}}}\right]\varepsilon_1^{k+2} \times \sup_{\theta'\in\Theta} E\big[K_{h_l}(\theta' - \theta(\widetilde{g}))\big]^{k+1} \frac{1}{(k-1)!} \Bigg/ \left[\frac{\delta}{4A\beta}\varepsilon_0 \inf_\theta g(\theta)\right]^{k+1}.$$

Next, we consider the second term of the RHS of equation (A24). We first prove the following claim.

CLAIM 2: *For any* $t(l-1) < t \leq t(l)$, *either* $[t(l-1)+1-\widetilde{N}(l-1)/2, t(l-1)+1] \subseteq [t-N(t), t]$ *or* $[t(l-1)+1, t(l-1)+\widetilde{N}(l-1)/2+1] \subseteq [t-N(t), t]$ *or both*.

PROOF: First, we show that for t satisfying $t(l-1) < t \leq t(l-1) + \widetilde{N}(l-1)/2$,

$$\text{(A33)}\quad [t(l-1)+1-\widetilde{N}(l-1)/2, t(l-1)+1] \subseteq [t-N(t), t].$$

Because $N(\cdot)$ is a nondecreasing function, $N(t) \geq \widetilde{N}(l-1)$. Hence,

$$t - t(l-1) \leq \widetilde{N}(l-1)/2 = \widetilde{N}(l-1) - \widetilde{N}(l-1)/2 \leq N(t) - \widetilde{N}(l-1)/2.$$

Thus,

$$t - N(t) \leq t(l-1) - \widetilde{N}(l-1)/2.$$

Since $t(l-1)+1 \leq t$, (A33) holds.

Next, we show that for t satisfying $t(l-1)+\widetilde{N}(l-1)/2 < t \le t(l)$,

$$[t(l-1)+1, t(l-1)+\widetilde{N}(l-1)/2+1] \subseteq [t-N(t), t]. \tag{A34}$$

From the definition of $\widetilde{N}(\cdot)$,

$$t(l)-\widetilde{N}(l) = t(l-1)+1.$$

Furthermore, because $N(s)$ is increasing at most by 1 with unit increase in s, $s-N(s)$ is nondecreasing in s. Hence,

$$t-N(t) \le t(l)-\widetilde{N}(l) = t(l-1)+1.$$

Furthermore, $t \ge t(l-1)+\widetilde{N}(l-1)/2+1$. Therefore, (A34) holds. Hence, Claim 2 is proved. *Q.E.D.*

Now, following an argument very similar to the one used in deriving (A9), it is straightforward to show that for any $\eta_3 > 0$, there exists L such that for any $l > L$, $t_1 = t(l-1)+1$, and for $t_2 = t(l-1)+\widetilde{N}(l-1)/2+1$,

$$\Pr\left[\frac{1}{\widetilde{N}(l)/2}\sum_{k=1}^{\widetilde{N}(l)/2} K_{h(N(t_i))}\big(\theta-\theta^{*(t_i-k)}\big) \le \frac{1}{2}\varepsilon_0 \inf_{\theta\in\Theta} g(\theta)\right]$$
$$\le \eta_3, \quad i=1,2.$$

Therefore,

$$\begin{aligned}
&\Pr\left[\left\{\frac{1}{\widetilde{N}(l)/2}\sum_{k=1}^{\widetilde{N}(l)/2} K_h\big(\theta-\theta^{*(t_1-k)}\big) \le \frac{1}{2}\varepsilon_0 \inf_{\theta\in\Theta} g(\theta)\right\}\right.\\
&\qquad \left.\cup\left\{\frac{1}{\widetilde{N}(l)/2}\sum_{k=1}^{\widetilde{N}(l)/2} K_h\big(\theta-\theta^{*(t_2-k)}\big) \le \frac{1}{2}\varepsilon_0 \inf_{\theta\in\Theta} g(\theta)\right\}\right]\\
&\quad \le \Pr\left[\frac{1}{\widetilde{N}(l)/2}\sum_{k=1}^{\widetilde{N}(l)/2} K_h\big(\theta-\theta^{*(t_1-k)}\big) \le \frac{1}{2}\varepsilon_0 \inf_{\theta\in\Theta} g(\theta)\right]\\
&\qquad + \Pr\left[\frac{1}{\widetilde{N}(l)/2}\sum_{k=1}^{\widetilde{N}(l)/2} K_h\big(\theta-\theta^{*(t_2-k)}\big) \le \frac{1}{2}\varepsilon_0 \inf_{\theta\in\Theta} g(\theta)\right]\\
&\quad \le 2\eta_3.
\end{aligned}$$

Therefore,

$$\Pr\left[\left\{\frac{1}{\widetilde{N}(l)/2}\sum_{k=1}^{\widetilde{N}(l)/2} K_h\big(\theta-\theta^{*(t_1-k)}\big) > \frac{1}{2}\varepsilon_0 \inf_{\theta\in\Theta} g(\theta)\right\}\right.$$
$$\left.\cap\left\{\frac{1}{\widetilde{N}(l)/2}\sum_{k=1}^{\widetilde{N}(l)/2} K_h\big(\theta-\theta^{*(t_2-k)}\big) > \frac{1}{2}\varepsilon_0 \inf_{\theta\in\Theta} g(\theta)\right\}\right]$$
$$> 1-2\eta_3.$$

Now, since from Assumptions 6 and 7, $h(N(t))$ is the same for any $t(l-1) < t \leq N(l)$, from Claim 2, for any t such that $t(l-1) < t \leq t(l)$,

$$\frac{1}{\widetilde{N}(l)}\sum_{k=1}^{N(t)} K_h\big(\theta-\theta^{*(t-k)}\big) \geq \frac{\widetilde{N}(l-1)/2}{\widetilde{N}(l)}\frac{1}{\widetilde{N}(l-1)/2}$$
$$\times \sum_{k=1}^{\widetilde{N}(l-1)/2} K_h\big(\theta-\theta^{*(s-k)}\big),$$

where either $s=t_1=t(l-1)+1$ or $s=t_2=t(l-1)+\widetilde{N}(l-1)/2+1$ or both. Furthermore, notice that $\frac{\widetilde{N}(l-1)/2}{\widetilde{N}(l)} \geq \frac{1}{2A}$. Therefore,

$$\Pr\left[\inf_{t(l-1)<t\leq t(l)} \frac{1}{\widetilde{N}(l)}\sum_{n=1}^{N(t)} K_h\big(\theta-\theta^{*(t-n)}\big) \geq \frac{1}{4A}\varepsilon_0 \inf_{\theta\in\Theta} g(\theta)\right] > 1-2\eta_3.$$

Thus,

$$\Pr\left[\inf_{t(l-1)<t\leq t(l)}\left[\frac{1}{\widetilde{N}(l)}\sum_{n=1}^{N(t)} \widetilde{K}(t,t-n)\right] < \frac{1}{4A}\varepsilon_0 \inf_{\theta} g(\theta)\right] \leq 2\eta_3. \tag{A35}$$

By (A32) and (A35),

RHS of (A24)

$$\leq \sum_{k=1}^{\widetilde{N}(l)}\left[C_1 h_{l-1}^{(3/2)J} + \frac{\overline{A}C_2}{\sqrt{N(l-1)h_l^{5J}}}\right]\varepsilon_1^{k+2}$$
$$\times \sup_{\theta'\in\Theta} E_\theta\big[K_{h_{l-1}}(\theta'-\theta(\widetilde{g}))\big]^{k+1}\frac{1}{(k-1)!}$$

$$\Big/ \left[\frac{\delta}{4A\beta}\varepsilon_0 \inf_\theta g(\theta)\right]^{k+1} + 2\eta_3$$

$$= \varepsilon_1 \left[C_1 h_{l-1}^{(3/2)J} + \frac{\overline{A}C_2}{\sqrt{N(l-1)h_l^{5J}}}\right] e^\lambda \lambda^2 \sum_{k=1}^{\widetilde{N}(l)} \left[e^{-\lambda}\frac{\lambda^{(k-1)}}{(k-1)!}\right] + 2\eta_3,$$

where

$$\lambda = \frac{4A\beta\varepsilon_1 \sup_{\theta'\in\Theta} E_\theta[K_{h_{l-1}}(\theta' - \theta(\widetilde{g}))]}{\delta\varepsilon_0 \inf_\theta g(\theta)} > 0.$$

Notice that $e^{-\lambda}(\lambda^k/k!)$ is the formula for the distribution function of the Poisson distribution. Hence,

$$\sum_{k=1}^{\widetilde{N}(l)} e^{-\lambda}\frac{\lambda^{(k-1)}}{(k-1)!} \leq \sum_{k=1}^{\infty} e^{-\lambda}\frac{\lambda^{(k-1)}}{(k-1)!} = 1.$$

Together, we have shown that

$$\text{(A36)} \qquad \text{LHS of (A24)} \leq \varepsilon_1 \left[C_1 h_{l-1}^{(3/2)J} + \frac{\overline{A}C_2}{\sqrt{N(l-1)h_l^{5J}}}\right]\lambda^2 \exp(\lambda) + 2\eta_3.$$

Now,

$$E_\theta\big[K_h(\theta', \theta(\widetilde{g}))\big] \to \widetilde{g}(\theta') \quad \text{as } h \to 0.$$

Hence, for any $B > \sup_{\theta\in\Theta}[\widetilde{g}(\theta)]$, there exists $H > 0$ such that for any positive $h < H$,

$$E_\theta\big\{K_h(\theta', \theta(\widetilde{g}))\big\} < B.$$

Therefore, for $h < H$, λ is uniformly bounded. Hence, for any $\delta > 0$, the RHS of (A36) can be made arbitrarily small by choosing $l - 1$ large enough to make h_{l-1} arbitrarily small, $N(l-1)h_l^{5J}$ arbitrarily large, and η_3 arbitrarily small. Hence, we have shown that $B_2(l, l) \xrightarrow{P} 0$.

