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

Jean-Jacques Forneron*

February 4, 2019

This Supplemental Material consists of Appendices F and G to the main text.

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

Appendix F Additional Asymptotic Results

This Appendix provides general results for Sieve-SMM estimates for other sieve bases and bounded moment functions. It adapts existing results from the sieve literature to a continuum of bounded complex-valued moments and extends them to a more general class of dynamic models. The following definition gives the two measures of dependence used in the results.

Definition F4 (α-Mixing and Uniform α-Mixing)**.** *For the sample observations $(y_t)_{t\geq 1}$, the α-mixing coefficients are defined as:*

$$\alpha(m) = 2 \sup_{t\geq 1} \sup_{y_1,y_2 \in \mathbb{R}^{d_y}} \Big| \mathbb{P}\left(y_t \geq y_1, y_{t+m} \geq y_2\right) - \mathbb{P}\left(y_t \geq y_1\right) \mathbb{P}\left(y_{t+m} \geq y_2\right) \Big|.$$

$(y_t)_{t\geq 1}$ is α-mixing if $\alpha(m) \to 0$ when $m \to \infty$.

For the simulated samples $(y(\beta)_t^s)_{t\geq 1}$ indexed by $\beta \in \mathcal{B}$ the uniform α-mixing coefficients are defined as:

$$\alpha^*(m) = 2 \sup_{t\geq 1, \beta \in \mathcal{B}} \sup_{y_1,y_2 \in \mathbb{R}^{d_y}} \Big| \mathbb{P}\left(y_t^s(\beta) \geq y_1, y_{t+m}^s(\beta) \geq y_2\right) - \mathbb{P}\left(y_t^s(\beta) \geq y_1\right) \mathbb{P}\left(y(\beta)_{t+m}^s \geq y_2\right) \Big|.$$

$(y_t^s(\beta))_{t\geq 1}$ is uniformly α-mixing if $\alpha^*(m) \to 0$ when $m \to \infty$.

F.1 Consistency

Recall that the Sieve-SMM estimator $\hat{\beta}_n$ satisfies:

$$\hat{Q}_n^S(\hat{\beta}_n) \leq \inf_{\beta \in \mathcal{B}_{k(n)}} \hat{Q}_n^S(\beta) + O_p(\eta_n)$$

where $\eta_n = o(1)$. The sample objective function is:

$$\hat{Q}_n^S(\beta) = \int \Big| B\hat{\psi}_n(\tau) - B\hat{\psi}_n^S(\tau,\beta) \Big|^2 \pi(\tau) d\tau$$

As in the main results, there is a sequence of population objective functions:

$$Q_n(\beta) = \int \Big| \mathbb{E}\left(B\hat{\psi}_n(\tau) - B\hat{\psi}_n^S(\tau,\beta) \right) \Big|^2 \pi(\tau) d\tau.$$

Q_n may depend on n when y_t^s is non-stationary. The following three assumptions are adapted from the high-level conditions in Chen (2007, 2011) and Chen & Pouzo (2012) to a continuum of of moments (Carrasco & Florens, 2000; Carrasco et al., 2007).

Assumption F8 (Sieves)**.** *$\{\mathcal{B}_k,\, k \geq 1\}$ is a sequence of non-empty compact subsets of $\mathcal{B}$ such that $\mathcal{B}_k \subseteq \mathcal{B}_{k+1} \subseteq \mathcal{B}$, $\forall k \geq 1$. There exists an approximating sequence $\Pi_k \beta_0 \in \mathcal{B}_k$ such that $\|\Pi_{k(n)}\beta_0 - \beta_0\|_{\mathcal{B}} = o(1)$ as $k(n) \to \infty$.*

Assumption F9 (Identification)**.** *i) $\lim_{n\to\infty} \mathbb{E}\left(\hat{\psi}_n^S(\tau,\beta) - \hat{\psi}_n(\tau)\right) = 0 \quad \pi$ a.s. $\Leftrightarrow \|\beta - \beta_0\|_{\mathcal{B}} = 0$. The null space of B is the singleton $\{0\}$. ii) $Q_n(\Pi_{k(n)}\beta_0) = o(1)$ as $n \to \infty$. iii) There exists a function g such that for all $\varepsilon > 0$: $g(k(n), n, \varepsilon) = \inf_{\beta \in \mathcal{B}_{k(n)},\, \|\beta-\beta_0\|_{\mathcal{B}} \geq \varepsilon} Q_n(\beta)$, g is decreasing in the first and last argument and $g(k(n), n, \varepsilon) > 0$ for all $k(n), n, \varepsilon > 0$.*

Assumption F10 (Convergence Rate over Sieves)**.** *There exists two constants $C_1, C_2 > 0$ such that, uniformly over $h \in \mathcal{B}_{k(n)}$: $\hat{Q}_n^S(\beta) \leq C_1 Q_n(\beta) + O_p(\delta_n^2)$, $Q_n(\beta) \leq C_2 \hat{Q}_n(\beta) + O_p(\delta_n^2)$ and $\delta_n^2 = o(1)$.*

Theorem F4 (Consistency)**.** *Suppose Assumptions F8-F10 hold. Furthermore, suppose that $h \to Q_n(\beta)$ is continuous on $\left(\mathcal{B}_{k(n)}, \|\cdot\|_{\mathcal{B}}\right)$. If $k(n) \overset{n\to\infty}{\longrightarrow} \infty$ and for all $\varepsilon > 0$ the following holds:*

$$\max\left(\eta_n, Q_n(\Pi_{k(n)}\beta_0), \delta_n^2\right) = o\left(g\left(k(n), \varepsilon\right)\right).$$

Then the estimator $\hat{\beta}_n$ is consistent: $\|\hat{\beta}_n - \beta_0\|_{\mathcal{B}} = o_p(1)$.

Theorem F4 is a direct consequence of the general consistency Lemma in Chen & Pouzo (2012) reproduced as Lemma G12 in the next appendix. Assumption F8 is standard and satisfied by the mixture sieve, the Hermite polynomial basis of Gallant & Nychka (1987) or the cosine basis as in Bierens & Song (2012). See e.g. Chen (2007) for further examples of sieve bases and their approximation properties. The choice of moments $\hat{\psi}_n$ and the restrictions on the parameter space $\mathcal{B}$ are assumed to ensure identification in Assumption F9. Verifying Assumption F10 is more challenging in this setting because of the dynamics and the continuum of moments. Furthermore, the rate for $Q_n(\Pi_{k(n)}\beta_0)$ needs to be derived. The following proposition derives the rate for iid data under an additional restriction.[1]

Proposition F1. *If y_t^s is iid and depends on f only through e_t^s, i.e. $y_t^s = g_{obs}(x_t, \theta, e_t^s)$ with $e_t^s \sim f$, then for Q_n based on the CF:*

$$Q_n(\Pi_{k(n)}\beta_0) \leq 2M_B^2 \|\Pi_{k(n)} f_0 - f_0\|_{TV}^2$$

where TV is the total variation norm: $\|\Pi_{k(n)} f_0 - f_0\|_{TV} = \int |\Pi_{k(n)} f_0(\varepsilon) - f_0(\varepsilon)| d\varepsilon$.

[1] A more general rate for $Q_n(\Pi_{k(n)}\beta_0)$ will be given in Proposition F3.

Remark F17. *Proposition F1 can be restated in terms of Hellinger distance by the inequality* $\|\Pi_{k(n)}f_0 - f_0\|_{TV} \leq 2d_H(\Pi_{k(n)}f_0, f_0)$. *Pinsker's inequality gives a similar relationship for the Kullback-Leibler divergence:* $\|\Pi_{k(n)}f_0 - f_0\|_{TV} \leq \sqrt{2KL(\Pi_{k(n)}f_0|f_0)}$.

Assumption F11 (Smoothness, Dependence, Complexity)**.** *Suppose that:*

i. (Smoothness) For $P \geq 2$, $\beta \to \psi_t^s(\tau, \beta)$ *is* L^p*-smooth. That is, there exists* $C > 0, \eta > 0$ *and* $\gamma \in (0, 1]$ *such that for all* $\tau \in \mathbb{R}^d$ *and all* $\delta > 0$*:*

$$\sup_{t\geq 1}\left[\mathbb{E}\left(\sup_{\beta_1,\beta_2\in\mathcal{B},\,\|\beta_1-\beta_2\|_{\mathcal{B}}\leq\delta}\left|[\psi_t^s(\tau,\beta_1) - \psi_t^s(\tau,\beta_2)]\pi(\tau)^{1/(2+\eta)}\right|^P\right)\right]^{1/P} \leq C\delta^\gamma$$

and $\int \pi(\tau)^{1-2/(2+\eta)}d\tau < \infty$.

ii. (Dependence) (y_t^s, x_t) *and* (y_t, x_t) *are either iid or uniformly* $\alpha-$*mixing with* $\alpha^*(m) \leq C\exp(-am)$ *for all* $m \geq 1$ *with* $C, a > 0$.

iii. (Complexity) The moment function is uniformly bounded: $|\hat{\psi}_t^s(\tau, \beta)| \leq M$ *for all* τ, β *and some* $M > 0$. *One of the following holds:*

a. if (y_t, x_t) *is iid, the integral*

$$\sqrt{C_n} \coloneqq \int_0^1 \sqrt{1 + \log N\left(x^{1/\gamma}, \mathcal{B}_{k(n)}, \|\cdot\|_{\mathcal{B}}\right)}dx$$

is such that $C_n/n \to 0$.

b. if (y_t^s, x_t) *is dependent, the integral*

$$\sqrt{C_n} \coloneqq \int_0^1 \left(x^{-\vartheta/2}\sqrt{\log N\left(x^{1/\gamma}, \mathcal{B}_{k(n)}, \|\cdot\|_{\mathcal{B}}\right)} + \log^2 N\left(x^{1/\gamma}, \mathcal{B}_{k(n)}, \|\cdot\|_{\mathcal{B}}\right)\right)dx$$

is such that $C_n/n \to 0$.

Where the covering number $N(x, \mathcal{B}_{k(n)}, \|\cdot\|_{\mathcal{B}})$ *is the minimal number of balls of radius* x *in* $\|\cdot\|_{\mathcal{B}}$ *norm needed to cover the space* $\mathcal{B}_{k(n)}$.

Assumption F11 provides conditions on the moments $\hat{\psi}_n^s$, the weights π, the dependence and the sieve space to ensure Assumption F10 holds. Condition *i.* assumes that the moments are L^p-smooth. Note that the condition involves π, the moments themselves need not be uniformly L^p-smooth. An additional requirement is given for π to handle the continuum of moments. Giving the condition on the moments rather than the DGP itself as in the main results in more common (Duffie & Singleton, 1993, see e.g.) in the literature. The two are actually related, as shown in the following remark.

Remark F18 (L^p-Smoothness of the Moments and the DGP)**.** *For the empirical CF, smoothness of the moment function directly relates to smoothness of the data generating process: i.e. L^p-smoothness of $\beta \to y_t^s(\beta)$ implies Assumption F11 i. It is a direct implication of the sine and cosine functions being uniformly Lipschitz on the real line:*

$$\begin{aligned}\left|\psi_t^s(\tau,\beta_1) - \psi_t^s(\tau,\beta_2)\right|\pi(\tau)^{1/(2+\eta)} &\leq 2\|\tau'(\mathbf{y}_t^s(\beta_1),\mathbf{x}_t) - \tau'(\mathbf{y}_t^s(\beta_2),\mathbf{x}_t)\|\pi(\tau)^{1/(2+\eta)} \\ &\leq 2\sup_{\tau\in\mathbb{R}^{d_\tau}}\left(\|\tau\|_\infty\pi(\tau)^{1/(2+\eta)}\right)\times\|\mathbf{y}_t^s(\beta_1) - \mathbf{y}_t^s(\beta_2)\|.\end{aligned}$$

This is the basis for the main results presented in Section 3.

Examples of DGPs and moments satisfying condition *i.* are given in Appendix F.4.

Assumption F11, condition *ii.* is satisfied under the geometric ergodicity condition of Duffie & Singleton (1993) as shown in Liebscher (2005)'s Propositions 2 and 4. Note that Liebscher's result holds whether (y_t, x_t) is stationary or not.

Assumption F11, condition *iii.* hold for linear sieves with $k(n)/n \to 0$ in the iid case and $k(n)^4/n \to 0$ in the dependent case. For non-linear sieves such as mixtures and neural networks the condition becomes $k(n)\log[k(n)]/n \to 0$ in the iid case and $(k(n)\log[k(n)])^4/n \to 0$ in the dependent case. The following Proposition F2 relates the low-level conditions in Assumption F11 to Assumption F10.

Proposition F2. *Suppose that Assumption F11 holds, then Assumption F10 holds with $\delta_n^2 = C_n/n$.*

Given this proposition, Corollary F3 is a direct consequence of Theorem F4.

Corollary F3. *Suppose Assumptions F8-F9 and F11 hold. Furthermore, suppose that $\beta \to Q_n(\beta)$ is continuous on $\left(\mathcal{B}_{k(n)}, \|\cdot\|_{\mathcal{B}}\right)$. If $k(n) \overset{n\to\infty}{\longrightarrow} \infty$ and for all $\varepsilon > 0$ the following holds:*

$$\max\left(\eta_n, Q_n(\Pi_{k(n)}\beta_0), \delta_n^2\right) = o\left(g\left(k(n),\varepsilon\right)\right)$$

then the estimator $\hat{\beta}_n$ is consistent:

$$\|\hat{\beta}_n - \beta_0\|_{\mathcal{B}} = o_p(1).$$

Proposition F3. *Suppose that the L^p-smoothness in Assumption F11 i. is satisfied, then there exists $K > 0$ which only depends on C and η, defined in Assumption F11 i., M_B and π such that:*

$$Q_n(\Pi_{k(n)}\beta_0) \leq K\left(\|\Pi_{k(n)}\beta_0 - \beta_0\|_{\mathcal{B}}^{2\gamma} + Q_n(\beta_0)\right).$$

The rate in Proposition F3 is different from the main results because the L^p-smoothness assumption is given on the moments rather than the DGP itself. Also, in Assumption F3 the L^p-smoothness constant does not increase with $k(n)$ so that the decay condition is not required to derive the rate.

For iid and stationary $(y_t^s)_{t\geq 1}$, $Q_n(\beta_0) = 0$ should generally hold so the rate at which $Q_n(\Pi_{k(n)}\beta_0)$ goes to zero only depends on the smoothness γ and the approximation rate of β_0. When the L^p-smoothness coefficient is $\gamma = 1$, the rate is similar to Proposition F1 while for $\gamma \in (0,1)$ the rate is slower. In the non-stationary case $Q_n(\beta_0)$ will depend on the rate at which $f_{y_t^s,x_t}$ convergences to the stationary distribution.

F.2 Rate of Convergence

This section establishes the rate of convergence of the estimator in the weak norm of Ai & Chen (2003) and the strong norm $\|\cdot\|_{\mathcal{B}}$. As in Chen & Pouzo (2012), assuming consistency holds, the parameter space can be restricted to a local neighborhood $\mathcal{B}_{os} = \{\beta \in \mathcal{B}, \|\beta - \beta_0\|_{\mathcal{B}} \leq \varepsilon\}$ with $\varepsilon > 0$ small such that $\mathbb{P}\left(\hat{\beta}_n \notin \mathcal{B}_\varepsilon\right) < \varepsilon$. Similarly let $\mathcal{B}_{osn} = \mathcal{B}_{os} \cap \mathcal{B}_{k(n)}$.

Assumption F12 (Differentiability)**.** *Suppose that for all $\beta_1, \beta_2 \in \mathcal{B}_{os}$, the pathwise derivative:*

$$\lim_{\varepsilon\in(0,1),\varepsilon\to 0} \int \left| B\frac{\mathbb{E}\left(\hat{\psi}_n^S(\tau,(1-\varepsilon)\beta_1+\varepsilon\beta_2) - \hat{\psi}_n^S(\tau,\beta_1)\right)}{\varepsilon}\right|^2 \pi(\tau)d\tau$$
$$= \int \left| B\frac{d\mathbb{E}\left(\hat{\psi}_n^S(\tau,\beta_1)\right)}{d\beta}[\beta_2]\right|^2 \pi(\tau)d\tau$$

exists and is finite.

