Online Supplement to "A Note on Generalized Empirical Likelihood Estimation of Semiparametric Conditional Moment Restriction Models"

Naoya Sueishi

*Kobe University**

Abstract

This supplement contains definitions, assumptions, and detailed proofs of the lemmas and theorems in the main paper. Section 1 gives definitions. Section 2 lists assumptions and main theorems. Section A provides proofs.

1 Definitions

First, we define a Hölder class. A real-valued function f on $\mathcal{X} \subset \mathbb{R}^{d_x}$ is said to satisfy a Hölder condition with exponent $\gamma \in (0,1)$ if there is a constant c such that $|f(x) - f(y)| \leq c\|x - y\|^{\gamma}$ for all $x, y \in \mathcal{X}$. Let $a = (a_1, \ldots, a_{d_x})'$ and $[a] = a_1 + \cdots + a_{d_x}$, we then define the differential operator ∇^a by

$$\nabla^a f(x) = \frac{\partial^{[a]} f(x)}{\partial x_1^{a_1} \cdots \partial x_{d_x}^{a_{d_x}}}.$$

Let m be a nonnegative integer and set $p = m + \gamma$. A real-valued function f is said to be p-smooth if it is m times continuously differentiable and $\nabla^a f$ satisfies a Hölder condition with exponent γ for all a with $[a] = m$.

Denote by $\Lambda^p(\mathcal{X})$ the class of all p-smooth real-valued functions on $\mathcal{X}$. $\Lambda^p(\mathcal{X})$ is called a Hölder class. Also, denote by $C^m(\mathcal{X})$ the space of all m-times continuously differentiable

*Graduate School of Economics, Rokkodai-cho, Nada-ku, Kobe, 657-8501, Japan. Email: sueishi@econ.kobe-u.ac.jp.

real-valued functions on $\mathcal{X}$. Define a Hölder ball with radius c and smoothness $p = m + \gamma$ as

$$\Lambda_c^p(\mathcal{X}) = \left\{ f \in C^m(\mathcal{X}) : \sup_{[a] \leq m} \sup_{x \in \mathcal{X}} |\nabla^a f(x)| \leq c, \sup_{[a]=m} \sup_{x,y \in \mathcal{X}, x \neq y} \frac{|\nabla^a f(x) - \nabla^a f(y)|}{\|x-y\|^\gamma} \leq c \right\}.$$

The Hölder class functions can be approximated well by various linear sieves such as power series, Fourier series, and splines. For details, see Chen (2007).

Let $\mathcal{A} = \Theta \times \mathcal{H}$ be the parameter space and let $\| \cdot \|_s$ be a metric on $\mathcal{A}$. The following definitions are borrowed from Ai and Chen (2003) (hereafter AC, 2003).

Definition 1.1 *A real-valued measurable function $f(Z, \alpha)$ is Hölder continuous in $\alpha \in \mathcal{A}$ if there exists a constant $\kappa \in (0, 1]$ and a measurable function $c(Z)$ with $E[c(Z)^2|X]$ bounded, such that $|f(Z, \alpha_1) - f(Z, \alpha_2)| \leq c(Z)\|\alpha_1 - \alpha_2\|_s^\kappa$ for all $Z \in \mathcal{Z}$ and $\alpha_1, \alpha_2 \in \mathcal{A}$.*

Definition 1.2 *A real-valued measurable function $f(Z, \alpha)$ satisfies an envelope condition over $\alpha \in \mathcal{A}$ if there exists a measurable function $c(Z)$ with $E[c(Z)^4] < \infty$ such that $|f(Z, \alpha)| \leq c(Z)$ for all $Z \in \mathcal{Z}$ and $\alpha \in \mathcal{A}$.*