We now proceed to show uniform convergence of $B_2(l, l)$. As before, we use a Fourier transform. Let

$$Y_m \equiv \sum_{i=k-1}^{\widetilde{N}(l)-m} K^*(t(l) - m, t(l) - m - i, t(l) - m - i - 1, k) \times \sup_{s'\in S}\big|A_1^{(t(l)-m-i-1)}\big|.$$

Then

$$
\begin{aligned}
\text{(A37)}\quad & E\left[\sup_{\theta\in\Theta}\sum_{m=1}^{\widetilde{N}(l)}\frac{1}{\widetilde{N}(l)}K_{h(\widetilde{N}(l))}\big(\theta,\theta^{*(t(l)-m)}(q)\big)Y_m\right] \\
&\le \varepsilon_1 E\left[\sup_{\theta\in\Theta}\sum_{m=1}^{\widetilde{N}(l)}\frac{1}{\widetilde{N}(l)}K_{h(\widetilde{N}(l))}\big(\theta,\theta^{*(t(l)-m)}(\widetilde{g})\big)Y_m\right] \\
&\le \varepsilon_1\left(\frac{1}{2\pi}\right)^J\left\{4E\left[\frac{1}{\widetilde{N}(l)}\sum_{m=1}^{\widetilde{N}(l)}Y_m^2\right]\right\}^{1/2}\int|\psi(hz)|\,dz.
\end{aligned}
$$

Now, using Claim 1,

$$
\begin{aligned}
&\text{RHS of (A37)} \\
&\le \frac{2\sqrt{2}\varepsilon_1^{k+2}}{(k-1)!}\Big[\sup_{\theta'\in\Theta}E\big[K_{h_{l-1}}(\theta'-\theta(\widetilde{g}))\big]\Big]^k \\
&\quad\times\sqrt{C_1^2h_{l-1}^{3J}+\frac{\overline{A}^2C_2}{\sqrt{N(l-1)h_l^{5J}}}}\left(\frac{1}{2\pi}\right)^J\int|\psi(hz)|\,dz \\
&\le \frac{2\sqrt{2}\varepsilon_1^{k+2}}{(k-1)!}\Big[\sup_{\theta'\in\Theta}E\big[K_{h_{l-1}}(\theta'-\theta(\widetilde{g}))\big]\Big]^k \\
&\quad\times\sqrt{C_1^2h_{l-1}^{J}+\frac{\overline{A}^2C_2}{\sqrt{N(l-1)h_l^{9J}}}}\left(\frac{1}{2\pi}\right)^J\int|\psi(z)|\,dz.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
&E\left[\sup_{\theta\in\Theta}\sum_{k=1}^{\widetilde{N}(l)}\sum_{m=1}^{\widetilde{N}(l)}K_{h(\widetilde{N}(l))}\big(\theta,\theta^{*(t(l)-m)}\big)\right. \\
&\quad\times\sum_{i=k-1}^{\widetilde{N}(l)-m}K^*(t(l)-m,t(l)-m-i,t(l)-m-i-1,k) \\
&\quad\left.\times\sup_{s'\in S}\big|A_1^{(t(l)-m-i-1)}\big|\right] \\
&\le\sum_{k=1}^{\widetilde{N}(l)}\frac{2\sqrt{2}\varepsilon_1^{k+2}}{(k-1)!}\Big[\sup_{\theta'\in\Theta}E\big[K_{h_{l-1}}(\theta'-\theta(\widetilde{g}))\big]\Big]^k
\end{aligned}
$$

$$\times \sqrt{C_1^2 h_{l-1}^J + \frac{\overline{A}^2 C_2}{\sqrt{N(l-1)h_l^{9J}}}} \left(\frac{1}{2\pi}\right)^J \int |\psi(z)|\, dz$$

$$\leq 2\sqrt{2}\varepsilon_1^2 \sqrt{\left[C_1^2 h_{l-1}^J + \frac{\overline{A}^2 C_2}{\sqrt{N(l-1)h_l^{9J}}}\right]}$$

$$\times e^{\lambda}\lambda \sum_{k=1}^{\widetilde{N}(l)} \left[e^{-\lambda}\frac{\lambda^{(k-1)}}{(k-1)!}\right]\left(\frac{1}{2\pi}\right)^J \int |\psi(z)|\, dz,$$

where

$$\lambda = \varepsilon_1 \sup_{\theta' \in \Theta} E_\theta\big[K_{h_{l-1}}(\theta' - \theta(\widetilde{g}))\big] \to \varepsilon_1 \sup_{\theta' \in \Theta} \widetilde{g}(\theta').$$

Because, $C_1^2 h_{l-1}^J + (\overline{A}^2 C_2/\sqrt{N(l-1)h_l^{9J}})$ can be made arbitrarily small by increasing l, we have shown that the numerator of $B_2(l,l)$ converges to zero uniformly. Now, consider the denominator. We can follow the earlier arguments (see (A9)) to prove that for any $\eta > 0$, there exists $l > 0$, $\overline{N} \equiv N(\bar{t}_\eta)$ such that for any $t > \bar{t}_\eta$, that is, $N(l) > \overline{N}$,

$$\Pr\left[\inf_{\theta \in \Theta} \frac{1}{\widetilde{N}(l-1)/2} \sum_{k=1}^{\widetilde{N}(l-1)/2} K_{h_{l-1}}\big(\theta - \theta^{*(t-k)}(q)\big) > \frac{1}{2}\varepsilon_0 \inf_{\theta\in\Theta} g(\theta)\right]$$
$$> 1 - \eta.$$

Therefore, for sufficiently large l, the denominator is bounded away from zero uniformly in Θ with probability arbitrarily close to 1. Together, we have shown that $B_2(l,l) \stackrel{P}{\to} 0$ uniformly in Θ. Thus, Lemma 5 is proved. That is, we have shown that

$$A(l,l) \stackrel{P}{\to} 0 \quad \text{as } l \to \infty,$$

uniformly in Θ. *Q.E.D.*

To prove Theorem 1, we now return to (A16). Before we proceed further, we introduce the following notation. Let

$$\Xi(l, l_1+1) \equiv \big\{(t_l, t_{l-1}, \ldots, t_{l_1+1}):$$
$$t(l_1) < t_{l_1+1} \leq t(l_1+1), \ldots, t_{l-1} \leq t(l-1) < t_l \leq t(l)\big\}.$$

Now, define $\overrightarrow{W}(t(l), t(l_1), t_{l_1})$ as follows: For $l_1 = l$,

$$\overrightarrow{W}(t(l), t(l), t_l) \equiv W_{\widetilde{N}(l),h}\big(\theta, \theta^{*(t_l)}\big).$$

For $l_1 = l - 1$,

$$\overrightarrow{W}\big(t(l), t(l-1), t_{l-1}\big) = \sum_{m=1}^{\widetilde{N}(l)} W_{\widetilde{N}(l),h}\big(\theta, \theta^{*(t(l)-m)}\big)\widehat{W}\big(t(l)-m, t(l-1), t_{l-1}\big).$$

For $l_1 \le l - 2$,

$$\begin{aligned}\overrightarrow{W}\big(t(l), t(l_1), t_{l_1}\big) &\equiv \sum_{(t_l, t_{t-1}, \dots, t_{l_1+1}) \in \Xi(l, l_1+1)} W_{\widetilde{N}(l),h}\big(\theta, \theta^{*(t_l)}\big) \\ &\quad \times \left\{\prod_{j=l_1+1}^{l-1} \widehat{W}(t_{j+1}, t(j), t_j)\right\}\widehat{W}\big(t_{l_1+1}, t(l_1), t_{l_1}\big).\end{aligned}$$

Recursively, we can express for $l_1 < l$,

$$\begin{aligned}\overrightarrow{W}\big(t(l), t(l_1), t_{l_1}\big) = \sum_{m=1}^{\widetilde{N}(l_1+1)} &\overrightarrow{W}(l, t(l_1+1), t(l_1+1)-m) \\ &\times \widehat{W}\big(t(l_1+1)-m, t(l_1), t_{l_1}\big).\end{aligned}$$

Hence, (A16) can be written as

$$\begin{aligned}&\sum_{m=1}^{\widetilde{N}(l)} \Delta V\big(\epsilon^{(t(l)-m)}, \theta^{*(t(l)-m)}\big)\overrightarrow{W}(t(l), t(l), t(l)-m) \\ &\quad \le \sum_{m=1}^{\widetilde{N}(l)} \overrightarrow{W}(t(l), t(l), t(l)-m) \sup_{s' \in S}\big| A_1^{(t(l)-m)}\big(\theta^{*(t(l)-m)}\big)\big| \\ &\qquad + \sum_{m=1}^{\widetilde{N}(l)} \overrightarrow{W}(t(l), t(l), t(l)-m)\end{aligned} \tag{A38}$$

$$
\begin{aligned}
&\times \sum_{i=0}^{N(l)-m} \widehat{W}(t(l)-m, t(l)-m-i, t(l)-m-i-1) \\
&\times \sup_{s' \in S} \bigl| A_1^{(t(l)-m-i-1)} \bigr| \\
&+ \sum_{m=1}^{\widetilde{N}(l-1)} \Delta V^{(t(l-1)-m)}\bigl(\epsilon^{(t(l-1)-m)}, \theta^{(t(l-1)-m)}\bigr) \\
&\times \overrightarrow{W}(t(l), t(l-1), t(l-1)-m).
\end{aligned}
$$

Furthermore, by Lemma 4,

$$
\sum_{m=1}^{\widetilde{N}(l_1)} \widehat{W}\bigl(t_{l_1+1}, t(l_1), t(l_1)-m\bigr) \le \beta.
$$

Applying these inequalities to $\overrightarrow{W}$ yields

$$
\sum_{m=1}^{\widetilde{N}(l_1)} \overrightarrow{W}(t(l), t(l_1), t(l_1)-m) \le \beta^{(l-l_1+1)}. \tag{A39}
$$

Now, let

$$
A(l, l_1) \equiv B_1(l, l_1) + B_2(l, l_1),
$$

where

$$
B_1(l, l_1) \equiv \sum_{m=1}^{\widetilde{N}(l_1)} \overrightarrow{W}(t(l), t(l_1), t(l_1)-m) \sup_{s' \in S} \bigl| A_1^{(t(l_1)-m)} \bigr|
$$

and

$$
\begin{aligned}
B_2(l, l_1) \equiv \sum_{m=1}^{\widetilde{N}(l_1)} &\overrightarrow{W}(t(l), t(l_1), t(l_1)-m) \\
&\times \sum_{j=0}^{N(l_1)-m} \Bigl\{ \widehat{W}(t(l_1)-m, t(l_1)-m-j, t(l_1)-m-j-1) \\
&\times \sup_{s' \in S} \bigl| A_1^{(t(l_1)-m-j-1)} \bigr| \Bigr\}.
\end{aligned}
$$

Then, for $l_1 \leq l$,

$$\sum_{m=1}^{\widetilde{N}(l_1)} \Delta V^{(t(l_1)-m)}\big(\epsilon^{(t(l_1)-m)}, \theta^{(t(l_1)-m)}\big)\overrightarrow{W}(t(l), t(l_1), t(l_1)-m) \leq A(l, l_1) + \sum_{m=1}^{\widetilde{N}(l_1-1)} \Delta V^{(t(l_1-1)-m)}\big(\epsilon^{(t(l_1-1)-m)}, \theta^{(t(l_1-1)-m)}\big) \times \overrightarrow{W}(t(l), t(l_1-1), t(l_1-1)-m). \tag{A40}$$

LEMMA 6: *Given* $\Delta = l - l_1 \geq 0$,

$$A(l, l-\Delta) \xrightarrow{P} 0 \quad \text{as } l \to \infty.$$

PROOF: Lemma 5 proves Lemma 6 with $\Delta = 0$. Consider $\Delta > 0$. We prove convergence of $B_1(l, l-\Delta)$. By definition of $\overrightarrow{W}$,

$$\begin{aligned}
&\overrightarrow{W}(t(l), t(l_1), t(l_1)-m) \\
&= \Bigg[\sum_{t(l-1)<t_l\leq t(l)} W_{\widetilde{N}(l),h}\big(\theta, \theta^{*(t_l)}\big)\Bigg\{\sum_{t(l-2)<t_{l-1}\leq t(l-1)} \widehat{W}(t_l, t(l-1), t_{l-1})\cdots \\
&\quad \times \Bigg\{\sum_{t(l_1)<t_{l_1+1}\leq t(l_1+1)} \widehat{W}\big(t_{l_1+2}, t(l_1+1), t_{l_1+1}\big) \\
&\quad \times \widehat{W}\big(t_{l_1+1}, t(l_1), t(l_1)-m\big)\Bigg\}\Bigg\}\Bigg].
\end{aligned}$$

