Supplement to
"A Sieve-SMM Estimator for Dynamic Models"

Jean-Jacques Forneron[*]

January 17, 2023

This Supplemental Material consists of Appendices C, D, E, F and G to the main text.

[*]Department of Economics, Boston University, 270 Bay State Road, Boston, MA 02215.
Email: jjmf@bu.edu.

Appendix C Proofs for the Preliminary Results

Proof of Lemma A1. The proof proceeds by recursion. Denote $\Pi_{k(n)}f_j \in \mathcal{F}_{k(n)}$ the mixture approximation of f_j from Lemma D7. For $d_e = 1$, Lemma D7 implies $\|f_1 - \Pi_{k(n)}f_1\|_{TV} = O(\frac{\log[k(n)]^{r/b}}{k(n)^r})$ and $\|f_1 - \Pi_{k(n)}f_1\|_\infty = O(\frac{\log[k(n)]^{r/b}}{k(n)^r})$. Suppose the result holds for $f_1 \times \cdots \times f_{d_e}$. Let $f = f_1 \times \cdots \times f_{d_e} \times f_{d_e+1}$; let:

$$d_{t+1} = f_1 \times \cdots \times f_{d_e} \times f_{d_e+1} - \Pi_{k(n)}f_1 \times \cdots \times \Pi_{k(n)}f_{d_e} \times \Pi_{k(n)}f_{d_e+1}$$
$$d_t = f_1 \times \cdots \times f_{d_e} - \Pi_{k(n)}f_1 \times \cdots \times \Pi_{k(n)}f_{d_e}.$$

The difference can be re-written recursively:

$$d_{t+1} = d_t f_{d_e+1} + \Pi_{k(n)}f_1 \times \cdots \times \Pi_{k(n)}f_{d_e}\left(f_{d_e+1} - \Pi_{k(n)}f_{d_e+1}\right).$$

Since $\int f_{d_e+1} = \int \Pi_{k(n)}f_1 \times \cdots \times \Pi_{k(n)}f_{d_e} = 1$, the total variation distance is: $\|d_{t+1}\|_{TV} \leq \|d_t\|_{TV} + \|f_{d_e+1} - \Pi_{k(n)}f_{d_e+1}\|_{TV} = O(\frac{\log[k(n)]^{r/b}}{k(n)^r})$. And the supremum distance is:

$$\|d_{t+1}\|_\infty \leq \|d_t\|_\infty \|f_{d_e+1}\|_\infty + \|\Pi_{k(n)}f_1 \times \cdots \times \Pi_{k(n)}f_{d_e}\|_\infty \|f_{d_e+1} - \Pi_{k(n)}f_{d_e+1}\|_\infty$$
$$\leq \|d_t\|_\infty \left(\|f_{d_e+1}\|_\infty + \|f_1 \times \cdots \times f_{d_e}\|_\infty \|f_{d_e+1} - \Pi_{k(n)}f_{d_e+1}\|_\infty\right) = O\left(\frac{\log[k(n)]^{r/b}}{k(n)^r}\right).$$

□

Proof of Lemma A2. :
To reduce notation, the t and s subscripts will be dropped in the following. The proof is similar for both e_1 and e_2 so the proof is only given for e_1.

First, the densities of e_1 and e_2 are derived, the first two results follow. Noting that the draws are defined using quantile functions, inverting the formula yields: $\nu_1 = \frac{1}{1-e_1^{2+\xi_1}}$. This is a proper CDF on $(-\infty, 0]$ since $e_1 \to \frac{1}{1-e_1^{2+\xi_1}}$ is increasing and has limits 0 at $-\infty$ and 1 at 0. Its derivative is the density function: $(2+\xi_1)\frac{e_1^{1+\xi_1}}{(1-e_1^{2+\xi_1})^2}$. It is continuous on $(-\infty, 0]$ and has an asymptote at $-\infty$: $(2+\xi_1)\frac{e_1^{1+\xi_1}}{(1-e_1^{2+\xi_1})^2} \times e_1^{3+\xi_1} \to (2+\xi_1)$ as $e_1 \to -\infty$. Since $\xi_1 \in [\underline{\xi}, \bar{\xi}]$ with $0 < \underline{\xi}$ then $\mathbb{E}|e_1|^2 \leq C < \infty$ for some finite $C > 0$. Similar results hold for e_2 which has density $(2+\xi_2)\frac{e_2^{1+\xi_2}}{(1+e_2^{2+\xi_2})^2}$ on $[0, +\infty)$.

Second, $\xi_1 \to e_1(\xi_1)$ is shown to be L^2-smooth. Let $|\xi_1 - \tilde{\xi}_1| \leq \delta$, using the mean value theorem, for each ν_1 there exists an intermediate value $\check{\xi}_1 \in [\xi_1, \tilde{\xi}_1]$ such that:

$$\left(\frac{1}{\nu_1} - 1\right)^{\frac{1}{2+\xi_1}} - \left(\frac{1}{\nu_1} - 1\right)^{\frac{1}{2+\tilde{\xi}_1}} = \frac{1}{2+\check{\xi}_1}\log(\frac{1}{\nu_1} - 1)\left(\frac{1}{\nu_1} - 1\right)^{\frac{1}{2+\check{\xi}_1}}(\xi_1 - \tilde{\xi}_1).$$

The first term is bounded by $1/(2+\underline{\xi})$, the second is bounded by $\log(\frac{1}{\nu_1}+1)\left(\frac{1}{\nu_1}+1\right)^{\frac{1}{2+\underline{\xi}}}$, and the last term is bounded above, in absolute value, by δ.

Finally, in order to conclude the proof, the integral $\int_0^1 \log(\frac{1}{\nu_1}+1)\left(\frac{1}{\nu_1}+1\right)^{\frac{2}{2+\underline{\xi}}} d\nu_1$ needs to be finite. By a change of variables, it can be re-written as: $\int_2^\infty \log(\nu)\nu^{\frac{2}{2+\underline{\xi}}-2}d\nu$. Since $\frac{2}{2+\underline{\xi}}-2<-1$, the integral is always finite and thus:

$$\left[\mathbb{E}\left(\sup_{|\xi_1-\tilde{\xi}_1|\leq\delta}|e^s_{t,1}(\xi_1)-e^s_{t,1}(\tilde{\xi}_1)|^2\right)\right]^{1/2} \leq \frac{\delta}{2+\underline{\xi}}\sqrt{\int_2^\infty \log(\nu)\nu^{\frac{2}{2+\underline{\xi}}-2}d\nu}.$$

□

Proof of Lemma A3: Since $\mathcal{B}_{k(n)}$ is contained in a ball of radius $\max(\overline{\mu}_{k(n)},\overline{\sigma},\|\theta\|_\infty)$ in $\mathbb{R}^{3[k(n)+2]+d_\theta}$ under $\|\cdot\|_m$, the covering number for $\mathcal{B}_{k(n)}$ can be computed under the $\|\cdot\|_m$ norm using a result from Kolmogorov and Tikhomirov (1959). As a result, the covering number $N(x,\mathcal{B}_{k(n)},\|\cdot\|_m)$ satisfies: $N(x,\mathcal{B}_{k(n)},\|\cdot\|_m)\leq 2\left(3[k(n)+2]+d_\theta\right)\left(\frac{2\max(\bar{\mu}_{k(n)},\bar{\sigma})}{x}+1\right)^{3[k(n)+2]+d_\theta}$. The rest follows from Lemmas 2 and D11. □

Proof of Lemma A4: First, using the assumption that B is a bounded linear operator:

$$\begin{aligned}
&Q_n(\Pi_{k(n)}\beta_0)\\
&\leq M_B^2\int\left|\mathbb{E}\left(\hat{\psi}_n(\tau)-\hat{\psi}_n^S(\tau,\Pi_{k(n)}\beta_0)\right)\right|^2\pi(\tau)d\tau\\
&\leq 3M_B^2\left(\int\left|\mathbb{E}\left(\hat{\psi}_n(\tau)-\hat{\psi}_n^S(\tau,\beta_0)\right)\right|^2\pi(\tau)d\tau+\int\left|\mathbb{E}\left(\hat{\psi}_n^S(\tau,\beta_0)-\hat{\psi}_n^S(\tau,\Pi_{k(n)}\beta_0)\right)\right|^2\pi(\tau)d\tau\right)
\end{aligned}$$

Each term can be bounded above individually. Re-write the first term in terms of distribution: $\left|\mathbb{E}\left(\hat{\psi}_n(\tau)-\hat{\psi}_n^S(\tau,\beta_0)\right)\right|=\left|\frac{1}{n}\sum_{t=1}^n\int e^{i\tau'(\mathbf{y}_t,\mathbf{x_t})}[f_t^*(\mathbf{y}_t,\mathbf{x_t})-f_t(\mathbf{y}_t,\mathbf{x_t})]d\mathbf{y}_t d\mathbf{x}_t\right|$, where f_t is the distribution of $(\mathbf{y}_t(\beta_0),\mathbf{x}_t)$ and f_t the stationary distribution of $(\mathbf{y}_t(\beta_0),\mathbf{x}_t)$. Using the geometric ergodicity assumption, for all τ:

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{t=1}^n\int e^{i\tau'(\mathbf{y}_t,\mathbf{x_t})}[f_t^*(\mathbf{y}_t,\mathbf{x_t})-f_t(\mathbf{y}_t,\mathbf{x_t})]d\mathbf{y}_t d\mathbf{x}_t\right|\leq\frac{1}{n}\sum_{t=1}^n\int\left|f_t^*(\mathbf{y}_t,\mathbf{x_t})-f_t(\mathbf{y}_t,\mathbf{x_t})\right|d\mathbf{y}_t d\mathbf{x}_t\\
&=\frac{2}{n}\sum_{t=1}^n\|f_t^*-f_t\|_{TV}\leq\frac{2C_\rho}{n}\sum_{t=1}^n\rho^t\leq\frac{2C_\rho}{(1-\rho)n}
\end{aligned}$$

for some $\rho\in(0,1)$ and $C_\rho>0$. This yields a first bound:

$$\int\left|\mathbb{E}\left(\hat{\psi}_n(\tau)-\hat{\psi}_n^S(\tau,\beta_0)\right)\right|^2\pi(\tau)d\tau\leq\frac{4C_\rho^2}{(1-\rho)^2}\frac{1}{n^2}=O\left(\frac{1}{n^2}\right).$$

The mixture norm $\|\cdot\|_m$ is not needed here to bound the second term since it involves population CFs. Some changes to the proof of Lemma 2 allows to find bounds in terms of $\|\cdot\|_{\mathcal{B}}$ and $\|\cdot\|_{TV}$ for which Lemma A1 gives the approximation rates.

To bound the second term, re-write the simulated data as:

$$y_t^s = g_{obs,t}(\mathbf{x}_{t:1}, \beta, \mathbf{e}_{t:1}^s), \quad u_t^s = g_{latent,t}(\beta, \mathbf{e}_{t:1}^s)$$

with $\beta = (\theta, f)$, $e_t^s \sim f$ and $\mathbf{x}_{t:1} = (x_t, \dots, x_1), \mathbf{e}_{t:1}^s = (e_t^s, \dots, e_1^s)$. Under Assumption 2 or 2′, using the same sequence of shocks (e_t^s): $\mathbb{E}\left(\left\|g_{obs,t}(\mathbf{x}_{t:1}, \beta_0, \mathbf{e}_{t:1}^s) - g_{obs,t}(\mathbf{x}_{t:1}, \Pi_{k(n)}\beta_0, \mathbf{e}_{t:1}^s)\right\|\right) \leq \overline{C}\|\Pi_{k(n)}f_0 - f_0\|_{\mathcal{B}}^{\gamma}$. This is similar to the proof of Lemma 2, first re-write the difference as:

$$\mathbb{E}\Big(\Big\|g_{obs}(g_{obs,t-1}(\mathbf{x}_{t-1:1}, \beta_0, \mathbf{e}_{t-1:1}^s), x_t, \beta_0, g_{latent}(g_{latent,t-1}(\beta_0, \mathbf{e}_{t-1:1}^s), \beta_0, e_t^s))$$
$$- g_{obs}(g_{obs,t-1}(\mathbf{x}_{t-1:1}, \Pi_{k(n)}\beta_0, \mathbf{e}_{t-1:1}^s), x_t, \Pi_{k(n)}\beta_0, g_{latent}(g_{latent,t-1}(\Pi_{k(n)}\beta_0, \mathbf{e}_{t-1:1}^s), \Pi_{k(n)}\beta_0, e_t^s\Big\|\Big).$$

Using Assumptions 2-2′, the following recursive relationship holds:

$$\begin{aligned}
&\mathbb{E}\Big(\Big\|g_{obs}(g_{obs,t-1}(\mathbf{x}_{t-1:1}, \beta_0, \mathbf{e}_{t-1:1}^s), x_t, \beta_0, g_{latent}(g_{latent,t-1}(\beta_0, \mathbf{e}_{t-1:1}^s), \beta_0, e_t^s)) \\
&\quad - g_{obs}(g_{obs,t-1}(\mathbf{x}_{t-1:1}, \Pi_{k(n)}\beta_0, \mathbf{e}_{t-1:1}^s), x_t, \Pi_{k(n)}\beta_0, \\
&\qquad g_{latent}(g_{latent,t-1}(\Pi_{k(n)}\beta_0, \mathbf{e}_{t-1:1}^s), \Pi_{k(n)}\beta_0, e_t^s))\Big\|\Big) \\
&\leq \Big[\mathbb{E}\Big(\Big\|g_{obs}(g_{obs,t-1}(\mathbf{x}_{t-1:1}, \beta_0, \mathbf{e}_{t-1:1}^s), x_t, \beta_0, g_{latent}(g_{latent,t-1}(\beta_0, \mathbf{e}_{t-1:1}^s), \beta_0, e_t^s)) \\
&\quad - g_{obs}(g_{obs,t-1}(\mathbf{x}_{t-1:1}, \Pi_{k(n)}\beta_0, \mathbf{e}_{t-1:1}^s), x_t, \Pi_{k(n)}\beta_0, \\
&\qquad g_{latent}(g_{latent,t-1}(\Pi_{k(n)}\beta_0, \mathbf{e}_{t-1:1}^s), \Pi_{k(n)}\beta_0, e_t^s))\Big\|^2\Big)\Big]^{1/2} \\
&\leq \overline{C}_1\Big[\mathbb{E}\Big(\Big\|g_{obs,t-1}(\mathbf{x}_{t-1:1}, \beta_0, \mathbf{e}_{t-1:1}^s) - g_{obs,t-1}(x_{t-1}, \dots, x_1, \Pi_{k(n)}\beta_0, \mathbf{e}_{t-1:1}^s)\Big\|^2\Big)\Big]^{1/2} \\
&\quad + \overline{C_2}\|\beta_0 - \Pi_{k(n)}\beta_0\|_{\mathcal{B}}^{\gamma} + \overline{C}_3\Big[\mathbb{E}\Big(\Big\|g_{latent,t}(\beta_0, \mathbf{e}_{t:1}^s) - g_{latent,t}(\Pi_{k(n)}\beta_0, \mathbf{e}_{t:1}^s)\Big\|^2\Big)\Big]^{\gamma/2}.
\end{aligned}$$

The last term also has a recursive structure:

$$\begin{aligned}
&\Big[\mathbb{E}\Big(\Big\|g_{latent,t}(\beta_0, \mathbf{e}_{t:1}^s) - g_{latent,t}(\Pi_{k(n)}\beta_0, \mathbf{e}_{t:1}^s)\Big\|^2\Big)\Big]^{1/2} \\
&\leq \overline{C}_4\left[\mathbb{E}\Big(\Big\|g_{latent,t-1}(\beta_0, \mathbf{e}_{t-1:1}^s) - g_{latent,t-1}(\Pi_{k(n)}\beta_0, \mathbf{e}_{t-1:1}^s)\Big\|^2\Big)\right]^{1/2} + \overline{C}_5\|\beta_0 - \Pi_{k(n)}\beta_0\|_{\mathcal{B}}^{\gamma}.
\end{aligned}$$

Together these inequalities imply:

$$\mathbb{E}\Big(\Big\|g_{obs}(g_{obs,t-1}(x_{t-1},\dots,x_1,\beta_0,\mathbf{e}^s_{t-1:1}),x_t,\beta_0,g_{latent}(g_{latent,t-1}(\beta_0,\mathbf{e}^s_{t-1:1}),\beta_0,e^s_t))$$
$$- g_{obs}(g_{obs,t-1}(x_{t-1},\dots,x_1,\Pi_{k(n)}\beta_0,\mathbf{e}^s_{t-1:1}),x_t,\Pi_{k(n)}\beta_0,$$
$$g_{latent}(g_{latent,t-1}(\Pi_{k(n)}\beta_0,\mathbf{e}^s_{t-1:1}),\Pi_{k(n)}\beta_0,e^s_t\Big\|\Big)$$
$$\leq \frac{1}{1-\overline{C}_1}\left(\overline{C}_2\|\beta_0-\Pi_{k(n)}\beta_0\|^\gamma_{\mathcal{B}} + \overline{C}_3\frac{\overline{C}_5^\gamma}{(1-\overline{C}_4)^\gamma}\|\beta_0-\Pi_{k(n)}\beta_0\|^{\gamma^2}_{\mathcal{B}}\right).$$

Recall that $\|\tau\|_\infty\sqrt{\pi(\tau)}$ is bounded above and $\pi(\tau)^{1/4}$ is integrable so that:

$$\int\left|\mathbb{E}\left(e^{i\tau'(\mathbf{y}_t(\beta_0,\mathbf{x}_{t:1}))}-e^{i\tau'(\mathbf{y}_t(\Pi_{k(n)}\beta_0,\mathbf{x}_{t:1}))}\right)\right|^2\pi(\tau)d\tau$$
$$\leq\left(\overline{C}_2\|\beta_0-\Pi_{k(n)}\beta_0\|^\gamma_{\mathcal{B}} + \overline{C}_3\frac{\overline{C}_5^\gamma}{(1-\overline{C}_4)^\gamma}\|\beta_0-\Pi_{k(n)}\beta_0\|^{\gamma^2}_{\mathcal{B}}\right)\frac{\sup_\tau[\|\tau\|_\infty\sqrt{\pi(\tau)}]\int\pi(\tau)^{1/4}d\tau}{1-\overline{C}_1}.$$

To conclude the proof, the difference due to e^s_t needs to be bounded. In order to do so, it suffice to bound the following integral:

$$\int e^{i\tau'(\mathbf{y}_t(y_0,u_0,\mathbf{x}_{t:1},\beta_0,\mathbf{e}^s_{t:1}),\mathbf{x}_t)}\left(\prod_{j=1}^t f_0(e^s_j)-\prod_{j=1}^t\Pi_{k(n)}f_0(e^s_j)\right)f_{\mathbf{x}}(\mathbf{x}_{t:1})d\mathbf{e}^s_{t:1}d\mathbf{x}_{t:1}.$$

A direct bound on this integral yields a term of order of $t\|f_0-\Pi_{k(n)}f_0\|_{TV}$ which increases with t, which is too fast to generate useful rates. Rather than using a direct bound, consider Assumptions 2-2′. The time-series y^s_t can be approximated by another time-series term which only depends on a fixed and finite $(e^s_t,\dots,e^s_{t-m})$ for a given integer $m\geq 1$. Making m grow with n at an appropriate rate allows to balance the bias $m\|f_0-\Pi_{k(n)}f_0\|_{TV}$ (computed from a direct bound) and the approximation due to $m<t$.