Following Ai & Chen (2003), the weak norm measure uses the pathwise derivative of the moments at β_0:

$$\|\beta_1 - \beta_2\|_{weak} = \left(\int \left| B\frac{d\mathbb{E}[\psi_n^s(\tau,\beta)}{d\beta}\Big|_{\beta=\beta_0}[\beta_1-\beta_2]\right|^2 \pi(\tau)d\tau\right)^{1/2}.$$

Suppose that there exists a $C > 0$ such that for all $\beta \in \mathcal{B}_{os}$ and all $n \geq 1$:

$$\|\beta - \beta_0\|_{weak}^2 \leq C Q_n(\beta).$$

Assumption F12 implies that $\|\cdot\|_{weak}$ is Lipschitz continuous with respect to the population criterion Q_n as in Chen & Pouzo (2012)'s Assumption 4.1. Under Assumption F12, the rate of convergence is easier to derive in $\|\cdot\|_{weak}$ than in the stronger norm $\|\cdot\|_{\mathcal{B}}$. However, a sufficiently fast rate of convergence in the stronger norm will be required for the stochastic equicontinuity results, since the strong norm $\|\cdot\|_{\mathcal{B}}$ appears in L^p-smoothness Assumption F11. The two convergence rates are related by the local measure of ill-posedness of Blundell et al. (2007).

Definition F5 (Local Measure of Ill-Posedness of Blundell et al. (2007))**.** *The local measure of ill-posedness τ_n is:*

$$\tau_n = \sup_{\beta \in \mathcal{B}_{osn}: \|\beta - \Pi_{k(n)}\beta_0\| \neq 0} \frac{\|\beta - \Pi_{k(n)}\beta_0\|_{\mathcal{B}}}{\|\beta - \Pi_{k(n)}\beta_0\|}.$$

The following theorem adapts the results of Chen & Pouzo (2012) to the continuum of moments with simulated data.

Theorem F5 (Rate of Convergence)**.** *Suppose that Assumptions F8, F9, F11 and F12 are satisfied and suppose that $\eta_n = o(\delta_n^2)$. Let $\beta_0, \Pi_{k(n)}\beta_0 \in \mathcal{B}_{os}$, then we have the rate of convergence in weak and strong norm:*

$$\|\hat{\beta}_n - \beta_0\|_{weak} = O_p\left(\max\left(\delta_n, \|\Pi_{k(n)}\beta_0 - \beta_0\|_{\mathcal{B}}^{\gamma}, \sqrt{Q_n(\beta_0)}\right)\right) \text{ and}$$

$$\|\hat{\beta}_n - \beta_0\|_{\mathcal{B}} = O_p\left(\|\Pi_{k(n)}\beta_0 - \beta_0\|_{\mathcal{B}} + \tau_n \max\left(\delta_n, \|\Pi_{k(n)}\beta_0 - \beta_0\|_{\mathcal{B}}^{\gamma}, \sqrt{Q_n(\beta_0)}\right)\right).$$

The rate δ_n is derived in Proposition F2: for linear sieves with iid data $\delta_n = \sqrt{k(n)/n}$ and $\delta_n = k(n)^2/\sqrt{n}$ in the dependent case. The rate $\|\Pi_{k(n)}\beta_0 - \beta_0\|_{\mathcal{B}}^{\gamma}$ depends on the approximation rate $\|\beta_0 - \Pi_{k(n)}\beta_0\|_{\mathcal{B}}$ and the L^p-smoothness of the objective function. In the iid and stationary case, $Q_n(\beta_0) = 0$ is not a concern for the rate of convergence.

Proposition F4. *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 $Q_n(\beta_0) = O(1/n^2)$.*

Proposition F4 shows that $Q_n(\beta_0)$ is negligible under the geometric ergodicity condition of Duffie & Singleton (1993): since δ_n is typically larger than a $O(1/\sqrt{n})$ term, $Q_n(\beta_0) = o(\delta_n^2)$.

Corollary F4. *Suppose that the assumptions of Theorem F5 and the (y_t^s, x_t) are iid, stationary or geometrically ergodic then the rate of convergence is:*

$$\|\hat{\beta}_n - \beta_0\|_{\mathcal{B}} = O_p\left(\|\Pi_{k(n)}\beta_0 - \beta_0\|_{\mathcal{B}} + \tau_n \max\left(\delta_n, \|\Pi_{k(n)}\beta_0 - \beta_0\|_{\mathcal{B}}^{\gamma}\right)\right).$$

The rate of convergence can be further improved for static models with iid data under the assumptions of Proposition F1, as shown in the Corollary below.

Corollary F5. *Suppose that the assumptions of Theorem F5 and Proposition F1 are satisfied then:*

$$\|\hat{\beta}_n - \beta_0\|_{\mathcal{B}} = O_p\left(\|\Pi_{k(n)}\beta_0 - \beta_0\|_{\mathcal{B}} + \tau_n \max\left(\delta_n, \|\Pi_{k(n)}f_0 - f_0\|_{TV}\right)\right).$$

F.3 Asymptotic Normality

As in Chen & Pouzo (2015), this section gives asymptotic normality results for functionals ϕ of the estimates $\hat{\beta}_n$. In order to conduct inferences, standard errors σ_n^* are derived such that:

$$\frac{\sqrt{n}}{\sigma_n^*}\left(\phi(\hat{\beta}_n)-\phi(\beta_0)\right)\overset{d}{\to}\mathcal{N}(0,1). \tag{F.1}$$

As in the main results, to reduce notation the following will be used:

$$\psi_\beta(\tau,v)=\frac{d\mathbb{E}(\hat{\psi}_n^S(\tau,\beta_0))}{d\beta}[v],\quad Z_t^S(\tau)=\hat{\psi}_t(\tau)-\frac{1}{S}\sum_{s=1}^S\hat{\psi}_t^s(\tau,\beta_0),\quad Z_n^S(\tau)=\hat{\psi}_n(\tau)-\hat{\psi}_n^S(\tau,\beta_0)$$

where v is a vector in $\bar{V}$ or $\bar{V}_n$ defined as in the main results. The sieve representer v_n^* is also defined as in the main results.

Definition F6 (Sieve Score, Sieve Variance, Scaled Sieve Representer)**.** *The Sieve Score S_n^* is defined as:*

$$S_n^*=\frac{1}{2}\int\left[B\psi_\beta(\tau,v_n^*)\overline{BZ_n^S(\tau)}+\overline{B\psi_\beta(\tau,v_n^*)}BZ_n^S(\tau)\right]\pi(\tau)d\tau.$$

The sieve variance is $\sigma_n^{*2}=n\mathbb{E}\left(|S_n^*|^2\right)$. The scaled sieve representer is $u_n^*=\frac{v_n^*}{\sigma_n^*}$.

As in the main results, the equivalence condition below is required.

Assumption F13 (Equivalence Condition)**.** *There exists $\underline{a}>0$ such that $\forall n\geq 1$:*

$$\underline{a}\|v_n^*\|_{weak}\leq\sigma_n^*.$$

Furthermore assume that $\sigma_n^*=o(\sqrt{n})$.

An discussion of this condition is given in Appendix F.5. The last part imposes that $k(n)$ does not increase too fast with n to control the variance of the sieve score.

Remark F19 (On the equivalence condition)**.** *Since $\hat{\psi}_t$ is bounded, the data is α-mixing and the simulations are geometrically ergodic there also exists a $\overline{a}>0$ such that $\sigma_n^*\leq\overline{a}\|v_n^*\|_{weak}$. Hence under Assumption F13 the following holds $\sigma_n^*\asymp\|v_n^*\|_{weak}$. To prove this statement, note that the Cauchy-Schwarz inequality implies:*

$$\begin{aligned}\sigma_n^*&\leq\sqrt{2n}\left[\mathbb{E}\left(\left[\int\left|B\psi_\beta(\tau,v_n^*)\right|\times\left|B\hat{\psi}_n^S(\tau,\beta_0)-B\hat{\psi}_n(\beta_0)\right|\pi(\tau)d\tau\right]^2\right)\right]^{1/2}\\&\leq\sqrt{2}\left[\int\left|B\psi_\beta(\tau,v_n^*)\right|^2\pi(\tau)d\tau\right]^{1/2}\left[\mathbb{E}\left(\left[\int n\left|B\hat{\psi}_n^S(\tau,\beta_0)-B\hat{\psi}_n(\beta_0)\right|^2\pi(\tau)d\tau\right]\right)\right]^{1/2}\end{aligned}$$

The first term in the product is $\|v_n^*\|_{weak}$. The second term is bounded by noting that for all $\tau \in \mathbb{R}^{d_\tau}$:

$$n\mathbb{E}\left|B\hat{\psi}_n^S(\tau,\beta_0) - \mathbb{E}\left(B\hat{\psi}_n^S(\tau,\beta_0)\right)\right|^2 \leq 1 + 24\sum_{m\geq 1}\alpha(m)^{1/p} < \infty.$$

for any $p > 1$ by Lemma G13, picking $p = 1/2$ implies:

$$\mathbb{E}\left(\int n\left|B\hat{\psi}_n^S(\tau,\beta_0) - B\hat{\psi}_n(\beta_0)\right|^2\pi(\tau)d\tau\right) \leq \left(1 + 24\sum_{m\geq 1}\sqrt{\alpha(m)}\right)$$

which yields $\overline{a} = \sqrt{4 + 96\sum_{m\geq 1}\sqrt{\alpha(m)}}$.

Assumption F14 (Undersmoothing, Convergence Rate)**.** *Let $\delta_{sn} = \|\hat{\beta}_n - \beta_0\|_{\mathcal{B}}$ the convergence rate in strong norm.*

i. Undersmoothing: $\|\hat{\beta}_n - \beta_0\|_{weak} = O_p(\delta_n)$ and $\delta_{sn} = \|\Pi_{k(n)}\beta - \beta_0\|_{\mathcal{B}} + \tau_n\delta_n$.

ii. Sufficient Rate: $\delta_n = o(n^{-1/4})$.

iii. The convergence rate in weak norm δ_n and in strong norm δ_{sn} are such that:

$$(M_n\delta_{sn})^\gamma\sqrt{C_{sn}} = o(1) \tag{F.2}$$

$$\sqrt{n}M_n^{1+\gamma}\delta_n^\gamma\sqrt{C_{sn}}\max\left(M_n\delta_n, \frac{1}{\sqrt{n}}\right) = o(1) \tag{F.3}$$

where

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

and $M_n = \log\log(n+1)$ for all all $n \geq 1$.

Assumptions F14 *i.*, *ii.* are common in the (semi)-nonparametric literature. Assumption F14 *iii.* ensures that a stochastic equicontinuity holds. It is needed several time throughout the proofs (see Lemma G17), in most cases the less demanding condition (F.2) is sufficient. Condition (F.3) is similar to Chen & Pouzo (2015)'s Assumption A.5 *(iii)*, it ensures that when $\hat{\psi}_n^S(\tau,\beta) - \hat{\psi}_n^S(\tau,\beta_0)$ is substituted under the integral with its smoothed version, the difference is negligible for $\sqrt{n}$-asymptotics.

Assumption F15 (Local Linear Expansion of ϕ)**.** *ϕ is continuously differentiable and $\frac{d\phi(\beta_0)}{d\beta}[\cdot]$ is a non-zero linear functional such that as $n \to \infty$:*

i. A linear expansion is locally uniformly valid

$$\sup_{\|\beta-\beta_0\|_{weak}\leq M_n\delta_n} \frac{\sqrt{n}}{\sigma_n^*}\Big|\phi(\beta)-\phi(\beta_0)-\frac{d\phi(\beta_0)}{d\beta}[\beta-\beta_0]\Big| \to 0.$$

ii. The approximation bias is negligible

$$\frac{\sqrt{n}}{\sigma_n^*}\frac{d\phi(\beta_0)}{d\beta}[\beta_{0,n}-\beta_0] \to 0.$$

Remark F20 (Sufficient Conditions for Assumption F15 *i.*)**.** *If ϕ is twice continuously differentiable then for some $v \in \bar{V}$ and $h \in [-1,1]$, $\beta = \beta_0 + hM_n\delta_n v$. Using a Mean Value Expansion:*

$$\begin{aligned}\Big|\phi(\beta_0+hM_n\delta_n v)-\phi(\beta_0)-\frac{d\phi(\beta_0)}{d\beta}[hM_n\delta_n v]\Big| &= \Big|\frac{1}{2}\frac{d^2\phi(\beta_0+\tilde{h}M_n\delta_n v)}{d\beta d\beta}[hM_n\delta_n v, hM_n\delta_n v]\Big| \\ &= h^2(M_n\delta_n)^2\Big|\frac{1}{2}\frac{d^2\phi(\beta_0+\tilde{h}M_n\delta_n v)}{d\beta d\beta}[v,v]\Big|.\end{aligned}$$

Hence Assumption F15 i. holds under the following two conditions:

i. The second derivative is locally uniformly bounded:

$$\sup_{\|v\|_{weak}=1, h\in(-1,1)}\Big|\frac{1}{2}\frac{d^2\phi(\beta_0+hM_n\delta_n v)}{d\beta d\beta}[v,v]\Big| = O(1).$$

ii. The rate of convergence satisfies:

$$\frac{\sqrt{n}}{\sigma_n^*}(M_n\delta_n)^2 = o(1).$$

This condition holds if $\delta_n = o(M_n^{-1}n^{-1/4})$ which is slightly stronger than Assumption F14 ii.

Remark F21 (Sufficient Conditions for Assumption F15 *ii.*)**.** *By definition of $\beta_{0,n}$, Assumptions F11, F12 and under geometric ergodicity:*

$$\|\beta_{0,n}-\beta_0\|_{weak} \leq \|\Pi_{k(n)}\beta_0-\beta_0\|_{weak} \leq C\sqrt{Q_n(\Pi_{k(n)}\beta_0)} \leq \tilde{C}\|\Pi_{k(n)}\beta_0-\beta_0\|_{\mathcal{B}}^{\gamma}.$$

The approximation rate is typically $\|\Pi_{k(n)}\beta_0-\beta_0\|_{\mathcal{B}}^{\gamma} = O(k(n)^{-r})$ where r is the smoothness of the density f_0 to be estimated. Rewriting $\beta_{0,n} = \beta_0 + h_n k(n)^{-r}v_n$ with $\|v_n\|_{weak} = 1$, $|h_n| \leq \bar{h}$ bounded, the expression can be bounded using:

$$\frac{\sqrt{n}}{\sigma_n^*}\Big|\frac{d\phi(\beta_0)}{d\beta}[\beta_{0,n}-\beta_0]\Big| \leq \bar{h}\frac{\sqrt{n}}{\sigma_n^*}k(n)^{-r}\Big|\frac{d\phi(\beta_0)}{d\beta}[v_n]\Big|.$$

Hence Assumption F15 ii. is satisfied under the following two conditions:

i. The first derivative is uniformly bounded on the unit circle:

$$\sup_{\|v\|_{weak}=1}\left|\frac{1}{2}\frac{d\phi(\beta_0)}{d\beta}[v]\right| < +\infty.$$

ii. The approximation rate satisfies:

$$\frac{\sqrt{n}}{\sigma_n^*}k(n)^{-\gamma r} = o(1).$$

With the undersmoothing assumption the $k(n)$ *must satisfy* $k(n)^{-\gamma r} = o(\delta_n) = o(n^{-1/4})$. *A sufficient condition on the bias/variance relation is* $k(n)^{-\gamma r} = o(\delta_n^2\sigma_n^*)$.