Using this expression, we write the numerator of $B_1(l, l_1)$ divided by $\widetilde{N}(l)$ as

$$E\Bigg[\sum_{t(l-1)<t_l\leq t(l)} \frac{1}{\widetilde{N}(l)} K_{h(\widetilde{N}(l))}\big(\theta, \theta^{*(t_l)}\big) \times \Bigg\{\sum_{t(l-2)<t_{l-1}\leq t(l-1)} \sum_k K^*(t_l, t(l-1), t_{l-1}, k)\cdots \times \Bigg\{\sum_{t(l_1)<t_{l_1+1}\leq t(l_1+1)} \sum_k K^*\big(t_{l_1+2}, t(l_1+1), t_{l_1+1}, k\big) \tag{A41}$$

$$\times \sum_{m=1}^{\widetilde{N}(l)} K^*\big(t_{l_1+1}, t(l_1), t(l_1)-m\big)\bigg\}\bigg\}$$

$$\times \sup_{s'\in S}\big|A_1^{(t(l_1)-m)}\big|\Bigg].$$

Then, using the arguments similar to the derivation of (A36), we can show that

$$\begin{aligned}
&\text{RHS of (A41)}\\
&\quad\le \varepsilon_1 \sup_{\theta'\in\Theta} E_\theta\big[K_{h_{l-1}}(\theta'-\theta(\widetilde{g}))\big]\\
&\qquad\times \prod_{j=l_1}^{l-1}\left\{\sum_k \frac{\varepsilon_1^{k+1}}{[(k-1)!]}\left[\prod_{s=0}^{k-1}\sup_{\theta'\in\Theta} E\big[K_{h_{j-1}}(\theta'-\theta(\widetilde{g}))\big]\right]\right\}\\
&\qquad\times\left[C_1 h_{l_1}^{(3/2)J}+\frac{\overline{A}C_2}{\sqrt{N(l_1)h_{l_1+1}^{5J}}}\right]\\
&\quad\le \lambda[\varepsilon_1\lambda e^{\lambda}]^{\Delta-1}\left[C_1 h_{l_1}^{(3/2)J}+\frac{\overline{A}C_2}{\sqrt{N(l_1)h_{l_1+1}^{5J}}}\right]\to 0\quad \text{as } l_1\to\infty,
\end{aligned}$$

where

$$\lambda=\varepsilon_1 \sup_{\theta'\in\Theta, l_1\le j\le l-1} E_\theta\big[K_{h_{j-1}}(\theta'-\theta(\widetilde{g}))\big]>0.$$

Next, we consider the denominator of $B_1(l,l_1)$ divided by $\widetilde{N}(l)$. Let $t_1(l)\equiv t(l-1)+1$ and $t_2(l)=t(l-1)+\widetilde{N}(l-1)/2+1$. Then arguments similar to ones used to derive equation (A35) can be used to derive the inequality

$$\begin{aligned}
&\inf_{t(l_1-1)<t\le t(l)}\left[\frac{1}{\widetilde{N}(l)}\sum_{i=1}^{N(t)} K_h\big(\theta-\theta^{*(t-i)}\big)\right]\\
&\quad\ge \min_{l_1-1<\widetilde{l}\le l}\left\{\frac{\widetilde{N}(\widetilde{l})/2}{\widetilde{N}(l)}\frac{1}{\widetilde{N}(\widetilde{l})/2}\right.\\
&\qquad\left.\times\min\left\{\sum_{k=1}^{\widetilde{N}(\widetilde{l})/2} K_{h_{\widetilde{l}-1}}\big(\theta-\theta^{*(t_1(\widetilde{l})-k)}\big),\ \sum_{k=1}^{\widetilde{N}(\widetilde{l})/2} K_{h_{\widetilde{l}-1}}\big(\theta-\theta^{*(t_2(\widetilde{l})-k)}\big)\right\}\right\}
\end{aligned}$$

$$\geq \frac{1}{2A^{l+1-l_1}} \frac{1}{\widetilde{N}(l^*)/2} \times \min\left\{ \sum_{k=1}^{\widetilde{N}(l^*)/2} K_{h_{l^*-1}}\big(\theta - \theta^{*(t_1(l^*)-k)}\big), \sum_{k=1}^{\widetilde{N}(l^*)/2} K_{h_{l^*-1}}\big(\theta - \theta^{*(t_2(l^*)-k)}\big) \right\},$$

where

$$l^* \equiv \arg \min_{\widetilde{l}: l_1-1<\widetilde{l}\leq l} \left\{ \frac{1}{2A^{l+1-\widetilde{l}}} \frac{1}{\widetilde{N}(\widetilde{l})/2} \times \min\left\{ \sum_{k=1}^{\widetilde{N}(\widetilde{l})/2} K_{h_{\widetilde{l}-1}}\big(\theta - \theta^{*(t_1(\widetilde{l})-k)}\big), \sum_{k=1}^{\widetilde{N}(\widetilde{l})/2} K_{h_{\widetilde{l}-1}}\big(\theta - \theta^{*(t_2(\widetilde{l})-k)}\big) \right\} \right\}.$$

Hence,

$$\inf_{t(l_1-1)<t\leq t(l)} \left[\frac{1}{\widetilde{N}(l)} \sum_{i=1}^{N(t)} K_h\big(\theta - \theta^{*(t-i)}\big) \right]$$
$$\geq \min_{l_1-1<\widetilde{l}\leq l} \left\{ \frac{\widetilde{N}(\widetilde{l})/2}{\widetilde{N}(l)} \frac{1}{\widetilde{N}(\widetilde{l})/2} \times \min\left\{ \sum_{k=1}^{\widetilde{N}(\widetilde{l})/2} K_{h_{\widetilde{l}-1}}\big(\theta - \theta^{*(t_1(\widetilde{l})-k)}\big), \sum_{k=1}^{\widetilde{N}(\widetilde{l})/2} K_{h_{\widetilde{l}-1}}\big(\theta - \theta^{*(t_2(\widetilde{l})-k)}\big) \right\} \right\}.$$

Hence, similarly to (A35), we can show that the denominator is positively bounded away from zero with probability arbitrarily close to 1 by choosing l to be large enough. Therefore, given that $\Delta = l - l_1$ is a constant,

$$B_1(l, l-\Delta) \xrightarrow{P} 0$$

as $l \to \infty$.

Next, we prove uniform convergence by using the arguments in Bierens (1994). Let

$$Y_m = \sum_{t_{l_1}=t(l_1-1)+1}^{t(l_1)} \sum_k K^*\big(t(l)-m, t(l_1), t_{l_1}, k\big) \sup_{s'\in S} \big|A^{(t_{l_1})}\big|.$$

Then

$$
\begin{aligned}
\text{(A42)}\quad & E\left[\sup_{\theta\in\Theta}\sum_{m=1}^{\widetilde{N}(l)}\frac{1}{\widetilde{N}(l)}K_{h(\widetilde{N}(l))}\big(\theta,\theta^{*(t(l)-m)}(q)\big)Y_m\right] \\
& \le \varepsilon_1\left(\frac{1}{2\pi}\right)^J\left\{4E\left[\frac{1}{\widetilde{N}(l)}\sum_{m=1}^{\widetilde{N}(l)}Y_m^2\right]\right\}^{1/2}\int|\psi(hz)|\,dz, \\
& E[Y_m^2] \le E\left[\sum_{t_{l_1}=t(l_1-1)+1}^{t(l_1)}\sum_k K^*\big(t(l)-m,t(l_1),t_{l_1},k\big)\sup_{s'\in S}\big|A_1^{(t_{l_1})}\big|\right]^2 \\
& \qquad = E\Bigg[\Bigg\{\sum_{t_{l-1}}\sum_k K^*(t(l)-m,t(l-1),t_{l-1},k) \\
& \qquad\quad \times\Bigg\{\sum_{t_{l-2}}\sum_k K^*(t_{l-1},t(l-2),t_{l-2},k)\cdots \\
& \qquad\quad \times\Bigg\{\sum_{t_{l_1}}\sum_k K^*\big(t_{l_1+1},t(l_1),t_{l_1},k\big)\sup_{s'\in S}\big|A_1^{(t_{l_1})}\big|\Bigg\}\cdots\Bigg\}\Bigg\}\Bigg]^2.
\end{aligned}
$$

Using the results in Claim 1, we can show that

$$
\begin{aligned}
\text{(A43)}\quad & \text{RHS of (A42)} \\
& \le\Bigg[\prod_{j=l_1}^{l}\sum_k\left(\frac{2\varepsilon_1^{k+1}}{(k-1)!}\right)^2 \\
& \quad\times\Big\{\sup_{\theta'\in\Theta,h_l\le\widetilde{h}\le h_{l_1-1}}E_\theta\big[K_{\widetilde{h}}(\theta'-\theta^*(\widetilde{g}))\big]\Big\}^{2k}\Bigg] \\
& \quad\times\left[C_1^2h_{l_1-1}^{3J}+\frac{\overline{A}^2C_2}{\sqrt{N(l)h_{l_1-1}^{5J}}}\right] \\
& \le\left[\prod_{j=l_1}^{l}\sum_k\frac{4\varepsilon_1^{2k+2}}{(k-1)!}\Big\{\sup_{\theta'\in\Theta,h_l\le\widetilde{h}\le h_{l_1-1}}E_\theta\big[K_{\widetilde{h}}(\theta'-\theta^*(\widetilde{g}))\big]\Big\}^{2k}\right]
\end{aligned}
$$

$$\times \left[C_1^2 h_{l_1-1}^{3J} + \frac{\overline{A}^2 C_2}{\sqrt{N(l_1-1)h_{l_1}^{5J}}} \right]$$

$$\le [4\varepsilon_1^4 e^{\lambda}\lambda]^{\Delta} \left[C_1^2 h_{l_1-1}^{3J} + \frac{\overline{A}^2 C_2}{\sqrt{N(l_1-1)h_{l_1}^{5J}}} \right],$$

where

$$\lambda = \varepsilon_1^2 \Big\{ \sup_{\theta' \in \Theta, h_{l_1-1} \le h \le h_l} E_\theta\big[K_h(\theta' - \theta(\widetilde{g}))\big] \Big\}^2$$

$$\to \varepsilon_1^2 \sup_{\theta' \in \Theta} \widetilde{g}(\theta')^2 \quad \text{as } l_1 \to \infty.$$

Hence,

$$E\left[\sup_{\theta \in \Theta} \sum_{m=1}^{\widetilde{N}(l)} \frac{1}{\widetilde{N}(l)} K_{h(\widetilde{N}(l))}\big(\theta, \theta^{*(t(l)-m)}(q)\big) \right.$$

$$\left. \times \sum_k K^*\big(t(l)-m, t(l_1), t_{l_1}, k\big) \sup_{s' \in S} \big| A^{(t_{l_1})} \big| \right]$$

$$\le \varepsilon_1 [4\varepsilon_1^4 e^{\lambda}\lambda]^{\Delta/2} \left[C_1^2 h_{l_1-1}^{3J} + \frac{\overline{A}^2 C_2}{\sqrt{N(l_1-1)h_{l_1}^{5J}}} \right] \left(\frac{1}{2\pi}\right)^J \int |\psi(hz)|\, dz$$

$$= \varepsilon_1 [4\varepsilon_1^4 e^{\lambda}\lambda]^{\Delta/2} \left[C_1^2 h_{l_1-1}^{J} + \frac{\overline{A}^2 C_2}{\sqrt{N(l_1-1)h_{l_1}^{9J}}} \right]$$

$$\times \left(\frac{1}{2\pi}\right)^J \int |\psi(z)|\, dz \to 0 \quad \text{as } l_1 \to \infty.$$

Therefore, the numerator divided by $\widetilde{N}(l)$ converges to zero. Because the denominator divided by $\widetilde{N}(l)$ can be shown to be bounded away from zero uniformly in Θ with probability arbitrarily close to 1 given sufficiently large l, we have shown that

$$B_1(l, l-\Delta) \xrightarrow{P} 0$$

uniformly in Θ.

Next, we prove convergence of $B_2(l, l_1)$.