The m-approximation rate of y_t is now derived. Let $\beta=(\theta,f)\in\mathcal{B}$, $e^s_t,\dots,e^s_1\sim f$ and $\tilde{y}^s_t$ such that $\tilde{y}^s_{t-m}=0,\tilde{u}^s_{t-m}=0$ and then $\tilde{y}^s_j=g_{obs}(\tilde{y}^s_{j-1},x_j,\beta,\tilde{u}^s_j),\tilde{u}^s_j=g_{latent}(\tilde{u}^s_{j-1},\beta,e^s_j)$ for $t-m+1\leq j\leq t$. Each observation t is approximated by its own time-series. For observation $t-m$, by construction: $\mathbb{E}\left(\left\|y^s_{t-m}-\tilde{y}^s_{t-m}\right\|\right)=\mathbb{E}\left(\left\|y^s_{t-m}\right\|\right)\leq\left[\mathbb{E}\left(\left\|y^s_{t-m}\right\|^2\right)\right]^{1/2}$ and $\mathbb{E}\left(\left\|u^s_{t-m}-\tilde{u}^s_{t-m}\right\|\right)=\mathbb{E}\left(\left\|u^s_{t-m}\right\|\right)\leq\left[\mathbb{E}\left(\left\|u^s_{t-m}\right\|^2\right)\right]^{1/2}$. Then, for any $t\geq\tilde{t}\geq t-m$:

$$\mathbb{E}\left(\left\|u^s_{\tilde{t}}-\tilde{u}^s_{\tilde{t}}\right\|\right)\leq\overline{C}_4\left[\mathbb{E}\left(\left\|u^s_{\tilde{t}-1}-\tilde{u}^s_{\tilde{t}-1}\right\|^2\right)\right]^{1/2}$$
$$\mathbb{E}\left(\left\|y^s_{\tilde{t}}-\tilde{y}^s_{\tilde{t}}\right\|\right)\leq\overline{C}_3\overline{C}_4^\gamma\left[\mathbb{E}\left(\left\|u^s_{\tilde{t}-1}-\tilde{u}^s_{\tilde{t}-1}\right\|^2\right)\right]^{\gamma/2}+\overline{C}_1\left[\mathbb{E}\left(\left\|y^s_{\tilde{t}-1}-\tilde{y}^s_{\tilde{t}-1}\right\|^2\right)\right]^{1/2}.$$

The previous two results and a recursion arguments leads to the following inequality:

$$\mathbb{E}\left(\left\|u_t^s - \tilde{u}_t^s\right\|\right) \leq \overline{C}_4^m \left[\mathbb{E}\left(\left\|u_{t-m}^s\right\|^2\right)\right]^{1/2} \tag{C.19}$$

$$\mathbb{E}\left(\left\|y_t^s - \tilde{y}_t^s\right\|\right) \leq \overline{C}_3\overline{C}_4^{\gamma m} \left[\mathbb{E}\left(\left\|u_{t-m}^s\right\|^2\right)\right]^{\gamma/2} + \overline{C}_1^m \left[\mathbb{E}\left(\left\|y_{t-m}^s\right\|^2\right)\right]^{1/2}. \tag{C.20}$$

For $\beta = \beta_0, \Pi_{k(n)}\beta_0$ since the expectations are finite and bounded by assumption, $\mathbb{E}\left(\left\|y_t^s - \tilde{y}_t^s\right\|\right) \leq \overline{C}\max(\overline{C}_1, \overline{C}_4)^{\gamma m}$ with $0 \leq \max(\overline{C}_1, \overline{C}_4) < 1$ and some $\overline{C} > 0$. For the first observations $t \leq m$ the data is unchanged, $y_t^s = \tilde{y}_t^s$, so that the bound still holds. The integral can be split and bounded:

$$\begin{aligned}
&\Big| \int e^{i\tau'(\mathbf{y}_t(y_0,u_0,\mathbf{x}_{t:1},\beta_0,\mathbf{e}_{t:1}^s),\mathbf{x}_t)} \left(\prod_{j=1}^t f_0(e_j^s) - \prod_{j=1}^t \Pi_{k(n)} f_0(e_j^s)\right) f_{\mathbf{x}}(\mathbf{x}_{t:1}) d\mathbf{e}_{t:1}^s d\mathbf{x}_{t:1} \Big| \\
&\leq \left| \mathbb{E}\left([\hat{\psi}_n^S(\tau,\beta_0) - \hat{\psi}_n^S(\tau,\Pi_{k(n)}\beta_0)] - [\tilde{\psi}_n^S(\tau,\beta_0) - \tilde{\psi}_n^S(\tau,\Pi_{k(n)}\beta_0)] \right) \right| \\
&+ \int \Big| \left(\prod_{j=t-m+1}^t f_0(e_j^s) - \prod_{j=t-m+1}^t \Pi_{k(n)} f_0(e_j^s) \right) d\mathbf{e}_{t:t-m+1}^s \Big| \\
&\leq 4\overline{C}\max(\overline{C}_1, \overline{C}_4)^{\gamma m} + 2m\|\Pi_{k(n)}f_0 - f_0\|_{TV}.
\end{aligned}$$

The last inequality is due to the cosine and sine functions being uniformly Lipschitz continuous and equations (C.19)-(C.20). Recall that $\|\Pi_{k(n)}f_0 - f_0\|_{TV} = O(\frac{\log[k(n)]^{2r/b}}{k(n)^r})$. To balance the two terms, pick: $m = -\frac{r}{\gamma\log[\max(\overline{C}_1,\overline{C}_4)]}\log[k(n)] > 0$. Then $\max(\overline{C}_1, \overline{C}_4)^{\gamma m} = k(n)^{-r}$ and

$$\overline{C}\max(\overline{C}_1, \overline{C}_4)^{\gamma m} + 2m\|\Pi_{k(n)}f_0 - f_0\|_{TV} = O\left(\frac{\log[k(n)]^{2r/b+1}}{k(n)^r}\right).$$

Combining all the bounds above yields:

$$Q_n(\Pi_{k(n)}\beta_0) = O\left(\max\left[\frac{\log[k(n)]^{4r/b+2}}{k(n)^{2r}}, \frac{\log[k(n)]^{4\gamma^2 r/b}}{k(n)^{2\gamma^2 r}}, \frac{1}{n^2}\right]\right)$$

where $\|\cdot\|_{\mathcal{B}} = \|\cdot\|_\infty$ or $\|\cdot\|_{TV}$ so that $\|\beta_0 - \Pi_{k(n)}\beta_0\|_{\mathcal{B}}^{\gamma^2} = O(\frac{\log[k(n)]^{4\gamma^2 r/b}}{k(n)^{2\gamma^2 r}})$. The term due to the non-stationarity is of order $1/n^2 = o\left(\max\left[\frac{\log[k(n)]^{4r/b+2}}{k(n)^{2r}}, \frac{\log[k(n)]^{4\gamma^2 r/b}}{k(n)^{2\gamma^2 r}}\right]\right)$ so it can be ignored. This concludes the proof. □

Proof of Lemma A5: Using the inequality $1/2|a|^2 \leq |a-b|^2 + |b|^2$ for any $a, b \in \mathbb{R}$:

$$
\begin{aligned}
0 &\leq 1/2 \int \left| B \frac{d\mathbb{E}(\hat{\psi}_n^S(\tau, \Pi_{k(n)}\beta_0))}{d\beta}[\hat{\beta}_n - \Pi_{k(n)}\beta_0] \right|^2 \pi(\tau) d\tau \\
&\leq \int \left| B \frac{d\mathbb{E}(\hat{\psi}_n^S(\tau, \beta_0))}{d\beta}[\hat{\beta}_n - \beta_0] \right|^2 \pi(\tau) d\tau \\
&+ \int \left| B \frac{d\mathbb{E}(\hat{\psi}_n^S(\tau, \beta_0))}{d\beta}[\hat{\beta}_n - \beta_0] - B \frac{d\mathbb{E}(\hat{\psi}_n^S(\tau, \Pi_{k(n)}\beta_0))}{d\beta}[\hat{\beta}_n - \Pi_{k(n)}\beta_0] \right|^2 \pi(\tau) d\tau \\
&\leq \int \left| B \frac{d\mathbb{E}(\hat{\psi}_n^S(\tau, \beta_0))}{d\beta}[\hat{\beta}_n - \beta_0] \right|^2 \pi(\tau) d\tau + \int \left| B \frac{d\mathbb{E}(\hat{\psi}_n^S(\tau, \Pi_{k(n)}\beta_0))}{d\beta}[\Pi_{k(n)}\beta_0 - \beta_0] \right|^2 \pi(\tau) d\tau \\
&+ \int \left| B \frac{d\mathbb{E}(\hat{\psi}_n^S(\tau, \beta_0))}{d\beta}[\hat{\beta}_n - \beta_0] - B \frac{d\mathbb{E}(\hat{\psi}_n^S(\tau, \beta_0))}{d\beta}[\hat{\beta}_n - \Pi_{k(n)}\beta_0] \right|^2 \pi(\tau) d\tau.
\end{aligned}
$$

By assumption the term on the left is $O_p(\delta_n^2)$, by condition ii. the middle term is $O_p(\delta_n^2)$ and condition i. implies that the term on the right is also $O_p(\delta_n^2)$. It follows that:

$$
\int \left| B \frac{d\mathbb{E}(\hat{\psi}_n^S(\tau, \Pi_{k(n)}\beta_0))}{d\beta}[\hat{\beta}_n - \Pi_{k(n)}\beta_0] \right|^2 \pi(\tau) d\tau = O_p(\delta_n^2). \tag{C.21}
$$

Now note that both $\hat{\beta}_n$ and $\Pi_{k(n)}\beta_0$ belong to the finite dimensional space $\mathcal{B}_{k(n)}$ parameterized by $(\theta, \omega, \mu, \sigma)$. To save space, $\hat{\beta}_n$ will be represented by $\hat{\varphi}_n = (\hat{\theta}_n, \hat{\omega}_n, \hat{\mu}_n, \hat{\sigma}_n)$ and $\Pi_{k(n)}\beta_0$ by $\varphi_{k(n)} = (\theta_{k(n)}, \omega_{k(n)}, \mu_{k(n)}, \sigma_{k(n)})$. Using this notation, equation (C.21) becomes:

$$
\begin{aligned}
&\int \left| B \frac{d\mathbb{E}(\hat{\psi}_n^S(\tau, \Pi_{k(n)}\beta_0))}{d\beta}[\hat{\beta}_n - \Pi_{k(n)}\beta_0] \right|^2 \pi(\tau) d\tau \\
&= \int \left| B \frac{d\mathbb{E}(\hat{\psi}_n^S(\tau, \Pi_{k(n)}\beta_0))}{d(\theta, \omega, \mu, \sigma)}[\hat{\varphi}_n - \varphi_{k(n)}] \right|^2 \pi(\tau) d\tau \\
&= trace \left([\hat{\varphi}_n - \varphi_{k(n)}]' \int B \frac{d\mathbb{E}(\hat{\psi}_n^S(\tau, \Pi_{k(n)}\beta_0))}{d(\theta, \omega, \mu, \sigma)}' \overline{B \frac{d\mathbb{E}(\hat{\psi}_n^S(\tau, \Pi_{k(n)}\beta_0))}{d(\theta, \omega, \mu, \sigma)}} \pi(\tau) d\tau [\hat{\varphi}_n - \varphi_{k(n)}] \right) \\
&\geq \underline{\lambda}_n \|\hat{\varphi}_n - \varphi_{k(n)}\|^2 = \underline{\lambda}_n \|\hat{\beta}_n - \Pi_{k(n)}\beta_0\|_m^2.
\end{aligned}
$$

It follows that $0 \leq \underline{\lambda}_n \|\hat{\beta}_n - \Pi_{k(n)}\beta_0\|_m^2 \leq O_p(\delta_n^2)$ so that the rate of convergence in mixture norm is: $\|\hat{\beta}_n - \Pi_{k(n)}\beta_0\|_m = O_p\left(\delta_n \underline{\lambda}_n^{-1/2}\right)$. □

Proof of Lemma A6. Using the rate assumptions and Lemma D13 implies the desired result.

□

Appendix D Intermediate Results

Lemma D7 (Kruijer, Rousseau and van der Vaart, 2010)**.** *Suppose that f is a continuous univariate density satisfying: i) Smoothness: f is r-times continuously differentiable with* *bounded r-th derivative. ii) Tails: f has exponential tails, i.e. there exists* $\bar{e}, M_{f_1}, a, b > 0$ *such that:* $f_1(e) \leq M_{f_1} e^{-a|e|^b}, \forall |e| \geq \bar{e}$. *iii) Monotonicity in the Tails: f is strictly positive and there exists* $\underline{e} < \bar{e}$ *such that* f_S *is weakly decreasing on* $(-\infty, \underline{e}]$ *and weakly increasing on* $[\bar{e}, \infty)$. *Let* $\mathcal{F}_k$ *be the sieve space consisting of Gaussian mixtures with the following restrictions. iv) Bandwidth:* $\sigma_j \geq \underline{\sigma}_k = O(\frac{\log[k(n)]^{2/b}}{k})$. *v) Location Parameter Bounds:* $\mu_j \in [-\bar{\mu}_k, \bar{\mu}_k]$. *vi) Growth Rate of Bounds:* $\bar{\mu}_k = O\left(\log[k]^{1/b}\right)$. *Then there exists a mixture sieve approximation of f,* $\Pi_k f \in \mathcal{F}_k$, *such that as* $k \to \infty$: $\|f - \Pi_k f\|_{\mathcal{F}} = O\left(\frac{\log[k(n)]^{2r/b}}{k(n)^r}\right)$, *where* $\|\cdot\|_{\mathcal{F}} = \|\cdot\|_{TV}$ *or* $\|\cdot\|_\infty$.