The last condition is strong and can be weakened if for instance $\delta_n^2 \ll 1/\sqrt{n}$, *replacing* δ_n^2 *with* $1/\sqrt{n}$. *Sharper bounds on the bias can also be found in the iid case (see Corollary F5) or under assumptions on the DGP itself as in the main results (see Lemma 4).*

Assumption F16 (Local Behaviour of $\mathbb{E}(\hat{\psi}(\tau,\beta))$)**.** *The mapping* $\beta \to \mathbb{E}(\hat{\psi}_n^S(\tau,\beta)$ *is twice continuously differentiable for all* τ *and satisfies:*

i. A linear expansion is locally uniformly valid

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

ii. The second derivative in direction u_n^* *is locally uniformly bounded*

$$\sup_{\|\beta-\beta_0\|_{weak}\leq M_n\delta_n}\int\left|\frac{d^2\mathbb{E}(\hat{\psi}_n^S(\tau,\beta))}{d\beta d\beta}[u_n^*,u_n^*]\right|^2\pi(\tau)d\tau = O(1)$$

Remark F22 (Sufficient Conditions for Assumption F16)**.** *Assumption F16 i. holds if* $\mathbb{E}(\hat{\psi}_n^S(\tau,\cdot))$ *is twice continuously differentiable around* β_0 *with locally uniformly bounded second derivative since for some* $\|v\|_{weak} = 1$ *and* $h \in [-1,1]$: $\beta = \beta_0 + hM_n\delta_n v$. *A Mean Value Expansion yields:*

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

Since $\tilde{h} \in (-1,1)$, the expression above is $O\left((M_n\delta_n)^2\right)$ if:

$$\sup_{\|v\|_{weak}=1,\, h\in(-1,1)} \left(\int \left| \frac{d^2\mathbb{E}(\hat{\psi}_n^S(\tau, \beta_0 + \tilde{h}M_n\delta_n v))}{d\beta d\beta}[v,v] \right|^2 \pi(\tau)d\tau \right)^{1/2} = O(1).$$

Hence Assumptions F16 i. and ii. could be nested under the following condition:

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

The following theorem establishes the asymptotic normality of $\phi(\hat{\beta}_n) - \phi(\beta_0)$ under the assumptions given above. Note that when $\sigma_n^* \to \infty$ the estimates converge at a slower than $\sqrt{n}$-rate.

Theorem F6 (Asymptotic Normality)**.** *Suppose Assumptions F13-F16 hold then:*

$$\frac{\sqrt{n}}{\sigma_n^*}\left(\phi(\hat{\beta}_n) - \phi(\beta_0)\right) = \frac{\sqrt{n}}{\sigma_n^*}\left(\phi(\hat{\beta}_n) - \phi(\beta_0)\right) S_n^* + o_p(1).$$

Furthermore, if the data (y_t, x_t) is stationary α-mixing, the simulated data is geometrically ergodic, the moments are bounded $|\hat{\psi}_t^s(\tau,\beta)| \leq 1$ and B is bounded linear then S_n^*/σ_n^* satisfies a Central Limit Theorem so that:

$$\frac{\sqrt{n}}{\sigma_n^*}\left(\phi(\hat{\beta}_n) - \phi(\beta_0)\right) \xrightarrow{d} \mathcal{N}(0,1).$$

F.4 Examples of L^p-smooth models

The following provides examples of DGP and moment combinations which satisfy Assumption F11 condition *i.*

1. iid data without covariates: $y_t^s = u_t^s$, $u_t^s \sim F$. The moment function is the empirical CDF:

   $$\hat{\psi}_n(\tau) = \frac{1}{n}\sum_{t=1}^n \mathbb{1}_{y_t \leq \tau}.$$

   Using the supremum distance, $\|\beta_1 - \beta_2\|_{\mathcal{B}} = \sup_y |F(y) - \tilde{F}(y)|$, the following holds:

   $$\left[\mathbb{E}\left(\sup_{\|F_1-F_2\|_\infty \leq \delta} \left|\mathbb{1}_{y_t^s(F_1)\leq\tau} - \mathbb{1}_{y_t^s(F_2)\leq\tau}\right|^2\right)\right]^{1/2} \leq 2\delta^{1/2}.$$

   Assumption F11, condition *i.* is satisfied with π equal to the normal density function for any $\eta > 0$.

2. Single Index Model: $y_t^s = \mathbb{1}_{x_t'\theta + u_t^s \leq 0}$, $u_t^s \sim F$. The moment function is the empirical CF:
$$\hat{\psi}_n(\tau) = \frac{1}{n}\sum_{t=1}^n \exp\left(i\tau'\left(y_t, x_t\right)\right).$$
The metric is the supremum distance between CDFs $\|\beta_1 - \beta_2\|_{\mathcal{B}} = \sup_y |F_1(y) - F_2(y)|$ and $\mathcal{B} = \{\beta = (\theta, F), \|F'\|_\infty \leq C_1 < \infty, \|\theta\| \leq C_2 < \infty\}$, a space with CDFs having continuous and bounded densities. Also, suppose that $\mathbb{E}\|x_t\| < \infty$, then:
$$\left[\mathbb{E}\left(\sup_{\|\beta_1-\beta_2\|_\infty \leq \delta} \left|\mathbb{1}_{y_t^s(\beta_1)\leq\tau} - \mathbb{1}_{y_t^s(\beta_2)\leq\tau}\right|^2\right)\right]^{1/2} \leq 2\sqrt{1 + C_1\mathbb{E}\|x_t\|}\delta^{1/2}\|\tau\|_\infty.$$
Condition $i.$ is satisfied with π equal to the normal density function for any $\eta > 0$.

3. MA(1) model: $y_t^s = u_t^s + \theta u_{t-1}^s$, $u_t^s \sim F$. The moment function is the empirical CF:
$$\hat{\psi}_n(\tau) = \frac{1}{n}\sum_{t=1}^n \exp\left(i\tau'\left(y_t, y_{t-1}\right)\right).$$
The metric is the supremum distance between quantile functions:
$\|F_1^{-1} - F_2^{-1}\|_{\mathcal{B}} = \sup_{0\leq\nu\leq1} |F_1^{-1}(\nu) - F_2^{-1}(\nu)|$. The parameter space $\mathcal{B} = \{\beta = (\theta, F), \|F^{-1}\|_\infty \leq C_1 < \infty, |\theta| \leq C_2 < \infty\}$ is the space of distributions with bounded quantile functions. The following holds:
$$\left[\mathbb{E}\left(\sup_{\|\beta_1-\beta_2\|_\infty\leq\delta}\left|\exp\left(i\tau'\left(y_t^s(\beta_1), y_{t-1}^s(\beta_1)\right)\right) - \exp\left(i\tau'\left(y_t^s(\beta_2), y_{t-1}^s(\beta_2)\right)\right)\right|^2\right)\right]^{1/2}$$
$$\leq 2(1 + C_1 + C_2)\delta\|\tau\|_\infty.$$
Condition $i.$ is satisfied with π equal to the normal density function for any $\eta > 0$.

4. AR(1) model: $y_t^s = \theta y_{t-1}^s + u_t^s$, $u_t^s \sim F$. The moment function is the empirical CF:
$$\hat{\psi}_n(\tau) = \frac{1}{n}\sum_{t=1}^n \exp\left(i\tau'\left(y_t, y_{t-1}\right)\right).$$
The metric is the supremum distance between quantile functions:
$\|F^{-1} - \tilde{F}^{-1}\|_{\mathcal{B}} = \sup_{0\leq\nu\leq1} |F^{-1}(\nu) - \tilde{F}^{-1}(\nu)|$. The parameter space $\mathcal{B} = \{\beta = (\theta, F), \|F^{-1}\|_\infty \leq C_1 < \infty, |\theta| \leq C_2 < 1\}$ is the space of distributions with uniformly bounded support, for which the quantile functions are also bounded. The following

holds:

$$\left[\mathbb{E}\left(\sup_{\|\beta_1-\beta_2\|_\infty\leq\delta}\left|\exp\left(i\tau'\left(y_t^s(\beta_1),y_{t-1}^s(\beta_1)\right)\right)-\exp\left(i\tau'\left(y_t^s(\beta_2),y_{t-1}^s(\beta_2)\right)\right)\right|^2\right)\right]^{1/2}$$
$$\leq\frac{2}{1-C_2}\left(1+\frac{C_1}{1-C_2}\right)\delta\|\tau\|_\infty.$$

Condition $i.$ is satisfied with π equal to the normal density function for any $\eta>0$.

5. Non-linear autoregressive model: $y_t^s=g_{obs}(y_{t-1}^s,\theta)+u_t^s$, $u_t^s\sim F$. The moment function is the empirical CF:

$$\hat{\psi}_n(\tau)=\frac{1}{n}\sum_{t=1}^n\exp\left(i\tau'\left(y_t,y_{t-1}\right)\right).$$

The metric is the supremum distance between quantile functions: $\|F^{-1}-\tilde{F}^{-1}\|_\mathcal{B}=\sup_{0\leq\nu\leq1}|F^{-1}(\nu)-\tilde{F}^{-1}(\nu)|$. The parameter space $\mathcal{B}=\{\beta=(\theta,F),\|F^{-1}\|_\infty\leq C_1<\infty,|\theta|\leq C_2<\infty\}$ is the space of distributions with uniformly bounded support, for which the quantile functions are also bounded. Furthermore, suppose that $|g_{obs}(y,\theta)-g_{obs}(\tilde{y},\theta)|\leq C_3|y-\tilde{y}|<|y-\tilde{y}|$ for all θ and $|g_{obs}(y,\theta)-g_{obs}(y,\tilde{\theta})|\leq C_4|\theta-\tilde{\theta}|$, then:

$$\left[\mathbb{E}\left(\sup_{\|\beta_1-\beta_2\|_\infty\leq\delta}\left|\exp\left(i\tau'\left(y_t^s(\beta_1),y_{t-1}^s(\beta_1)\right)\right)-\exp\left(i\tau'\left(y_t^s(\beta_2),y_{t-1}^s(\beta_2)\right)\right)\right|^2\right)\right]^{1/2}$$
$$\leq 2\frac{1+C_4}{1-C_3}\delta\|\tau\|_\infty.$$

Condition $i.$ is satisfied with π equal to the normal density function for any $\eta>0$.

The derivations for these examples are given below.

1. iid data without covariates: $y_t^s=u_t^s$, $u_t^s\sim F$. The moment function is the empirical CDF:

$$\hat{\psi}_n(\tau)=\frac{1}{n}\sum_{t=1}^n\mathbb{1}_{y_t\leq\tau}.$$

The metric is the supremum distance between CDFs: $\|F_1-F_2\|_\mathcal{B}=\sup_y|F_1(y)-F_2(y)|$. If $\sup_y\|F_1(y)-F_2(y)\|_\mathcal{B}\leq\delta$ then $F_1(y)-\delta\leq F_2(y)\leq F_1(y)+\delta$. Hence for $\tau\in\mathbb{R}$:

$$\begin{aligned}|\mathbb{1}_{y_t^s\leq\tau}-\mathbb{1}_{\tilde{y}_t^s\leq\tau}|^2&\leq 2|\mathbb{1}_{y_t^s\leq\tau}-\mathbb{1}_{\tilde{y}_t^s\leq\tau}|=2|\mathbb{1}_{\nu_t^s\leq F(\tau)}-\mathbb{1}_{\nu_t^s\leq\tilde{F}(\tau)}|\\&\leq 2\left(\mathbb{1}_{\nu_t^s\leq F_1(\tau)+\delta}-\mathbb{1}_{\nu_t^s\leq F_1(\tau)-\delta}\right)\end{aligned}$$

Taking expectations with respect to $\nu_t^s\sim\mathcal{U}_{[0,1]}$, for all $\tau\in\mathbb{R}$:

$$\mathbb{E}\left(\sup_{\sup_y|F_1(y)-F_2(y)|\leq\delta}|\mathbb{1}_{y_t^s\leq\tau}-\mathbb{1}_{\tilde{y}_t^s\leq\tau}|^2\right)\leq 4\delta.$$

2. Single Index Model: $y_t^s = \mathbb{1}_{x_t'\theta + u_t^s \leq 0}$, $u_t^s \sim F$. The moment function is the empirical CF:

$$\hat{\psi}_n(\tau) = \frac{1}{n}\sum_{t=1}^n \exp\left(i\tau'\left(y_t, x_t\right)\right).$$

The metric is the supremum distance between CDFs: $\|\beta_1 - \beta_2\|_{\mathcal{B}} = \sup_y |F_1(y) - F_2(y)|$ and the parameter space is $\mathcal{B} = \{\beta = (\theta, F), \|F'\|_\infty \leq C_1 < \infty, \|\theta\| \leq C_2 < \infty\}$, a space with CDFs with continuous and bounded densities. Also assume that $\mathbb{E}\|x_t\| < \infty$.

Proceeding similarly to example i.:

$$\begin{aligned}
|\mathbb{1}_{y_t^s(\beta_1)\leq\tau} - \mathbb{1}_{y_t^s(\beta_2)\leq\tau}|^2 &\leq 2|\mathbb{1}_{y_t^s(\beta_1)\leq\tau} - \mathbb{1}_{y_t^s(\beta_2)\leq\tau}| \\
&= 2|\mathbb{1}_{\nu_t^s \leq F_1(\tau - x_t'\theta_1)} - \mathbb{1}_{\nu_t^s \leq F_2(\tau - x_t'\theta_2)}| \\
&\leq 2|\mathbb{1}_{\nu_t^s \leq F_1(\tau - x_t'\theta_1)} - \mathbb{1}_{\nu_t^s \leq F_2(\tau - x_t'\theta_1)}| \\
&+ 2|\mathbb{1}_{\nu_t^s \leq F_2(\tau - x_t'\theta_1)} - \mathbb{1}_{\nu_t^s \leq F_2(\tau - x_t'\theta_2)}| \\
&\leq 2\left(\mathbb{1}_{\nu_t^s \leq F_1(\tau - x_t'\theta_1)+\delta} - \mathbb{1}_{\nu_t^s \leq F_2(\tau - x_t'\theta_1)-\delta}\right) \\
&+ 2|\mathbb{1}_{\nu_t^s \leq F_2(\tau - x_t'\theta)} - \mathbb{1}_{\nu_t^s \leq F_2(\tau - x_t'\theta_2)}|
\end{aligned}$$

Without loss of generality, assume that $x_t \geq 0$ so that:

$$\begin{aligned}
|\mathbb{1}_{y_t^s(\beta_1)\leq\tau} - \mathbb{1}_{y_t^s(\beta_2)\leq\tau}|^2 &\leq 2|\mathbb{1}_{y_t^s(\beta_1)\leq\tau} - \mathbb{1}_{y_t^s(\beta_2)\leq\tau}| \\
&\leq 2\left(\mathbb{1}_{\nu_t^s \leq F_1(\tau - x_t'\theta_1)+\delta} - \mathbb{1}_{\nu_t^s \leq F_2(\tau - x_t'\theta_1)-\delta}\right) \\
&+ 2|\mathbb{1}_{\nu_t^s \leq F_2(\tau - x_t'[\theta_1-\delta])} - \mathbb{1}_{\nu_t^s \leq F_2(\tau - x_t'[\theta_1+\delta])}|.
\end{aligned}$$

Taking expectations with respect to $\nu_t^s \sim \mathcal{U}_{[0,1]}$, for all $\tau \in \mathbb{R}$:

$$\begin{aligned}
&\mathbb{E}\left(\sup_{\beta=(\theta,F),\,\|\beta_1-\beta_2\|\leq\delta} |\mathbb{1}_{y_t^s(\beta_1)\leq\tau} - \mathbb{1}_{y_t^s(\beta_2)\leq\tau}|^2 \Big| x_t\right) \\
&\leq 2\left([F_1(\tau - x_t'\theta_1) + \delta] - [F_1(\tau - x_t'\theta_1) - \delta]\right) \\
&+ 2\left(F_2(\tau - x_t'[\theta_1 - \delta]) - F_2(\tau - x_t'[\theta_1 + \delta])\right) \\
&\leq 4\delta + 4C_1\|x_t\|\delta.
\end{aligned}$$

And then, taking expectations with respect to x_t:

$$\mathbb{E}\left(\sup_{\beta=(\theta,F),\,\|\beta_1-\beta_2\|\leq\delta} |\mathbb{1}_{y_t^s(\beta_1)\leq\tau} - \mathbb{1}_{y_t^s(\beta_2)\leq\tau}|^2\right) \leq 4\left(1 + C_1\mathbb{E}\|x_t\|\right)\delta.$$

3. MA(1) model: $y_t^s = u_t^s + \theta u_{t-1}^s$, $u_t^s \sim F$. The moment function is the empirical CF:

$$\hat{\psi}_n(\tau) = \frac{1}{n}\sum_{t=1}^{n} \exp\left(i\tau'\left(y_t, y_{t-1}\right)\right).$$

The metric is the supremum distance on quantiles: $\|F^{-1} - \tilde{F}^{-1}\|_{\mathcal{B}} = \sup_{0\le\nu\le 1}|F^{-1}(\nu) - \tilde{F}^{-1}(\nu)|$. The parameter space is $\mathcal{B} = \{\beta = (\theta, F), \|F^{-1}\|_\infty \le C_1 < \infty, |\theta| \le C_2 < \infty\}$, a space with bounded quantile functions.