Next, we define a pseudo-metric $\| \cdot \|_w$, which is originally introduced by AC (2003). We assume that $\mathcal{A}$ is connected in the sense that for any $\alpha_1, \alpha_2 \in \mathcal{A}$, there exists a continuous path $\{\alpha(\tau) : \tau \in [0, 1]\}$ in $\mathcal{A}$ such that $\alpha(0) = \alpha_1$ and $\alpha(1) = \alpha_2$. Suppose that $\mathcal{A}$ is convex at α_0 in the sense that for any $\alpha \in \mathcal{A}$, $(1-\tau)\alpha_0 + \tau\alpha \in \mathcal{A}$ for small $\tau > 0$. Moreover, suppose that for almost all Z, $\rho(Z, (1-\tau)\alpha_0 + \tau\alpha)$ is continuously differentiable at $\tau = 0$. Under these assumptions, we define the pathwise derivative at the direction $[\alpha - \alpha_0]$ at α_0 by

$$\frac{d\rho(Z, \alpha_0)}{d\alpha}[\alpha - \alpha_0] = \left. \frac{d\rho(Z, (1-\tau)\alpha_0 + \tau\alpha)}{d\tau} \right|_{\tau=0} \quad \text{a.s. } Z.$$

Also, for $\alpha_1, \alpha_2 \in \mathcal{A}$, we denote

$$\frac{d\rho(Z, \alpha_0)}{d\alpha}[\alpha_1 - \alpha_2] = \frac{d\rho(Z, \alpha_0)}{d\alpha}[\alpha_1 - \alpha_0] - \frac{d\rho(Z, \alpha_0)}{d\alpha}[\alpha_2 - \alpha_0],$$
$$\frac{dm(X, \alpha_0)}{d\alpha}[\alpha_1 - \alpha_2] = E\left[\frac{d\rho(Z, \alpha_0)}{d\alpha}[\alpha_1 - \alpha_2]|X\right].$$

For $\alpha_1, \alpha_2 \in \mathcal{A}$, we define the pseudo-metric $\| \cdot \|_w$ as

$$\|\alpha_1 - \alpha_2\|_w = \sqrt{E\left[\left\{\frac{dm(X, \alpha_0)}{d\alpha}[\alpha_1 - \alpha_2]\right\}' \Sigma(X, \alpha_0)^{-1} \left\{\frac{dm(X, \alpha_0)}{d\alpha}[\alpha_1 - \alpha_2]\right\}\right]}.$$

2 Assumptions and main results

2.1 Consistency

We impose the following assumptions to prove consistency. Most of them are adopted from AC (2003) and Donald, Imbens, and Newey (2003) (hereafter DIN, 2003).

Assumption 3.1 *(i) The data* $\{(Y_i, X_i)_{i=1}^n\}$ *are i.i.d.; (ii)* $\mathcal{X}$ *is compact; (iii) the density of* X *is bounded above and away from zero.*

Assumption 3.2 *(i) For each* k_n *there is a constant* $\zeta(k_n)$ *and matrix* B *such that* $\tilde{p}^{k_n}(X) = Bp^{k_n}(X)$ *for all* $X \in \mathcal{X}$, $\sup_{X\in\mathcal{X}} \|\tilde{p}^{k_n}(X)\| \leq \zeta(k_n)$, $E[\tilde{p}^{k_n}(X)\tilde{p}^{k_n}(X)']$ *has smallest eigenvalue bounded away from zero, and* $\sqrt{k_n} \leq \zeta(k_n)$*; (ii) for any* $f(\cdot)$ *with* $E[f(X)^2] < \infty$*, there exists* $k_n \times 1$ *vector* π_{k_n} *such that* $E[\{f(X) - p^{k_n}(X)'\pi_{k_n}\}^2] = o(1)$.

Assumption 3.3 $\alpha_0 \in \mathcal{A}$ *is the only* $\alpha \in \mathcal{A}$ *satisfying* $E[\rho(Z, \alpha)|X] = 0$ *a.s.* X.

Assumption 3.4 $\Sigma(X, \alpha)$ *is finite positive definite uniformly over* $X \in \mathcal{X}$ *and* $\alpha \in \mathcal{A}$.

Assumption 3.5 *(i) There is a metric* $\|\cdot\|_s$ *such that* $\mathcal{A} = \Theta \times \mathcal{H}$ *is compact under* $\|\cdot\|_s$*; (ii) there is a constant* μ_1 *such that for any* $\alpha \in \mathcal{A}$*, there exists* $\Pi_n\alpha \in \mathcal{A}_n = \Theta \times \mathcal{H}_n$ *such that* $\|\Pi_n\alpha - \alpha\|_s^\kappa = O(k_{1n}^{-\mu_1})$ *with* $k_{1n}^{-\mu_1}\sqrt{k_n} \to 0$.