Lemma D8 (Chen and Pouzo, 2012)**.** *Let* $\hat{\beta}_n$ *be such that* $\hat{Q}_n(\hat{\beta}_n) \leq \inf_{\beta \in \mathcal{B}_{k(n)}} + O_{p^*}(\eta_n)$, *where* $(\eta_n)_{n \geq 1}$ *is a positive real-valued sequence such that* $\eta_n = o(1)$. *Let* $\bar{Q}_n : \mathcal{B} \to [0, +\infty)$ *be a sequence of non-random measurable functions and let the following conditions hold: a. i)* $0 \leq \bar{Q}_n(\beta_0) = o(1)$; *ii) there is a positive function* $g_0(n, k, \varepsilon)$ *such that:* $\inf_{h \in \mathcal{B}_k : \|\beta - \beta_0\|_{\mathcal{B}} > \varepsilon} \bar{Q}_n(\beta) \geq g_0(n, k, \varepsilon) > 0$ *for each* $n, k \geq 1$, *and* $\liminf_{n \to \infty} g_0(n, k(n), \varepsilon) \geq 0$ *for all* $\varepsilon > 0$. *b. i)* $\mathcal{B}$ *is an infinite dimensional, possibly non-compact subset of a Banach space* $(B, \|\|_{\mathcal{B}})$; *ii)* $\mathcal{B}_k \subseteq \mathcal{B}_{k+1} \subseteq \mathcal{B}$ *for all* $k \geq 1$, *and there is a sequence* $\{\Pi_{k(n)} \beta_0 \in \mathcal{B}_{k(n)}\}$ *such that* $\bar{Q}_n(\Pi_{k(n)} \beta_0) = o(1)$. *c.* $\hat{Q}_n(\beta)$ *is jointly measurable in the data* $(y_t, x_t)_{t \geq 1}$ *and the parameter* $h \in \mathcal{B}_{k(n)}$. *d. i)* $\hat{Q}_n(\Pi_{k(n)} \beta_0) \leq K_0 \bar{Q}_n(\Pi_{k(n)} \beta_0) + O_{p^*}(c_{0,n})$ *for some* $c_{0,n} = o(1)$ *and a finite constant* $K_0 > 0$; *ii)* $\hat{Q}_n(\beta) \geq K \bar{Q}_n(\beta) - O_{p^*}(c_n)$ *uniformly over* $h \in \mathcal{B}_{k(n)}$ *for some* $c_n = o(1)$ *and a finite constant* $K > 0$; *iii)* $\max(c_{0,n}, c_n, \bar{Q}_n(\Pi_{k(n)} \beta_0), \eta_n) = o(g_0(n, k(n), \varepsilon))$ *for all* $\varepsilon > 0$. *Then for all* $\varepsilon > 0$: $\mathbb{P}^*\left(\|\hat{\beta}_n - \beta_0\|_{\mathcal{B}} > \varepsilon\right) \to 0$ *as* $n \to \infty$.

Lemma D9. *Let* $(Y_t)_{t \geq 1}$ *mean zero,* α*-mixing with rate* $\alpha(m)$ *such that* $\sum_{m \geq 1} \alpha(m)^{1/p} < \infty$ *for some* $p > 1$, *and* $|Y_t| \leq 1$ *for all* $t \geq 1$. *Then we have* $\mathbb{E}\left(n |\bar{Y}_n|^2\right) \leq 1 + 24 \sum_{m \geq 1} \alpha(m)^{1/p}$.

Lemma D10. *Let* $(X_t)_{t > 0}$ *be a sequence of real-valued, centered random variables and* $(\alpha_m)_{m \geq 0}$ *be the sequence of strong mixing coefficients. Suppose that* X_t *is uniformly bounded and there exists* $A, C > 0$ *such that* $\alpha(m) \leq A \exp(-Cm)$ *then there exists* $K > 0$ *that depends only on the mixing coefficients such that for any* $p \geq 2$:

$$\mathbb{E}\left(|\sqrt{n} \bar{X}_n|^p\right)^{1/p} \leq K \left[\sqrt{p} \left(\int_0^1 \min(\alpha^{-1}(u), n) \sum_{t=1}^n \frac{Q_t^2(u)}{n}\right)^{1/2} + n^{1/p - 1/2} p^2 \|\sup_{t > 0} X_t\|_\infty\right]$$

where Q_t *is the quantile function of* X_t, $\min(\alpha^{-1}(u), n) = \sum_{i=k}^n \mathbb{1}_{u \leq \alpha_k}$.

Lemma D11. *Suppose that* $(X_t(\beta))_{t > 0}$ *is a real valued, mean zero random process for any* $\beta \in \mathcal{B}$. *Suppose that it is* α*-mixing with exponential decay:* $\alpha(m) \leq A \exp(-Cm)$ *for*

$A, C > 0$ *and bounded* $|X_t(\beta)| \leq 1$. *Let* $\mathcal{X} = \{X : \mathcal{B} \to \mathbb{C}, \beta \to X_t(\beta)\}$ *and suppose that* $\int_0^1 \log^2 N_{[]}(x, \mathcal{X}, \|\cdot\|)dx < \infty$ *then:* $\int_0^1 x^{\vartheta/2-1}\sqrt{\log N_{[]}(x, \mathcal{X}, \|\cdot\|)} + \log^2 N_{[]}(x, \mathcal{X}, \|\cdot\|) < \infty$ *for all* $\vartheta \in (0,1)$ *and:*

$$\mathbb{E}\big(\sup_{\beta\in\mathcal{B}} |\sqrt{n}[\hat{\psi}_t^S(\beta) - \mathbb{E}(\hat{\psi}_t^S(\beta))]|^2\big) \leq K\big(\int_0^1 x^{\vartheta/2-1}\sqrt{\log N_{[]}(x, \mathcal{X}, \|\cdot\|)} + \log^2 N_{[]}(x, \mathcal{X}, \|\cdot\|)dx\big).$$

Assumption 2′ (Data Generating Process - L^2-Smoothness). y_t^s *is simulated according to the dynamic model (1)-(2) where* g_{obs} *and* g_{latent} *satisfy the following* L^2*-smoothness conditions for some* $\gamma \in (0,1]$ *and any* $\delta \in (0,1)$*:*

$y(i)'$. *For some* $0 \leq \bar{C}_1 < 1$*:*
$\big[\mathbb{E}\big(\sup_{\|\beta_1-\beta_2\|_{\mathcal{B}}\leq\delta} \|g_{obs}(y_t^s(\beta_1), x_t, \beta_1, u_t^s(\beta_1)) - g_{obs}(y_t^s(\beta_2), x_t, \beta_1, u_t^s(\beta_1))\|^2 \big| y_t^s(\beta_1), y_t^s(\beta_2)\big)\big]^{1/2} \leq \bar{C}_1\|y_t^s(\beta_1) - y_t^s(\beta_2)\|$

$y(ii)'$. *For some* $0 \leq \bar{C}_2 < \infty$*:*
$\big[\mathbb{E}\big(\sup_{\|\beta_1-\beta_2\|_{\mathcal{B}}\leq\delta} \|g_{obs}(y_t^s(\beta_1), x_t, \beta_1, u_t^s(\beta_1)) - g_{obs}(y_t^s(\beta_1), x_t, \beta_2, u_t^s(\beta_1))\|^2\big)\big]^{1/2} \leq \bar{C}_2\delta^\gamma$

$y(iii)'$. *For some* $0 \leq \bar{C}_3 < \infty$*:*
$\big[\mathbb{E}\big(\sup_{\|\beta_1-\beta_2\|_{\mathcal{B}}\leq\delta} \|g_{obs}(y_t^s(\beta_1), x_t, \beta_1, u_t^s(\beta_1)) - g_{obs}(y_t^s(\beta_1), x_t, \beta_1, u_t^s(\beta_2))\|^2 \big| u_t^s(\beta_1), u_t^s(\beta_2)\big)\big]^{1/2} \leq \bar{C}_3\|u_t^s(\beta_1) - u_t^s(\beta_2)\|^\gamma$

$u(i)'$. *For some* $0 \leq \bar{C}_4 < 1$*:*
$\big[\mathbb{E}\big(\sup_{\|\beta_1-\beta_2\|_{\mathcal{B}}\leq\delta} \|g_{latent}(u_{t-1}^s(\beta_1), \beta, e_t^s(\beta_1)) - g_{latent}(u_{t-1}^s(\beta_2), \beta, e_t^s(\beta_1))\|^2\big)\big]^{1/2} \leq \bar{C}_4\|u_{t-1}^s(\beta_1) - u_{t-1}^s(\beta_2)\|$

$u(ii)'$. *For some* $0 \leq \bar{C}_5 < \infty$*:*
$\big[\mathbb{E}\big(\sup_{\|\beta_1-\beta_2\|_{\mathcal{B}}\leq\delta} \|g_{latent}(u_{t-1}^s(\beta_1), \beta_1, e_t^s(\beta_1)) - g_{latent}(u_{t-1}^s(\beta_1), \beta_2, e_t^s(\beta_1))\|^2\big)\big]^{1/2} \leq \bar{C}_5\delta^\gamma$

$u(iii)'$. *For some* $0 \leq \bar{C}_5 < \infty$*:*
$\big[\mathbb{E}\big(\sup_{\|\beta_1-\beta_2\|_{\mathcal{B}}\leq\delta} \|g_{latent}(u_{t-1}^s(\beta_1), \beta_1, e_t^s(\beta_1)) - g_{latent}(u_{t-1}^s(\beta_1), \beta_1, e_t^s(\beta_2))\|^2 \big| e_t^s(\beta_1), e_t^s(\beta_2)\big)\big]^{1/2} \leq \bar{C}_6\|e_t^s(\beta_1) - e_t^s(\beta_2)\|$

for $\|\beta_1 - \beta_2\|_{\mathcal{B}} = \|\theta_1 - \theta_2\| + \|f_1 - f_2\|_\infty$ *or* $\|\theta_1 - \theta_2\| + \|f_1 - f_2\|_{TV}$.

Lemma D12. *Suppose that* $(\mathbf{y}_t^s, \mathbf{x}_t)_{t\geq 1}$ *is geometrically ergodic for* $\beta = \beta_0$ *and the moments are bounded* $|\hat{\psi}_t^s(\tau, \beta_0)| \leq M$ *for all* τ *then* $Q_n(\beta_0) = O(1/n^2)$.

Lemma D13 (Stochastic Equicontinuity). *Let* $M_n = \log\log(n+1)$ *and* $\delta_{mn} = \delta_n/\sqrt{\underline{\lambda}_n}$. *Let* $\Delta_n^S(\tau,\beta) = \hat{\psi}_n^S(\tau,\beta) - \mathbb{E}(\hat{\psi}_n^S(\tau,\beta))$. *Suppose that the assumptions of Lemma A5 and the conditions for Theorem 3 hold then for any* $\eta > 0$, *uniformly over* $\beta \in \mathcal{B}_{k(n)}$ *:*

$$\left[\mathbb{E}\left(\sup_{\|\beta-\Pi_{k(n)}\beta_0\|_m \leq M_n\delta_{mn}} \left|\Delta_n^S(\tau,\beta) - \Delta_n^S(\tau, \Pi_{k(n)}\beta_0)\right|^2 \pi(\tau)^{\frac{2}{2+\eta}}\right)\right]^{1/2} \leq C\frac{(M_n\delta_{mn})^{\frac{\gamma^2}{2}}}{\sqrt{n}} I_{m,n}$$

Where $I_{m,n}$ is defined as:

$$I_{m,n} = \int_0^1 \left(x^{-\vartheta/2}\sqrt{\log N([xM_n\delta_{mn}]^{\frac{2}{\gamma^2}}, \mathcal{B}_{k(n)}, \|\cdot\|_m)} + \log^2 N([xM_n\delta_{mn}]^{\frac{2}{\gamma^2}}, \mathcal{B}_{k(n)}, \|\cdot\|_m) \right) dx$$

For the mixture sieve the integral is a $O(k(n)\log[k(n)] + k(n)|\log(M_n\delta_{mn})|)$ so that:

$$\begin{aligned}
&\left[\mathbb{E}\left(\int \sup_{\|\beta-\Pi_{k(n)}\beta_0\|_m \leq M_n\delta_{mn}} \left|\Delta_n^S(\tau,\beta) - \Delta_n^S(\tau,\Pi_{k(n)}\beta_0)\right|^2 \pi(\tau)d\tau\right)\right]^{1/2} \\
&= O\left((M_n\delta_{mn})^{\frac{\gamma^2}{2}} \max(\log[k(n)]^2, |\log[M_n\delta_{mn}]|^2)\frac{k(n)^2}{\sqrt{n}}\right)
\end{aligned}$$

Now suppose that $(M_n\delta_{mn})^{\frac{\gamma^2}{2}} \max(\log[k(n)]^2, |\log[M_n\delta_{mn}]|^2)k(n)^2 = o(1)$. The first stochastic equicontinuity result is:

$$\left[\mathbb{E}\left(\int \sup_{\|\beta-\Pi_{k(n)}\beta_0\|_m \leq M_n\delta_{mn}} \left|\Delta_n^S(\tau,\beta) - \Delta_n^S(\tau,\Pi_{k(n)}\beta_0)\right|^2 \pi(\tau)d\tau\right)\right]^{1/2} = o(1/\sqrt{n}).$$

Also, suppose that $\beta \to \int \mathbb{E}\left|\hat{\psi}_t^s(\tau,\beta_0) - \hat{\psi}_t^s(\tau,\beta)\right|^2 \pi(\tau)d\tau$ is continuous at $\beta = \beta_0$ under the norm $\|\cdot\|_{\mathcal{B}}$, uniformly in $t \geq 1$. Then, the second stochastic equicontinuity result is:

$$\left[\mathbb{E}\left(\int \sup_{\|\beta-\Pi_{k(n)}\beta_0\|_m \leq M_n\delta_{mn}} \left|\Delta_n^S(\tau,\beta) - \Delta_n^S(\tau,\beta_0)\right|^2 \pi(\tau)d\tau\right)\right]^{1/2} = o(1/\sqrt{n}).$$

Lemma D14. *Suppose that $\|\hat{\beta}_n - \beta_0\|_{weak} = O_p(\delta_n)$. Under the Assumptions of Theorem 3:*

a) $\int \psi_\beta(\tau,u_n^*)\left(\overline{B\mathbb{E}(\hat{\psi}_n^S(\tau,\hat{\beta}_n) - \hat{\psi}_n^S(\tau,\beta_0)) - B\frac{d\mathbb{E}(\hat{\psi}_n^S(\tau,\beta_0))}{d\beta}[\hat{\beta}_n - \beta_0]}\right)\pi(\tau)d\tau = o(1/\sqrt{n})$.

b) $\int \psi_\beta(\tau,u_n^*)\left(\overline{B\mathbb{E}(\hat{\psi}_n^S(\tau,\hat{\beta}_n) - \hat{\psi}_n^S(\tau,\beta_0)) - B[\hat{\psi}_n^S(\tau,\hat{\beta}_n) - \hat{\psi}_n^S(\tau,\beta_0)]}\right)\pi(\tau)d\tau = o(1/\sqrt{n})$.

c) $\int \left[\psi_\beta(\tau,u_n^*)\left(\overline{B[\hat{\psi}_n(\tau) - \hat{\psi}_n^S(\tau,\hat{\beta}_n)]}\right) + \overline{\psi_\beta(\tau,u_n^*)}\left(B[\hat{\psi}_n(\tau) - \hat{\psi}_n^S(\tau,\hat{\beta}_n)]\right)\right]\pi(\tau)d\tau = o(1/\sqrt{n})$.