As discussed in Section F.1, because the sine and cosine functions are Lipschitz continuous, the following holds for all $\tau = (\tau_1, \tau_2) \in \mathbb{R}^2$:

$$\begin{aligned}
&\Big| \exp\left(i\tau'\left(y_t^s(\beta_1), y_{t-1}^s(\beta_1)\right)\right) - \exp\left(i\tau'\left(y_t^s(\beta_2), y_{t-1}^s(\beta_2)\right)\right) \Big| \\
&\le \|\tau\|_\infty \left(|y_t^s(\beta_1) - y_t^s(\beta_2)| + |y_{t-1}^s(\beta_1) - y_{t-1}^s(\beta_2)|\right).
\end{aligned}$$

Consider the case of $|y_t^s(\beta_1) - y_t^s(\beta_2)|$:

$$\begin{aligned}
|y_t^s(\beta_1) - y_t^s(\beta_2)| &= |[F_1^{-1}(\nu_t^s) + \theta_1 F_1^{-1}(\nu_{t-1}^s)] - [F_2^{-1}(\nu_t^s) + \theta_2 F_2^{-1}(\nu_{t-1}^s)]| \\
&\le |[F_1^{-1}(\nu_t^s) - F_2^{-1}(\nu_t^s)| + |\theta_1||F_1^{-1}(\nu_{t-1}^s) - F_2^{-1}(\nu_{t-1}^s)| \\
&+ |\theta_1 - \theta_2||F_2^{-1}(\nu_{t-1}^s)| \\
&\le (1 + C_2 + C_1)\delta.
\end{aligned}$$

The same bound applies for $|y_{t-1}^s(\beta_1) - y_{t-1}^s(\beta_2)|$. Together with the previous inequalities it implies:

$$\begin{aligned}
&\Big| \exp\left(i\tau'\left(y_t^s(\beta_1), y_{t-1}^s(\beta_1)\right)\right) - \exp\left(i\tau'\left(y_t^s(\beta_2), y_{t-1}^s(\beta_2)\right)\right) \Big|^2 \\
&\le [2(1 + C_2 + C_1)\delta\|\tau\|_\infty]^2.
\end{aligned}$$

4. AR(1) model: $y_t^s = \theta y_{t-1}^s + u_t^s$, $u_t^s \sim F$. The moment function is the empirical CF:

$$\hat{\psi}_n(\tau) = \frac{1}{n}\sum_{t=1}^{n} \exp\left(i\tau'\left(y_t, y_{t-1}\right)\right).$$

The metric is the supremum distance on quantile functions:
$\|F_1^{-1} - F_2^{-1}\|_{\mathcal{B}} = \sup_{0\le\nu\le 1}|F_1^{-1}(\nu) - F_2^{-1}(\nu)|$. The parameter space is $\mathcal{B} = \{\beta = (\theta, F), \|F^{-1}\|_\infty \le C_1 < \infty, |\theta| \le C_2 < 1\}$, a space with bounded quantile functions.
Similarly to the MA(1), only $|y_t^s(\beta) - y_t^s(\tilde{\beta})|$ needs to be bounded:

$$\begin{aligned}
|y_t^s(\beta_1) - y_t^s(\beta_2)| &= |[\theta_1 y_{t-1}^s(\beta_1) + F_1^{-1}(\nu_t^s)] - [\theta_2 y_{t-1}^s(\beta_2) + F_2^{-1}(\nu_t^s)]| \\
&\le |F_1^{-1}(\nu_t^s) - F_2^{-1}(\nu_t^s)| + |\theta_1||y_{t-1}^s(\beta_1) - y_{t-1}^s(\beta_2)| + |\theta_1 - \theta_2||y_{t-1}^s(\beta_2)| \\
&\le \delta\left(1 + \frac{C_1}{1 - C_2}\right) + |C_2||y_{t-1}^s(\beta_1) - y_{t-1}^s(\beta_2)|.
\end{aligned}$$

The last inequality comes from the fact that $|\theta_1| \leq C_2 < 1$ and $|F_1^{-1}| \leq C_2$ combined with the fact that $y_t^s(\beta) = \sum_{k=0}^{t-1} \theta^k F^{-1}(\nu_t^s) + \theta^t y_0$. The initial condition y_0 is fixed, so by iterating the previous inequality:

$$|y_t^s(\beta_1) - y_t^s(\beta_2)| \leq \delta \left(1 + \frac{C_1}{1-C_2}\right) \frac{1}{1-C_2}.$$

Applying this inequality and the Lipschitz continuity of the sine and cosine functions:

$$\begin{aligned}
&\left| \exp\left(i\tau'\left(y_t^s(\beta_1), y_{t-1}^s(\beta_1)\right)\right) - \exp\left(i\tau'\left(y_t^s(\beta_2), y_{t-1}^s(\beta_2)\right)\right) \right|^2 \\
&\leq [2\left(1 + \frac{C_1}{1-C_2}\right) \frac{1}{1-C_2} \delta \|\tau\|_\infty]^2.
\end{aligned}$$

5. Non-linear autoregressive model: $y_t^s = g_{obs}(y_{t-1}^s, \theta) + u_t^s$, $u_t^s \sim F$. The moment function is the empirical CF:

$$\hat{\psi}_n(\tau) = \frac{1}{n} \sum_{t=1}^n \exp\left(i\tau'\left(y_t, y_{t-1}\right)\right).$$

The metric is the supremum distance on quantile functions:
$\|F_1^{-1} - F_2^{-1}\|_{\mathcal{B}} = \sup_{0 \leq \nu \leq 1} |F_1^{-1}(\nu) - F_2^{-1}(\nu)|$. The parameter space is $\mathcal{B} = \{\beta = (\theta, F), \|F^{-1}\|_\infty \leq C_1 < \infty, |\theta| \leq C_2 < \infty\}$, a space with bounded quantile functions. Furthermore, assume $|g_{obs}(y,\theta) - g_{obs}(\tilde{y},\theta)| \leq C_3|y - \tilde{y}| < |y - \tilde{y}|$ for all θ and $|g_{obs}(y,\theta) - g_{obs}(y,\tilde{\theta})| \leq C_4|\theta - \tilde{\theta}|$.
The proof is similar to the AR(1) example, first y_t^s needs to be bounded:

$$\begin{aligned}
|y_t^s(\beta_1) - y_t^s(\beta_2)| &= |[g_{obs}(y_{t-1}^s(\beta_1), \theta) + F_1^{-1}(\nu_t^s)] - [g_{obs}(y_{t-1}^s(\beta_2), \theta_2) + F_2^{-1}(\nu_t^s)]| \\
&\leq |F_1^{-1}(\nu_t^s) - F_2^{-1}(\nu_t^s)| + |g_{obs}(y_{t-1}^s(\beta_1), \theta_1) - g_{obs}(y_{t-1}^s(\beta_2), \theta_1)| \\
&+ |g_{obs}(y_{t-1}^s(\beta_1), \theta_2) - g_{obs}(y_{t-1}^s(\beta_2), \theta_2)| \\
&\leq (1 + C_4)\delta + C_3|y_{t-1}^s(h) - y_{t-1}^s(\beta_2)|.
\end{aligned}$$

Iterating this inequality up to $t = 0$ where the initial condition is fixed implies:

$$|y_t^s(\beta_1) - y_t^s(\beta_2)| \leq \frac{1 + C_4}{1 - C_3}\delta.$$

Similarly to the MA(1) and AR(1) models:

$$\begin{aligned}
&\left| \exp\left(i\tau'\left(y_t^s(\beta_1), y_{t-1}^s(\beta_1)\right)\right) - \exp\left(i\tau'\left(y_t^s(\beta_2), y_{t-1}^s(\beta_2)\right)\right) \right|^2 \\
&\leq \left(2\frac{1 + C_4}{1 - C_3}\delta\|\tau\|_\infty\right)^2.
\end{aligned}$$

F.5 Interpretation of the Equivalence Conditions

To prove the existence of an $\underline{a} > 0$ in Assumption F13, Chen & Pouzo (2015) use an eigenvalue condition on the variance of the moments. Since they have a bounded support the smallest eigenvalue can be bounded below. Here, the variance operator is infinite dimensional (see Carrasco & Florens, 2000, for a discussion) so that the eigenvalues may not be bounded below. However, an interpretation in terms of the eigenvalues and eigenvectors of the variance opertator is still possible. First, note that σ_n^*, $\|v_n^*\|_{weak}$ can be written as:

$$\|v_n^*\|_{weak}^2 = \int \left[Real\left(B\psi_\beta(\tau, v_n^*)\right)^2 + Im\left(B\psi_\beta(\tau, v_n^*)\right)^2\right] \pi(\tau) d\tau$$
$$\sigma_n^{*2} = \mathbb{E}\left[\int Real\left(B\psi_\beta(\tau, v_n^*)\right) Real\left(BZ_n^S(\tau)\right) + Im\left(B\psi_\beta(\tau, v_n^*)\right) Im\left(BZ_n^S(\tau)\right) \pi(\tau) d\tau\right]^2$$

The sieve variance can be expanded into three terms:

$$\begin{aligned}
&\sigma_n^{*2} = \\
&\int Real\left(B\psi_\beta(\tau_1, v_n^*)\right) Real\left(B\psi_\beta(\tau_2, v_n^*)\right) \mathbb{E}\left[Real\left(BZ_n^S(\tau_1)\right) Real\left(BZ_n^S(\tau_2)\right)\right] \pi(\tau_1)\pi(\tau_2) d\tau_1 d\tau_2 \\
&+ \int Im\left(B\psi_\beta(\tau_1, v_n^*)\right) Im\left(B\psi_\beta(\tau_2, v_n^*)\right) \mathbb{E}\left[Im\left(BZ_n^S(\tau_1)\right) Im\left(BZ_n^S(\tau_2)\right)\right] \pi(\tau_1)\pi(\tau_2) d\tau_1 d\tau_2 \\
&+ 2\int Real\left(B\psi_\beta(\tau_1, v_n^*)\right) Im\left(B\psi_\beta(\tau_2, v_n^*)\right) \mathbb{E}\left[Real\left(BZ_n^S(\tau_1)\right) Im\left(BZ_n^S(\tau_2)\right)\right] \pi(\tau_1)\pi(\tau_2) d\tau_1 d\tau_2.
\end{aligned}$$

This expansion can be re-written more compactly in matrix form:

$$\sigma_n^{*2} = \int \begin{pmatrix} Real\left(B\psi_\beta(\tau_1, v_n^*)\right) \\ Im\left(B\psi_\beta(\tau_1, v_n^*)\right) \end{pmatrix}' \Sigma_n(\tau_1, \tau_2) \begin{pmatrix} Real\left(B\psi_\beta(\tau_2, v_n^*)\right) \\ Im\left(B\psi_\beta(\tau_2, v_n^*)\right) \end{pmatrix} \pi(\tau_1)\pi(\tau_2) d\tau_1 d\tau_2$$

where

$$\Sigma_n(\tau_1, \tau_2) = n\mathbb{E}\begin{pmatrix} Real\left(BZ_n^S(\tau_1)\right) Real\left(BZ_n^S(\tau_2)\right) & Real\left(BZ_n^S(\tau_1)\right) Im\left(BZ_n^S(\tau_2)\right) \\ Im\left(BZ_n^S(\tau_2)\right) Im\left(BZ_n^S(\tau_1)\right) & Im\left(BZ_n^S(\tau_1)\right) Im\left(BZ_n^S(\tau_2)\right) \end{pmatrix}.$$

Before comparing this expression with $\|v_n^*\|_{weak}$ further simplifications are possible. Let K_n be the operator satisfying:

$$K_n B\psi_\beta(\tau, v_n^*) = \int \Sigma_n(\tau, \tau_2) \begin{pmatrix} Real\left(B\psi_\beta(\tau_2, v_n^*)\right) \\ Im\left(B\psi_\beta(\tau_2, v_n^*)\right) \end{pmatrix} \pi(\tau_2) d\tau_2$$

Then the sieve variance can be expressed in terms of the operator K_n:

$$\sigma_n^{*2} = \int B\psi_\beta(\tau, v_n^*) K_n \begin{pmatrix} Real\,(B\psi_\beta(\tau, v_n^*)) \\ Im\,(B\psi_\beta(\tau, v_n^*)) \end{pmatrix} \pi(\tau) d\tau$$

The term $\|v_n^*\|_{weak}$ can also be re-written in a similar notation:

$$\|v_n^*\|_{weak}^2 = \int \begin{pmatrix} Real\,(B\psi_\beta(\tau, v_n^*)) \\ Im\,(B\psi_\beta(\tau, v_n^*)) \end{pmatrix}' \begin{pmatrix} Real\,(B\psi_\beta(\tau, v_n^*)) \\ Im\,(B\psi_\beta(\tau, v_n^*)) \end{pmatrix} \pi(\tau) d\tau$$

Now note that these integrals are associated with an inner product in the Hilbert space $\left(L^2(\pi), \langle \cdot, \cdot \rangle_{L^2(\pi)}\right)$ with for all complex valued $\varphi_1, \varphi_2 \in L^2(\pi)$:

$$\langle \varphi_1, \varphi_2 \rangle_{L^2(\pi)} = \int \begin{pmatrix} Real(\varphi_1(\tau)) \\ Im(\varphi_1(\tau)) \end{pmatrix}' \begin{pmatrix} Real(\varphi_1(\tau)) \\ Im(\varphi_1(\tau)) \end{pmatrix} \pi(\tau) d\tau.$$

As a result, Assumption F13 can be re-written in terms of the covariance operator K_n:

$$\underline{a} \langle \psi_\beta(\cdot, v_n^*), \psi_\beta(\cdot, v_n^*) \rangle_{L^2(\pi)} \leq \langle \psi_\beta(\cdot, v_n^*), K_n \psi_\beta(\cdot, v_n^*) \rangle_{L^2(\pi)}.$$

Since $\sigma_n^* > 0$ by construction, K_n has positive eigenvalues. Let $(\varphi_{1,n}, \varphi_{2,n}, \dots)$ be the eigenvector associated with K_n and $(\lambda_{1,n}, \lambda_{2,n}, \dots)$ the associated eigenvalues (in decreasing modulus). Then $B\psi_\beta(\cdot, v_n^*) = \sum_{j \geq 1} a_{j,n} \varphi_{j,n}$ and

$$\begin{aligned} \langle \psi_\beta(\cdot, v_n^*), K_n \psi_\beta(\cdot, v_n^*) \rangle_{L^2(\pi)} &= \sum_{j \geq 1} a_{j,n}^2 \lambda_{j,n} \\ \langle \psi_\beta(\cdot, v_n^*), \psi_\beta(\cdot, v_n^*) \rangle_{L^2(\pi)} &= \sum_{j \geq 1} a_{j,n}^2. \end{aligned}$$

To go further, there are two cases:

i. $\|v_n^*\|_{weak} \to \infty$ (slower than $\sqrt{n}$ convergence rate): assume that there exists a pair $(a_{j,n}, \lambda_{j,n})$ such that $\lambda_{j,n} \geq \underline{\lambda}_j > 0$ and $a_{j,n} \to \infty$ at the same rate as $\|v_n^*\|_{weak}$: $\frac{a_{j,n}}{\|v_n^*\|_{weak}} \geq \underline{a}_j > 0$. In this case:

$$\begin{aligned} \langle \psi_\beta(\cdot, v_n^*), K_n \psi_\beta(\cdot, v_n^*) \rangle_{L^2(\pi)} \geq a_{j,n}^2 \underline{\lambda}_j &\geq \frac{a_{j,n}^2 \underline{\lambda}_j}{\langle \psi_\beta(\cdot, v_n^*), \psi_\beta(\cdot, v_n^*) \rangle_{L^2(\pi)}} \langle \psi_\beta(\cdot, v_n^*), \psi_\beta(\cdot, v_n^*) \rangle_{L^2(\pi)} \\ &\geq \underline{a}_j \langle \psi_\beta(\cdot, v_n^*), \psi_\beta(\cdot, v_n^*) \rangle_{L^2(\pi)}. \end{aligned}$$

Take for instance $\underline{a} = \underline{a}_j > 0$.

ii. $\|v_n^*\|_{weak} \not\to \infty$ ($\sqrt{n}$ convergence rate): it suffice that there exist a pair $(a_{j,n}, \lambda_{j,n})$ such that $\lambda_{j,n} \geq \underline{\lambda}_j > 0$ and $a_{j,n} \geq \underline{a}_j > 0$. In this case:

$$\langle \psi_\beta(\cdot, v_n^*), K_n \psi_\beta(\cdot, v_n^*)\rangle_{L^2(\pi)} \geq \underline{a}_j^2 \underline{\lambda}_j \geq \frac{\underline{a}_j^2 \underline{\lambda}_j}{\langle \psi_\beta(\cdot, v_n^*), \psi_\beta(\cdot, v_n^*)\rangle_{L^2(\pi)}} \langle \psi_\beta(\cdot, v_n^*), \psi_\beta(\cdot, v_n^*)\rangle_{L^2(\pi)}.$$

Let $\underline{a} = \inf_{n\geq 1} \frac{\underline{a}_j^2 \underline{\lambda}_j}{\langle \psi_\beta(\cdot, v_n^*), \psi_\beta(\cdot, v_n^*)\rangle_{L^2(\pi)}} > 0$ by assumption.