Assumption 3.6 *(i)* $E[\sup_{\alpha\in\mathcal{A}} \|\rho(Z, \alpha)\|^4|X] < \infty$*; (ii)* $\rho(Z, \alpha)$ *is Hölder continuous in* $\alpha \in \mathcal{A}$ *with respect to the metric given in Assumption 3.5.*

Assumption 3.7 *(i)* $d_\rho k_n \geq d_\theta + k_{1n}$ *and* $k_n/n = o(1)$.

Assumption 3.8 *(i)* $s(v)$ *is twice continuously differentiable with Lipschitz second derivative in a neighborhood of 0; (ii) there exists* $m > 2$ *such that* $E[\sup_{\alpha\in\mathcal{A}} \|\rho(Z, \alpha)\|^m] < \infty$ *and* $\zeta(k_n)^2 k_n/n^{1-2/m} \to 0$*; (iii)* $n^{1/m}\zeta(k_n)\sqrt{k_n}k_{1n}^{-\mu_1} \to 0$.

Theorem 3.1 *Suppose that Assumptions 3.1-3.8 hold. Then, the SGEL estimator satisfies* $\|\hat{\alpha}_n - \alpha_0\|_s = o_p(1)$.

2.2 Rate of convergence

Let $N(\delta, \mathcal{A}_n, \|\cdot\|_s)$ be the covering number of radius δ balls of $\mathcal{A}_n$ under $\|\cdot\|_s$. To obtain the convergence rate of the SGEL estimator, we impose additional assumptions.

Assumption 3.2 *(iii) For any $f(\cdot) \in \Lambda_c^p(\mathcal{X})$ with $p > d_x/2$, there exists $p^{k_n}(\cdot)'\pi_{k_n} \in \Lambda_c^p(\mathcal{X})$ such that $\sup_{X\in\mathcal{X}} |f(X) - p^{k_n}(X)'\pi_{k_n}| = O(k_n^{-p/d_x})$ and $k_n^{-p/d_x} = o(n^{-1/4})$.*

Assumption 3.5 *(iii) There is a constant $\mu_2 > 0$ such that for any $\alpha \in \mathcal{A}$, there is $\Pi_n\alpha \in \mathcal{A}_n$ satisfying $\|\Pi_n\alpha - \alpha\|_w = O(k_{1n}^{-\mu_2})$ and $k_{1n}^{-\mu_2} = o(n^{-1/4})$.*

Assumption 3.6 *(iii) Each element of $\rho(Z, \alpha)$ satisfies an envelope condition over $\alpha \in \mathcal{A}_n$; (iv) each element of $m(\cdot, \alpha)$ is in $\Lambda_c^p(\mathcal{X})$ with $p > d_x/2$ for all $\alpha \in \mathcal{A}_n$.*

Assumption 3.7 *(ii) $k_{1n}\ln(n)\zeta(k_n)^2 n^{-1/2} = o(1)$.*

Assumption 3.8 *(iv) $n^{-1/4+1/m}\zeta(k_n) = o(1)$; (v) $n^{-1/2}k_n^{1/2}\zeta(k_n) = o(n^{-1/4})$.*

Assumption 3.9 *(i) $\|\Sigma(X, \alpha_1)^{1/2} - \Sigma(X, \alpha_2)^{1/2}\| \leq c\|\alpha_1 - \alpha_2\|_s^\kappa$ for all $\alpha_1, \alpha_2 \in \mathcal{A}_{0n}$ and $X \in \mathcal{X}$ with some constant $c < \infty$; (ii) each element of $\Sigma(\cdot, \alpha)^{-1/2}$ is in $\Lambda_c^p(\mathcal{X})$ with $p > d_x/2$ for all $\alpha \in \mathcal{A}_n$.*

Assumption 3.10 $\ln N(\epsilon^{1/\kappa}, \mathcal{A}_n, \|\cdot\|_s) \leq$ *const.* $\times\, k_{1n}\ln(k_{1n}/\epsilon)$.

Assumption 3.11 *(i) $\mathcal{A}$ is convex in α_0 and $\rho(Z, \alpha)$ is pathwise differentiable at α_0; (ii) for some $c_1, c_2 > 0$,*

$$c_1 E[m(X,\alpha)'\Sigma(X,\alpha_0)^{-1}m(X,\alpha)] \leq \|\alpha - \alpha_0\|_w^2 \leq c_2 E[m(X,\alpha)'\Sigma(X,\alpha_0)^{-1}m(X,\alpha)]$$

for all $\alpha \in \mathcal{A}_n$ with $\|\alpha - \alpha_0\|_s = o(1)$.