Appendix E Proofs for the Intermediate Results

Proof of Lemma D9: The proof follows from Davydov (1968)'s inequality: let $p,q,r \geq 0, 1/p + 1/q + 1/r = 1$, for any random variables X,Y: $|cov(X,Y)| \leq 12\alpha(\sigma(X),\sigma(Y))^{1/p}\mathbb{E}(|X|^q)^{1/q}\mathbb{E}(|Y|^r)^{1/r}$, where $\alpha(\sigma(X), \sigma(Y))$ is the mixing coefficient between X and Y. As a result:

$$\begin{aligned}
\mathbb{E}\left(n|\bar{Y}_n|^2\right) &= \frac{1}{n}\sum_{t=1}^{n}\mathbb{E}(|X_n|^2) + \frac{1}{n}\sum_{t\neq t'} cov(Y_t, Y_{t'}) \leq 1 + 2\times\frac{1}{n}\sum_{t>t'} cov(Y_t, Y_{t'}) \\
&\leq 1 + 24\times\frac{1}{n}\sum_{t>t'}\alpha(\sigma(Y_t), \sigma(Y_{t'}))^{1/p}(\mathbb{E}|Y_t|^q)^{1/q}(\mathbb{E}|Y_{t'}|^r)^{1/r} \\
&= 1 + 24\sum_{m=1}^{n}\frac{n-m}{n}\alpha(m)^{1/p} \leq 1 + 24\sum_{m=1}^{\infty}\alpha(m)^{1/p}.
\end{aligned}$$

□

Proof of Lemma D10: Theorem 6.3 Rio (2000) implies the following inequality:

$$\mathbb{E}\left(|\sum_{t=1}^{n} X_t|^p\right) \leq a_p s_n^p + n b_p \int_0^1 \min(\alpha^{-1}(u), n)^{p-1} Q^p(u) du$$

where $a_p = p4^{p+1}(p+1)^{p/2}$ and $b_p = \frac{p}{p-1}4^{p+1}(p+1)^{p-1}$, $Q = \sup_{t>0} Q_t$ and $s_n^2 = \sum_{t=1}^n \sum_{t'=1}^n |\text{cov}(X_t, X_{t'})|$. Since X_t is uniformly bounded, using the results from Appendix C in Rio (2000): $\int_0^1 \min(\alpha^{-1}(u), n)^{p-1} Q^p(u) du \leq 2\left[\sum_{k=0}^{n-1}(k+1)^{p-1}\alpha_k\right] \|\sup_{t>0} X_t\|_\infty$. Because the strong-mixing coefficients are exponentially decreasing, it implies:

$$\sum_{k=0}^{n-1}(k+1)^{p-1}\alpha_k \leq A\exp(C)\sum_{k\geq 1} k^{p-1}\exp(-Ck) \leq A\exp(C)(p-1)^{p-1}\frac{1}{(1-\exp(-C))^{p-1}}$$

And Corollary 1.1 of Rio (2000) yields: $s_n^2 \leq 4\int_0^1 \min(\alpha^{-1}(u), n)\sum_{t=1}^n Q_k^2(u) du$. Altogether:

$$\begin{aligned}
\mathbb{E}\left(|\sqrt{n}\bar{X}_n|^p\right)^{1/p} &\leq K_1(p+1)^{1/2}\left(\int_0^1 \min(\alpha^{-1}(u), n)\sum_{t=1}^{n}\frac{Q_t^2(u)}{n}\right)^{1/2} \\
&+ K_2 n^{1/p-1/2}(p-1)^{(p-1)/p}(p+1)^{(p-1)/p}\|\sup_{t>0} X_t\|_\infty \\
&\leq K\left(\sqrt{p}\left(\int_0^1 \min(\alpha^{-1}(u), n)\sum_{t=1}^{n}\frac{Q_t^2(u)}{n}\right)^{1/2} + n^{1/p-1/2}p^2\|\sup_{t>0} X_t\|_\infty\right).
\end{aligned}$$

with $K_1 \geq 2^{1/p}p^{1/p}4^{(p+1)/p}$, $K_2 \geq (p/[p-1])^{1/p}4^{(p+1)/p}2^{1/p}A\exp(C)\frac{1}{(1-\exp(-C))^{(p-1)/p}}$. Note that since $p \geq 2$, $2^{1/p} \leq \sqrt{2}, p^{1/p} \leq 1, 4^{(p+1)/p} \leq 16$, etc. The constants K_1, K_2 do not depend on p. K only depends on the constants A and C. □

Proof of Lemma D11: Let $Z_n(\beta) = \frac{1}{\sqrt{n}}\sum_{t=1}^n X_t(\beta)$, by Lemma D10:

$$\|Z_n(\beta)\|_p = \mathbb{E}\left(|Z_n(\beta)|^p\right)^{1/p} \leq K\left(\sqrt{p}\frac{1}{n}\sum_{t=1}^{n}\|X_t(\beta)\|^{\vartheta/2} + p^2 n^{-1/2+1/p}\|\sup_{t>0} X_t(\beta)\|_\infty\right).$$

The term $\frac{1}{n}\sum_{t=1}^{n}\|X_t(\beta)\|^{\vartheta}$ comes from Hölder's inequality, for any $\vartheta \in (0,1)$:

$$\Big|\int_0^1 \min(\alpha^{-1}(u),n)\sum_{t=1}^{n}\frac{Q_t^2(u)}{n}\Big|^{1/2} \leq \left(\int_0^1 \min(\alpha^{-1}(u),n)^{1/(1-\vartheta)}\right)^{\frac{1-\vartheta}{2}}\left(\int_0^1 |\frac{1}{n}\sum_{t=1}^{n} Q_t(u)^2|^{1/\vartheta}\right)^{\frac{\vartheta}{2}}$$
$$\leq \left(\frac{1}{1-\vartheta}\sum_{j=1}^{n}(1+j)^{1/(1-\vartheta)}\alpha(j)\right)^{\frac{1-\vartheta}{2}}\frac{1}{n}\sum_{t=1}^{n}\left(\int_0^1 |Q_t(u)|^{2/\vartheta}du\right)^{\frac{\vartheta}{2}}$$
$$\leq \left(\frac{1}{1-\vartheta}\sum_{j=1}^{n}(1+j)^{1/(1-\vartheta)}\alpha(j)\right)^{\frac{1-\vartheta}{2}}\frac{1}{n}\sum_{t=1}^{n}\|Q_t\|_1^{\vartheta/2}.$$

The last inequality follows from assuming $|Q_t| \leq 1$. To simplify notation, use $\frac{1}{n}\sum_{t=1}^{n}\|Q_t\|_1^{\vartheta}$ rather than $\frac{1}{n}\sum_{t=1}^{n}\|Q_t\|_1^{\vartheta/2}$. Also since $\alpha(j)$ has exponential decay, $\sum_{j=1}^{\infty}(1+j)^{1/(1-\vartheta)}\alpha(j) < \infty$ so the first term is a constant which only depends on $(\alpha(j))_j$ and ϑ. To derive the inequality, construct bracketing pairs $(\beta_j^k,\Delta_j^k)_{1\leq j\leq N(k)}$ with $N(k) = N_{[]}(2^{-k},\mathcal{X},\|\cdot\|_2)$ the minimal number of brackets needed to cover $\mathcal{X}$. By definition of $N(k)$ there exists brackets $(\Delta_{t,j}^k)_{j=1,\dots,N(k)}$ such that: 1) $\mathbb{E}\left(|\Delta_{t,j}^k|^2\right)^{1/2} \leq 2^{-k}$ for all t,j,k. 2) For all $\beta \in \mathcal{B}$ and $k \geq 1$, there exists an index j such that $|X_t(\beta) - X_t(\beta_j^k) \leq \Delta_{t,j}^k$. Note that brackets constructed the usual way need not be α-mixing, a construction which preserve the dependence properties is given at the end of the proof.

Assume that, without loss of generality, $|\Delta_j^k| \leq 1$ for all j,k. Let $(\pi_k(\beta),\Delta_k(\beta))$ be a bracketing pair for $\beta \in \mathcal{B}$. Let q_0,k,q be positive integers such that $q_0 \leq k \leq q$ and let $T_k(\beta) = \pi_k \circ \pi_{k+1} \circ \cdots \circ \pi_q(\beta)$. Using the following identity:

$$\left[\mathbb{E}\left(\sup_{\beta\in\mathcal{B}}|Z_n(\beta)|^2\right)\right]^{1/2}$$
$$= \left[\mathbb{E}\left(\sup_{\beta\in\mathcal{B}}|Z_n(\beta) - Z_n(T_q(\beta)) + \sum_{k=q_0+1}^{q}[Z_n(T_k(\beta)) - Z_n(T_{k-1}(\beta))] + Z_n(T_{q_0}(\beta))|^2\right)\right]^{1/2}$$

and the triangle inequality, decompose the identity into three groups:

$$\left[\mathbb{E}\left(\sup_{\beta\in\mathcal{B}}|Z_n(\beta)|^2\right)\right]^{1/2} \leq \left[\mathbb{E}\left(\sup_{\beta\in\mathcal{B}}|Z_n(\beta) - Z_n(T_q(\beta))|^2\right)\right]^{1/2}$$
$$+ \sum_{k=q_0+1}^{q}\left[\mathbb{E}\left(\sup_{h\in\mathcal{B}}|Z_n(T_k(\beta)) - Z_n(T_{k-1}(\beta))|^2\right)\right]^{1/2}$$
$$+ \left[\mathbb{E}\left(\sup_{\beta\in\mathcal{B}}|Z_n(T_{q_0}(\beta))|^2\right)\right]^{1/2} \leq E_{q+1} + \sum_{k=q_0+1}^{q}E_k + E_{q_0}.$$

The following inequality is due to Pisier (1983), for any $X_1, \dots, X_N$ random variables: $\left[\mathbb{E}\left(\max_{1\leq t\leq N}|X_t|^p\right)\right]^{1/p} \leq N^{1/p}\max_{1\leq t\leq N}\left[\mathbb{E}\left(|X_t|^p\right)\right]^{1/p}$. Now that $\{T_k(\beta), \beta \in \mathcal{B}\}$ has at most $N(k)$ elements by construction. Some terms can be simplified: $E_k = \mathbb{E}\left(\max_{g\in T_k(\mathcal{B})}|Z_n(g) - Z_n(T_{k-1}(g))|^2\right)^{1/2}$ for $q_0 + 1 \leq k \leq q$. For $p \geq 2$ using both Hölder and Pisier's inequalities:

$$E_k \leq \left[\mathbb{E}\left(\sup_{\beta\in T_k(\mathcal{B})}|Z_n(\beta) - Z_n(T_{k-1}(\beta))|^p\right)\right]^{1/p} \leq N(k)^{1/p}\max_{g\in T_k(\mathcal{B})}\left[\mathbb{E}\left(|Z_n(g) - Z_n(T_{k-1}(g))|^p\right)\right]^{1/p}.$$

By the definition of Δ_j^k: $E_k \leq N(k)^{1/p}\max_{1\leq j\leq N(k)}\left[\mathbb{E}\left(|\Delta_j^k(g)|^p\right)\right]^{1/p}$. This is also valid for E_{q+1}. Using Rio's inequality for α-mixing dependent processes:

$$\begin{aligned}
E_k &\leq KN(k)^{1/p}\left(\sqrt{p}\max_{g\in T_k(\mathcal{B})}\|\Delta^k(g)\|_1^{\vartheta/2} + p^2 n^{-1/2+1/p}\max_{g\in T_k(\mathcal{B})}\|\Delta^k(g)\|_\infty\right)\\
&\leq KN(k)^{1/p}\left(\sqrt{p}2^{-\vartheta/2k} + p^2 n^{-1/2+1/p}\right)\\
&\leq KN(k)^{1/p}2^{-k}\left(\sqrt{p}2^{k-\vartheta/2k} + p^2[n^{-1/2}2^k]^{1-2/p}2^{2k/p}\right).
\end{aligned}$$

For $p > 2$ and $2^q/\sqrt{n} \geq 1$, the inequality becomes:

$$E_k \leq KN(k)^{1/p}2^{-k}\left(\sqrt{p}2^{k-\vartheta/2k} + p^2[n^{-1/2}2^q]2^{2k/p}\right).$$

Choosing $p = k + \log N(k)$ implies:

$$N(k)^{1/p} \leq \exp(1),\ \sqrt{p} \leq \sqrt{k} + \sqrt{\log N(k)},\ p^2 \leq 4[k^2 + \log^2 N(k)],\ 2^{2k/p} \leq 4.$$

Applying these bounds to the previous inequality:

$$\begin{aligned}
E_k &\leq 16K\exp(1)2^{-k}\left([\sqrt{k} + \sqrt{\log N(k)}]2^{k-\vartheta/2k} + [k^2 + \log(N(k))^2]\frac{2^q}{\sqrt{n}}\right)\\
&\leq \frac{2^q}{\sqrt{n}}16K\exp(1)2^{-k}\left([\sqrt{k} + \sqrt{\log N(k)}]2^{k-\vartheta/2k} + k^2 + \log(N(k))^2\right).
\end{aligned}$$

Note that $\sum_{k\geq 1}(\sqrt{k} + k^2)2^{-k} \leq 2\sum_{k\geq 1}k^2 2^{-k} = 12$. Hence:

$$\sum_{k=q_0+1}^{q+1}E_k \leq \frac{2^{q+1}}{\sqrt{n}}16K\exp(1)\left(12 + \int_0^1[x^{\vartheta/2-1}\sqrt{\log N_{[]}(x, \mathcal{X}, \|\cdot\|)} + \log^2 N_{[]}(x, \mathcal{X}, \|\cdot\|)]dx\right).$$

Pick the smallest integer q such that $q \geq \log(n)/(2\log 2) - 1$ so that $4\sqrt{n} \geq 2^q \geq \sqrt{n}/2$ and $2^q/\sqrt{n} \in [1/2, 4]$. Only E_{q_0} remains to be bounded, using Rio's inequality again:

$$\left[\mathbb{E}\left(\sup_{\beta\in\mathcal{B}}|Z_n(T_{q_0}(\beta))|^2\right)\right]^{1/2} \leq KN(q_0)^{1/p}\left(\sqrt{p}\max_{h\in T_{q_0}(\mathcal{B})}\|X_1(\beta)\|^\vartheta + p^2 n^{-1/2+1/p}\|X_1(\beta)\|_\infty\right).$$

For any $\varepsilon > 0$ pick $p = \max(2+\varepsilon, q_0 + \log N(q_0))$ then: $N(q_0)^{1/p} \leq \exp(1)$, $n^{-1/2+1/p} \leq n^{-1/2+1/(2+\varepsilon)} \leq 1$. Then conclude that:

$$\left[\mathbb{E}\left(\sup_{\beta\in\mathcal{B}} |Z_n(T_{q_0}(\beta))|^2\right)\right]^{1/2} \leq 4\exp(1)K\left(\sqrt{q_0} + \sqrt{\log N(q_0)} + q_0^2 + \log N(q_0)^2\right)$$
$$\leq K' \log N(q_0)^2 \leq K' \int_0^1 \log^2 N_{[]}(x, \mathcal{X}, \|\cdot\|)dx$$

Hence, there exists a constant $K > 0$ which only depends on $(\alpha(m))_{m>0}$ such that:

$$\left[\mathbb{E}\left(\sup_{\beta\in\mathcal{B}} |Z_n(\beta)|^2\right)\right]^{1/2} \leq K \int_0^1 [x^{\vartheta/2-1}\sqrt{\log N_{[]}(x, \mathcal{X}, \|\cdot\|)} + \log^2 N_{[]}(x, \mathcal{X}, \|\cdot\|)]dx.$$

Let $\sqrt{C_n} = K\int_0^1 [x^{\vartheta/2-1}\sqrt{\log N_{[]}(x, \mathcal{X}, \|\cdot\|)} + \log^2 N_{[]}(x, \mathcal{X}, \|\cdot\|)]dx$, then $\mathbb{E}\left(\sup_{\beta\in\mathcal{B}} |Z_n(\beta)|^2\right) \leq C_n$ for all $n \geq 1$.

Bracketing: Because of the dynamics, the dependence of X_t can vary with β, which is not the case in Ben Hariz (2005) or Andrews and Pollard (1994). The following details the construction of the brackets $(\Delta_{t,j}^k)$ in the current setting. Suppose that $\beta \to X_t(\beta)$ is L^p-smooth. Let $\beta_1^k, \dots, \beta_{N(k)}^k$ be such that $\mathcal{B}_{k_n} \subseteq \cup_{j=1}^{N(k)} B_{[\delta/C]^\gamma}(\beta_j^k)$ then for $j \leq N(k)$ and some $Q \geq 2$: $\left[\mathbb{E}\left(\sup_{\|\beta-\beta_j^k\|_{\mathcal{B}}\leq[\delta/C]^\gamma} |X_t(\beta) - X_t(\beta_j^k)|^Q\right)\right]^{1/Q} \leq \delta$. Let $\Delta_{t,j}^k = \sup_{\|\beta-\beta_j^k\|_{\mathcal{B}}\leq[\delta/C]^\gamma} |X_t(\beta) - X_t(\beta_j^k)|$ then $\left[\mathbb{E}\left(\Delta_{t,j}^{2k}\right)\right]^{1/2} \leq \left[\mathbb{E}\left(\Delta_{t,j}^{Qk}\right)\right]^{1/Q}$ by Hölder's inequality which is smaller than δ by construction. $\left[\mathbb{E}(|\Delta_{t,j}^k|^2)\right]^{1/2} \leq \delta = 2^{-k}$ by construction. However, there is no guarantee that $(\Delta_{t,j}^k)_{t\geq 1}$ as constructed above is α-mixing. Another construction for the bracket which preserves the mixing property is now suggested. Let $B \subseteq \mathcal{B}$ a non-empty compact set in $\mathcal{B}$. Note that since the (β_j^k) cover $\mathcal{B}$, they also cover B. Let $\tilde{\Delta}_{t,j}^k$ be such that $|\frac{1}{n}\sum_{t=1}^n \tilde{\Delta}_{t,j}^k| = \sup_{\beta\in B, \|\beta-\beta_j^k\|\leq[\delta/C]^\gamma} |\frac{1}{n}\sum_{t=1}^n X_t(\beta) - X_t(\beta_j^k)|$. Because B is compact, the supremum is attained at some $\tilde{\beta}_j^k \in B$. For all $t = 1, \dots, n$, take $\tilde{\Delta}_{t,j}^k = X_t(\tilde{\beta}_j^k) - X_t(\beta_j^k)$. For each (j,k) the sequence $(\tilde{\Delta}_{t,j}^k)_{t\geq 0}$ is α-mixing by construction. Furthermore, by construction: $|\tilde{\Delta}_{t,j}^k| \leq |\Delta_{t,j}^k|$ and thus $\left[\mathbb{E}(|\tilde{\Delta}_{t,j}^k|^Q)\right]^{1/Q} \leq 2^{-k}$. These brackets, built in B rather than $\mathcal{B}$, preserve the mixing properties. The rest of the proof applied to B implies:

$$\mathbb{E}\left(\sup_{\beta\in B} |\sqrt{n}[\hat{\psi}_t^S(\beta) - \mathbb{E}(\hat{\psi}_t^S(\beta))]|^2\right)$$
$$\leq K\left(\int_0^1 x^{\vartheta/2-1}\sqrt{\log N_{[]}(x^{1/\gamma}, B, \|\cdot\|)} + \log^2 N_{[]}(x^{1/\gamma}, B, \|\cdot\|)dx\right)$$
$$\leq K\left(\int_0^1 x^{\vartheta/2-1}\sqrt{\log N_{[]}(x^{1/\gamma}, \mathcal{B}, \|\cdot\|)} + \log^2 N_{[]}(x^{1/\gamma}, \mathcal{B}, \|\cdot\|)dx\right).$$