In summary: to satisfy the equivalence condition, the moments ψ_β must project on the covariance operator in directions where the variance increases at least as fast as the weak norm.

Appendix G Proofs for the Additional Asymptotic Results

G.1 Consistency

The following lemma, taken from Chen & Pouzo (2012),[2] gives sufficient conditions for the consistency of the estimators.

Lemma G12. *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:

$$\inf_{h \in \mathcal{B}_k \colon \|\beta - \beta_0\|_{\mathcal{B}} > \varepsilon} \bar{Q}_n(\beta) \geq g_0(n, k, \varepsilon) > 0 \text{ for each } n, k \geq 1$$

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 $h \in \mathcal{B}_{k(n)}$.

[2]The notation here is adapted for this paper's setting.

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 $\varepsilon > 0$*:*

$$\mathbb{P}^*\left(\|\hat{\beta}_n - \beta_0\|_{\mathcal{B}} > \varepsilon\right) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Remark G23. *Condition a. is an identification conditions. Condition b. requires the sieve approximation to be valid for the objective function. Condition d. gives an asymptotic equivalence between* $\hat{Q}_n$ *and* $\bar{Q}_n$ *up to a* $O_{p^*}(\max(c_n, c_{0,n}))$ *term; if one is close to zero, the other must be as well. It also requires that the sieve approximation rate, the rate at which* $\bar{Q}_n$ *and* $\hat{Q}_n$ *become equivalent and the approximation error goes to zero faster than the ill-posedness of the problem as measured by* g_0.

Proof of Proposition F1. :
In the iid case, if y_t^s depends on f only via the shocks e_t^s, i.e. $y_t^s = g_{obs}(x_t, \theta, e_t^s)$, then $\mathbb{E}(\hat{\psi}_t^s(\tau, \beta)) = \int \mathbb{E}\left(\exp(i\tau'(g_{obs}(x_t, \theta, \varepsilon), x_t))f(\varepsilon)d\varepsilon\right)$ for each τ. First note that $\Pi_{k(n)}\beta_0 = (\theta_0, \Pi_{k(n)}f_0)$ and:

$$\begin{aligned}\left|\mathbb{E}[\hat{\psi}_t^s(\tau, \Pi_{k(n)}\beta_0) - \hat{\psi}_t(\tau)]\right| &= \left|\mathbb{E}[\hat{\psi}_t^s(\tau, \Pi_{k(n)}\beta_0) - \hat{\psi}_t^s(\tau, \beta_0)]\right| \\ &= \left|\int \mathbb{E}\left(\exp(i\tau'(g_{obs}(x_t, \theta_0, u), x_t))\right)[\Pi_{k(n)}f_0(u) - f_0(u)]du\right| \\ &\leq \int |\Pi_{k(n)}f_0(u) - f_0(u)|du = \|\Pi_{k(n)}f_0 - f_0\|_{TV}.\end{aligned}$$

Taking squares on both sides and integrating:

$$\int \left|\mathbb{E}[(\hat{\psi}_t^s(\tau, \Pi_{k(n)}\beta_0) - \hat{\psi}_t(\tau))]\right|^2 \pi(\tau)d\tau \leq \|\Pi_{k(n)}f_0 - f_0\|_{TV}^2.$$

To conclude the proof, use the assumption that B is bounded linear so that:

$$Q_n(\Pi_{k(n)}\beta_0) \leq M_B^2 \int \left|\mathbb{E}[(\hat{\psi}_t^s(\tau, \Pi_{k(n)}\beta_0) - \hat{\psi}_t(\tau))]\right|^2 \pi(\tau)d\tau \leq M_B^2\|\Pi_{k(n)}f_0 - f_0\|_{TV}^2.$$

□

Proof of Proposition F2. :
To prove the proposition, proceed in four steps:

1. First, Assumption F11 implies:

$$\int |\hat{\psi}_n(\tau) - \mathbb{E}(\hat{\psi}_n(\tau))|^2 \pi(\tau) d\tau = O_p(1/n)$$

2. It also implies that, uniformly over $\beta \in \mathcal{B}_{k(n)}$:

$$\int |\hat{\psi}_n^S(\tau, \beta) - \mathbb{E}(\hat{\psi}_n^S(\tau, \beta))|^2 \pi(\tau) d\tau = O_p(C_n/n)$$

3. By the triangular inequality, the previous two results imply that, uniformly over $\beta \in \mathcal{B}_{k(n)}$:

$$\int \left| [\hat{\psi}_n^S(\tau, \beta) - \hat{\psi}_n(\tau)] - \mathbb{E}[\hat{\psi}_n^S(\tau, \beta) - \hat{\psi}_n(\tau)] \right|^2 \pi(\tau) d\tau = O_p(\max(1, C_n)/n).$$

   And, because B is a bounded linear operator:

$$\begin{aligned} &\int \left| [B\hat{\psi}_n^S(\tau, \beta) - B\hat{\psi}_n(\tau)] - \mathbb{E}[B\hat{\psi}_n^S(\tau, \beta) - B\hat{\psi}_n(\tau)] \right|^2 \pi(\tau) d\tau \\ &\leq M_B^2 \int \left| [\hat{\psi}_n^S(\tau, \beta) - \hat{\psi}_n(\tau)] - \mathbb{E}[\hat{\psi}_n^S(\tau, \beta) - \hat{\psi}_n(\tau)] \right|^2 \pi(\tau) d\tau = O_p(\max(1, C_n)/n). \end{aligned}$$

4. Using the inequality $|a-b|^2 \geq 1/2|a|^2 + |b|^2$ and the previous result, uniformly over $\beta \in \mathcal{B}_{k(n)}$:

$$\begin{aligned} &1/2 \int |B\hat{\psi}_n^S(\tau, \beta) - B\hat{\psi}_n(\tau)|^2 \pi(\tau) d\tau \\ &\leq \int |\mathbb{E}(B\hat{\psi}_n^S(\tau, \beta) - B\hat{\psi}_n(\tau))|^2 \pi(\tau) d\tau + O_p(\max(1, C_n)/n) \end{aligned}$$

   and

$$\begin{aligned} &1/2 \int |\mathbb{E}(B\hat{\psi}_n^S(\tau, \beta) - B\hat{\psi}_n(\tau))|^2 \pi(\tau) d\tau \\ &\leq \int |B\hat{\psi}_n^S(\tau, \beta) - B\hat{\psi}_n(\tau)|^2 \pi(\tau) d\tau + O_p(\max(1, C_n)/n). \end{aligned}$$

The last step concludes the proof of the proposition with $\delta_n^2 = \max(1, C_n)/n = o(1)$ if $C_n/n \to 0$ as $n \to \infty$.

First, consider steps *1.* and *2*:

Step 1.: For $M > 0$, a convergence rate r_n and Markov's inequality:

$$\begin{aligned}
\mathbb{P}\left(\int |\hat{\psi}_n(\tau) - \mathbb{E}(\hat{\psi}_n(\tau))|^2 \pi(\tau) d\tau \geq M r_n\right) &\leq \frac{1}{M r_n}\mathbb{E}\left(\int |\hat{\psi}_n(\tau) - \mathbb{E}(\hat{\psi}_n(\tau))|^2 \pi(\tau) d\tau\right) \\
&= \frac{1}{M r_n}\int \mathbb{E}\left(|\hat{\psi}_n(\tau) - \mathbb{E}(\hat{\psi}_n(\tau))|^2\right) \pi(\tau) d\tau \\
&\leq \frac{2}{M r_n}\frac{1 + 24\sum_{m \geq 0}\alpha(m)^{1/p}}{n}\int \pi(\tau) d\tau \\
&\leq \frac{C_{\alpha,p}}{M r_n n}.
\end{aligned}$$

The last two inequalities come from Lemma G13. If the data is iid then the mixing coefficients $\alpha(m) = 0$ for all $m \geq 1$. $C_{\alpha,p}$ is a constant that only depends on the mixing rate α, p and the bound on $|\hat{\psi}_t(\tau) - \mathbb{E}(\hat{\psi}_t(\tau))| \leq 2$. For $r_n = 1/n$ and $M \to \infty$ the probability goes to zero. As a result: $\int |\hat{\psi}_n(\tau) - \mathbb{E}(\hat{\psi}_n(\tau))|^2 \pi(\tau) d\tau = O_p(1/n)$.

Step 2.: The proof is similar to the proof of Lemma C.1 in Chen & Pouzo (2012). It also begins similarly to *Step 1*, for $M > 0$, a convergence rate r_n; using Markov's inequality:

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\beta \in \mathcal{B}_{k(n)}}\int |\hat{\psi}^S_n(\tau,\beta) - \mathbb{E}(\hat{\psi}^S_n(\tau,\beta))|^2 \pi(\tau) d\tau \geq M r_n\right) \\
&\leq \frac{1}{M r_n}\mathbb{E}\left(\sup_{\beta \in \mathcal{B}_{k(n)}}\int |\hat{\psi}^S_n(\tau,\beta) - \mathbb{E}(\hat{\psi}^S_n(\tau,\beta))|^2 \pi(\tau) d\tau\right) \\
&\leq \frac{1}{M r_n}\int \mathbb{E}\left(\sup_{\beta \in \mathcal{B}_{k(n)}} |\hat{\psi}^S_n(\tau,\beta) - \mathbb{E}(\hat{\psi}^S_n(\tau))|^2\right) \pi(\tau) d\tau \\
&\leq \frac{1}{M r_n}\int \mathbb{E}\left(\sup_{\beta \in \mathcal{B}_{k(n)}} |\hat{\psi}^s_n(\tau,\beta) - \mathbb{E}(\hat{\psi}^s_n(\tau))|^2\right) \pi(\tau) d\tau
\end{aligned}$$

Suppose that there is an upper bound C_n such that for all τ:

$$\mathbb{E}\left(\sup_{\beta \in \mathcal{B}_{k(n)}} |[\hat{\psi}^s_n(\tau,\beta) - \mathbb{E}(\hat{\psi}^s_n(\tau,\beta))]\pi(\tau)^{1/(2+\eta)}|^2\right) \leq C_n/n$$

If the following also holds $\int \pi(\tau)^{1-2/(2+\eta)} d\tau = C_\eta < \infty$ then:

$$\frac{1}{M r_n}\int \mathbb{E}\left(\sup_{h \in \mathcal{B}_{k(n)}} |\hat{\psi}^s_n(\tau,\beta) - \mathbb{E}(\hat{\psi}^s_n(\tau,\beta))|^2\right) \pi(\tau) d\tau \leq \frac{C_\eta C_n}{M r_n n}.$$

Take $r_n = C_n/n = o(1)$, then for $M \to \infty$ the probability goes to zero. As a result:

$$\sup_{\beta \in \mathcal{B}_{k(n)}}\int |\hat{\psi}^S_n(\tau,\beta) - \mathbb{E}(\hat{\psi}^S_n(\tau,\beta))|^2 \pi(\tau) d\tau = O_p(C_n/n).$$

The bounds C_n are now computed, first in the iid case. By theorem 2.14.5 of van der Vaart & Wellner (1996):

$$\mathbb{E}\left(\sup_{\beta\in\mathcal{B}_{k(n)}}\left|\sqrt{n}[\hat{\psi}_n^s(\tau,\beta)-\mathbb{E}(\hat{\psi}_n^s(\tau,\beta))]\pi(\tau)^{1/(2+\eta)}\right|^2\right)$$
$$\leq\left(1+\mathbb{E}\left(\sup_{\beta\in\mathcal{B}_{k(n)}}\left|\sqrt{n}[\hat{\psi}_n^s(\tau,\beta)-\mathbb{E}(\hat{\psi}_n^s(\tau,\beta))]\pi(\tau)^{1/(2+\eta)}\right|\right)\right)^2.$$

Also, by theorem 2.14.2 of van der Vaart & Wellner (1996) there exists a universal constant $K>0$ such that for each $\tau\in\mathbb{R}^{d_\tau}$:

$$\mathbb{E}\left(\sup_{\beta\in\mathcal{B}_{k(n)}}\left|\sqrt{n}[\hat{\psi}_n^s(\tau,\beta)-\mathbb{E}(\hat{\psi}_n^s(\tau,\beta))]\pi(\tau)^{1/(2+\eta)}\right|\right)\leq K\int_0^1\sqrt{1+\log N_{[]}(x,\Psi_{k(n)},\|\cdot\|)}dx$$

with $\Psi_{k(n)}=\left\{\psi:\mathcal{B}_{k(n)}\to\mathbb{C},\beta\to\psi_t^S(\tau,\beta)\pi(\tau)^{1/(2+\eta)}\right\}$, $N_{[]}$ is the covering number with bracketing. Because of the L^p-smoothness, it is bounded above by:

$$N_{[]}(x,\Psi_{k(n)},\|\cdot\|)\leq N_{[]}(\frac{x^{1/\gamma}}{C^{1/\gamma}},\mathcal{B}_{k(n)},\|\cdot\|)\leq C'N_{[]}(x^{1/\gamma},\mathcal{B}_{k(n)},\|\cdot\|).$$

Let $\sqrt{C_n}=\sqrt{1+\log N_{[]}(x^{1/\gamma},\mathcal{B}_{k(n)},\|\cdot\|)}dx$, together with the previous inequality, it implies:

$$\mathbb{E}\left(\sup_{\beta\in\mathcal{B}_{k(n)}}\left|\sqrt{n}[\hat{\psi}_n^s(\tau,\beta)-\mathbb{E}(\hat{\psi}_n^s(\tau,\beta))]\pi(\tau)^{1/(2+\eta)}\right|^2\right)\leq\left(1+K\sqrt{C_n}\right)^2\leq 4(1+K^2)C_n.$$

To conclude, divide by n on both sides to get the bound:

$$\mathbb{E}\left(\sup_{\beta\in\mathcal{B}_{k(n)}}\left|[\hat{\psi}_n^s(\tau,\beta)-\mathbb{E}(\hat{\psi}_n^s(\tau,\beta))]\pi(\tau)^{1/(2+\eta)}\right|^2\right)\leq 4(1+K^2)C_n/n.$$

For the dependent case, Lemma G15 implies that if $\hat{\psi}_t^s(\tau,\beta)$ is α-mixing at an exponential rate, the moments are bounded and the sieve spaces are compact:

$$\mathbb{E}\left(\sup_{\beta\in\mathcal{B}_{k(n)}}\left|\sqrt{n}[\hat{\psi}_n^s(\tau,\beta)-\mathbb{E}(\hat{\psi}_n^s(\tau,\beta))]\pi(\tau)^{1/(2+\eta)}\right|^2\right)\leq\left(1+K\sqrt{C_n}\right)^2\leq KC_n$$

with, for any $\vartheta\in(0,1)$ such that the integral exists:

$$C_n=\int_0^1\left(x^{\vartheta/2-1}\sqrt{\log N_{[]}(x^{1/\gamma},\mathcal{B}_{k(n)},\|\cdot\|_{\mathcal{B}})}+\log^2 N_{[]}(x^{1/\gamma},\mathcal{B}_{k(n)},\|\cdot\|_{\mathcal{B}})\right)dx$$

Step 3.: follows from the triangular inequality and the assumption that B is a bounded linear operator.