Theorem 3.2 *Suppose that Assumptions 3.1-3.11 hold. Then, the SGEL estimator satisfies $\|\hat{\alpha}_n - \alpha_0\|_w = o_p(n^{-1/4})$.*

3.3 Asymptotic normality

Let $\mathcal{N}_{0n} = \{\alpha \in \mathcal{A}_n : \|\alpha - \alpha_0\|_s = o(1), \|\alpha - \alpha_0\|_w = o(n^{-1/4})\}$ and $\mathcal{N}_0 = \{\alpha \in \mathcal{A} : \|\alpha - \alpha_0\|_s = o(1), \|\alpha - \alpha_0\|_w = o(n^{-1/4})\}$. Following additional assumptions are required for the asymptotic normality.

Assumption 4.1 *(i)* $E[D_{w^*}(X)'\Sigma(X,\alpha_0)^{-1}D_{w^*}(X)]$ *is positive definite; (ii)* $\theta_0 \in int(\Theta)$.

Assumption 4.2 *There is a* $v_n^* = (v_\theta^*, -\Pi_n w^* \times v_\theta^*) \in \mathcal{A}_n - \alpha_0$ *such that* $\|v_n^* - v^*\|_w = O(n^{-1/4})$.

Assumption 4.3 *For all* $\alpha \in \mathcal{N}_0$*, the pathwise first derivative* $(d\rho(Z,\alpha(t))/d\alpha)[v]$ *exists a.s.* $Z \in \mathcal{Z}$*. Also, (i) each element of* $(d\rho(Z,\alpha(t))/d\alpha)[v_n^*]$ *satisfies an envelope condition and is Hölder continuous in* $\alpha \in \mathcal{N}_{0n}$*; (ii) each element of* $(dm(\cdot,\alpha)/d\alpha)[v_n^*]$ *is in* $\Lambda_c^p(\mathcal{X})$*,* $p > d_x/2$ *for all* $\alpha \in \mathcal{N}_0$.

Assumption 4.4 *Uniformly over* $\alpha \in \mathcal{N}_{0n}$*,*

$$E\left[\left\|\Sigma(X,\alpha)^{-1/2}\frac{dm(X,\alpha)}{d\alpha}[v_n^*] - \Sigma(X,\alpha_0)^{-1/2}\frac{dm(X,\alpha_0)}{d\alpha}[v_n^*]\right\|^2\right] = o(n^{-1/2}).$$

Assumption 4.5 *Uniformly over* $\alpha \in \mathcal{N}_0$ *and* $\bar{\alpha} \in \mathcal{N}_{0n}$*,*

$$E\left[\left\{\frac{dm(X,\alpha_0)}{d\alpha}\right\}'\Sigma(X,\alpha_0)^{-1}\left\{\frac{dm(X,\alpha)}{d\alpha}[\bar{\alpha}-\alpha_0] - \frac{dm(X,\alpha_0)}{d\alpha}[\bar{\alpha}-\alpha_0]\right\}\right] = o(n^{-1/2}).$$

Assumption 4.6 *For all* $\alpha \in \mathcal{N}_{0n}$*, the pathwise second derivative* $d^2\rho(Z,\alpha+\tau v_n^*)/d\tau^2|_{\tau=0}$ *exists a.s.* $Z \in \mathcal{Z}$*, and is bounded by a measurable function* $c(Z)$ *with* $E[c(Z)^2] < \infty$.

Assumption 4.7 $s(v)$ *is three times continuously differentiable with Lipschitz third derivative in a neighborhood of 0.*

Assumption 4.8 $n^{-1/4+1/m}\zeta(k_n) = o(n^{-1/8})$.

Theorem 4.1 *Under Assumptions 3.1-3.11 and 4.1-4.8, the SGEL estimator satisfies* $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, V^{-1})$*, where* $V = E[D_{w^*}(X)'\Sigma(X,\alpha_0)^{-1}D_{w^*}(X)]$.