For an increasing sequence of compact sets $B_k \subseteq B_{k+1} \subseteq \mathcal{B}$ dense in $\mathcal{B}$, there is an increasing and bounded sequence:

$$\begin{aligned}
&\mathbb{E}\left(\sup_{\beta \in B_k} |\sqrt{n}[\hat{\psi}_t^S(\beta) - \mathbb{E}(\hat{\psi}_t^S(\beta))]|^2\right) \leq \mathbb{E}\left(\sup_{\beta \in B_{k+1}} |\sqrt{n}[\hat{\psi}_t^S(\beta) - \mathbb{E}(\hat{\psi}_t^S(\beta))]|^2\right) \\
&\leq K\left(\int_0^1 x^{\vartheta/2-1}\sqrt{\log N_{[]}(x^{1/\gamma}, \mathcal{B}, \|\cdot\|)} + \log^2 N_{[]}(x^{1/\gamma}, \mathcal{B}, \|\cdot\|)dx\right).
\end{aligned}$$

This sequence is thus convergent with limit less or equal than the upper-bound. Hence, it must be that the supremum over $\mathcal{B}$ is also bounded. It can thus be assumed that $(\Delta_{t,j}^k)_{t \geq 1}$ are α-mixing. □

Proof of Lemma D12: Since $(\mathbf{y}_t^s, \mathbf{x}_t)$ is geometrically ergodic, the joint density converges to the stationary distribution at a geometric rate: $\|f_t(y,x) - f_t^*(y,x)\|_{TV} \leq C\rho^t$, $\rho < 1$. Because B is bounded linear and the moments $\hat{\psi}_n, \hat{\psi}_n^s$ are bounded above by M, uniformly in τ:

$$\begin{aligned}
Q_n(\beta_0) &\leq M_B^2 \int \left|\mathbb{E}\left(\hat{\psi}_n^S(\tau, \beta_0)\right) - \lim_{n\to\infty} \mathbb{E}\left(\hat{\psi}_n(\tau)\right)\right|^2 \pi(\tau)d\tau \\
&\leq M^2 M_B^2 \int \left|\frac{1}{n}\sum_{t=1}^n \int [f_t(y,x) - f_t^*(y,x)]dydx\right|^2 \pi(\tau)d\tau \\
&\leq M^2 M_B^2 \left(\frac{1}{n}\sum_{t=1}^n \int |f_t(y,x) - f_t^*(y,x)|\, dydx\right)^2 \\
&\leq CM^2 M_B^2 \left(\frac{1}{n}\sum_{t=1}^n \rho^t\right)^2 \leq \frac{CM^2M_B^2}{(1-\rho)^2} \times \frac{1}{n^2} = O(1/n^2).
\end{aligned}$$

□

Proof of Lemma D13. Lemma D11 implies that for some $C > 0$:

$$\begin{aligned}
&\left[\mathbb{E}\left(\sup_{\|\beta_1-\beta_2\|_m \leq \delta, \|\beta_j - \Pi_{k(n)}\beta_0\|_m \leq M_n\delta_{m,n}, j=1,2} \left|\hat{\psi}_t^s(\tau,\beta_1) - \hat{\psi}_t^s(\tau,\beta_2)\right|^2\right)\right]^{1/2} \frac{\sqrt{\pi(\tau)}}{(M_n\delta_{m,n})^{\gamma^2/2}} \\
&\leq Ck(n)^{2\gamma^2}\left(\frac{\delta}{M_n\delta_{m,n}}\right)^{\gamma^2/2}.
\end{aligned}$$

Next, apply the inequality of Lemma D11 to generate the bound:

$$\left[\mathbb{E}\left(\sup_{\|\beta - \Pi_{k(n)}\beta_0\|_m \leq M_n\delta_{m,n}} \left|\Delta_n^S(\tau,\beta) - \Delta_n^S(\tau, \Pi_{k(n)}\beta_0)\right|^2\right)\right]^{1/2} \sqrt{\pi(\tau)} \leq \overline{C}\frac{(M_n\delta_{m,n})^{\gamma^2/2}}{\sqrt{n}} J_{m,n}$$

for some $\overline{C} > 0, \vartheta \in (0,1)$ and

$$J_{m,n} = \int_0^1 \left(x^{-\vartheta/2} \sqrt{\log N(\left[\frac{xM_n\delta_{mn}}{k(n)^{2\gamma^2}}\right]^{\frac{2}{\gamma^2}}, \mathcal{B}_{k(n)}, \|\cdot\|_m) + \log^2 N(\left[\frac{xM_n\delta_{mn}}{k(n)^{2\gamma^2}}\right]^{\frac{2}{\gamma^2}}, \mathcal{B}_{k(n)}, \|\cdot\|_m)} \right) dx.$$

Since $\int \sqrt{\pi(\tau)} d\tau < \infty$, the term on the left-hand side of the inequality can be squared and multiplied by $\sqrt{\pi(\tau)}$. Then, taking the integral:

$$\left[\mathbb{E} \left(\int \sup_{\|\beta - \Pi_{k(n)}\beta_0\|_m \leq M_n\delta_{m,n}} \left| \Delta_n^S(\tau,\beta) - \Delta_n^S(\tau, \Pi_{k(n)}\beta_0) \right|^2 \pi(\tau) d\tau \right) \right]^{1/2} \leq \overline{C}_\pi \frac{(M_n\delta_{m,n})^{\gamma^2/2}}{\sqrt{n}} J_{m,n}$$

where $\overline{C}_\pi = \overline{C} \int \sqrt{\pi(\tau)} d\tau$. Note that $J_{m,n} = O(k(n)^2 \max(\log[k(n)]^2, \log[M_n\delta_{m,n}]^2))$.

To prove the final statement, notation will be shortened using $\Delta\hat{\psi}_t^s(\tau,\beta) = \hat{\psi}_t^s(\tau,\beta_0) - \hat{\psi}_t^s(\tau,\beta)$. Note that, by applying Davydov (1968)'s inequality:

$$\begin{aligned}
n\mathbb{E}\left|\Delta\hat{\psi}_n^S(\tau,\Pi_{k(n)}\beta_0) - \mathbb{E}[\Delta\hat{\psi}_n^S(\tau,\Pi_{k(n)}\beta_0)]\right|^2 &\leq \frac{1}{n}\sum_{t=1}^n \mathbb{E}\left|\Delta\hat{\psi}_t^s(\tau,\Pi_{k(n)}\beta_0) - \mathbb{E}[\Delta\hat{\psi}_t^s(\tau,\Pi_{k(n)}\beta_0)]\right|^2 \\
&+ \frac{24}{n}\sum_{m=1}^n (n-m)\alpha(m)^{1/3} \max_{1\leq t\leq n} \left(\mathbb{E}\left|\Delta\hat{\psi}_t^s(\tau,\Pi_{k(n)}\beta_0) - \mathbb{E}[\Delta\hat{\psi}_t^s(\tau,\Pi_{k(n)}\beta_0)]\right|^6 \right)^{2/3} \\
&\leq \left(1 + 24\sum_{m\geq 1}\alpha(m)^{1/3}\right) \max_{1\leq t\leq n} \left(\mathbb{E}\left|\Delta\hat{\psi}_t^s(\tau,\Pi_{k(n)}\beta_0) - \mathbb{E}[\Delta\hat{\psi}_t^s(\tau,\Pi_{k(n)}\beta_0)]\right|^6 \right)^{2/3} \\
&\leq 4^{8/3}\left(1 + 24\sum_{m\geq 1}\alpha(m)^{1/3}\right) \max_{1\leq t\leq n} \left(\mathbb{E}\left|\Delta\hat{\psi}_t^s(\tau,\Pi_{k(n)}\beta_0) - \mathbb{E}[\Delta\hat{\psi}_t^s(\tau,\Pi_{k(n)}\beta_0)]\right|^2 \right)^{2/3}.
\end{aligned}$$

The last inequality is due to $|\Delta\hat{\psi}_t^s(\tau,\beta)| \leq 2$. By the continuity assumption the last term is a $o(1)$ when $\|\beta_0 - \Pi_{k(n)}\|_{\mathcal{B}} \to 0$. As a result: $\int \mathbb{E}\left|\Delta\hat{\psi}_n^S(\tau,\Pi_{k(n)}\beta_0) - \mathbb{E}[\Delta\hat{\psi}_n^S(\tau,\Pi_{k(n)}\beta_0)]\right|^2 \pi(\tau)d\tau = o(1/n)$. To conclude the proof, apply a triangle inequality and the results above:

$$\begin{aligned}
&\left[\mathbb{E} \left(\int \sup_{\|\beta - \Pi_{k(n)}\beta_0\|_m \leq M_n\delta_{mn}} \left| \Delta_n^S(\tau,\beta) - \Delta_n^S(\tau,\beta_0) \right|^2 \pi(\tau) d\tau \right) \right]^{1/2} \\
&\leq \left[\mathbb{E} \left(\int \sup_{\|\beta - \Pi_{k(n)}\beta_0\|_m \leq M_n\delta_{mn}} \left| \Delta_n^S(\tau,\beta) - \Delta_n^S(\tau,\Pi_{k(n)}\beta_0) \right|^2 \pi(\tau) d\tau \right) \right]^{1/2} \\
&+ \left[\int \mathbb{E} \left(\left| \Delta\hat{\psi}_n^S(\tau,\Pi_{k(n)}\beta_0) - \mathbb{E}[\Delta\hat{\psi}_n^S(\tau,\Pi_{k(n)}\beta_0)] \right|^2 \pi(\tau) d\tau \right) \right]^{1/2} = o(1/\sqrt{n}).
\end{aligned}$$

□

Proof of Lemma D14: Let $R_n(\beta,\beta_0) = \mathbb{E}(\hat{\psi}_n^S(\tau,\beta) - \hat{\psi}_n^S(\tau,\beta_0)) - \frac{d\mathbb{E}(\hat{\psi}_n^S(\tau,\beta_0))}{d\beta}[\beta-\beta_0]$.

a) Since B bounded linear, the Cauchy-Schwarz inequality implies:

$$\Big|\int \psi_\beta(\tau,u_n^*)\big(\overline{B\mathbb{E}(\hat{\psi}_n^S(\tau,\hat{\beta}_n) - \hat{\psi}_n^S(\tau,\beta_0)) - B\frac{d\mathbb{E}(\hat{\psi}_n^S(\tau,\beta_0))}{d\beta}[\hat{\beta}_n-\beta_0]}\big)\pi(\tau)d\tau\Big|$$

$$= \Big|\int \psi_\beta(\tau,u_n^*)\big(\overline{BR_n(\hat{\beta}_n,\beta_0)}\big)\pi(\tau)d\tau\Big| \leq M_B\big(\int|\psi_\beta(\tau,u_n^*)|^2\pi(\tau)d\tau\big)^{1/2}\big(\int\Big|R_n(\hat{\beta}_n,\beta_0)\Big|^2\pi(\tau)d\tau\big)^{1/2}$$

By definition of M_n and the inequality above:

$$\begin{aligned}
&\mathbb{P}\left(\Big|\int \psi_\beta(\tau,u_n^*)\left(\overline{BR_n(\hat{\beta}_n,\beta_0)}\right)\pi(\tau)d\tau\Big| > \frac{\varepsilon}{\sqrt{n}}\right)\\
&\leq \mathbb{P}\left[M_B^2\left(\int|\psi_\beta(\tau,u_n^*)|^2\pi(\tau)d\tau\right)\sup_{\|\beta-\beta_0\|_{weak}\leq M_n\delta_n}\left(\int\Big|R_n(\beta,\beta_0)\Big|^2\pi(\tau)d\tau\right) > \frac{\varepsilon^2}{n}\right]\\
&+\mathbb{P}\left(\|\hat{\beta}_n-\beta_0\|_{\mathcal{B}} > M_n\delta_n\right)
\end{aligned}$$

$\mathbb{P}\left(\|\hat{\beta}_n-\beta_0\|_{\mathcal{B}} > M_n\delta_n\right) \to 0$ regardless of ε. Furthermore, Assumption 5 *ii.* implies:

$$\begin{aligned}
&\sup_{\|\beta-\beta_0\|_{weak}\leq M_n\delta_n}\left(\int\Big|\mathbb{E}(\hat{\psi}_n^S(\tau,\beta) - \hat{\psi}_n^S(\tau,\beta_0)) - \frac{d\mathbb{E}(\hat{\psi}_n^S(\tau,\beta_0))}{d\beta}[\beta-\beta_0]\Big|^2\pi(\tau)d\tau\right)^{1/2}\\
&= \sup_{\|\beta-\beta_0\|_{weak}\leq M_n\delta_n}\left(\int\Big|R_n(\beta,\beta_0)\Big|^2\pi(\tau)d\tau\right)^{1/2} = O\left((M_n\delta_n)^2\right).
\end{aligned}$$

Assumption 5 *i.* implies that $(M_n\delta_n)^2 = o(\frac{1}{\sqrt{n}})$, and thus: $\mathbb{P}\big(\Big|\int\psi_\beta(\tau,u_n^*)\big(\overline{BR_n(\hat{\beta}_n,\beta_0)}\big)\pi(\tau)d\tau\Big| > \frac{\varepsilon}{\sqrt{n}}\big) = o(1)$ regardless of $\varepsilon > 0$. Hence: $\int\psi_\beta(\tau,u_n^*)\big(\overline{BR_n(\hat{\beta}_n,\beta_0)}\big)\pi(\tau)d\tau = o_p(1/\sqrt{n})$.

b) Let $\Delta_n^S(\tau,\beta) = \hat{\psi}_n^S(\tau,\beta) - \mathbb{E}[\hat{\psi}_n^S(\tau,\beta)]$. By the second stochastic equicontinuity result of Lemma D13 and the Cauchy-Schwarz inequality:

$$\begin{aligned}
&\Big|\int\psi_\beta(\tau,u_n^*)\left(\overline{B[\Delta_n^S(\hat{\beta}_n)-\Delta_n^S(\beta_0)]}\right)\pi(\tau)d\tau\Big|\\
&\leq\left(\int|\psi_\beta(\tau,u_n^*)|^2\pi(\tau)d\tau\right)^{1/2}\left(\int\Big|B[\Delta_n^S(\hat{\beta}_n)-\Delta_n^S(\beta_0)]\Big|^2\pi(\tau)d\tau\right)^{1/2}\\
&\leq M_B\left(\int|\psi_\beta(\tau,u_n^*)|^2\pi(\tau)d\tau\right)^{1/2}\left(\int\Big|[\Delta_n^S(\hat{\beta}_n)-\Delta_n^S(\beta_0)]\Big|^2\pi(\tau)d\tau\right)^{1/2}\\
&\leq M_B\left(\int|\psi_\beta(\tau,u_n^*)|^2\pi(\tau)d\tau\right)^{1/2}\left(\sup_{\|\beta-\Pi_{k(n)}\beta_0\|\leq M_n\delta_{mn}}\int\Big|[\Delta_n^S(\beta)-\Delta_n^S(\beta_0)]\Big|^2\pi(\tau)d\tau\right)^{1/2}\\
&= o_p(1/\sqrt{n}),
\end{aligned}$$

where the last inequality holds with probability going to 1 by definition of $M_n\delta_{mn}$.

c) Let $\varepsilon_n = \pm\frac{1}{\sqrt{n}M_n} = o(\frac{1}{\sqrt{n}})$. For $h \in (0,1)$ define $\hat\beta(h) = \hat\beta_n + h\varepsilon_n u_n^*$. Since $\hat\beta_n = \hat\beta(0)$. Recall that $\hat\beta_n$ is the approximate minimizer of $\hat Q_n^s$ so that: $0 \le \hat Q_n^S(\hat\beta_n) \le \inf_{\beta\in\mathcal{B}_{k(n)}} \hat Q_n^S(\beta) + O_p(\eta_n)$. Hence the following holds:

$$
\begin{aligned}
0 &\le \frac{1}{2}\left(\hat Q_n^S(\hat\beta(1)) - \hat Q_n^S(\hat\beta(0))\right) + O_p(\eta_n) &\text{(E.22)}\\
&= \frac{1}{2}\Big[\int B\left(\hat\psi_n(\tau) - \hat\psi_n^S(\tau,\hat\beta(0))\right)\overline{B\left(\hat\psi_n^S(\tau,\hat\beta(0)) - \hat\psi_n^S(\tau,\hat\beta(1))\right)}\pi(\tau)d\tau &\text{(E.23)}\\
&+ \int \overline{B\left(\hat\psi_n(\tau) - \hat\psi_n^S(\tau,\hat\beta(0))\right)} B\left(\hat\psi_n^S(\tau,\hat\beta(0)) - \hat\psi_n^S(\tau,\hat\beta(1))\right)\pi(\tau)d\tau &\text{(E.24)}\\
&+ \int \left|B\left(\hat\psi_n^S(\tau,\hat\beta(0)) - \hat\psi_n^S(\tau,\hat\beta(1))\right)\right|^2\pi(\tau)d\tau\Big] + O_p(\eta_n). &\text{(E.25)}
\end{aligned}
$$