Again, the arguments are very similar to these of Lemma 5. That is, we consider the numerator divided by $\widetilde{N}(l)$. It can be written as

$$E\left[\frac{1}{\widetilde{N}(l)}K_{h(\widetilde{N}(l))}\big(\theta,\theta^{*(t(l)-m)}\big)\left\{\sum_{t_{l-1}}\sum_{k}K^*(t(l)-m,t(l-1),t_{l-1},k)\right.\right.$$
$$\times\left\{\sum_{t_{l-2}}\sum_{k}K^*(t_{l-1},t(l-2),t_{l-2},k)\cdots\right.$$
$$\times\left\{\sum_{k=1}^{\widetilde{N}(l_1)}\sum_{j=k-1}^{t_{l_1}-t(l_1-1)-1}K^*\big(t_{l_1},t_{l_1}-j,t_{l_1}-j-1,k\big)\right.$$
$$\left.\left.\left.\left.\times\sup_{s'\in S}\big|A_1^{(t_{l_1}-j-1)}\big|\right\}\cdots\right\}\right\}\right]. \tag{A44}$$

Then, again by using Claim 1,

$$\begin{aligned}
&\text{RHS of (A44)}\\
&\le \varepsilon_1\sup_{\theta'\in\Theta,h_l\le h\le h_{l_1-1}}E_\theta\big[K_h(\theta'-\theta(\widetilde{g}))\big]\\
&\quad\times\prod_{j=l_1}^{l-1}\left\{\sum_k\frac{\varepsilon_1^{k+1}}{[(k-1)!]}\left[\prod_{s=0}^{k-1}\sup_{\theta'\in\Theta,h_l\le h\le h_{l_1-1}}E\big[K_h(\theta'-\theta(\widetilde{g}))\big]\right]\right\}\\
&\quad\times\left[C_1h_{l_1-1}^{(3/2)J}+\frac{\overline{A}C_2}{\sqrt{N(l_1-1)h_{l_1}^{5J}}}\right]\\
&\le\lambda[\varepsilon_1e^\lambda\lambda]^{\Delta-1}\left[C_1h_{l_1-1}^{(3/2)J}+\frac{\overline{A}C_2}{\sqrt{N(l)h_{l_1}^{5J}}}\right]\to0\quad\text{as }l_1\to\infty,
\end{aligned}$$

where

$$\lambda=\varepsilon_1\sup_{\theta'\in\Theta,h_l\le h\le h_{l_1-1}}E_\theta\big[K_h(\theta'-\theta(\widetilde{g}))\big]\to\varepsilon_1\sup_{\theta'\in\Theta}\widetilde{g}(\theta')\quad\text{as }l_1\to\infty.$$

Next, we consider the denominator of $B_2(l,l_1)$ divided by $\widetilde{N}(l)$. Let $t_1(l)\equiv t(l-1)+1$ and $t_2(l)=t(l-1)+\widetilde{N}(l-1)/2+1$. Then, arguments similar to those used to derive equation (A35) can be used to derive the inequality

$$\begin{aligned}
&\inf_{t(l_1-1)<t\le t(l)}\left[\frac{1}{\widetilde{N}(l)}\sum_{i=1}^{N(t)}\widetilde{K}(t,t-i)\right]\\
&\ge\min_{l_1-1\le\widetilde{l}<l}\left\{\frac{\widetilde{N}(\widetilde{l})/2}{\widetilde{N}(l)}\frac{1}{\widetilde{N}(\widetilde{l})/2}\right.
\end{aligned}$$

$$\times \min\left\{\sum_{k=1}^{\widetilde{N}(\widetilde{l})/2} K_{h_{\widetilde{l}}}\big(\theta - \theta^{*(t_1(\widetilde{l})-k)}\big), \sum_{k=1}^{\widetilde{N}(\widetilde{l})/2} K_{h_{\widetilde{l}}}\big(\theta - \theta^{*(t_2(\widetilde{l})-k)}\big)\right\}\Bigg\}$$

$$\geq \frac{1}{2A^{l+1-l_1}} \frac{1}{\widetilde{N}(l^*)/2}$$

$$\times \min\left\{\sum_{k=1}^{\widetilde{N}(l^*)/2} K_{h_{l^*}}\big(\theta - \theta^{*(t_1(l^*)-k)}\big), \sum_{k=1}^{\widetilde{N}(l^*)/2} K_{h_{l^*}}\big(\theta - \theta^{*(t_2(l^*)-k)}\big)\right\},$$

where

$$l^* \equiv \arg \min_{\widetilde{l}:\, l_1-1\leq \widetilde{l}<l} \Bigg\{ \frac{1}{2A^{l+1-\widetilde{l}}} \frac{1}{\widetilde{N}(\widetilde{l})/2}$$

$$\times \min\left\{\sum_{k=1}^{\widetilde{N}(\widetilde{l})/2} K_{h_{\widetilde{l}}}\big(\theta - \theta^{*(t_1(\widetilde{l})-k)}\big), \sum_{k=1}^{\widetilde{N}(\widetilde{l})/2} K_{h_{\widetilde{l}}}\big(\theta - \theta^{*(t_2(\widetilde{l})-k)}\big)\right\}\Bigg\}.$$

Hence, we can show that the denominator divided by $\widetilde{N}(l)$ is bounded away from zero uniformly in Θ with probability arbitrarily close to 1 with sufficiently large l. Because the numerator divided by $\widetilde{N}(l)$ converges to zero uniformly in Θ, we have shown convergence. Therefore,

$$B_2(l, l-\Delta) \xrightarrow{P} 0$$

as $l \to \infty$. Next, we prove uniform convergence. Let

$$Y_m \equiv \sum_{j=0}^{t_{l_1}-t(l_1-1)} \sum_k K^*\big(t(l)-m, t(l_1)-j, t_{l_1}-j-1, k\big)$$

$$\times \sup_{s'\in S} \big|A_1^{(t(l)-m-i-1)}\big|.$$

Then

$$E\left[\sup_{\theta\in\Theta} \sum_{m=1}^{\widetilde{N}(l)} \frac{1}{\widetilde{N}(l)} K_{h(\widetilde{N}(l))}\big(\theta, \theta^{*(t(l)-m)}(q)\big) Y_m\right] \tag{A45}$$

$$\leq \varepsilon_1 \left(\frac{1}{2\pi}\right)^J \left\{4E\left[\frac{1}{\widetilde{N}(l)} \sum_{m=1}^{\widetilde{N}(l)} Y_m^2\right]\right\}^{1/2} \int |\psi(hz)|\, dz$$

and

$$\text{(A46)}\quad E[Y_m^2] \le E\Bigg[\sum_{j=1}^{t_{l_1}-t(l_1-1)} \sum_k K^*\big(t(l)-m, t(l_1)-j, t_{l_1}-j-1, k\big) \times \sup_{s'\in S}\big|A_1^{(t(l)-m-i-1)}\big|\Bigg]^2$$

$$= E\Bigg[\Bigg\{\sum_{t_{l-1}}\sum_k K^*(t(l)-m, t(l-1), t_{l-1}, k) \times \Bigg\{\sum_{t_{l-2}}\sum_k K^*(t_{l-1}, t(l-2), t_{l-2}, k)\cdots \times \Bigg\{\sum_{j=1}^{t_{l_1}-t(l_1-1)}\sum_k K^*\big(t_{l_1}, t_{l_1}-j, t_{l_1}-j-1, k\big) \times \sup_{s'\in S}\big|A_1^{(t_{l_1}-j-1)}\big|\Bigg\}\cdots\Bigg\}\Bigg\}\Bigg]^2.$$

Using the results in Claim 1, we can show that

$$\text{(A47)}\quad \text{RHS of (A46)}$$

$$\le \Bigg[\prod_{j=l_1}^{l}\sum_k \bigg(\frac{2\varepsilon_1^{k+1}}{(k-1)!}\bigg)^2 \Big\{\sup_{\theta'\in\Theta, h_l\le h\le h_{l_1-1}} E_\theta\big[K_h(\theta'-\theta^*(\widetilde{g}))\big]\Big\}^{2k}\Bigg] \times \Bigg[C_1^2 h_{l_1-1}^{3J} + \frac{\overline{A}^2 C_2}{\sqrt{N(l_1-1)h_{l_1}^{5J}}}\Bigg]$$

$$\le \Bigg[\prod_{j=l_1}^{l}\sum_k \frac{4\varepsilon_1^{2k+2}}{(k-1)!} \Big\{\sup_{\theta'\in\Theta, h_l\le h\le h_{l_1-1}} E_\theta\big[K_h(\theta'-\theta^*(\widetilde{g}))\big]\Big\}^{2k}\Bigg] \times \Bigg[C_1^2 h_{l_1-1}^{3J} + \frac{\overline{A}^2 C_2}{\sqrt{N(l_1-1)h_{l_1}^{5J}}}\Bigg]$$

$$\le [4\varepsilon_1^4 e^\lambda \lambda]^\Delta \Bigg[C_1^2 h_{l_1-1}^{3J} + \frac{\overline{A}^2 C_2}{\sqrt{N(l_1-1)h_{l_1}^{5J}}}\Bigg],$$

where

$$\lambda = \varepsilon_1^2 \sup_{\theta' \in \Theta, h_l \le h \le h_{l_1 - 1}} E_\theta\big[K_h(\theta' - \theta^*(\widetilde{g}))\big]^2.$$

Hence, substituting (A47) into (A45), we obtain

$$\begin{aligned} &\text{RHS of (A45)} \\ &\quad \le \varepsilon_1 \left(\frac{1}{2\pi}\right)^J 2[4\varepsilon_1^4 e^\lambda \lambda]^{\Delta/2} \sqrt{C_1^2 h_{l_1-1}^J + \frac{\overline{A}^2 C_2}{\sqrt{N(l_1 - 1) h_{l_1}^{9J}}}} \\ &\qquad \times \int |\psi(z)|\, dz \to 0 \quad \text{as } l_1 \to \infty. \end{aligned}$$

Therefore, the numerator of $B_2(l, l - \Delta)$ divided by $\widetilde{N}(l)$ converges to zero as $l \to \infty$. Since the denominator divided by $\widetilde{N}(l)$ is bounded away from zero uniformly in Θ with probability arbitrarily close to 0 for sufficiently large l, we have shown that $B_2(l, l - \Delta) \overset{P}{\to} 0$ uniformly in Θ. Thus, we have proved Lemma 6. *Q.E.D.*

Now, let

$$\begin{aligned} &\Delta V(m, n) \equiv \Delta V\big(\epsilon^{(t(m)-n)}, \theta^{*(t(m)-n)}\big), \\ &\Delta V(m) \equiv \big[\Delta V(m, 1), \dots, \Delta V(m, \widetilde{N}(m))\big], \\ &\overline{W}(l, k) \equiv \big[\overrightarrow{W}(l, t(l + 1 - k), t(l + 1 - k) - m)\big]_{m=1}^{\widetilde{N}(l+1-k)}. \end{aligned}$$

Then, iterating on (A37) and (A40), we obtain

$$\begin{aligned} \Delta V(l)'\overline{W}(l, 1) &\le A(l, l) + \Delta V(l - 1)'\overline{W}(l, 2) \\ &\le \cdots \le \sum_{i=0}^{k-1} A(l, l - i) + \Delta V(l - k)'\overline{W}(l, k + 1). \end{aligned}$$

By Lemma 6, given k, the first term on the RHS, $\sum_{i=0}^{k-1} A(l, l - i)$ converges to 0 in probability as $l \to \infty$ uniformly in Θ, and since $\Delta V(l + 1 - k)$ is uniformly bounded and $\overline{W}(l, k)'\iota \le \beta^k$ from (A39), the second term can be made arbitrarily small by choosing a large enough k. Therefore, $\Delta V(l)'\overline{W}(l, 1)$ converges to zero in probability as $l \to \infty$ uniformly in Θ. Hence, we have shown that

$$\big|A_2^{(t(l))}(\theta)\big| \overset{P}{\to} 0$$

uniformly in Θ. Therefore,

$$\begin{aligned}
&\left|EV(s,\epsilon,\theta)-\widehat{E}^{(t(l))}[V(s,\epsilon,\theta)]\right| \\
&\quad\le\left|\int V(s',\epsilon',\theta)\,dF_{\epsilon'}(\epsilon',\theta)\right. \\
&\qquad\left.-\sum_{n=1}^{N(t)}V^{(t-n)}\big(s',\epsilon^{(t-n)},\theta^{*(t-n)}\big)W_{N(t),h}\big(\theta,\theta^{*(t-n)}\big)\right| \\
&\quad\le\big[\big|A_1^{(t(l))}\big|+\big|A_2^{(t(l))}\big|\big]/\beta\xrightarrow{P}0
\end{aligned}$$

uniformly in Θ.