A Proofs

Throughout this section, C denotes a generic positive constant which may be different in different uses. The qualifier "with probability approaching one" will be abbreviated as w.p.a.1. $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote the minimum and maximum eigenvalues of a matrix A. Also, let $p_i = p^{k_n}(X_i)$.

A.1 Consistency

The outline of the proof is the same as that of Theorem 5.5 in DIN (2003). There are two main differences: (1) our parameter of interest is infinite dimensional; and (2) the minimization problem is solved over the sieve space $\mathcal{A}_n$ rather than the original parameter space $\mathcal{A}$.

Lemma A.1 *Suppose that $\alpha \in \mathcal{A}_n$ satisfies $\|\alpha - \alpha_0\|_s = o(1)$. Let*

$$\begin{aligned}\hat{\Omega}(\alpha) &= \frac{1}{n}\sum_{i=1}^{n} g_i(\alpha)g_i(\alpha)', \ \bar{\Omega}(\alpha) = \frac{1}{n}\sum_{i=1}^{n} \Sigma(X_i,\alpha)\otimes p_i p_i',\\ \Omega(\alpha) &= E\left[g_i(\alpha)g_i(\alpha)'\right].\end{aligned}$$

Suppose that Assumptions 3.1 (i), 3.2 (i), 3.4, and 3.6 (i) are satisfied. Then we have

$$\left\|\hat{\Omega}(\alpha) - \bar{\Omega}(\alpha)\right\| = O_p(\zeta(k_n)\sqrt{k_n/n}), \quad \left\|\bar{\Omega}(\alpha) - \Omega(\alpha)\right\| = O_p(\zeta(k_n)\sqrt{k_n/n}).$$

Also, we obtain $1/C \le \lambda_{min}(\Omega(\alpha)) \le \lambda_{\max}(\Omega(\alpha)) \le C$. Moreover, if $\zeta(k_n)\sqrt{k_n/n} \to 0$, then $1/C \le \lambda_{min}(\hat{\Omega}(\alpha)) \le \lambda_{\max}(\hat{\Omega}(\alpha)) \le C$ w.p.a.1.

Proof. The result is obtained from Lemma A.6 of DIN (2003). In their lemma, $\tilde{\Omega}$, $\bar{\Omega}$, and Ω are evaluated at the true parameter value β_0, while $\hat{\Omega}(\alpha)$, $\bar{\Omega}(\alpha)$, and $\Omega(\alpha)$ depend on general α, which can be different from α_0. Because of this, we impose Assumptions 3.4 and 3.6 (i), which are stronger than the assumptions in DIN (2003). Then the proof is almost the same as that of DIN (2003). ■

Lemma A.2 *Suppose that Assumptions 3.1 (i) and 3.8 (ii) hold. Then for $\delta_n = o(n^{-1/m}\zeta(k_n)^{-1})$ and $\Lambda_n = \{\lambda : \|\lambda\| \le \delta_n\}$, we have*

$$\max_{1\le i\le n}\sup_{\alpha\in\mathcal{A}}\sup_{\lambda\in\Lambda_n}|\lambda' g_i(\alpha)| \xrightarrow{p} 0.$$

Also, w.p.a.1 we have $\Lambda_n \subset \hat{\Lambda}(\alpha)$ for all $\alpha \in \mathcal{A}$.

Proof. See Lemma A.10 of DIN (2003). ■

Hereafter, let $\delta_n = o(n^{-1/m}\zeta(k_n)^{-1})$ and $\alpha_{n0} = \Pi_n\alpha_0$. Also, let $\hat{g}(\alpha) = n^{-1}\sum_{i=1}^n g_i(\alpha)$ and $\hat{S}(\alpha,\lambda) = n^{-1}\sum_{i=1}^n s(\lambda' g_i(\alpha))$.

Lemma A.3 *Suppose that Assumptions 3.1 (i), 3.2 (i), 3.4, 3.5 (ii), 3.6 (i), (ii), and 3.8 hold. Then,* $\sup_{\lambda\in\hat{\Lambda}(\alpha_{n0})}\hat{S}(\alpha_{n0},\lambda) = o_p(\delta_n^2)$, $\tilde{\lambda} = \arg\max_{\lambda\in\hat{\Lambda}(\alpha_{n0})}\hat{S}(\alpha_{n0},\lambda)$ *exists w.p.a.1, and* $\|\tilde{\lambda}\| = o_p(\delta_n)$.