To prove Lemma D14 c), (E.23)-(E.24) are expanded individually and shown to be $o_p(1/\sqrt{n})$ and (E.25) is bounded, shown to be negligible under the assumptions.
The first step deals with (E.25):

$$
\begin{aligned}
&\left(\int \left|B\left(\hat\psi_n^S(\tau,\hat\beta(0)) - \hat\psi_n^S(\tau,\hat\beta(1))\right)\right|^2\pi(\tau)d\tau\right)^{1/2} \le M_B\left(\int\left|\hat\psi_n^S(\tau,\hat\beta(0)) - \hat\psi_n^S(\tau,\hat\beta(1))\right|^2\pi(\tau)d\tau\right)^{1/2}\\
&\le \left(\int\left|[\hat\psi_n^S(\tau,\hat\beta(0)) - \hat\psi_n^S(\tau,\hat\beta(1))] - \mathbb{E}[\hat\psi_n^S(\tau,\hat\beta(0)) - \hat\psi_n^S(\tau,\hat\beta(1))]\right|^2\pi(\tau)d\tau\right)^{1/2}\\
&+ \left(\int\left|\mathbb{E}[\hat\psi_n^S(\tau,\hat\beta(0)) - \hat\psi_n^S(\tau,\hat\beta(1))]\right|^2\pi(\tau)d\tau\right)^{1/2}
\end{aligned}
$$

By the triangle inequality and the stochastic equicontinuity results from Lemma D13:

$$
\begin{aligned}
&\left(\int\left|[\hat\psi_n^S(\tau,\hat\beta(0)) - \hat\psi_n^S(\tau,\hat\beta(1))] - \mathbb{E}[\hat\psi_n^S(\tau,\hat\beta(0)) - \hat\psi_n^S(\tau,\hat\beta(1))]\right|^2\pi(\tau)d\tau\right)^{1/2}\\
&= O_p\left(\frac{I_{m,n}(M_n\delta_{mn})^{\gamma^2/2}}{\sqrt{n}}\right).
\end{aligned}
$$

Also, note that $\hat\beta(1) = \hat\beta(0) + \varepsilon_n u_n^*$, so that the Mean Value Theorem applies to last term:

$$
\left(\int\left|\mathbb{E}[\hat\psi_n^S(\tau,\hat\beta(0)) - \hat\psi_n^S(\tau,\hat\beta(1))]\right|^2\pi(\tau)d\tau\right) = \left(\int\left|\frac{d\mathbb{E}[\hat\psi_n^S(\tau,\hat\beta(\tilde h))}{d\beta}[\varepsilon_n u_n^*]\right|^2\pi(\tau)d\tau\right)
$$

for some intermediate value $\tilde h \in (0,1)$. Also, by assumption: $\left(\int\left|\frac{d\mathbb{E}[\hat\psi_n^S(\tau,\hat\beta(\tilde h))}{d\beta}[u_n^*]\right|^2\pi(\tau)d\tau\right)^{1/2} = O_p(1)$. Together these two imply: $\left(\int\left|\mathbb{E}[\hat\psi_n^S(\tau,\hat\beta(0)) - \hat\psi_n^S(\tau,\hat\beta(1))]\right|^2\pi(\tau)d\tau\right)^{1/2} = O(\varepsilon_n)$. This yields the bound for (E.25):

$$
\int\left|B\left(\hat\psi_n^S(\tau,\hat\beta(0)) - \hat\psi_n^S(\tau,\hat\beta(1))\right)\right|^2\pi(\tau)d\tau \le O_p(\varepsilon_n^2) + O_p\left(\frac{(M_n\delta_{mn})^{\gamma^2}I_{m,n}^2}{n}\right).
$$

The remaining terms, (E.23)-(E.24), are conjugates of each other. A bound for (E.23) is also valid for (E.24). Expanding (E.23) yields:

$$\int B\left(\hat{\psi}_n(\tau)-\hat{\psi}_n^S(\tau,\hat{\beta}(0))\right)\overline{B\left(\hat{\psi}_n^S(\tau,\hat{\beta}(0))-\hat{\psi}_n^S(\tau,\hat{\beta}(1))\right)}\pi(\tau)d\tau \tag{E.23}$$

$$=\int B\left(\hat{\psi}_n(\tau)-\hat{\psi}_n^S(\tau,\hat{\beta}(0))\right)\left[\overline{B\left(\Delta_n^S(\tau,\hat{\beta}(0))-\Delta_n^S(\tau,\hat{\beta}(1))\right)}\right]\pi(\tau)d\tau \tag{E.26}$$

$$+\int B\left(\hat{\psi}_n(\tau)-\hat{\psi}_n^S(\tau,\hat{\beta}(0))\right)\overline{B\mathbb{E}\left(\hat{\psi}_n^S(\tau,\hat{\beta}(0))-\hat{\psi}_n^S(\tau,\hat{\beta}(1))\right)}\pi(\tau)d\tau. \tag{E.27}$$

Applying the Cauchy-Schwarz inequality to (E.26) implies:

$$\Big|\int B\left(\hat{\psi}_n(\tau)-\hat{\psi}_n^S(\tau,\hat{\beta}(0))\right)\left[\overline{B\left(\Delta_n^S(\tau,\hat{\beta}(0))-\Delta_n^S(\tau,\hat{\beta}(1))\right)}\right]\pi(\tau)d\tau\Big| \tag{E.26}$$

$$\leq M_B\left(\int\left|B\hat{\psi}_n(\tau)-B\hat{\psi}_n^S(\tau,\hat{\beta}(0))\right|^2\pi(\tau)d\tau\right)^{1/2} \tag{E.28}$$

$$\times\left(\int\left|\Delta_n^S(\tau,\hat{\beta}(0))-\Delta_n^S(\tau,\hat{\beta}(1))\right|^2\pi(\tau)d\tau\right)^{1/2} \tag{E.29}$$

The term (E.28) can be bounded above using the triangle inequality:

$$\Big(\int\left|B\hat{\psi}_n(\tau)-B\hat{\psi}_n^S(\tau,\hat{\beta}(0))\right|^2\pi(\tau)d\tau\Big)^{1/2}$$
$$\leq M_B\Big(\int\left|\hat{\psi}_n(\tau)-\hat{\psi}_n^S(\tau,\beta_0)\right|^2\pi(\tau)d\tau\Big)^{1/2}+\Big(\int\left|B\hat{\psi}_n^S(\tau,\beta_0)-B\hat{\psi}_n^S(\tau,\hat{\beta}(0))\right|^2\pi(\tau)d\tau\Big)^{1/2}.$$

An application of Lemma D9 and the geometric ergodicity of $(\mathbf{y}_t^s,\mathbf{x}_t)$ yields: $\left(\int\left|\hat{\psi}_n(\tau)-\hat{\psi}_n^S(\tau,\beta_0)\right|^2\pi(\tau)d\tau\right)^{1/2}=O_p(1/\sqrt{n})$. Then, expanding the term in $\hat{\psi}_n^s$:

$$\left(\int\left|B\hat{\psi}_n^S(\tau,\beta_0)-B\hat{\psi}_n^S(\tau,\hat{\beta}(0))\right|^2\pi(\tau)d\tau\right)^{1/2}\leq\left(\int\left|B\mathbb{E}[\hat{\psi}_n^S(\tau,\beta_0)-\hat{\psi}_n^S(\tau,\hat{\beta}(0))]\right|^2\pi(\tau)d\tau\right)^{1/2}$$
$$+M_B\left(\int\left|[\hat{\psi}_n^S(\tau,\beta_0)-\hat{\psi}_n^S(\tau,\hat{\beta}(0))]-\mathbb{E}[\hat{\psi}_n^S(\tau,\beta_0)-\hat{\psi}_n^S(\tau,\hat{\beta}(0))]\right|^2\pi(\tau)d\tau\right)^{1/2}$$
$$\leq\left(\int\left|B\mathbb{E}[\hat{\psi}_n^S(\tau,\beta_0)-\hat{\psi}_n^S(\tau,\hat{\beta}(0))]\right|^2\pi(\tau)d\tau\right)^{1/2}+O_p\left(\frac{(M_n\delta_{mn})^{\gamma^2/2}I_{m,n}}{\sqrt{n}}\right)$$
$$\leq M_B\left(\int\left|\mathbb{E}[\hat{\psi}_n^S(\tau,\beta_0)-\hat{\psi}_n^S(\tau,\hat{\beta}(0))]-\frac{d\mathbb{E}(\hat{\psi}_n^S(\tau,\beta_0))}{d\beta}[\beta_0-\hat{\beta}(0)]\right|^2\pi(\tau)d\tau\right)^{1/2}$$
$$+\left(\int\left|B\frac{d\mathbb{E}(\hat{\psi}_n^S(\tau,\beta_0))}{d\beta}[\beta_0-\hat{\beta}(0)]\right|^2\pi(\tau)d\tau\right)^{1/2}+O_p\left(\frac{(M_n\delta_{mn})^{\gamma^2/2}I_{m,n}}{\sqrt{n}}\right).$$

Note that Assumption 5 ii. implies that:

$$\left(\int \left|\mathbb{E}[\hat{\psi}_n^S(\tau,\beta_0) - \hat{\psi}_n^S(\tau,\hat{\beta}(0))] - \frac{d\mathbb{E}(\hat{\psi}_n^S(\tau,\beta_0))}{d\beta}[\beta_0 - \hat{\beta}(0)]\right|^2 \pi(\tau)d\tau\right)^{1/2} = O_p(M_n\delta_n).$$

By definition of the weak norm: $\left(\int \left|B\frac{d\mathbb{E}(\hat{\psi}_n^S(\tau,\beta_0))}{d\beta}[\beta_0 - \hat{\beta}(0)]\right|^2 \pi(\tau)d\tau\right)^{1/2} = \|\hat{\beta}_n - \beta_0\|_{weak}$. Furthermore, $\|\hat{\beta}_n - \beta_0\|_{weak} = O_p(\delta_n)$ by assumption. Overall, the following bound holds for (E.27): $\left(\int \left|B\hat{\psi}_n(\tau) - B\hat{\psi}_n^S(\tau,\hat{\beta}(0))\right|^2 \pi(\tau)d\tau\right)^{1/2} \leq O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\delta_n\right) + O_p\left(\frac{(M_n\delta_{mn})^{\gamma^2/2}I_{m,n}}{\sqrt{n}}\right)$.
Re-arranging (E.29) to apply the stochastic equicontinuity result again yields:

$$\left(\int \left|\Delta_n^S(\tau,\hat{\beta}(0)) - \Delta_n^S(\tau,\hat{\beta}(1))\right|^2 \pi(\tau)d\tau\right)^{1/2} \leq \left(\int \left|\Delta_n^S(\tau,\beta_0) - \Delta_n^S(\tau,\hat{\beta}(1))\right|^2 \pi(\tau)d\tau\right)^{1/2}$$
$$+ \left(\int \left|\Delta_n^S(\tau,\beta_0) - \Delta_n^S(\tau,\hat{\beta}(0))\right|^2 \pi(\tau)d\tau\right)^{1/2} = O_p\left(\frac{(M_n\delta_{mn})^{\gamma^2/2}I_{m,n}}{\sqrt{n}}\right).$$

Using the bounds for (E.27) and (E.29) yields the bound for (E.26):

$$\left|\int B\left(\hat{\psi}_n(\tau) - \hat{\psi}_n^S(\tau,\hat{\beta}(0))\right)\left[\overline{B\left(\Delta_n^S(\tau,\hat{\beta}(0)) - \Delta_n^S(\tau,\hat{\beta}(1))\right)}\right]\pi(\tau)d\tau\right|$$
$$\leq O_p\left(\frac{(M_n\delta_{mn})^{\gamma^2/2}I_{m,n}}{\sqrt{n}}\right) O_p\left(\max\left(M_n\delta_n, \frac{1}{\sqrt{n}}, \frac{(M_n\delta_{mn})^{\gamma^2/2}I_{m,n}}{\sqrt{n}}\right)\right).$$

To bound (E.27), apply the Mean Value theorem up to the second order:

$$\int B\left(\hat{\psi}_n(\tau) - \hat{\psi}_n^S(\tau,\hat{\beta}(0))\right)\overline{B\mathbb{E}\left(\hat{\psi}_n^S(\tau,\hat{\beta}(0)) - \hat{\psi}_n^S(\tau,\hat{\beta}(1))\right)}\pi(\tau)d\tau$$
$$= -\int B\left(\hat{\psi}_n(\tau) - \hat{\psi}_n^S(\tau,\hat{\beta}(0))\right)\overline{B\frac{d\mathbb{E}(\hat{\psi}_n^S(\tau,\hat{\beta}(0)))}{d\beta}[\varepsilon_n u_n^*]}\pi(\tau)d\tau$$
$$+ \frac{1}{2}\int B\left(\hat{\psi}_n(\tau) - \hat{\psi}_n^S(\tau,\hat{\beta}(0))\right)\overline{B\frac{d^2\mathbb{E}(\hat{\psi}_n^S(\tau,\hat{\beta}(\tilde{h})))}{d\beta d\beta}[\varepsilon_n u_n^*, \varepsilon_n u_n^*]}\pi(\tau)d\tau$$
$$= -\int B\left(\hat{\psi}_n(\tau) - \hat{\psi}_n^S(\tau,\hat{\beta}(0))\right)\overline{B\frac{d\mathbb{E}(\hat{\psi}_n^S(\tau,\beta_0))}{d\beta}[\varepsilon_n u_n^*]}\pi(\tau)d\tau + O_p(\varepsilon_n^2)$$
$$+ \int B\left(\hat{\psi}_n(\tau) - \hat{\psi}_n^S(\tau,\hat{\beta}(0))\right)\overline{B\left[\frac{d\mathbb{E}(\hat{\psi}_n^S(\tau,\hat{\beta}(0)))}{d\beta}[\varepsilon_n u_n^*] - \frac{d\mathbb{E}(\hat{\psi}_n^S(\tau,\beta_0))}{d\beta}[\varepsilon_n u_n^*]\right]}\pi(\tau)d\tau.$$

Where the $O_p(\varepsilon_n^2)$ term is due to the Cauchy-Schwarz inequality and Assumption 5 ii.:

$$\Big|\int B\left(\hat{\psi}_n(\tau)-\hat{\psi}_n^S(\tau,\hat{\beta}(0))\right)\overline{\frac{1}{2}B\frac{d^2\mathbb{E}(\hat{\psi}_n^S(\tau,\hat{\beta}(\tilde{t})))}{d\beta d\beta}[\varepsilon_n u_n^*,\varepsilon_n u_n^*]}\pi(\tau)d\tau\Big|^2$$
$$\leq \frac{\varepsilon_n^2}{2}\left(\int\left|B\left(\hat{\psi}_n(\tau)-\hat{\psi}_n^S(\tau,\hat{\beta}(0))\right)\right|^2\pi(\tau)d\tau\right)\int\left|B\frac{d^2\mathbb{E}(\hat{\psi}_n^S(\tau,\hat{\beta}(\tilde{t})))}{d\beta d\beta}[u_n^*,u_n^*]\right|^2\pi(\tau)d\tau.$$

It was shown above that:

$$\left(\int\left|B\left(\hat{\psi}_n(\tau)-\hat{\psi}_n^S(\tau,\hat{\beta}(0))\right)\right|^2\pi(\tau)d\tau\right)=O_p\left(\max\left(M_n\delta_n,\frac{1}{\sqrt{n}},\frac{(M_n\delta_{mn})^{\gamma^2/2}I_{m,n}}{\sqrt{n}}\right)^2\right).$$

Also, by Assumption 5 *ii.*: $\left(\int\left|B\frac{d^2\mathbb{E}(\hat{\psi}_n^S(\tau,\hat{\beta}(\tilde{t})))}{d\beta d\beta}[u_n^*,u_n^*]\right|^2\pi(\tau)d\tau\right)=O_p(1)$.
Finally, applying the Cauchy-Schwarz inequality to the last term of the expansion of (E.27) yields:

$$\int B\left(\hat{\psi}_n(\tau)-\hat{\psi}_n^S(\tau,\hat{\beta}(0))\right)\left[\overline{B\frac{d\mathbb{E}(\hat{\psi}_n^S(\tau,\hat{\beta}(0)))}{d\beta}[\varepsilon_n u_n^*]-B\frac{d\mathbb{E}(\hat{\psi}_n^S(\tau,\beta_0))}{d\beta}[\varepsilon_n u_n^*]}\right]\pi(\tau)d\tau$$
$$\leq\left(\int\left|B\hat{\psi}_n(\tau)-\hat{\psi}_n^S(\tau,\hat{\beta}(0))\right|^2\pi(\tau)d\tau\right)^{1/2}$$
$$\times\,\varepsilon_n\left(\int\left|B\frac{d\mathbb{E}(\hat{\psi}_n^S(\tau,\hat{\beta}(0)))}{d\beta}[u_n^*]-B\frac{d\mathbb{E}(\hat{\psi}_n^S(\tau,\beta_0))}{d\beta}[u_n^*]\right|^2\pi(\tau)d\tau\right)^{1/2}$$
$$=O_p\left(\varepsilon_n\max\left(M_n\delta_n,\frac{1}{\sqrt{n}},\frac{(M_n\delta_{mn})^{\gamma^2/2}I_{m,n}}{\sqrt{n}}\right)\delta_n\right).$$