LEMMA 7: *We have*

$$\left|V(s,\epsilon,\theta)-V^{(t)}(s,\epsilon,\theta)\right|\xrightarrow{P}0\quad\textit{as } t\to\infty \textit{ uniformly in } \Theta.$$

PROOF: Because both $A_1^{(t(l))}(\theta)$ and $\Delta V(l)'\overline{W}(l,1)$ converge to zero in probability as $l\to\infty$ uniformly in Θ, then

$$\begin{aligned}
&\left|V(s,\epsilon,\theta)-V^{(t(l))}(s,\epsilon,\theta)\right| \\
&\quad\le\sup_{s'\in S}\left|A_1^{(t(l))}(\theta)\right| \\
&\qquad+\beta\sum_{n=1}^{N(t(l))}\Delta V\big(\epsilon^{(t(l)-n)},\theta^{*(t(l)-n)}\big)W_{N(t(l)),h}\big(\theta,\theta^{*(t(l)-n)}\big)\xrightarrow{P}0.
\end{aligned}$$

From Lemma 3,

$$\sup_{s'\in S}\left|A_1^{(t)}(\theta)\right|\to 0$$

uniformly in Θ in probability as $t\to\infty$. That is, for any $\delta>0$ and $\eta>0$, there exists L such that for any $l\ge L$,

$$(*)\qquad \mathrm{Prob}\Big(\sup_{\theta\in\Theta}\left|V(s,\epsilon,\theta)-V^{(t(l))}(s,\epsilon,\theta)\right|>\delta\Big)<\eta,$$

$$\mathrm{Prob}\left(\sum_{n=1}^{N(t(l))}\Delta V\big(\epsilon^{(t(l)-n)},\theta^{*(t(l)-n)}\big)W_{N(t(l)),h}\big(\theta,\theta^{*(t(l)-n)}\big)>\delta\right)<\eta,$$

and, for any $t\ge t(l)$,

$$\mathrm{Prob}\Big(\sup_{s'\in S,\theta\in\Theta}\left|A_1^{(t)}(\theta)\right|>\delta\Big)<\eta.$$

Now, for $t(l)+1$,

$$\begin{aligned}
&\left|V(s,\epsilon,\theta)-V^{(t(l)+1)}(s,\epsilon,\theta)\right| \\
&\quad\leq \sup_{s'\in S}\left|A_1^{(t(l)+1)}(\theta)\right| \\
&\qquad + \beta \sum_{n=1}^{N(t(l)+1)} \Delta V\big(\epsilon^{(t(l)+1-n)}, \theta^{*(t(l)+1-n)}\big) W_{N(t(l)+1),h}\big(\theta, \theta^{*(t(l)+1-n)}\big) \\
&\quad\leq \sup_{s'\in S}\left|A_1^{(t(l)+1)}(\theta)\right| + \beta \Delta V^{t(l)}\big(\epsilon^{(t(l))}, \theta^{(t(l))}\big)\widetilde{W}(t(l)+1, t(l)) \\
&\qquad + \beta \sum_{n=1}^{N(t(l))} \Delta V\big(\epsilon^{(t(l)-n)}, \theta^{*(t(l)-n)}\big) W_{N(t(l)+1),h}\big(\theta, \theta^{*(t(l)-n)}\big).
\end{aligned}$$

Now, for large enough $L' \geq L$, for any $l \geq L'$,

$$h(\widetilde{N}(l)+1)^{9J}\widetilde{N}(l) \geq h(\widetilde{N}(L))^{9J}\widetilde{N}(L).$$

This is true from the assumption $h(\widetilde{N}(k+1))^{9J}\widetilde{N}(k+1) \to \infty$ and

$$\frac{\widetilde{N}(k)}{\widetilde{N}(k)+1} \to 1$$

as $k \to \infty$. Furthermore, because $h(N)$ is nonincreasing in N,

$$h(\widetilde{N}(l)+1)^{kJ}\widetilde{N}(l) \geq h(\widetilde{N}(L))^{kJ}\widetilde{N}(L)$$

for all $0 \leq k \leq 9$. Now define the new bandwidth $\widehat{h}$, which is $\widehat{h}(\widetilde{N}(j)) = h(\widetilde{N}(j))$ for $j = 1, \ldots, l-1, l+1, \ldots$ and $\widehat{h}(\widetilde{N}(l)) = h(\widetilde{N}(l+1))$. Then, because the bandwidth is the same as before until $j = l$ and also because the new bandwidth satisfies the same assumptions as before, Lemma 6 holds with the new bandwidth. Therefore,

$$\mathrm{Prob}\Big(\sup_{\theta\in\Theta}\left|V(s,\epsilon,\theta)-V_{\widehat{h}}^{(t(l))}(s,\epsilon,\theta)\right| > \delta\Big) > \eta$$

and, similarly,

$$\mathrm{Prob}\left(\sum_{n=1}^{N(t(l))} \Delta V\big(\epsilon^{(t(l)-n)}, \theta^{*(t(l)-n)}\big) W_{N(t(l)),\widehat{h}}\big(\theta, \theta^{*(t(l)-n)}\big) > \delta\right) < \eta.$$

Now, by construction

$$
\begin{aligned}
&W_{N(t(l)+1),h}\big(\theta-\theta^{*(t(l)-n)}\big)\\
&\quad=\frac{K_{h(\widetilde{N}(l)+1)}(\theta-\theta^{*(t(l)-n)})}{\sum_{k=1}^{\widetilde{N}(l)+1}K_{h(\widetilde{N}(l)+1)}(\theta-\theta^{*(t(l)+1-k)})}\\
&\quad=\frac{K_{h(\widetilde{N}(l)+1)}(\theta-\theta^{*(t(l)-n)})}{\sum_{k=1}^{\widetilde{N}(l)}K_{h(\widetilde{N}(l)+1)}(\theta-\theta^{*(t(l)-k)})+K_{h(\widetilde{N}(l)+1)}(\theta-\theta^{*(t(l))})}.
\end{aligned}
$$

Hence,

$$
0\le W_{N(t(l)+1),h}\big(\theta,\theta^{*(t(l)-n)}\big)\le W_{N(t(l)),\widehat{h}}\big(\theta,\theta^{*(t(l)-n)}\big).
$$

Therefore, for the original bandwidth h,

$$
\begin{aligned}
&\operatorname{Prob}\Big(\sup_{\theta\in\Theta}\big|V(s,\epsilon,\theta)-V_{\widehat{h}}^{(t(l)+1)}(s,\epsilon,\theta)\big|>2\delta\Big)\\
&\quad\le\operatorname{Prob}\Big(\sup_{s'\in S}\big|A_1^{(t(l)+1)}(\theta)\big|>\delta\Big)\\
&\qquad+\operatorname{Prob}\Bigg(\beta\sum_{n=1}^{N(t(l))}\Delta V\big(\epsilon^{(t(l)-n)},\theta^{*(t(l)-n)}\big)\\
&\qquad\times W_{N(t(l)+1),h}\big(\theta,\theta^{*(t(l)-n)}\big)>\delta\Bigg)\\
&\quad<2\eta.
\end{aligned}
$$

Therefore, $\Delta V(\epsilon^{(t(l)+1)},\theta)$ also converges to zero in probability uniformly in $\theta\in\Theta$.

Let $t=t(l)+1$. Now, from Lemma 4,

$$
\begin{aligned}
&\Delta V^{(t+\underline{N})}(\epsilon,\theta)\\
&\quad\le\sup_{s'\in S}\big|A_1^{(t+\underline{N})}(\theta)\big|\\
&\qquad+\sum_{m=0}^{\underline{N}-1}\widehat{W}(t+\underline{N},t+\underline{N}-m,t+\underline{N}-m-1)
\end{aligned}
$$

$$\times \sup_{s' \in S} \left| A^{(t+\underline{N}-m-1)}\left(\theta^{*(t+\underline{N}-m-1)}\right)\right|$$
$$+ \sum_{n=1}^{N(t)} \Delta V^{(t-n)}\left(\epsilon^{(t-n)}, \theta^{*(t-n)}\right) \widehat{W}(t+\underline{N}, t, t-n).$$

Then, for any $M > 0$,

$$\begin{aligned}
&\widehat{W}(t+M, t, t-n) \\
&\quad = \widehat{W}(t+M, t+1, t)\widetilde{W}(t, t-n) + \widehat{W}(t+M, t+1, t-n) \\
&\quad = \widehat{W}(t+M, t+1, t)\widetilde{W}(t, t-n) \\
&\qquad + \widehat{W}(t+M, t+2, t+1)\widetilde{W}(t+1, t-n) \\
&\qquad + \widehat{W}(t+M, t+2, t-n) \\
&\quad = \cdots = \sum_{m=0}^{M} \widehat{W}(t+M, t+m+1, t+m)\widetilde{W}(t+m, t-n),
\end{aligned}$$

where

$$\widehat{W}(t+\underline{N}, t, t-n) = \sum_{m=0}^{\underline{N}} \widehat{W}(t+\underline{N}, t+m+1, t+m) \times \widetilde{W}(t+m, t-n).$$

We have already shown that

$$\sup_{s' \in S} \left| A_1^{(t+\underline{N})}(\theta)\right| + \sum_{m=0}^{\underline{N}-1} \widehat{W}(t+\underline{N}, t+\underline{N}-m, t+\underline{N}-m-1) \times \sup_{s' \in S} \left| A^{(t+\underline{N}-m-1)}\left(\theta^{*(t+\underline{N}-m-1)}\right)\right| \to 0$$

in probability as $t \to \infty$. We know from the proof of Lemma 5 that

$$\sum_{m=0}^{\underline{N}} \widehat{W}(t+\underline{N}, t+m+1, t+m)$$

is uniformly bounded with probability that can be made arbitrarily close to 1 with sufficiently large t.