Step 4.: The following two inequalities can be derived from the inequality $|a-b|^2 \geq 1/2|a|^2 + |b|^2$, which is symmetric in a and b:

$$\begin{aligned} &\left|[B\hat{\psi}_n^S(\tau,\beta) - B\hat{\psi}_n(\tau)] - \mathbb{E}[B\hat{\psi}_n^S(\tau,\beta) - B\hat{\psi}_n(\tau)]\right|^2 \\ &\geq 1/2\left|B\hat{\psi}_n^S(\tau,\beta) - B\hat{\psi}_n(\tau)\right|^2 + \left|\mathbb{E}[B\hat{\psi}_n^S(\tau,\beta) - B\hat{\psi}_n(\tau)]\right|^2 \end{aligned}$$

and

$$\begin{aligned} &\left|[B\hat{\psi}_n^S(\tau,\beta) - B\hat{\psi}_n(\tau)] - \mathbb{E}[B\hat{\psi}_n^S(\tau,\beta) - B\hat{\psi}_n(\tau)]\right|^2 \\ &\geq \left|B\hat{\psi}_n^S(\tau,\beta) - B\hat{\psi}_n(\tau)\right|^2 + 1/2\left|\mathbb{E}[B\hat{\psi}_n^S(\tau,\beta) - B\hat{\psi}_n(\tau)]\right|^2. \end{aligned}$$

Taking integrals on both sides and given that

$$\int \left|[B\hat{\psi}_n^S(\tau,\beta) - B\hat{\psi}_n(\tau)] - \mathbb{E}[B\hat{\psi}_n^S(\tau,\beta) - B\hat{\psi}_n(\tau)]\right|^2 \pi(\tau)d\tau = O_p(C_n/n)$$

uniformly in $h \in \mathcal{B}_{k(n)}$, the desired result follows:

$$\begin{aligned} 1/2\hat{Q}_n^S(\beta) &\leq Q_n(\beta) + O_p(C_n/n) \\ 1/2Q_n(\beta) &\leq \hat{Q}_n^S(\beta) + O_p(C_n/n). \end{aligned}$$

With this, it follows that Assumption F10 is satisfied. □

Lemma G13. *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:*

$$\mathbb{E}\left(n|\bar{Y}_n|^2\right) \leq 1 + 24\sum_{m\geq 1}\alpha(m)^{1/p}$$

Proof of Lemma G13: 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:

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

□

The following Lemma gives a Rosenthal type inequality for possibly non-stationary α-mixing random variables. As shown in van der Vaart & Wellner (1996) and Dedecker & Louhichi (2002) these inequalities are very important to bound the expected value of the supremum of an empirical process.

Lemma G14. *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, $\min(\alpha^{-1}(u), n) = \sum_{i=k}^{n}\mathbb{1}_{u\leq\alpha_k}$.

Proof of Lemma G14: 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:

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

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. $\square$

Lemma G15. *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* $\mathcal{X} = \left\{X : \mathcal{B} \to \mathbb{C}, \beta \to X_t(\beta)\right\}$ *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:*

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

Proof of Lemma G15: The method of proof is adapted from the proof of Theorem 3 in Ben Hariz (2005); he only considers the stationary case, the non-stationary case is permitted here. Let $Z_n(\beta) = \frac{1}{\sqrt{n}}\sum_{t=1}^n X_t(\beta)$, by Lemma G14:

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

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

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

The last inequality follows from assuming $|Q_t| \le 1$. 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 $(\beta_j^k, \Delta_j^k)_{1\le j\le N(k)}$ 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. $\mathbb{E}\left(|\Delta_{t,j}^k|^2\right)^{1/2} \le 2^{-k}$ for all t, j, k.
2. For all $\beta \in \mathcal{B}$ and $k \ge 1$, there exists an index j such that $|X_t(\beta) - X_t(\beta_j^k) \le \Delta_{t,j}^k$.

Remark G24. *Because of the dynamics, the dependence of X_t can vary with β, which is not the case in Ben Hariz (2005) or Andrews & Pollard (1994). This remark, 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 as in Assumption F11. 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 $j \le N(k)$ and some $Q \ge 2$:*

$$\left[\mathbb{E}\left(\sup_{\|\beta-\beta_j^k\|_{\mathcal{B}}\le[\delta/C]^\gamma}|X_t(\beta) - X_t(\beta_j^k)|^Q\right)\right]^{1/Q} \le \delta.$$

Let $\Delta_{t,j}^k = \sup_{\|\beta-\beta_j^k\|_{\mathcal{B}}\le[\delta/C]^\gamma}|X_t(\beta) - X_t(\beta_j^k)|$ then $\left[\mathbb{E}\left(\Delta_{t,j}^{2k}\right)\right]^{1/2} \le \left[\mathbb{E}\left(\Delta_{t,j}^{Qk}\right)\right]^{1/Q}$ by Hölder's inequality which is smaller than δ by construction. $\left[\mathbb{E}(|\Delta_{t,j}^k|^2)\right]^{1/2} \le \delta = 2^{-k}$ by construction.

However, there is no guarantee that $(\Delta_{t,j}^k)_{t\ge1}$ 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:

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

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.

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:

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

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

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

The following inequality is due to Pisier (1983), for any $X_1,\ldots,X_N$ random variable:

$$\left[\mathbb{E}\left(\max_{1\leq t\leq N}|X_t|^p\right)\right]^{1/p}\leq N^{1/p}\max_{1\leq t\leq N}\left[\mathbb{E}\left(|X_t|^p\right)\right]^{1/p}.$$

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:

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

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:

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

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:

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

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:

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

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:

$$\begin{aligned}
&N(q_0)^{1/p} \leq \exp(1)\\
&n^{-1/2+1/p} \leq n^{-1/2+1/(2+\varepsilon)} \leq 1.
\end{aligned}$$

Then conclude that:

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

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 $n \geq 1$. □

G.2 Rate of Convergence

Proof of Proposition F3. : By Hölder's inequality and the L^p-smoothness assumption:

$$\left|\mathbb{E}\left(\hat{\psi}_n^s(\tau,\Pi_{k(n)}\beta_0) - \hat{\psi}_n^s(\tau,\beta_0)\right)\right|^2 \pi(\tau)^{1/(1+\eta/2)} \leq C^2\|\Pi_{k(n)}\beta_0 - \beta_0\|_{\mathcal{B}}^{2\gamma}.$$

Using the fact that $|a+b|^2 \leq 3[|a|^2+|b|^2]$:

$$\begin{aligned}
Q_n(\Pi_{k(n)}\beta_0) &\leq 3\left[Q_n(\beta_0) + \int |B\mathbb{E}\left(\hat{\psi}_n^S(\tau,\Pi_{k(n)}\beta_0) - \hat{\psi}_n^S(\tau,\beta_0)\right)|^2\pi(\tau)d\tau\right] \\
&\leq 3\left[Q_n(\beta_0) + M_B^2\int |\mathbb{E}\left(\hat{\psi}_n^S(\tau,\Pi_{k(n)}\beta_0) - \hat{\psi}_n^S(\tau,\beta_0)\right)|^2\pi(\tau)d\tau\right] \\
&\leq 3\left[Q_n(\beta_0) + \left(C^2M_B^2\int \pi^{1-\frac{2}{2+\eta}}(\tau)d\tau\right)\|\Pi_{k(n)}\beta_0 - \beta_0\|_{\mathcal{B}}^{2\gamma}\right].
\end{aligned}$$

The last inequality comes from taking integrals on both sides of the first inequality. The integral on the right-hand side is finite by assumption. To conclude the proof, take $K = 3[1 + C^2M_B^2\int \pi^{1-\frac{2}{2+\eta}}(\tau)d\tau]$. □

Proof of Theorem F5: Let $\varepsilon > 0$ and $r_n = \max(\delta_n, \sqrt{\eta_n}, \|\Pi_{k(n)}\beta_0 - \beta_0\|_{\mathcal{B}}^{\gamma}, \sqrt{Q_n(\beta_0)})$. To prove the result, it will be shown that there exists some $M > 0$ and $N > 0$ such that for all $n \geq N$:

$$\mathbb{P}\left(\|\hat{\beta}_n - \beta_0\|_{weak} \geq Mr_n\right) < \varepsilon. \tag{G.4}$$

The approach to prove existence is similar to the proof of Lemma B.1 in Chen & Pouzo (2012). First, under the stated assumptions, the following inequalities hold:

1. $\hat{Q}_n^S(\beta) \leq 2Q_n(\beta) + O_p(\delta_n^2)$

2. $Q_n(\beta) \leq K\left(\|\beta-\beta_0\|^{2\gamma} + Q_n(\beta_0)\right)$

3. $\|\beta-\beta_0\|_{weak}^2 \leq CQ_n(\beta)$

Applying them in the same order, equation (G.4) can be bounded above:

$$\begin{aligned}
&\mathbb{P}\left(\|\hat{\beta}_n - \beta_0\|_{weak} \geq M r_n\right) \\
&\leq \mathbb{P}\left(\inf_{\beta \in \mathcal{B}_{osn},\, \|\beta-\beta_0\|_{weak} \geq M r_n} \hat{Q}_n^s(\beta) \leq \inf_{\beta \in \mathcal{B}_{osn}} \hat{Q}_n^S(\beta) + O_p(\eta_n)\right) \\
&\leq \mathbb{P}\left(\inf_{\beta \in \mathcal{B}_{osn},\, \|\beta-\beta_0\|_{weak} \geq M r_n} Q_n(\beta) \leq \inf_{\beta \in \mathcal{B}_{osn}} Q_n(\beta) + O_p(\max(\delta_n^s, \eta_n))\right) \\
&\leq \mathbb{P}\left(\inf_{\beta \in \mathcal{B}_{osn},\, \|\beta-\beta_0\|_{weak} \geq M r_n} Q_n(\beta) \leq Q_n(\Pi_{k(n)}\beta_0) + O_p(\max(\delta_n^s, \eta_n))\right) \\
&\leq \mathbb{P}\left(\inf_{\beta \in \mathcal{B}_{osn},\, \|\beta-\beta_0\|_{weak} \geq M r_n} Q_n(\beta) \leq O_p(\max(\|\Pi_{k(n)}\beta_0 - \beta_0\|_{\mathcal{B}}^{2\gamma}, Q_n(\beta_0), \delta_n^s, \eta_n))\right) \\
&\leq \mathbb{P}\left(M^2 r_n^2 \leq O_p(\max(\|\Pi_{k(n)}\beta_0 - \beta_0\|_{\mathcal{B}}^{2\gamma}, Q_n(\beta_0), \delta_n^s, \eta_n))\right)
\end{aligned}$$

For r_n defined above, this probability becomes:

$$\mathbb{P}\left(M^2 \leq O_p(1)\right) \to 0 \text{ as } M \to \infty.$$

This concludes the first part of the proof. Finally:

$$\begin{aligned}
&\|\hat{\beta}_n - \beta_0\|_{\mathcal{B}} \\
&\leq \|\Pi_{k(n)}\beta_0 - \beta_0\|_{\mathcal{B}} + \|\hat{\beta}_n - \Pi_{k(n)}\beta_0\|_{\mathcal{B}} \frac{\|\hat{\beta}_n - \Pi_{k(n)}\beta_0\|_{weak}}{\|\hat{\beta}_n - \Pi_{k(n)}\beta_0\|_{weak}} \\
&\leq \|\Pi_{k(n)}\beta_0 - \beta_0\|_{\mathcal{B}} + \tau_n \|\hat{\beta}_n - \Pi_{k(n)}\beta_0\|_{weak} \\
&\leq \|\Pi_{k(n)}\beta_0 - \beta_0\|_{\mathcal{B}} + \tau_n \left(\|\hat{\beta}_n - \beta_0\|_{weak} + \|\beta_0 - \Pi_{k(n)}\beta_0\|_{weak}\right) \\
&\leq \|\Pi_{k(n)}\beta_0 - \beta_0\|_{\mathcal{B}} + \tau_n \left(\|\hat{\beta}_n - \beta_0\|_{weak} + C Q_n(\Pi_{k(n)}\beta_0)\right) \\
&\leq \|\Pi_{k(n)}\beta_0 - \beta_0\|_{\mathcal{B}} \\
&+ \tau_n \left(O_p\left(\max(\delta_n, \sqrt{\eta_n}, \|\Pi_{k(n)}\beta_0 - \beta_0\|_{\mathcal{B}}^{\gamma}, \sqrt{Q_n(\beta_0)}, \|\Pi_{k(n)}\beta_0 - \beta_0\|^{2\gamma}, Q_n(\beta_0)\right)\right) \\
&= \|\Pi_{k(n)}\beta_0 - \beta_0\|_{\mathcal{B}} + \tau_n \left(O_p\left(\max(\delta_n, \sqrt{\eta_n}, \|\Pi_{k(n)}\beta_0 - \beta_0\|_{\mathcal{B}}^{\gamma}, \sqrt{Q_n(\beta_0)})\right)\right).
\end{aligned}$$

This concludes the proof. □

Proof of Proposition F4: 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:

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

□

G.3 Asymptotic Normality

Lemma G16 (Stochastic Equicontinuity)**.** *Let* $M_n = \log\log(n+1)$ *as defined in Assumption F14. Also,* $\|\hat{\beta}_n - \beta_0\|_{\mathcal{B}} = O_p(\delta_{sn})$*. Suppose Assumption F11 holds then for any* $\vartheta \in (0,1)$*, there exists a* $C > 0$ *such that:*

$$
\left[\mathbb{E}\left(\sup_{\|\beta-\beta_0\|_{\mathcal{B}} \leq M_n \delta_{sn}} \left| [\hat{\psi}_n^S(\tau,\beta) - \hat{\psi}_n^S(\tau,\beta_0)] - \mathbb{E}[\hat{\psi}_n^S(\tau,\beta) - \hat{\psi}_n^S(\tau,\beta_0)] \right|^2 \pi(\tau)^{\frac{2}{2+\eta}} \right) \right]^{1/2}
$$
$$
\leq \frac{(M_n \delta_{sn})^\gamma}{\sqrt{n}} \sqrt{C_{sn}}
$$

where

$$
\sqrt{C_{sn}} :=
$$
$$
\int_0^1 \left(x^{-\vartheta/2} \sqrt{\log N([x M_n \delta_{sn}]^{1/\gamma}, \mathcal{B}_{k(n)}, \|\cdot\|_{\mathcal{B}})} + \log^2 N([x M_n \delta_{sn}]^{1/\gamma}, \mathcal{B}_{k(n)}, \|\cdot\|_{\mathcal{B}}) \right) dx.
$$