Using inequality (E.22) together with the bounds above and the expansions of (E.23) and (E.24) yields:

$$0\leq-\varepsilon_n\int B\left(\hat{\psi}_n(\tau)-\hat{\psi}_n^S(\tau,\hat{\beta}(0))\right)\overline{B\frac{d\mathbb{E}(\hat{\psi}_n^S(\tau,\beta_0))}{d\beta}[u_n^*]}\pi(\tau)d\tau$$
$$-\varepsilon_n\int\overline{B\left(\hat{\psi}_n(\tau)-\hat{\psi}_n^S(\tau,\hat{\beta}(0))\right)}B\frac{d\mathbb{E}(\hat{\psi}_n^S(\tau,\beta_0))}{d\beta}[u_n^*]\pi(\tau)d\tau$$
$$+O_p\left(\varepsilon_n^2\right)+O_p\left(\frac{(M_n\delta_{mn})^{\gamma^2/2}I_{m,n}}{\sqrt{n}}\max(M_n\delta_n,\frac{1}{\sqrt{n}},\frac{(M_n\delta_{mn})^{\gamma^2/2}I_{m,n}}{\sqrt{n}}))\right)$$
$$+O_p\left(\varepsilon_n M_n\delta_n\max(M_n\delta_n,\frac{1}{\sqrt{n}},\frac{(M_n\delta_{mn})^{\gamma^2/2}I_{m,n}}{\sqrt{n}})\right)+O_p\left(\frac{[(M_n\delta_{mn})^{\gamma^2/2}I_{m,n}]^2}{n}\right)$$

Since $\varepsilon_n = \pm \frac{1}{\sqrt{n}M_n}$, dividing by ε_n both keeps and flips the inequality so that:

$$\int B\left(\hat{\psi}_n(\tau) - \hat{\psi}_n^S(\tau, \hat{\beta}_n)\right) \overline{B \frac{d\mathbb{E}(\hat{\psi}_n^S(\tau, \beta_0))}{d\beta}[u_n^*]\pi(\tau)d\tau}$$
$$+ \int \overline{B\left(\hat{\psi}_n(\tau) - \hat{\psi}_n^S(\tau, \hat{\beta}_n)\right)} B \frac{d\mathbb{E}(\hat{\psi}_n^S(\tau, \beta_0))}{d\beta}[u_n^*]\pi(\tau)d\tau$$
$$= O_p(\varepsilon_n) + O_p\left(\frac{(M_n\delta_{mn})^{\gamma^2/2} I_{m,n}}{\varepsilon_n \sqrt{n}} \max\left(M_n\delta_n, \frac{1}{\sqrt{n}}, \frac{(M_n\delta_{mn})^{\gamma^2/2} I_{m,n}}{\sqrt{n}})\right)\right)$$
$$+ O_p\left(\max\left(M_n\delta_n, \frac{1}{\sqrt{n}}, \frac{(M_n\delta_{mn})^{\gamma^2/2} I_{m,n}}{\sqrt{n}}\right)\delta_n\right) + O_p\left(\frac{[(M_n\delta_{mn})^{\gamma^2/2} I_{m,n}]^2}{\varepsilon_n n}\right).$$

By construction, $\varepsilon_n = o_p(1/\sqrt{n})$ and Assumption 5 i. implies that $(M_n\delta_{mn})^{\gamma^2/2} I_{m,n} = o(1)$ so that all terms above are $o(1/\sqrt{n})$. To conclude the proof, note that:

$$\int B\left(\hat{\psi}_n(\tau) - \hat{\psi}_n^S(\tau, \hat{\beta}_n)\right) \overline{B \frac{d\mathbb{E}(\hat{\psi}_n^S(\tau, \beta_0))}{d\beta}[u_n^*]\pi(\tau)d\tau}$$
$$+ \int \overline{B\left(\hat{\psi}_n(\tau) - \hat{\psi}_n^S(\tau, \hat{\beta}_n)\right)} B \frac{d\mathbb{E}(\hat{\psi}_n^S(\tau, \beta_0))}{d\beta}[u_n^*]\pi(\tau)d\tau$$
$$= \int [\psi_\beta(\tau, u_n^*)\left(\overline{B[\hat{\psi}_n(\tau) - \hat{\psi}_n^S(\tau, \hat{\beta}_n)]}\right) + \overline{\psi_\beta(\tau, u_n^*)}\left(B[\hat{\psi}_n(\tau) - \hat{\psi}_n^S(\tau, \hat{\beta}_n)]\right)] = o_p(1/\sqrt{n}).$$

□

Appendix F Additional Results for the Applications

F.1 Verifying the Primitive Conditions in the First Application

Recall the data generating process used in Sections 4 and 5:

$$y_t = \mu_y + \rho_y(y_{t-1} - \mu_y) + \sigma_t(e_{1,t} + \vartheta_y e_{1,t-1}),\ \sigma_t^2 = \mu_\sigma + \rho_\sigma \sigma_{t-1}^2 + \kappa_\sigma e_{2,t}, \qquad (10)$$

The following verifies 1) the identification condition, that is for any $L \geq \underline{L}$, to be determined, Assumption 1 ii holds if f has sub-exponential tails, as required in Assumption 1 i, and 2) that Assumption 2 is satisfied. Geometric ergodicity can be verified by checking if Assumption 2.1 and the additional condition in Theorem 3.1 of Cline and Pu (1999) hold. Using their notation, $\alpha(\cdot)$ is linear and $\gamma(\cdot)$ is a product so the required conditions are verified.

Identification: Assume $e_{1,t} \sim f$ with $\mathbb{E}(e_{1,t}) = 0, \mathbb{E}(e_{1,t}^2) = 1$ and $e_{2,t} \sim f_2$ a non-negative, known distribution with finite moment of order p for any $p \geq 1$, and $\mathbb{E}(e_{2,t}) = \text{var}(e_{2,t}) = 1$. Assume $\rho_\sigma \in [0,1)$, $\mu_\sigma \geq 0$ and $\kappa_\sigma > 0$. For $L \geq 1$, let $\mathbf{y}_t = (y_t, \dots, y_{t-L})$ and $\psi(\tau, \theta, f) = \int \exp(i\tau'\mathbf{y}_t) f(\mathbf{y}_t, \theta, f) d\mathbf{y}_t$, note that $\partial_\tau \psi(0, \theta, f) = i\mathbb{E}(\mathbf{y}_t) = i(\mu_y, \dots, \mu_y)$ so that μ_y is identified. Similar any joint moments of $\mathbf{y}_t$ can be recovered from the CF ψ. It suffices to show that moments spanned by $\mathbf{y}_t$ can be used to identify (θ, f). The coefficient ρ_y is identified by the moment condition $\mathbb{E}([y_t - \mu_y - \rho_y(y_{t-1} - \mu_y)]y_{t-2}) = 0$. Take $\tilde{y}_t = y_t - \mu_y - \rho_y(y_{t-1} - \mu_y)$, we have: $\tilde{y}_t = \sigma_t[e_t + \vartheta_y e_{t-1}]$.

Compute two more moments: $\mathbb{E}(\tilde{y}_t^2) = \mathbb{E}(\sigma_t^2)(1 + \vartheta^2)$, and $E(\tilde{y}_t\tilde{y}_{t-1}) = \vartheta\mathbb{E}(\sigma_t\sigma_{t-1})$. Unlike the MA(1) with time-invariant volatility, these two moments alone are not sufficient to identify ϑ because $|\mathbb{E}(\sigma_t\sigma_{t-1})| \leq \mathbb{E}(\sigma_t^2)$, strictly with time-varying volatility.

Consider three additional moments: $\mathbb{E}(\tilde{y}_t^2\tilde{y}_{t-2}^2) = \mathbb{E}(\sigma_t^2\sigma_{t-2}^2)(1+\vartheta^2)^2$, $\mathbb{E}(\tilde{y}_t^2\tilde{y}_{t-4}^2) = \mathbb{E}(\sigma_t^2\sigma_{t-4}^2)(1+\vartheta^2)^2$, and $\mathbb{E}(\tilde{y}_t^2\tilde{y}_{t-2}^2\tilde{y}_{t-4}^2) = \mathbb{E}(\sigma_t^2\sigma_{t-2}^2\sigma_{t-4}^2)(1+\vartheta^2)^3$, the main idea here is to lag twice each time to only measure dependence in σ_t^2, lagging once would pick-up autocorrelations due to the MA(1) component. Let $\overline{\sigma}^2 = \mathbb{E}(\sigma_t^2)$, we have: $\mathbb{E}(\sigma_t^2) = \frac{\mu_\sigma + \kappa_\sigma}{1-\rho_\sigma}$, $\mathbb{E}([\sigma_t^2 - \bar{\sigma}^2][\sigma_{t-2}^2 - \bar{\sigma}^2]) = \rho_\sigma^2 \frac{\kappa_\sigma^2 \text{var}(u_t)}{1-\rho_\sigma^2}$, and $\mathbb{E}([\sigma_t^2 - \bar{\sigma}^2][\sigma_{t-4}^2 - \bar{\sigma}^2]) = \rho_\sigma^4 \frac{\kappa_\sigma^2 \text{var}(u_t)}{1-\rho_\sigma^2}$. Taking a ratio, we can identify $\rho_\sigma \geq 0$ by assumption: $\frac{\mathbb{E}(\tilde{y}_t^2\tilde{y}_{t-2}^2) - \mathbb{E}(\tilde{y}_t^2)^2}{\mathbb{E}(\tilde{y}_t^2\tilde{y}_{t-4}^2) - \mathbb{E}(\tilde{y}_t^2)^2} = \frac{\mathbb{E}(\sigma_t^2\sigma_{t-2}^2) - \mathbb{E}(\sigma_t^2)^2}{\mathbb{E}(\sigma_t^2\sigma_{t-4}^2) - \mathbb{E}(\sigma_t^2)^2} = \rho_\sigma^2$. We will assume $\rho_\sigma > 0$ in the following. Similarly, using moments of $\tilde{y}_t$ can compute: $\frac{\mathbb{E}(\sigma_t^2)^2}{\mathbb{E}(\sigma_t^2\sigma_{t-2}^2) - \mathbb{E}(\sigma_t^2)^2} = \frac{(\mu_\sigma + \kappa_\sigma)^2}{\kappa_\sigma^2} \frac{1-\rho_\sigma^2}{(1-\rho_\sigma)^2} \rho_\sigma^2 \text{var}(e_{2,t})$, since f_2 is known, this identifies the ratio $(\kappa_\sigma + \mu_\sigma)/\kappa_\sigma$ since the indivial terms are non-negative. Now: $\mathbb{E}(\tilde{y}_t^2) = \frac{\mu_\sigma + \kappa_\sigma}{\kappa_\sigma(1-\rho_\sigma)} \kappa_\sigma(1 + \vartheta^2)$, identifies the product $\kappa_\sigma(1 + \vartheta^2)$. The moment $\mathbb{E}(\tilde{y}_t\tilde{y}_{t-1})$ does not have a closed-form expression but can be approximated by expanding $\sqrt{\sigma_t}$ around the mean $\overline{\sigma}^2 = (\mu_\sigma + \kappa_\sigma)/(1-\rho_\sigma)$: $\mathbb{E}(\sigma_t\sigma_{t-1}) \simeq \mathbb{E}\left([\bar{\sigma} + \frac{1}{2\bar{\sigma}}(\sigma_t^2 - \bar{\sigma}^2)][\bar{\sigma} + \frac{1}{2\bar{\sigma}}(\sigma_{t-1}^2 - \bar{\sigma}^2)]\right) = \frac{1}{4\bar{\sigma}^2}\mathbb{E}([\sigma_t^2 - \bar{\sigma}^2][\sigma_{t-1}^2 - \bar{\sigma}^2])$. The coefficients κ_σ, ϑ are then separately identified using the system of equation: $\mathbb{E}(y_t y_{t-1}) = \vartheta \frac{1}{4\bar{\sigma}^2}\mathbb{E}([\sigma_t^2 - \bar{\sigma}^2][\sigma_{t-1}^2 - \bar{\sigma}^2])$, $\mathbb{E}(y_t^2) = \bar{\sigma}^2(1 + \vartheta^2)$, and $\mathbb{E}(y_t^2 y_{t-2}^2) - [\mathbb{E}(y_t^2)]^2 = \rho_\sigma(1+\vartheta^2)\mathbb{E}([\sigma_t^2 - \bar{\sigma}^2][\sigma_{t-1}^2 - \bar{\sigma}^2])$, using the same approach as for identifying the parameters of an MA(1) model with time-invariant volatility. This implies that $\underline{L} = 5$ lags are sufficient to identify $\theta = (\mu_y, \rho_y, \vartheta_y, \mu_\sigma, \rho_\sigma, \kappa_\sigma)$. If the unknown distribution f has sub-exponential tails, then its moment generating function is analytic on some interval and the distribution is determined by its moments. The idea is to solve for the moments of $e_{1,t}$ recursively from moments of y_t. We already assume that $\mathbb{E}(e_{1,t}) = 0, \mathbb{E}(e_{1,t}^2) = 1$. The third moment $\mathbb{E}(\tilde{y}_t^3) = \mathbb{E}(e_{1,t}^3)\mathbb{E}(\sigma^3)(1 + \vartheta^3)$, where the last two terms can be computed from knowledge of θ. Using the Binomial Theorem: $\mathbb{E}(\tilde{y}_t^k) = \mathbb{E}(\sigma_t^k) \sum_{j=0}^k C_{k-j}^j \mathbb{E}(e_{1,t}^{k-j})\mathbb{E}(e_{1,t}^j)\vartheta^j$. With $k = 3$, this pins down the third moments, for $k = 4$ the only unknown is the fourth moment, etc. Hence, once θ is known $(\mathbb{E}(\tilde{y}_t^3), \dots, \mathbb{E}(\tilde{y}_t^k))$ identifies $(\mathbb{E}(e_{1,t}^3), \dots, \mathbb{E}(e_{1,t}^k))$ for any

$k \geq 3$. Since f is determined by its moments, it uniquely determines the distribution itself so that (θ, f) is jointly identified. With ergodicity, this implies $\lim_{n\to\infty} \mathbb{E}(\hat{\psi}_n(\tau) - \hat{\psi}_n^s(\tau,\beta)) = 0$, $\forall \tau$ if, and only if, $\beta = \beta_0$.

Data Generating Process: Condition y(i): $\|g_{\text{obs}}(y_1, \beta_1, \sigma) - g_{\text{obs}}(y_2, \beta_1, \sigma)\| = |\rho_y|\|y_1 - y_2\| \leq \bar{\rho}_y\|y_1 - y_2\|$, which implies the strict contraction property if $|\rho_y| \leq \bar{\rho}_y < 1$. For condition y(ii), $\|g_{\text{obs}}(y_1, \mu_1, \rho_1, \vartheta_1, \sigma) - g_{\text{obs}}(y_1, \mu_2, \rho_2, \vartheta_2, \sigma)\| \leq |\mu_1 - \mu_2| + |\rho_1 - \rho_2| \times |y_1| + \sigma|\vartheta_1 - \vartheta_2| \times |e_1|$ which satisfies the desired bound if $|y_{t-1}|$, σ_t, and $|e_{t-1}|$ have bounded second moments. This is implied by restrictions on the parameters θ and the distribution f. For condition y(iii), note that the $\sqrt{\cdot}$ function is Hölder continuous with exponent 1/2 so that $\|g_{\text{obs}}(y_1, \beta, \sigma_1) - g_{\text{obs}}(y_1, \beta_1, \sigma)\| \leq |e_t + \vartheta e_{t-1}| \times \sqrt{|\sigma_1 - \sigma_2|}$, and $\mathbb{E}(|e_t + \vartheta e_{t-1}|^2) \leq 3(1 + \overline{\vartheta}^2)$ if $|\vartheta| \leq \overline{\vartheta}$ and $\mathbb{E}(e_t^2) = 1$. Hence, the assumptions on the DGP are satisfied.

F.2 Additional Results for the Second Application

Table F6 below reports estimates for $1/\tau, 1/\gamma$ instead of τ, γ in Table 4. CIs are reported for τ, γ by transforming $[1/\hat{\tau}_n \pm 1.96\text{se}(1/\hat{\tau}_n)]$.