Furthermore, because $h(N(t+m)) = h(N(t))$,

$$\begin{aligned}
\widetilde{W}(t+m, t-n) &= \beta W_{N(t+m),h}\left(\theta^{*(t+m)} - \theta^{*(t-n)}\right) \\
&\leq \beta W_{N(t),h}\left(\theta^{*(t+m)} - \theta^{*(t-n)}\right).
\end{aligned}$$

Therefore,

$$\sum_{n=1}^{N(t)} \Delta V^{(t-n)}\big(\epsilon^{(t-n)}, \theta^{*(t-n)}\big)\widehat{W}(t+\underline{N}, t, t-n)$$
$$\leq \sum_{m=0}^{\underline{N}} \widehat{W}(t+\underline{N}, t+m+1, t+m)$$
$$\times \left\{\sum_{n=1}^{N(t)} \Delta V^{(t-n)}\big(\epsilon^{(t-n)}, \theta^{*(t-n)}\big)\beta W_{N(t),h}\big(\theta^{*(t+m)} - \theta^{*(t-n)}\big)\right\}.$$

Because we have shown that

$$\sum_{n=1}^{N(t)} \Delta V^{(t-n)}\big(\epsilon^{(t-n)}, \theta^{*(t-n)}\big)\beta W_{N(t),h}\big(\theta - \theta^{*(t-n)}\big),$$

$t = t(l) + 1$ converges to zero uniformly in probability in θ as $l \to \infty$, and because

$$\sum_{m=0}^{\underline{N}} \widehat{W}(t+\underline{N}, t+m+1, t+m)$$

is uniformly bounded with probability arbitrarily close to 1 for large enough t, we have shown that for any $\delta > 0$, $\eta > 0$, there exists sufficiently large L such that for any $l \geq L$,

$$\Pr\left(\sum_{n=1}^{N(t)} \Delta V^{(t-n)}\big(\epsilon^{(t-n)}, \theta^{*(t-n)}\big)\widehat{W}(t+\underline{N}, t, t-n) < \delta\right) > \eta$$

holds for any $\underline{N} : 0 \leq \underline{N} \leq \widetilde{N}(l+1)$. Therefore, for any $\delta > 0$, $\eta > 0$, there exists sufficiently large L such that for any $l \geq L$,

$$\Pr\Big(\sup_{\theta\in\Theta} \Delta V^{(t+\underline{N})}(\epsilon, \theta) < \delta\Big) > \eta$$

for any $\underline{N} : 0 \leq \underline{N} \leq \widetilde{N}(l+1)$ and we have proved Lemma 7. *Q.E.D.*

This also concludes the proof of Theorem 1. *Q.E.D.*

COROLLARY 1: *Suppose Assumptions* 1–8 *are satisfied. Then Theorem* 1 *implies that* $\lambda(\theta^{(t)}, \theta^{*(t)}|\mathcal{H}^{(t-1)})$ *converges to* $\lambda(\theta^{(t)}, \theta^{*(t)})$ *in probability uniformly.*

PROOF: From Assumption 5, $L(Y_{N^d,T^d}|\theta, V(\cdot,\theta))$ is uniformly continuous and strictly positive. Hence, $\lambda(\theta,\theta^*)$ is also uniformly continuous in θ, $\theta^* \in \Theta$. Since $V^{(t)} \to V$ in probability uniformly in s, Θ, then $L(Y_{N^d,T^d}|\theta, V^{(t)}(\cdot,\theta))$ is stochastically equicontinuous. To see this, note that

$$\begin{aligned}&\bigl|L\bigl(Y_{N^d,T^d}|\theta, V^{(t)}(\cdot,\theta)\bigr) - L\bigl(Y_{N^d,T^d}|\theta', V^{(t)}(\cdot,\theta')\bigr)\bigr|\\&\quad\le \bigl|L\bigl(Y_{N^d,T^d}|\theta,\bigl[V^{(t)}(\cdot,\theta)\bigr]\bigr) - L(Y_{N^d,T^d}|\theta', V(\cdot,\theta))\bigr|\\&\qquad + \bigl|L\bigl(Y_{N^d,T^d}|\theta,[V(\cdot,\theta)]\bigr) - L(Y_{N^d,T^d}|\theta', V(\cdot,\theta'))\bigr|\\&\qquad + \bigl|L\bigl(Y_{N^d,T^d}|\theta',[V(\cdot,\theta')]\bigr) - L\bigl(Y_{N^d,T^d}|\theta', V^{(t)}(\cdot,\theta')\bigr)\bigr|.\end{aligned}$$

Hence,

$$\begin{aligned}\text{(A51)}\quad&\Pr\bigl(\bigl|L\bigl(Y_{N^d,T^d}|\theta, V^{(t)}(\cdot,\theta)\bigr) - L\bigl(Y_{N^d,T^d}|\theta', V^{(t)}(\cdot,\theta')\bigr)\bigr| \ge \eta\bigr)\\&\quad\le \Pr\bigl(\bigl|L\bigl(Y_{N^d,T^d}|\theta,\bigl[V^{(t)}(\cdot,\theta)\bigr]\bigr) - L\bigl(Y_{N^d,T^d}|\theta, V(\cdot,\theta)\bigr)\bigr| \ge \eta/3\bigr)\\&\qquad + \Pr\bigl(\bigl|L\bigl(Y_{N^d,T^d}|\theta,[V(\cdot,\theta)]\bigr) - L(Y_{N^d,T^d}|\theta', V(\cdot,\theta'))\bigr| \ge \eta/3\bigr)\\&\qquad + \Pr\bigl(\bigl|L\bigl(Y_{N^d,T^d}|\theta',[V(\cdot,\theta')]\bigr) - L\bigl(Y_{N^d,T^d}|\theta', V^{(t)}(\cdot,\theta')\bigr)\bigr| \ge \eta/3\bigr).\end{aligned}$$

Furthermore, because of uniform continuity of the likelihood, there exists $\delta > 0$ such that for any θ, $\theta' \in \Theta$, $|\theta - \theta'| < \delta$,

$$\text{(A52)}\quad \bigl|L\bigl(Y_{N^d,T^d}|\theta,[V(\cdot,\theta)]\bigr) - L(Y_{N^d,T^d}|\theta', V(\cdot,\theta'))\bigr| < \eta/3.$$

Because $V^{(t)} \to V$ in probability uniformly, for any $\kappa > 0$ there exists positive T such that for any $t > T$,

$$\begin{aligned}\text{(A53)}\quad&\Pr\Bigl(\sup_{\theta\in\Theta}\bigl|L\bigl(Y_{N^d,T^d}|\theta,\bigl[V^{(t)}(\cdot,\theta)\bigr]\bigr) - L(Y_{N^d,T^d}|\theta, V(\cdot,\theta))\bigr| \ge \eta/3\Bigr)\\&\quad< \kappa/2.\end{aligned}$$

(A51), (A52), and (A53) imply that for any $\kappa > 0$, $\eta > 0$, there exist positive δ and T such that for any $t > T$, for any θ, $\theta' \in \Theta$, $|\theta - \theta'| < \delta$,

$$\Pr\bigl(\bigl|L\bigl(Y_{N^d,T^d}|\theta, V^{(t)}(\cdot,\theta)\bigr) - L\bigl(Y_{N^d,T^d}|\theta', V^{(t)}(\cdot,\theta')\bigr)\bigr| \ge \eta\bigr) < \kappa.$$

Therefore, $L(Y_{N^d,T^d}|\theta, V^{(t)}(\cdot,\theta))$ is stochastically equicontinuous. Because the likelihood is strictly positive in Θ, this implies that $\lambda(\theta,\theta^*|\mathcal{H}^{(t)})$ is also stochastically equicontinuous. Hence, all the conditions for Theorem 2.1 of Newey (1991) are satisfied. Therefore, $\lambda(\theta,\theta^*|\mathcal{H}^{(t)})$ converges to $\lambda(\theta,\theta^*)$ uniformly in θ, $\theta^* \in \Theta$. *Q.E.D.*

THEOREM 2: *Suppose Assumptions* 1–8 *are satisfied for* $V^{(t)}$, $t = 1,\dots,\pi$, L, ϵ, *and* θ. *Suppose* $\theta^{(t)}$, $t = 1,\dots,$ *is generated by the modified Metropolis–Hastings algorithm described earlier, where* $\lambda(\theta^{(t)},\theta^{*(t)}|\mathcal{H}^{(t-1)})$ *converges to*

$\lambda(\theta^{(t)}, \theta^{*(t)})$ in probability uniformly. Then $\theta^{(t)}$ converges to $\widetilde{\theta}^{(t)}$ in probability, where $\widetilde{\theta}^{(t)}$ is a Markov chain generated by the Metropolis–Hastings algorithm with proposal density $q(\theta, \theta^{(*)})$ and acceptance probability function $\lambda(\theta, \theta^{(*)})$.

PROOF: We are given a random process with transition probability $f^{(t)}(\cdot, \cdot)$ which is

$$f^{(t)}(\theta^{(t)}, \theta') = \lambda(\theta^{(t)}, \theta' | \mathcal{H}^{(t-1)}) q(\theta^{(t)}, \theta') + \left[1 - \int \lambda(\theta^{(t)}, \theta | \mathcal{H}^{(t-1)}) q(\theta^{(t)}, \theta) \, d\theta\right] \delta_{\theta^{(t)}}(\theta'),$$

where $\delta_{\theta^{(t)}}$ is the Dirac mass at $\theta^{(t)}$. Because $\lambda(\theta, \theta' | \mathcal{H}^{(t-1)})$ converges to $\lambda(\theta, \theta')$ in probability uniformly on $\theta, \theta' \in \Theta$, $f^{(t)}(\cdot, \cdot)$ converges to $f(\cdot, \cdot)$ in probability uniformly as $t \to \infty$. Because both $\lambda(\theta, \cdot | \mathcal{H}^{(t-1)})$ and $q(\theta, \cdot)$ are strictly positive functions for any $\theta \in \Theta$, using similar arguments as in the proof of Lemma 1, we can construct a density $g(\cdot)$ and a constant $\varepsilon_0 > 0$ such that for any $\theta \in \Theta$,

$$f^{(t)}(\theta, \cdot) \geq \varepsilon_0 g(\cdot),$$
$$f(\theta, \cdot) \geq \varepsilon_0 g(\cdot).$$

Define

$$v^{(t)} = \min\left\{\inf_{\theta' \in \Theta}\left\{\frac{f^{(t)}(\theta, \theta')}{f(\theta, \theta')}\right\}, 1\right\}.$$

Then

$$f^{(t)}(\theta, \cdot) \geq v^{(t)} f(\theta, \cdot),$$
$$f(\theta, \cdot) \geq v^{(t)} f(\theta, \cdot).$$

Then, because of uniform convergence of $f^{(t)}$ to f in probability, $v^{(t)}$ converges to 1 in probability.