Now suppose that $\sqrt{C_{sn}}(M_n \delta_{sn})^\gamma = o(1)$ *as in Assumption F14. For linear sieves,* $\sqrt{C_{sn}}$ *is proportional to:*

$$
\left(\log[M_n \delta_{sn}] k(n)\right)^2 .
$$

Hence, for linear sieves $\sqrt{C_{sn}}(M_n \delta_{sn})^\gamma = o(1)$ *is implied by* $(M_n \delta_{sn})^\gamma \log(M_n \delta_{sn})^2 = o(1/k(n)^2)$*. Together with the previous inequality, this assumption implies a stochastic equicontinuity result:*

$$
\left(\int \left| [\hat{\psi}_n^S(\tau,\hat{\beta}_n) - \hat{\psi}_n^S(\tau,\beta_0)] - \mathbb{E}[\hat{\psi}_n^S(\tau,\hat{\beta}_n) - \hat{\psi}_n^S(\tau,\beta_0)] \right|^2 \pi(\tau) d\tau \right)^{1/2} = o_p(1/\sqrt{n}).
$$

Proof of Lemma G16: Let $\Delta\hat{\psi}_t^s(\tau,\beta) = \hat{\psi}_t^s(\tau,\beta) - \hat{\psi}_t^s(\tau,\beta_0)$. Under Assumption F11:

$$\left[\mathbb{E}\left(\sup_{\|\beta-\beta_0\|_{\mathcal{B}}\leq M_n\delta_{sn}} \left|\Delta\hat{\psi}_t^s(\tau,\beta)\right|^2 \pi(\tau)^{\frac{2}{2+\eta}}\right)\right]^{1/2} \leq C(M_n\delta_{sn})^\gamma$$

and

$$\left[\mathbb{E}\left(\sup_{\|\beta_1-\beta_2\|_{\mathcal{B}}\leq\delta,\beta_1,\beta_2\in B_{M_n\delta_{sn}}(\beta_0)} \left|\Delta\hat{\psi}_t^s(\tau,\beta_1) - \Delta\hat{\psi}_t^s(\tau,\beta_2)\right|^2 \frac{\pi(\tau)^{\frac{2}{2+\eta}}}{(M_n\delta_{sn})^{2\gamma}}\right)\right]^{1/2} \leq C\left(\frac{\delta}{M_n\delta_{sn}}\right)^\gamma.$$

Applying Lemma G15 to the empirical process $\Delta\hat{\psi}_t^s(\tau,\beta)\frac{\pi(\tau)^{\frac{1}{2+\eta}}}{(M_n\delta_{sn})^\gamma}$ yields:

$$\begin{aligned}
&\left[\mathbb{E}\left(\sup_{\|\beta-\beta_0\|_{\mathcal{B}}\leq M_n\delta_{sn}} \left|\Delta\hat{\psi}_n^S(\tau,\beta) - \mathbb{E}\left(\Delta\hat{\psi}_n^S(\tau,\beta)\right)\right|^2 \frac{\pi(\tau)^{\frac{2}{2+\eta}}}{(M_n\delta_{sn})^{2\gamma}}\right)\right]^{1/2} \\
&\leq \frac{C}{\sqrt{n}}\int_0^1 \left(x^{-\vartheta/2}\sqrt{\log N([xM_n\delta_{sn}]^{1/\gamma}, \mathcal{B}_{k(n)}, \|\cdot\|_{\mathcal{B}})} + \log^2 N([xM_n\delta_{sn}]^{1/\gamma}, \mathcal{B}_{k(n)}, \|\cdot\|_{\mathcal{B}})\right) dx
\end{aligned}$$

for some constant $C > 0$ and any $\vartheta \in (0,1)$ such that the integral is finite. For finite dimensional linear sieves the integral is proportional to $k(n)^2\log(M_n\delta_{sn})^2$ and the bound becomes, after multiplying by $(M_n\delta_{sn})^\gamma$ on both sides:

$$\begin{aligned}
&\left[\mathbb{E}\left(\sup_{\|\beta-\beta_0\|_{\mathcal{B}}\leq M_n\delta_{sn}} \left|\Delta\hat{\psi}_n^S(\tau,\beta) - \mathbb{E}\left(\Delta\hat{\psi}_n^S(\tau,\beta)\right)\right|^2 \pi(\tau)^{\frac{2}{2+\eta}}\right)\right]^{1/2} \\
&\leq \frac{C}{\sqrt{n}}(M_n\delta_{sn})^\gamma[\log(M_n\delta_{sn})k(n)]^2.
\end{aligned}$$

Note that $\mathbb{P}\left(\|\hat{\beta}_n - \beta_0\|_{\mathcal{B}} \leq M_n\delta_{sn}\right) \to 1$ by construction of M_n and definition of δ_{sn}. Let $\Delta_n^S(\tau,\beta) = \hat{\psi}_n^S(\tau,\hat{\beta}_n) - \mathbb{E}[\hat{\psi}_n^S(\tau,\hat{\beta}_n)]$. The following inequalities can be used:

$$\begin{aligned}
&\mathbb{P}\left(\int \left|\Delta_n^S(\tau,\hat{\beta}_n) - \Delta_n^S(\tau,\beta_0)\right|^2 \pi(\tau)^{\frac{2}{2+\eta}}\pi(\tau)^{1-\frac{2}{2+\eta}}d\tau > \frac{\varepsilon}{n}\right) \\
&\leq \mathbb{P}\left(\sup_{\|\beta-\beta_0\|_{\mathcal{B}}\leq M_n\delta_{sn}} \int \left|\Delta_n^S(\tau,\beta) - \Delta_n^S(\tau,\beta_0)\right|^2 \pi(\tau)^{\frac{2}{2+\eta}}\pi(\tau)^{1-\frac{2}{2+\eta}}d\tau > \frac{\varepsilon}{n}\right) \\
&+ \mathbb{P}\left(\|\beta-\beta_0\|_{\mathcal{B}} > M_n\delta_{sn}\right) \\
&\leq \frac{n}{\varepsilon}\mathbb{E}\left(\int \left|\Delta_n^S(\tau,\beta) - \Delta_n^S(\tau,\beta_0)\right|^2 \pi(\tau)^{\frac{2}{2+\eta}}\pi(\tau)^{1-\frac{2}{2+\eta}}d\tau\right) + \mathbb{P}\left(\|\beta-\beta_0\|_{\mathcal{B}} > M_n\delta_{sn}\right) \\
&= \int \frac{n}{\varepsilon}\mathbb{E}\left(\left|\Delta\hat{\psi}_n^S(\tau,\beta) - \mathbb{E}[\Delta\hat{\psi}_n^S(\tau,\beta)]\right|^2 \pi(\tau)^{\frac{2}{2+\eta}}\right)\pi(\tau)^{1-\frac{2}{2+\eta}}d\tau + \mathbb{P}\left(\|\beta-\beta_0\|_{\mathcal{B}} > M_n\delta_{sn}\right) \\
&\leq C_{sn}(M_n\delta_{sn})^{2\gamma}\int \pi(\tau)^{1-\frac{2}{2+\eta}}d\tau + \mathbb{P}\left(\|\beta-\beta_0\|_{\mathcal{B}} > M_n\delta_{sn}\right) = o(1).
\end{aligned}$$

These inequalitites hold regardless of $\varepsilon > 0$ given the assumptions above and the definition of $M_n\delta_{sn}$. To conclude, the stochastic equicontinuity result holds:

$$\left(\int \left|[\hat{\psi}_n^S(\tau,\hat{\beta}_n) - \hat{\psi}_n^S(\tau,\beta_0)] - \mathbb{E}[\hat{\psi}_n^S(\tau,\hat{\beta}_n) - \hat{\psi}_n^S(\tau,\beta_0)]\right|^2 \pi(\tau)^{\frac{2}{2+\eta}}\pi(\tau)^{1-\frac{2}{2+\eta}}d\tau\right)^{1/2} = o_p(1/\sqrt{n}).$$

$\square$

Lemma G17. *Suppose that $\|\hat{\beta}_n - \beta_0\|_{weak} = O_p(\delta_n)$. Under Assumptions F11, F13, F14 and F16:*

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)

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

Proof of Lemma G17: 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:

$$\left|\int \psi_\beta(\tau,u_n^*)\left(\overline{B\mathbb{E}(\hat{\psi}_n^S(\tau,\hat{\beta}_n) - \hat{\psi}_n^S(\tau,\beta_0)) - B\frac{d\mathbb{E}(\hat{\psi}_n^S(\tau,\beta_0))}{d\beta}[\hat{\beta}_n - \beta_0]}\right)\pi(\tau)d\tau\right|$$
$$= \left|\int \psi_\beta(\tau,u_n^*)\left(\overline{BR_n(\hat{\beta}_n,\beta_0)}\right)\pi(\tau)d\tau\right|$$
$$\leq M_B\left(\int |\psi_\beta(\tau,u_n^*)|^2\pi(\tau)d\tau\right)^{1/2}\left(\int \left|R_n(\hat{\beta}_n,\beta_0)\right|^2\pi(\tau)d\tau\right)^{1/2}$$

By definition of M_n and the inequality above:

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

The term $\mathbb{P}\left(\|\hat\beta_n-\beta_0\|_{\mathcal{B}} > M_n\delta_n\right) \to 0$ regardless of ε. Furthermore, Assumption F16 *i.* implies that

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

Furthermore Assumption F14 *iii.*, condition (F.3) implies that $(M_n\delta_n)^{1+\gamma} = o(\frac{1}{\sqrt{nC_{sn}}})$. Since $\gamma\in(0,1]$ it implies $(M_n\delta_n)^2 = o(1/\sqrt{n})$ and thus:

$$
\mathbb{P}\left(\Big|\int \psi_\beta(\tau,u_n^*)\left(\overline{BR_n(\hat\beta_n,\beta_0)}\right)\pi(\tau)d\tau\Big| > \frac{\varepsilon}{\sqrt{n}}\right) = o(1)
$$

regardless of $\varepsilon > 0$. Finally:

$$
\int \psi_\beta(\tau,u_n^*)\left(\overline{BR_n(\hat\beta_n,\beta_0)}\right)\pi(\tau)d\tau = o_p(1/\sqrt{n}).
$$

b) Let $\Delta_n^S(\tau,\beta) = \hat\psi_n^S(\tau,\beta)-\mathbb{E}[\hat\psi_n^S(\tau,\beta)]$. By the stochastic equicontinuity result of Lemma G16 and the Cauchy-Schwarz inequality:

$$
\begin{aligned}
&\Big|\int \psi_\beta(\tau,u_n^*)\left(\overline{B[\Delta_n^S(\hat\beta_n)-\Delta_n^S(\beta_0)]}\right)\pi(\tau)d\tau\Big|\\
&\leq\left(\int|\psi_\beta(\tau,u_n^*)|^2\pi(\tau)d\tau\right)^{1/2}\left(\int\Big|B[\Delta_n^S(\hat\beta_n)-\Delta_n^S(\beta_0)]\Big|^2\pi(\tau)d\tau\right)^{1/2}\\
&\leq M_B\left(\int|\psi_\beta(\tau,u_n^*)|^2\pi(\tau)d\tau\right)^{1/2}\left(\int\Big|[\Delta_n^S(\hat\beta_n)-\Delta_n^S(\beta_0)]\Big|^2\pi(\tau)d\tau\right)^{1/2}\\
&\leq M_B\left(\int|\psi_\beta(\tau,u_n^*)|^2\pi(\tau)d\tau\right)^{1/2}\left(\int\pi(\tau)^{1-\frac{2}{2+\eta}}d\tau\right)^{1/2}o_p(1/\sqrt{n})\\
&= o_p(1/\sqrt{n}).
\end{aligned}
$$

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:

$$0 \leq \hat Q_n^S(\hat\beta_n) \leq \inf_{\beta\in\mathcal{B}_{k(n)}} \hat Q_n^S(\beta) + O_p(\eta_n).$$

Hence the following holds:

$$0 \leq \frac{1}{2}\left(\hat Q_n^S(\hat\beta(1)) - \hat Q_n^S(\hat\beta(0))\right) + O_p(\eta_n) \tag{G.5}$$

$$= \frac{1}{2}\Big[\int B\left(\hat\psi_n(\tau) - \hat\psi_n^S(\tau,\hat\beta(0))\right)\overline{B\left(\hat\psi_n^S(\tau,\hat\beta(0)) - \hat\psi_n^S(\tau,\hat\beta(1))\right)}\pi(\tau)d\tau \tag{G.6}$$

$$+ \int \overline{B\left(\hat\psi_n(\tau) - \hat\psi_n^S(\tau,\hat\beta(0))\right)} B\left(\hat\psi_n^S(\tau,\hat\beta(0)) - \hat\psi_n^S(\tau,\hat\beta(1))\right)\pi(\tau)d\tau \tag{G.7}$$

$$+ \int \left|B\left(\hat\psi_n^S(\tau,\hat\beta(0)) - \hat\psi_n^S(\tau,\hat\beta(1))\right)\right|^2\pi(\tau)d\tau\Big] + O_p(\eta_n). \tag{G.8}$$

To prove Lemma G17 c), (G.6)-(G.7) are expanded and shown to be $o_p(1/\sqrt{n})$ and (G.8) is bounded, shown to be negligible under the assumptions.

The first step deals with (G.8):

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

By the triangular inequality and the stochastic equicontinuity results from Lemma G16:

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

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

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

for some intermediate value $\tilde{h} \in (0,1)$. Also, by assumption:

$$\left(\int \Big| \frac{d\mathbb{E}[\hat{\psi}_n^S(\tau, \hat{\beta}(\tilde{t}))}{d\beta}[u_n^*] \Big|^2 \pi(\tau) d\tau \right)^{1/2} = O_p(1).$$

Together these two elements imply:

$$\left(\int \Big| \mathbb{E}[\hat{\psi}_n^S(\tau, \hat{\beta}(t)) - \hat{\psi}_n^S(\tau, \hat{\beta}(1))] \Big|^2 \pi(\tau) d\tau \right)^{1/2} = O(\varepsilon_n).$$

This yields the bound for (G.8):

$$\int \Big| B\left(\hat{\psi}_n^S(\tau, \hat{\beta}(0)) - \hat{\psi}_n^S(\tau, \hat{\beta}(1)) \right) \Big|^2 \pi(\tau) d\tau \le O_p(\varepsilon_n^2) + O_p(\frac{(M_n \delta_{sn})^{2\gamma} C_{sn}}{n}).$$

The remaining terms, (G.6)-(G.7), are conjugates of each other. A bound for (G.6) is also valid for (G.7). Expanding (G.6) yields:

$$\int B\left(\hat{\psi}_n(\tau) - \hat{\psi}_n^S(\tau, \hat{\beta}(0)) \right) \overline{B\left(\hat{\psi}_n^S(\tau, \hat{\beta}(0)) - \hat{\psi}_n^S(\tau, \hat{\beta}(1)) \right)} \pi(\tau) d\tau \tag{G.6}$$

$$= \int B\left(\hat{\psi}_n(\tau) - \hat{\psi}_n^S(\tau, \hat{\beta}(0)) \right) \left[\overline{B\left(\Delta_n^S(\tau, \hat{\beta}(0)) - \Delta_n^S(\tau, \hat{\beta}(1)) \right)} \right] \pi(\tau) d\tau \tag{G.9}$$

$$+ \int B\left(\hat{\psi}_n(\tau) - \hat{\psi}_n^S(\tau, \hat{\beta}(0)) \right) \overline{B\mathbb{E}\left(\hat{\psi}_n^S(\tau, \hat{\beta}(0)) - \hat{\psi}_n^S(\tau, \hat{\beta}(1)) \right)} \pi(\tau) d\tau. \tag{G.10}$$