Table F6: Estimates, Standard Errors, Confidence Intervals without the Delta-Method

	$1/\hat{\tau}_n$	$\text{se}(1/\hat{\tau}_n)$	95% CI for τ	$1/\hat{\gamma}_n$	$\text{se}(1/\hat{\gamma}_n)$	95% CI for γ
$k=1$	0.001	0.004	[128.35, $+\infty$)	0.029	0.013	[18.52, 266.65]
$k=2$	0.020	0.008	[28.81, 204.99]	0.050	0.012	[13.61, 38.78]
$k=3$	0.018	0.006	[32.95, 158.60]	0.079	0.021	[8.43, 26.19]
$k=4$	0.019	0.005	[34.17, 107.94]	0.096	0.025	[6.97, 21.11]
$k=5$	0.015	0.005	[38.77, 245.53]	0.084	0.022	[7.90, 24.69]

Appendix G Additional Results

G.1 Convergence rate in the MA(1) model

The following derives the rate of convergence for the MA(1) process: $y_t = e_t + \vartheta e_{t-1}$, $e_t \overset{iid}{\sim} f$, first when $S = +\infty$. Here $\beta = (\vartheta, f) \in [-1, 1] \times \mathcal{F}$. Take $L \geq 1$, then the joint distribution $\mathbf{y}_t = (y_t, y_{t-1})$ uniquely identifies β. Let $h(\tau, e, \vartheta) = e^{i\tau_1 e_1 + i\vartheta\tau_2 e_2 + i\tau_2 e_2 + i\vartheta\tau_2 e_3}$. The CF of $\mathbf{y}_t$ is: $\psi(\tau;\beta) = \int h(\tau, e, \vartheta) f(e_1) f(e_2) f(e_3) de_1 de_2 de_3$, for $L = 1$ where $\tau = (\tau_1, \tau_2)$. Let $\beta_k = (\vartheta_0, f_k)$ and $\hat{\beta}_n$ be an exact minimizer of Q_n, then by triangular inequalities in $\mathbb{L}^2(\pi)$:

$$\left(\int |\psi(\tau;\hat{\beta}_n) - \psi(\tau;\beta_0)|^2 \pi(\tau) d\tau\right)^{1/2} - \left(\int |\hat{\psi}_n(\tau) - \psi(\tau;\beta_0)|^2 \pi(\tau) d\tau\right)^{1/2}$$

$$\leq \sqrt{Q_n(\hat{\beta}_n)} \leq \sqrt{Q_n(\beta_k)} \leq \left(\int |\psi(\tau;\beta_k) - \psi(\tau;\beta_0)|^2 \pi(\tau) d\tau\right)^{1/2} + \left(\int |\hat{\psi}_n(\tau) - \psi(\tau;\beta_0)|^2 \pi(\tau) d\tau\right)^{1/2}.$$

The last term is $O_p(n^{-1/2})$ plus $(\int |\psi(\tau;\beta_k) - \psi(\tau;\beta_0)|^2\pi(\tau)d\tau)^{1/2} \leq (L+1)\|f_k - f_0\|_{TV}$ because the exponential has modulus 1 and the density f appears $L+1$ times in the CF. This is related to the bias accumulation discussed in the main text. From this we deduce the convergence rate under the distance implied by the CF:

$$\Big(\int |\psi(\tau;\hat\beta_n) - \psi(\tau;\beta_0)|^2\pi(\tau)d\tau\Big)^{1/2} \leq 2\Big(\int |\hat\psi_n(\tau) - \psi(\tau;\beta_0)|^2\pi(\tau)d\tau\Big)^{1/2} + (L+1)\|f_k - f_0\|_{TV},$$

which is a $O_p(\max[n^{-1/2}, \log[k]^{2r/b}k^{-r}])$, since $\|f_k - f_0\|_{TV} = O(\log[k]^{2r/b}k^{-r})$ under the smoothness and tails assumptions. Because here $S = +\infty$, we can use $k\log[k]^{-2/b} \asymp n^{-1/2r}$ which gives: $\int |\psi(\tau;\hat\beta_n) - \psi(\tau;\beta_0)|^2\pi(\tau)d\tau\Big)^{1/2} = O_p(n^{-1/2})$, in line with Corollary 1. For $r=2$, this implies $k \asymp n^{-1/4}$, up to log-terms. Asymptotically, $(\int |\psi(\tau;\hat\beta_n) - \psi(\tau;\beta_0)|^2\pi(\tau)d\tau)^{1/2} \asymp \|\hat\beta_n - \beta_0\|_{\text{weak}}$ which implies the convergence rate in weak norm. It involves the derivative $\psi_\beta(\tau,f)[v]$, i.e. $\psi_f(\tau,\beta)[v] = \int h(\tau,e,\vartheta)\{v(e_1)f(e_2)f(e_3) + f(e_1)v(e_2)f(e_3) + f(e_1)f(e_2)v(e_3)\}de_1de_2de_3$ and $\psi_\vartheta(\tau,\beta) = \int[\tau_1e_2 + \tau_2e_3]h(\tau,e,\vartheta)f(e_1)f(e_2)f(e_3)de_1de_2de_3$, for $L=1$. The local measure of ill-posedness τ_n is not closed-form, making the rate in stronger norm intractable. For $S < +\infty$, the term $\sup_{\beta\in\mathcal{B}_{k(n)}}(\int |\psi(\tau;\beta) - \hat\psi_n^S(\tau;\hat\beta_n)|^2\pi(\tau)d\tau)^{1/2} = O_p([k(n)\log[k(n)]]^2/\sqrt{nS})$ also affects the rate of convergence. Here geometric ergodicity automatically holds; an MA(1) being m-dependent regardless of the MA coefficient.

G.2 Sieve Long-Run Variance

The following derives the formula for the sieve long-run variance $\sigma_n^{\star 2}$. For brevity of notation, let $Z_t(\tau) = \hat\psi_t^S(\tau,\beta_0) - \hat\psi_t(\tau)$ and $Z_n(\tau) = \frac{1}{n}\sum_t Z_t(\tau)$. Let: $S_t^\star = \frac{1}{2}\int\{\psi_\beta(\tau,v_n^\star)\overline{Z_t(\tau)} + \overline{\psi_\beta(\tau,v_n^\star)}Z_t(\tau)\}\pi(\tau)d\tau$, the sieve score is $S_n^\star = \frac{1}{n}\sum_t S_t^\star$, and the sieve long-run variance is: $\sigma_n^{\star 2} = n\mathbb{E}(S_n^{\star 2}) = \mathbb{E}(S_t^{\star 2}) + 2\sum_{j=1}^{n-1}\frac{n-j}{n}\mathbb{E}(S_t^\star S_{t-j}^\star)$. For any $j \geq 0$, we have:

$$\mathbb{E}(S_t^\star S_{t-j}^\star) = \frac{1}{4}\int \Big\{\psi_\beta(\tau,v_n^\star)\mathbb{E}[\overline{Z_t(\tau_1)Z_{t-j}(\tau_2)}]\psi_\beta(\tau_2,v_n^\star) + \psi_\beta(\tau,v_n^\star)\mathbb{E}[\overline{Z_t(\tau_1)}Z_{t-j}(\tau_2)]\overline{\psi_\beta(\tau_2,v_n^\star)}$$
$$+ \overline{\psi_\beta(\tau,v_n^\star)}\mathbb{E}[Z_t(\tau_1)\overline{Z_{t-j}(\tau_2)}]\psi_\beta(\tau_2,v_n^\star) + \overline{\psi_\beta(\tau,v_n^\star)}\mathbb{E}[Z_t(\tau_1)Z_{t-j}(\tau_2)]\overline{\psi_\beta(\tau_2,v_n^\star)}\Big\}\pi(\tau_1)\pi(\tau_2)d\tau_1d\tau_2.$$

Let $K_j : \mathbb{L}^2(\pi) \to \mathbb{L}^2(\pi)$ be a linear operator such that: $K_jf(\tau_1) = \frac{1}{2}\int\Big\{\mathbb{E}[\overline{Z_t(\tau_1)Z_{t-j}(\tau_2)}]f(\tau_2) + \mathbb{E}[\overline{Z_t(\tau_1)}Z_{t-j}(\tau_2)]\overline{f(\tau_2)}\Big\}\pi(\tau_2)d\tau_2$, with the associated inner-product in $\mathbb{L}^2(\pi)$: $\langle f_1, f_2\rangle_\pi = \frac{1}{2}\int\{f_1(\tau)\overline{f_2(\tau)} + \overline{f_1(\tau)}f_2(\tau)\}\pi(\tau)d\tau$.[1] Compactly re-write the autocovariance: $\mathbb{E}(S_t^\star S_{t-j}^\star) = \langle\psi_\beta(\cdot,v_n^\star), K_j\psi_\beta(\cdot,v_n^\star)\rangle_\pi$. Then, by linearity: $\sigma_n^{\star 2} = \langle\psi_\beta(\cdot,v_n^\star), K_n\psi_\beta(\cdot,v_n^\star)\rangle_\pi$, where $K_n = K_0 + 2\sum_{j=1}^{n-1}\frac{n-j}{n}K_j$ is the long-run variance operator. Their sample counterparts are:

[1] Notice that $\langle v_1, v_2\rangle = 1/2\int\{\psi_\beta(\tau,v_1)\overline{\psi_\beta(\tau,v_2)}\} + \overline{\psi_\beta(\tau,v_1)}\psi_\beta(\tau,v_2)\}\pi(\tau)d\tau$ is also $\langle\psi_\beta(\cdot,v_1), \psi_\beta(\cdot,v_2)\rangle_\pi$.

$\hat{\psi}_\beta(\tau,v) = d_\beta \hat{\psi}^S_n(\tau,\hat{\beta}_n)[v]$, $\langle v_1, v_2\rangle_n = \frac{1}{2}\int\{\hat{\psi}_\beta(\tau,v_1)\overline{\hat{\psi}_\beta(\tau,v_2)} + \overline{\hat{\psi}_\beta(\tau,v_1)}\hat{\psi}_\beta(\tau,v_2)\}\pi(\tau)d\tau$, $\hat{v}^\star_n$ such that $\langle \hat{v}^\star_n, v\rangle_n = d_\beta\phi(\hat{\beta}_n)[v]$ for any v. Let $\hat{Z}_t(\tau) = \hat{\psi}^S_t(\tau,\hat{\beta}_n) - \hat{\psi}_t(\tau)$, $\hat{S}^\star_t = \frac{1}{2}\int\{\hat{\psi}_\beta(\tau,\hat{v}^\star_n)\overline{\hat{Z}_t(\tau)} + \overline{\hat{\psi}_\beta(\tau,\hat{v}^\star_n)}\hat{Z}_t(\tau)\}\pi(\tau)d\tau$, and $\hat{S}^\star_n = \frac{1}{n}\sum_t \hat{S}^\star_t$. Using an estimate $\hat{K}_n$ of K_n, we have: $\|\hat{v}^\star_{n,sd}\|^2 = \hat{\sigma}^{\star 2}_n = \langle \hat{\psi}_\beta(\cdot,\hat{v}^\star_n), \hat{K}_n\hat{\psi}_\beta(\cdot,\hat{v}^\star_n)\rangle_\pi = \langle \hat{v}^\star_n, \hat{v}^\star_n\rangle_{n,\hat{K}_n}$. Now, to estimate the long-run variance operator K_n, take $j \geq 0$ and let $\hat{K}_j$ be such that: $\hat{K}_j f(\tau_1) = \frac{1}{2}\int \Big\{ \frac{1}{n}\big[\sum_{t=j+1}^n \overline{\hat{Z}_t(\tau_1)\hat{Z}_{t-j}(\tau_2)}\big] f(\tau_2) + \frac{1}{n}\big[\sum_{t=j+1}^n \overline{\hat{Z}_t(\tau_1)}\hat{Z}_{t-j}(\tau_2)\big]\overline{f(\tau_2)}\Big\}\pi(\tau_2)d\tau_2$; then $\hat{K}_n = \hat{K}_0 + 2\sum_{j=1}^{n-1}\omega(j/T_n)\hat{K}_j$, where ω and T_n are the HAC kernel and bandwidth.

Assumption G6. *Suppose i.* $\sup_{\beta\in\mathcal{N}_{osn}}\sup_{v\in\overline{V}^1_{k(n)}} |d_\beta\phi(\beta)[v] - d_\beta\phi(\beta_0)[v]| = o(1)$*, ii. for each* $k(n)$*, any* $\beta\in\mathcal{N}_{osn}$*, and any* $v\in\overline{V}^1_{k(n)}$*,* $\hat{\psi}_\beta(\cdot,v)\in\mathbb{L}^2(\pi)$*,* $\sup_{v_1,v_2\in\overline{V}^1_{k(n)}}|\langle v_1,v_2\rangle_n - \langle v_1,v_2\rangle| = o_p(1)$*, iii.* $\sup_{v\in\overline{V}^1_{k(n)}}|\langle v,v\rangle_{n,K_n} - \langle v,v\rangle_{K_n}| = o_p(1)$*, iv.* $\|\hat{K}_n - K_n\|_{op} = o_p(1)$.

Where $\|\cdot\|_{op}$ is the operator norm in $(\mathbb{L}^2(\pi), \langle\cdot,\cdot\rangle_\pi)$. Assumption G6 i-iii is based on Assumption 4.1 in Chen and Pouzo (2015a). Given Assumption 1 iii, Proposition 3.3 in Carrasco et al. (2007) imply Assumption G6 iv holds under Assumption G7 below.

Assumption G7. *Suppose i.* $\omega:\mathbb{R}\to[0,1]$, $\omega(0)=1$, $\omega(-x)=\omega(x)$, $\forall x\in\mathbb{R}$, $\omega\in\mathbb{L}^2(\mathbb{R})$, *$\omega$ is continuous at 0 and all, but finitely many, values of x; ii.* $T_n^{2\nu+1}/n \to \gamma\in(0,\infty)$ *for some ν for which* $\|\omega^\nu\|<\infty$ *and* $\|f^\nu_Y\|<\infty$*,* ω^ν *and* f^ν_Z *are the ν-th derivative of ω and f_Y, the spectral density of* $(\mathbf{y}_t, \mathbf{y}^s_t)$ *at 0.*

Proposition G1. *Suppose Assumption G6 holds, then* $\left|\hat{\sigma}^\star_n/\sigma^\star_n - 1\right| = o_p(1)$.

Proposition G1 follows from Theorem 4.2 in Chen and Pouzo (2015a), where now Step 2A in their proof (Chen and Pouzo, 2015b, p9) requires $\|\hat{K}_n - K_n\|_{op} = o_p(1)$ as in Assumption G6 iv. The formula used in the main text is easier to implement, but equivalent. For each $j\geq 0$: $\int \text{real}\{\psi_\beta(\tau_1,v^\star_n)\mathbb{E}[\overline{Z_t(\tau_1)}\text{real}[Z_{t-j}(\tau_2)\overline{\psi_\beta(\tau_2,v^\star_n)}]]\}\pi(\tau_1)\pi(\tau_2)d\tau_1 d\tau_2 = \langle\psi_\beta(\cdot,v^\star_n), K_j\psi_\beta(\cdot,v^\star_n)\rangle_\pi$. Because $\mathbb{E}$, $\int$ and real are linear operators, they arrange into:

$$\begin{aligned}&\langle\psi_\beta(\cdot,v^\star_n), K_j\psi_\beta(\cdot,v^\star_n)\rangle_\pi\\ &= \mathbb{E}\Big\{\left(\int \text{real}[\psi_\beta(\tau_1,v^\star_n)\overline{Z_t(\tau_1)}]\pi(\tau_1)d\tau_1\right)\left(\int \text{real}[\psi_\beta(\tau_2,v^\star_n)\overline{Z_{t-j}(\tau_2)}]\pi(\tau_2)d\tau_2\right)\Big\}.\end{aligned}$$

Then replace $\text{real}[\psi_\beta(\tau_1,v^\star_n)\overline{Z_t(\tau_1)}] = \text{real}[\psi_\beta(\tau_1,v^\star_n)]\text{real}[Z_t(\tau_1)] + \text{im}[\psi_\beta(\tau_1,v^\star_n)]\text{im}[Z_t(\tau_1)]$. Next, let $\varphi = (\theta,\omega,\mu,\sigma)$ denote the parameter β in the sieve basis. For any v, $v' d_\varphi\phi(\beta_0) = \langle v, v^\star_n\rangle = v'\text{real}[\int \psi_{\varphi'}(\tau,\beta_0)\overline{\psi_{\varphi'}(\tau,\beta_0)}\pi(\tau)d\tau]v^\star_n$ so $v^\star_n = \text{real}[\int\psi_{\varphi'}(\tau,\beta_0)\overline{\psi_{\varphi'}(\tau,\beta_0)}\pi(\tau)d\tau]^{-1}d_\varphi\phi(\beta_0)$. Now, substitude $v^\star_n$ into $\langle\psi_\beta(\cdot,v^\star_n), K_n\psi_\beta(\cdot,v^\star_n)\rangle_\pi$ to get the sandwich formula. The same derivations applied to the sample quantities yield the formula in the main text.