Now, construct the following coupling scheme. Let $X^{(t)}$ be a random variable that follows the transition probability $f^{(t)}(x, \cdot)$ given $X^{(t-1)} = x$, and let $Y^{(t)}$ be a Markov process that follows the transition probability $f(y, \cdot)$ given $Y^{(t-1)} = y$. Suppose $X^{(t)} \neq Y^{(t)}$. With probability $\varepsilon_0 > 0$, let

$$X^{(t+1)} = Y^{(t+1)} = Z^{(t+1)} \sim g(\cdot),$$

and with probability $1 - \varepsilon_0$, let

$$X^{(t+1)} \sim \frac{1}{1 - \varepsilon_0}\left[f^{(t)}(X^{(t)}, \cdot) - \varepsilon_0 g(\cdot)\right],$$

$$Y^{(t+1)} \sim \frac{1}{1-\varepsilon_0}\big[f\big(Y^{(t)},\cdot\big) - \varepsilon g(\cdot)\big].$$

Suppose $X^{(t)} = Y^{(t)} = Z^{(t)}$. With probability $v^{(t)}$,

$$X^{(t+1)} = Y^{(t+1)} \sim f\big(Z^{(t)},\cdot\big),$$

and with probability $(1 - v^{(t)})$,

$$X^{(t+1)} \sim \frac{1}{1-v^{(t)}}\big[f^{(t)}\big(X^{(t)},\cdot\big) - v^{(t)} f\big(Z^{(t)},\cdot\big)\big],$$

$$Y^{(t+1)} \sim \frac{1}{1-v^{(t)}}\big[f\big(Y^{(t)},\cdot\big) - v^{(t)} f\big(Z^{(t)},\cdot\big)\big].$$

Let $w^{(t)} = 1 - v^{(t)}$. Then $w^{(t)} \stackrel{P}{\to} 0$ as $t \to \infty$. Let $S^{(t)} \in \{1, 2\}$ be the state at iteration t, where state 1 is assumed to be the state in which $X^{(t)} = Y^{(t)}$ and state 2 is the state in which $X^{(t)} \neq Y^{(t)}$. Then $S^{(t)}$ follows the Markov process with the transition matrix

$$P = \begin{bmatrix} 1 - w^{(t)} & w^{(t)} \\ \varepsilon_0 & 1 - \varepsilon_0 \end{bmatrix}.$$

Denote the unconditional probability of state 1 at time t as $\rho^{(t)}$. Then

$$\big[\rho^{(t+1)}, 1 - \rho^{(t+1)}\big] = \big[\rho^{(t)}, 1 - \rho^{(t)}\big] \begin{bmatrix} 1 - w^{(t)} & w^{(t)} \\ \varepsilon_0 & 1 - \varepsilon_0 \end{bmatrix}.$$

Hence,

$$\begin{aligned} \rho^{(t+1)} &= \rho^{(t)}\big[\big(1 - w^{(t)}\big) - \varepsilon_0\big] + \varepsilon_0 \\ &\geq \rho^{(t)}(1 - \varepsilon_0) + \varepsilon_0 - w^{(t)} \\ &\geq \rho^{(t-m)}(1 - \varepsilon_0)^{m+1} + 1 - (1 - \varepsilon_0)^{m+1} \\ &\quad - \big[w^{(t)} + (1 - \varepsilon_0)w^{(t-1)} + \cdots + (1 - \varepsilon_0)^m w^{(t-m)}\big]. \end{aligned}$$

We now prove that $\rho^{(t)} \stackrel{P}{\to} 1$. Define

$$W_{tm} = w^{(t)} + (1 - \varepsilon_0)w^{(t-1)} + \cdots + (1 - \varepsilon_0)^m w^{(t-m)}.$$

Because $w^{(t)} \stackrel{P}{\to} 0$, for any $\delta_1 > 0$, $\delta_2 > 0$, there exists $N > 0$ such that for any $t \geq N$,

$$\Pr\big[\big|w^{(t)} - 0\big| < \delta_1\big] > 1 - \delta_2.$$

Now, given any $\overline{\delta}_1 > 0$, $\overline{\delta}_2 > 0$, let m be such that,

$$(1-\varepsilon_0)^m < \frac{\overline{\delta}_1}{5}.$$

Also, let δ_1 satisfy $\delta_1 < \frac{\overline{\delta}_1}{5(m+1)}$ and let δ_2 satisfy $\delta_2 < \frac{\overline{\delta}_2}{m+1}$. Then

$$\text{(A54)}\quad \Pr\left\{|W_{tm}-0| < \frac{\overline{\delta}_1}{5}\right\} \geq \Pr\left\{\bigcap_{j=t-m}^{t} \left|w^{(j)}-0\right| < \delta_1\right\}$$
$$= 1-\Pr\left\{\bigcup_{j=t-m}^{t} \left|w^{(j)}-0\right| \geq \delta_1\right\}$$
$$\geq 1-\sum_{j=t-m}^{t}\Pr\left\{\left|w^{(j)}-0\right| \geq \delta_1\right\} \geq 1-\overline{\delta}_2.$$

Now, let $\overline{N}$ be defined as $\overline{N} = \max\{N, m\}$. Then, for each $k > \overline{N}$,

$$\text{(A55)}\quad \Pr\left[\left|\rho^{(t+1)}-1\right| < \overline{\delta}_1\right]$$
$$= \Pr\left[\left|\rho^{(t-m)}(1-\varepsilon_0)^m - (1-\varepsilon_0)^{m+1} + W_{tm}\right| < \overline{\delta}_1\right]$$
$$\geq \Pr\left[\left|\rho^{(t-m)}(1-\varepsilon_0)^m - (1-\varepsilon_0)^{m+1}\right| < \frac{2\overline{\delta}_1}{5}, |W_{tm}| < \frac{\overline{\delta}_1}{5}\right].$$

Because $0 \leq \rho^{(t-m)} \leq 1$,

$$\left|\rho^{(t-m)}(1-\varepsilon_0)^m - (1-\varepsilon_0)^{m+1}\right| \leq |2(1-\varepsilon_0)^m| \leq \frac{2\overline{\delta}_1}{5}.$$

Hence,

$$\text{RHS of (A55)} = \Pr\left[|W_{tm}| < \frac{\overline{\delta}_1}{5}\right] \geq 1-\overline{\delta}_2.$$

Therefore, $\rho^{(t)}$ converges to 1 in probability.

Therefore, for any $\delta > 0$, there exists M such that for any $t > M$,

$$\Pr\left[X^{(t)} = Y^{(t)}\right] > 1-\delta.$$

Since $Y^{(t)}$follows a stationary distribution, $X^{(t)}$ converges to a stationary process in probability and we have proved the theorem. *Q.E.D.*

COROLLARY 2: *The sequence of parameter simulations generated by the Metropolis–Hastings algorithm with proposal density* $q(\theta, \theta^*)$ *and acceptance probability* $\lambda(\theta, \theta^*)$ *converges to the true posterior in total variation norm. That is,*

$$\lim_{n\to\infty}\left\| \int K^n(\theta, \cdot)\mu_0(d\theta) - \mu \right\|_{\mathrm{TV}} = 0$$

for arbitrary initial distribution μ_0, *where* μ *is the true posterior distribution and* $K^n(\theta, \cdot)$ *is the transition kernel for n iterations.*

PROOF: Here, we use Corollary 7.7 of Robert and Casella (2004), which states the following: Assume the Metropolis–Hastings Markov chain has invariant probability density μ and the proposal density satisfies the property that there exist positive that ϵ and δ such that $q(x, y) > \epsilon$ if $|x - y| < \delta$. Then for any $h \in L^1(\mu)$,

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} h\big(\theta^{(t)}\big) = \int h(\theta)\mu(\theta)\, d\theta,$$

$$\lim_{n\to\infty}\left\| \int K^n(\theta, \cdot)\mu_0(d\theta) - \mu \right\|_{\mathrm{TV}} = 0$$

for arbitrary initial distribution μ_0, where $K^n(\theta, \cdot)$ is the transition kernel for n iterations, which in this case is defined by the proposal density q and the acceptance rate λ, and the norm is the total variation norm. By construction, the Metropolis–Hastings Markov chain has an invariant probability density, which is proportional to $\pi(\theta)L(Y_{N^d,T^d}|\theta)$, which is assumed to be bounded and positive on Θ. Since the proposal density is strictly positive over the parameter space, the condition for the proposal density is also satisfied. *Q.E.D.*

APPENDIX C: FIGURES

We present the MCMC plots of the estimation of the random effects model. In Figures 1–9, we plotted one of every ten MCMC iterations. We can see that there is sufficient movement of the parameters around the posterior mean and the serial correlations do not seem to be excessively large.

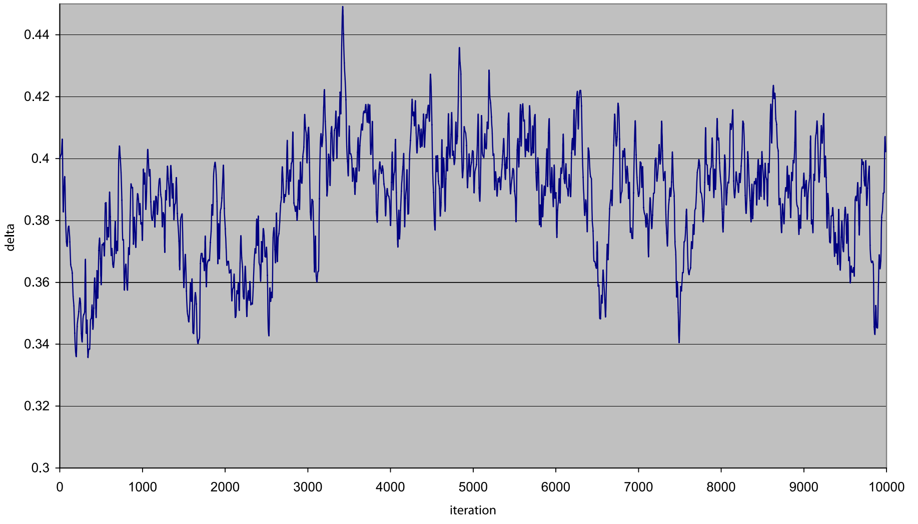

FIGURE 1.—Entry cost (true value = 0.4).

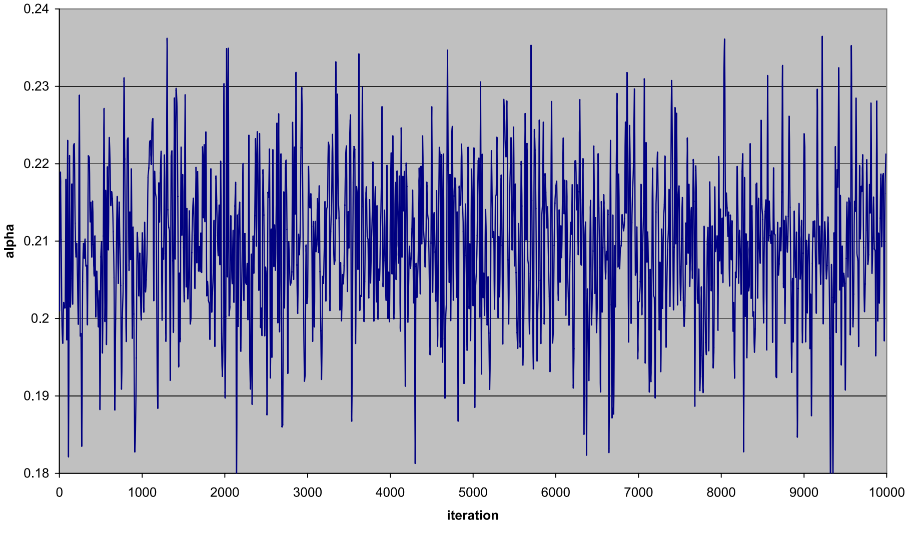

FIGURE 2.—Mean of capital coefficient (true value = 0.2).

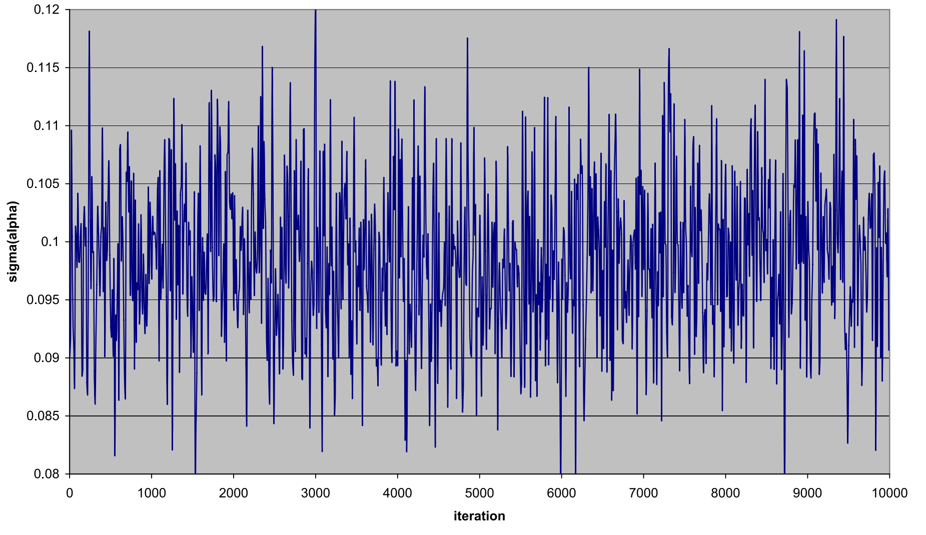

FIGURE 3.—Standard error of capital coefficient (true value = 0.1).