Applying the Cauchy-Schwarz inequality to (G.9) implies:

$$\Big| \int B\left(\hat{\psi}_n(\tau) - \hat{\psi}_n^S(\tau, \hat{\beta}(0)) \right) \left[\overline{B\left(\Delta_n^S(\tau, \hat{\beta}(0)) - \Delta_n^S(\tau, \hat{\beta}(1)) \right)} \right] \pi(\tau) d\tau \Big| \tag{G.9}$$

$$\le M_B \left(\int \Big| B\hat{\psi}_n(\tau) - B\hat{\psi}_n^S(\tau, \hat{\beta}(0)) \Big|^2 \pi(\tau) d\tau \right)^{1/2} \tag{G.11}$$

$$\times \left(\int \Big| \Delta_n^S(\tau, \hat{\beta}(0)) - \Delta_n^S(\tau, \hat{\beta}(1)) \Big|^2 \pi(\tau) d\tau \right)^{1/2} \tag{G.12}$$

The term (G.11) can be bounded above using the triangular inequality:

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

An application of Lemma G13 and the geometric ergodicity of $(\mathbf{y}_t^s, \mathbf{x}_t)$ yields:

$$\left(\int \left|\hat{\psi}_n(\tau) - \hat{\psi}_n^S(\tau, \beta_0)\right|^2 \pi(\tau) d\tau\right)^{1/2} = O_p(1/\sqrt{n}).$$

Expanding the term in $\hat{\psi}_n^s$ yields:

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

Note that:

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

by assumption and

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

by definition. Furthermore, the rate is $\|\hat{\beta}_n - \beta_0\|_{weak} = O_p(\delta_n)$ by assumption. Overall, the following bound holds for (G.10):

$$\begin{aligned}
&\left(\int \left|B\hat{\psi}_n(\tau) - B\hat{\psi}_n^S(\tau, \hat{\beta}(0))\right|^2 \pi(\tau) d\tau\right)^{1/2} \\
&\leq O_p\left(\frac{1}{\sqrt{n}}\right) + O_p(\delta_n) + O_p\left(\frac{(M_n\delta_n)^\gamma \sqrt{C_{sn}}}{\sqrt{n}}\right).
\end{aligned}$$

Re-arranging (G.12) to apply the stochastic equicontinuity result again yields:

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

Using the bounds for (G.10) and (G.12) yields the bound for (G.9):

$$\left|\int B\left(\hat\psi_n(\tau) - \hat\psi_n^S(\tau,\hat\beta(0))\right)\left[\overline{B\left(\Delta_n^S(\tau,\hat\beta(0)) - \Delta_n^S(\tau,\hat\beta(1))\right)}\right]\pi(\tau)d\tau\right|$$
$$\leq O_p\left(\frac{(M_n\delta_{sn})^\gamma\sqrt{C_{sn}}}{\sqrt{n}}\right) O_p\left(\max\left(M_n\delta_n, \frac{1}{\sqrt{n}}, \frac{(M_n\delta_{sn})^\gamma\sqrt{C_{sn}}}{\sqrt{n}}\right)\right).$$

To bound (G.10), 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 comes from the Cauchy-Schwarz inequality and the assumptions:

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

It was shown above that:

$$\left(\int \left|B\left(\hat\psi_n(\tau) - \hat\psi_n^S(\tau,\hat\beta(0))\right)\right|^2\pi(\tau)d\tau\right) = O_p\left(\max\left(M_n\delta_n, \frac{1}{\sqrt{n}}, \frac{(M_n\delta_{sn})^\gamma\sqrt{C_{sn}}}{\sqrt{n}}\right)^2\right).$$

Also, by Assumption F16 *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 (G.10) yields:

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

Using inequality (G.5) together with the bounds above and the expansions of (G.6) and (G.7) yields:

$$\begin{aligned}
0 \leq &-\varepsilon_n\int B\left(\hat{\psi}_n(\tau)-\hat{\psi}_n^S(\tau,\hat{\beta}(0))\right)\overline{B\frac{d\mathbb{E}(\hat{\psi}_n^S(\tau,\beta_0))}{d\beta}[u_n^*]}\pi(\tau)d\tau\\
&-\varepsilon_n\int \overline{B\left(\hat{\psi}_n(\tau)-\hat{\psi}_n^S(\tau,\hat{\beta}(0))\right)}B\frac{d\mathbb{E}(\hat{\psi}_n^S(\tau,\beta_0))}{d\beta}[u_n^*]\pi(\tau)d\tau\\
&+O_p\left(\varepsilon_n^2\right)+O_p\left(\frac{M_{sn}^\gamma C_{sn}}{\sqrt{n}}\max(\delta_{wn},\frac{1}{\sqrt{n}},\frac{M_{sn}^\gamma C_{sn}}{\sqrt{n}}))\right)\\
&+O_p\left(\varepsilon_n\delta_{wn}\max(\delta_{wn},\frac{1}{\sqrt{n}},\frac{M_{sn}^\gamma C_{sn}}{\sqrt{n}})\right)+O_p\left(\frac{M_{sn}^{2\gamma}C_{sn}^2}{n}\right)
\end{aligned}$$

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

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

By construction, $\varepsilon_n = o_p(1/\sqrt{n})$ and the assumptions imply that

$$M_n^{1+\gamma}\delta_{sn}^{\gamma}\sqrt{C_{sn}}\max\left(M_n\delta_n, \frac{1}{\sqrt{n}}, \frac{(M_n\delta_{sn})^{\gamma}\sqrt{C_{sn}}}{\sqrt{n}}\right) = o(1/\sqrt{n})$$
$$\text{and} \quad \frac{M_n^{2\gamma+1}\delta_{sn}^{2\gamma}C_{sn}}{\sqrt{n}} = o(1/\sqrt{n}).$$

To conclude the proof, note that:

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

□

Proof of Theorem F6: Using Assumption F15, the difference between ϕ evaluated at $\hat{\beta}_n$ and at the true value β_0 can be linearized:

$$\begin{aligned}
&\frac{\sqrt{n}}{\sigma_n^*}\left(\phi(\hat{\beta}_n) - \phi(\beta_0)\right) \\
&= \frac{\sqrt{n}}{\sigma_n^*}\frac{d\phi(\beta_0)}{d\beta}[\hat{\beta}_n - \beta_0] + o_p(1) \\
&= \frac{\sqrt{n}}{\sigma_n^*}\frac{d\phi(\beta_0)}{d\beta}[\hat{\beta}_n - \beta_{0,n}] + o_p(1) \\
&= \sqrt{n}\langle u_n^*, \hat{\beta}_n - \beta_{0,n}\rangle + o_p(1) \\
&= \sqrt{n}\langle u_n^*, \hat{\beta}_n - \beta_0\rangle + o_p(1) \\
&= \frac{\sqrt{n}}{2}\left(\int\left[B\psi_\beta(\tau,u_n^*)\overline{B\psi_\beta(\tau,\hat{\beta}_n - \beta_0)} + \overline{B\psi_\beta(\tau,u_n^*)}B\psi_\beta(\tau,\hat{\beta}_n - \beta_0)\right]\right)\pi(\tau)d\tau.
\end{aligned}$$

Using Lemma G16 *a)* and *b)*, replace the term $B\psi_\beta(\tau,\hat{\beta}_n - \beta_0)$ under the integral with $B\hat{\psi}_n^S(\tau,\hat{\beta}_n) - B\hat{\psi}_n^S(\tau,\beta_0)$ so that:

$$\begin{aligned}
\frac{\sqrt{n}}{\sigma_n^*}\left(\phi(\hat{\beta}_n) - \phi(\beta_0)\right) = \frac{1}{2}\Big(\int\Big[&B\psi_\beta(\tau,u_n^*)\overline{[B\hat{\psi}_n^S(\tau,\hat{\beta}_n) - B\hat{\psi}_n^S(\tau,\beta_0)]} \\
&+ \overline{B\psi_\beta(\tau,u_n^*)}[B\hat{\psi}_n^S(\tau,\hat{\beta}_n) - B\hat{\psi}_n^S(\tau,\beta_0)]\Big]\Big)\pi(\tau)d\tau + o_p(1).
\end{aligned}$$

Now Lemma G16 *c)* implies that $B\hat{\psi}_n^S(\tau,\hat{\beta}_n)$ can be replaced with $B\hat{\psi}_n(\tau)$ up to a $o_p(1/\sqrt{n})$ so that the above becomes:

$$\frac{\sqrt{n}}{\sigma_n^*}\left(\phi(\hat{\beta}_n)-\phi(\beta_0)\right)=\frac{\sqrt{n}}{2}\Big(\int\Big[B\psi_\beta(\tau,u_n^*)\overline{BZ_n^S(\tau)}+\overline{B\psi_\beta(\tau,u_n^*)}BZ_n^S(\tau)\Big]\Big)\pi(\tau)d\tau+o_p(1).$$

To conclude, apply a Central Limit Theorem to the scalar and real-valued random variable variable:

$$\frac{1}{2}\int[B\psi_\beta(\tau,u_n^*)\overline{BZ_t^S(\tau)}+\overline{B\psi_\beta(\tau,u_n^*)}BZ_t^S(\tau)]\pi(\tau)d\tau.$$

Because of u_n^* and the geometric ergodicity of the simulated data, a CLT for non-stationary mixing triangular arrays is required: results in Wooldridge & White (1988); de Jong (1997) can be applied. For any $\delta>0$:

$$\begin{aligned}
&\mathbb{E}\left(\Big|\int[\psi_\beta(\tau,u_n^*)\overline{Z_t^S(\tau)}+\overline{\psi_\beta(\tau,u_n^*)}Z_t^S(\tau)]\pi(\tau)d\tau\Big|^{2+\delta}\right)\\
&\leq 2^{2+\delta}\left[\mathbb{E}\left(\int\Big|\overline{\psi_\beta(\tau,u_n^*)}Z_t^S(\tau)\Big|\pi(\tau)d\tau\right)\right]^{2+\delta}\\
&\leq 2^{2+\delta}\left(\int\Big|B\psi_\beta(\tau,u_n^*)\Big|^2\pi(\tau)d\tau\right)^{\frac{2+\delta}{2}}\left[\mathbb{E}\left(\int\Big|BZ_t^S(\tau)\Big|^2\pi(\tau)d\tau\right)\right]^{\frac{2+\delta}{2}}.
\end{aligned}$$

By definition of u_n^* and $\|\cdot\|_{weak}$:

$$\left(\int\Big|B\psi_\beta(\tau,u_n^*)\Big|^2\pi(\tau)d\tau\right)^{1/2}=\|v_n^*\|_{weak}/\sigma_n^*\in[1/\overline{a},1/\underline{a}].$$

Because B is bounded linear and $|Z_t^S(\tau)|\leq 2$:

$$\left[\mathbb{E}\left(\int\Big|BZ_t^S(\tau)\Big|^2\pi(\tau)d\tau\right)\right]^{\frac{2+\delta}{2}}\leq[2M_B]^{2+\delta}.$$

Eventually, it implies:

$$\mathbb{E}\left(\Big|\int[\psi_\beta(\tau,u_n^*)\overline{Z_t^S(\tau)}+\overline{\psi_\beta(\tau,u_n^*)}Z_t^S(\tau)]\pi(\tau)d\tau\Big|^{2+\delta}\right)\leq\frac{[4M_B]^{2+\delta}}{\underline{a}}<\infty.$$

Given the mixing condition and the definition of σ_n^*:

$$\frac{\sqrt{n}}{2}\int[B\psi_\beta(\tau,u_n^*)[\overline{BZ_t^S(\tau)-B\mathbb{E}(Z_t^S(\tau))}]+\overline{B\psi_\beta(\tau,u_n^*)}[BZ_t^S(\tau)-B\mathbb{E}(Z_t^S(\tau))]]\pi(\tau)d\tau\xrightarrow{d}\mathcal{N}(0,1).$$

By geometric ergodicity and because the characteristic function is bounded $\sqrt{n}|\mathbb{E}(Z_t^S(\tau))|\leq C_\rho/\sqrt{n}=o(1)$, hence:

$$\frac{\sqrt{n}}{2}\int[B\psi_\beta(\tau,u_n^*)\overline{BZ_t^S(\tau)}+\overline{B\psi_\beta(\tau,u_n^*)}BZ_t^S(\tau)]\pi(\tau)d\tau\xrightarrow{d}\mathcal{N}(0,1).$$

This concludes the proof. □

References

Ai, C. & Chen, X. (2003). Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions. *Econometrica*, 71(6), 1795–1843.

Andrews, D. W. K. & Pollard, D. (1994). An Introduction to Functional Central Limit Theorems for Dependent Stochastic Processes. *International Statistical Review / Revue Internationale de Statistique*, 62(1), 119.

Ben Hariz, S. (2005). Uniform CLT for empirical process. *Stochastic Processes and their Applications*, 115(2), 339–358.

Bierens, H. J. & Song, H. (2012). Semi-nonparametric estimation of independently and identically repeated first-price auctions via an integrated simulated moments method. *Journal of Econometrics*, 168(1), 108–119.

Blundell, R., Chen, X., & Kristensen, D. (2007). Semi-nonparametric IV estimation of shape-invariant engel curves. *Econometrica*, 75(6), 1613–1669.

Carrasco, M., Chernov, M., Florens, J.-P., & Ghysels, E. (2007). Efficient estimation of general dynamic models with a continuum of moment conditions. *Journal of Econometrics*, 140(2), 529–573.

Carrasco, M. & Florens, J.-P. (2000). Generalization of GMM to a Continuum of Moment Conditions. *Econometric Theory*, 16(6), 797–834.

Chen, X. (2007). Chapter 76 Large Sample Sieve Estimation of Semi-Nonparametric Models. In *Handbook of Econometrics*, volume 6 (pp. 5549–5632).

Chen, X. (2011). Penalized Sieve Estimation and Inference of Seminonparametric Dynamic Models: A Selective Review. In D. Acemoglu, M. Arellano, & E. Dekel (Eds.), *Advances in Economics and Econometrics* (pp. 485–544). Cambridge: Cambridge University Press.

Chen, X. & Pouzo, D. (2012). Estimation of Nonparametric Conditional Moment Models With Possibly Nonsmooth Generalized Residuals. *Econometrica*, 80(1), 277–321.

Chen, X. & Pouzo, D. (2015). Sieve Wald and QLR Inferences on Semi/Nonparametric Conditional Moment Models. *Econometrica*, 83(3), 1013–1079.

Davydov, Y. A. (1968). Convergence of Distributions Generated by Stationary Stochastic Processes. *Theory of Probability & Its Applications*, 13(4), 691–696.

de Jong, R. M. (1997). Central Limit Theorems for Dependent Heterogeneous Random Variables. *Econometric Theory*, 13(03), 353.

Dedecker, J. & Louhichi, S. (2002). Maximal Inequalities and Empirical Central Limit Theorems. In *Empirical Process Techniques for Dependent Data* (pp. 137–159). Boston, MA: Birkhäuser Boston.

Duffie, D. & Singleton, K. J. (1993). Simulated Moments Estimation of Markov Models of Asset Prices. *Econometrica*, 61(4), 929.

Gallant, a. R. & Nychka, D. W. (1987). Semi-Nonparametric Maximum Likelihood Estimation. *Econometrica*, 55(2), 363–390.

Liebscher, E. (2005). Towards a Unified Approach for Proving Geometric Ergodicity and Mixing Properties of Nonlinear Autoregressive Processes. *Journal of Time Series Analysis*, 26(5), 669–689.

Pisier, G. (1983). Some applications of the metric entropy condition to harmonic analysis. In R. C. Blei & S. J. Sidney (Eds.), *Banach spaces, Harmonic analysis and Probability, Univ. of Connecticut 1980-81. Lecture Notes in Mathematics, 995* (pp. 123–154). Berlin, Heidelberg: Springer Berlin Heidelberg.

Rio, E. (2000). *Théorie Asymptotique des Processus Aléatoires Faiblement Dépendants*, volume 31 of *Mathématiques et Applications*. Springer Berlin Heidelberg.

van der Vaart, A. W. & Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer Series in Statistics. New York, NY: Springer New York.

Wooldridge, J. M. & White, H. (1988). Some Invariance Principles and Central Limit Theorems for Dependent Heterogeneous Processes. *Econometric Theory*, 4(02), 210–230.