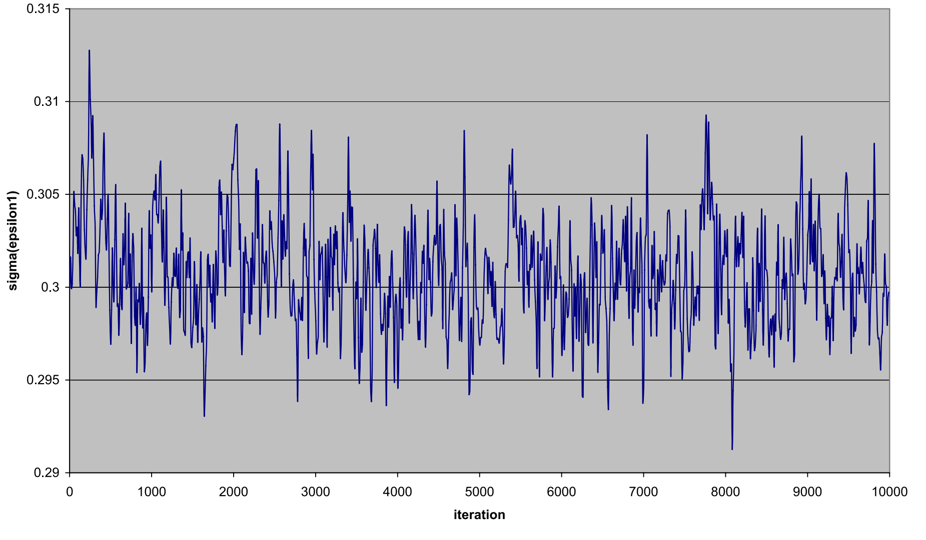

FIGURE 4.—Standard error of ϵ_1 (true value = 0.3).

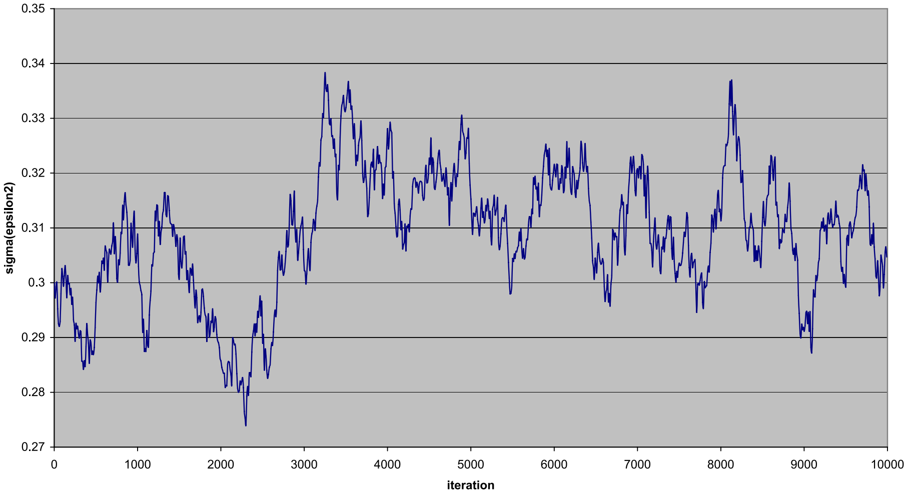

FIGURE 5.—Standard error of ϵ_2 (true value $= 0.3$).

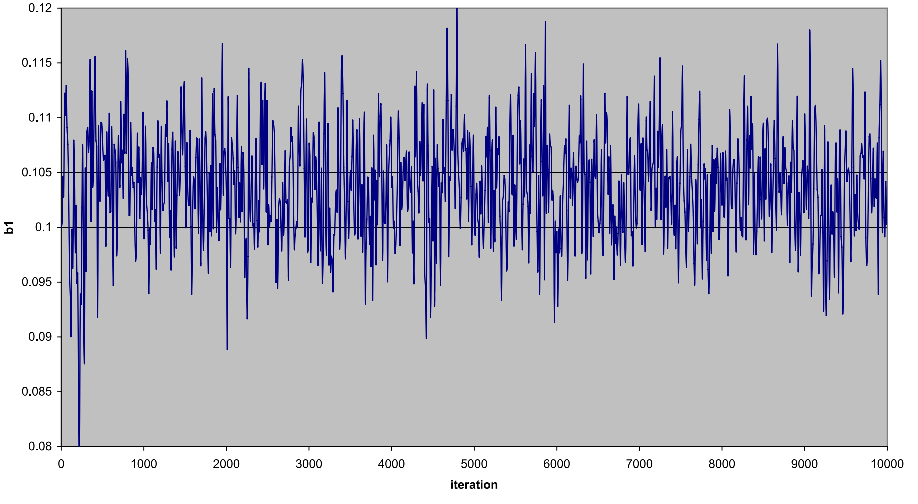

FIGURE 6.—Capital transition coefficient b_1 (true value $= 0.1$).

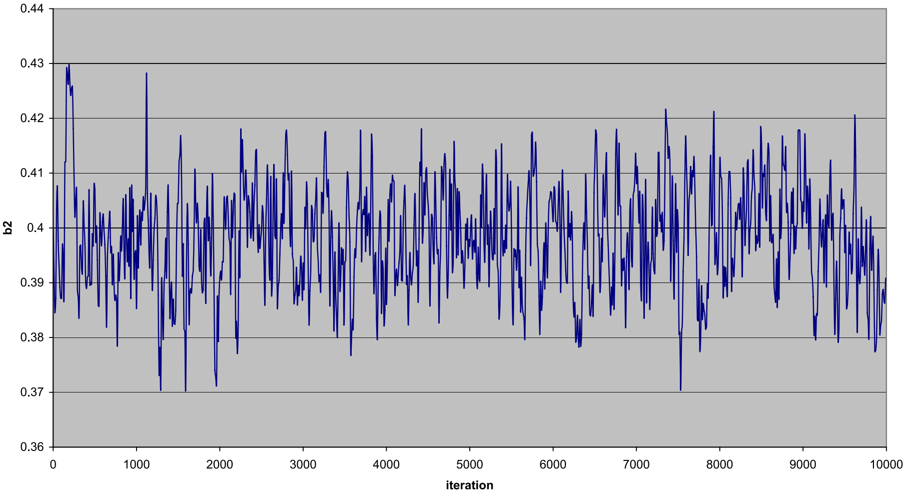

FIGURE 7.—Capital transition coefficient b_2 (true value = 0.4).

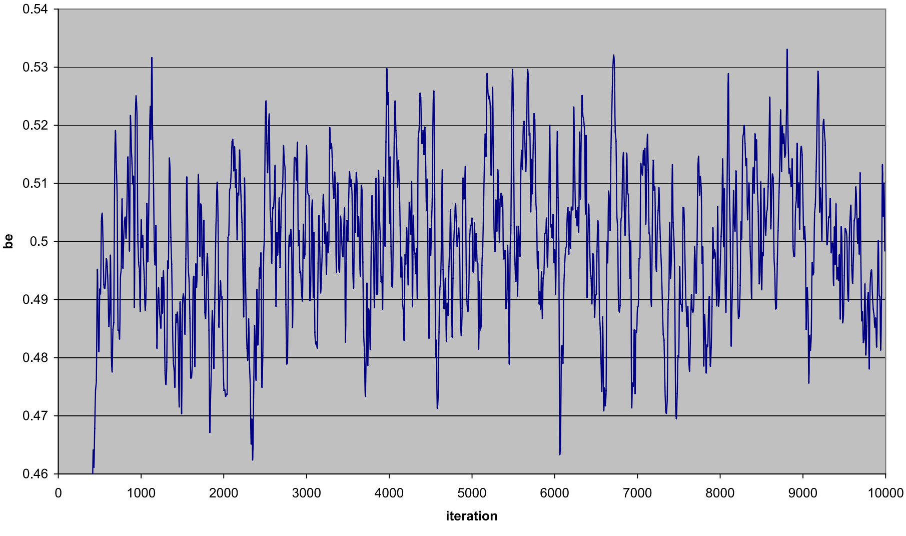

FIGURE 8.—Initial capital parameter (true value = 0.5).

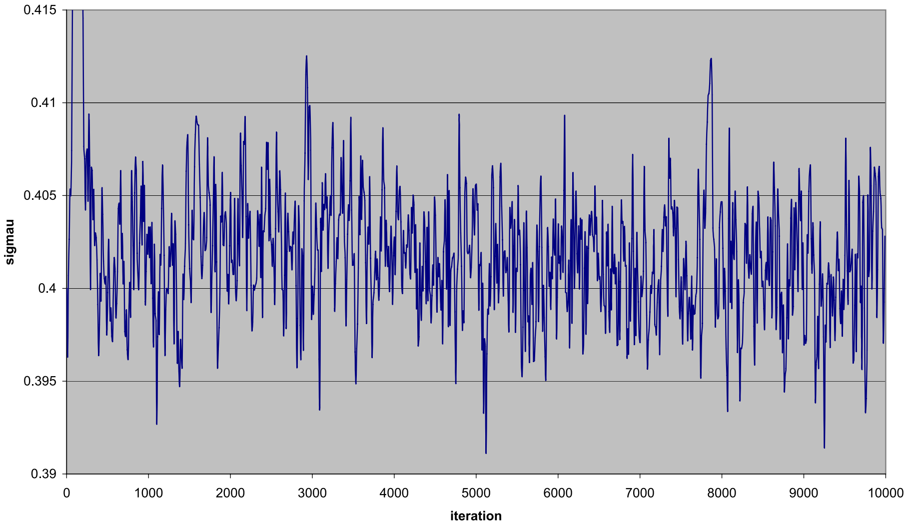

FIGURE 9.—Standard error of transition shock (true value = 0.4).

REFERENCES

BIERENS, H. J. (1994): *Topics in Advanced Econometrics, Estimation, Testing and Specification of Cross-Section and Time Series Models.* Cambridge, MA: Cambridge University Press. [10,13,24, 44]

NEWEY, W. K. (1991): "Uniform Convergence in Probability and Stochastic Equicontinuity," *Econometrica*, 59, 1161–1167. [56]

ROBERT, C. P., AND G. CASELLA (2004): *Monte Carlo Statistical Methods*. *Springer Texts in Statistics*. New York: Springer-Verlag. [60]

Dept. of Economics, Queen's University, 233 Dunning Hall, 94 University Avenue, Kingston, ON K7L 5M2, Canada; imais@econ.queensu.ca,

Dept. of Economics, City University London, Northampton Square, London EC1V 0HB, U.K. and Dept. of Economics, Northern Illinois University, 508 Zulauf Hall, DeKalb, IL 60115, U.S.A.; njain@niu.edu,

and

Rotman School of Management, University of Toronto, 105 St. George Street, Toronto, ON M5S 3E6, Canada; andrew.ching@rotman.utoronto.ca.

Manuscript received January, 2005; final revision received May, 2009.