Proof. We modify Lemma A.11 of DIN (2003) to take into account the difference in convergence rate between their and our estimators. By Lemma A.9 of DIN (2003), we have $\|\hat{g}(\alpha_0)\| = O_p(\sqrt{k_n/n})$. Also, by Assumptions 3.5 (ii) and 3.6 (ii),

$$\|\hat{g}(\alpha_{n0}) - \hat{g}(\alpha_0)\| \le \|\alpha_{n0} - \alpha_0\|_s^\kappa \frac{1}{n}\sum_{i=1}^n c(Z_i)\|p_i\| = O_p(k_{1n}^{-\mu_1}\sqrt{k_n}).$$

Thus by the triangular inequality, we have $\|\hat{g}(\alpha_{n0})\| = O_p(\sqrt{k_n/n} + k_{1n}^{-\mu_1}\sqrt{k_n})$.

It follows from Assumptions 3.8 (ii) and (iii) that we can choose $\sqrt{k_n/n} + k_{1n}^{-\mu_1}\sqrt{k_n} = o(\delta_n)$. Also, we choose Λ_n as in Lemma A.2. Then $\bar{\lambda} = \arg\max_{\lambda\in\Lambda_n}\hat{S}(\alpha_{n0},\lambda)$ exists w.p.a.1. Moreover, by Lemmas A.1 and A.2 and Assumptions 3.4 and 3.8, a Taylor expansion yields

$$0 = \hat{S}(\alpha_{n0},0) \le \hat{S}(\alpha_{n0},\bar{\lambda}) \le \|\bar{\lambda}\|\|\hat{g}(\alpha_{n0})\| - C\|\bar{\lambda}\|^2 \tag{A.1}$$

and hence $\|\bar{\lambda}\| = o_p(\delta_n)$. The remaining part of the proof follows DIN (2003). ■

Lemma A.4 *Suppose that Assumptions 3.1 (i), 3.2 (i), 3.4, 3.5 (ii), 3.6, and 3.8 (i)-(iii) hold. Then* $\|\hat{g}(\hat{\alpha}_n)\| = O_p(\delta_n)$.

Proof. We modify Lemmas A.13 and A.14 of DIN (2003). In their proof, they use the fact that $\sup_{\lambda\in\hat{\Lambda}(\hat{\beta})}\hat{S}(\hat{\beta},\lambda) \le \sup_{\lambda\in\hat{\Lambda}(\beta_0)} S(\beta_0,\lambda)$, which is obtained by the definition of $\hat{\beta}$. In contrast, we may not have $\sup_{\lambda\in\hat{\Lambda}(\hat{\alpha}_n)}\hat{S}(\hat{\alpha}_n,\lambda) \le \sup_{\lambda\in\hat{\Lambda}(\alpha_0)} S(\alpha_0,\lambda)$ because the minimization problem is solved over the sieve space. This requires a modification of the proof.

Choose $k_{1n}^{-\mu_1}\sqrt{k_n} + \sqrt{k_n/n} = o(\delta_n)$ and let $\Lambda_n = \{\lambda : \|\lambda\| \le \delta_n\}$. Let $\bar{\lambda} = -\delta_n\hat{g}(\hat{\alpha}_n)/\|\hat{g}(\hat{\alpha}_n)\|$. Then $\bar{\lambda}'\hat{g}(\hat{\alpha}_n) = -\delta_n\|\hat{g}(\hat{\alpha}_n)\|$ and $\bar{\lambda}\in\Lambda_n$. By Lemma A.12 of DIN (2003) and definition of $\hat{\alpha}_n$, a Taylor expansion yields

$$\delta_n\|\hat{g}(\hat{\alpha}_n)\| - C\delta_n^2 \le \hat{S}(\hat{\alpha}_n,\bar{\lambda}) \le \sup_{\lambda\in\hat{\Lambda}(\hat{\alpha}_n)}\hat{S}(\hat{\alpha}_n,\lambda) \le \sup_{\lambda\in\hat{\Lambda}(\alpha_{n0})}\hat{S}(\alpha_{n0},\lambda). \tag{A.2}$$

Then, it follows from Lemma A.3 that $\sup_{\lambda \in \hat{\Lambda}(\alpha_{n0})} \hat{S}(\alpha_{n0}, \lambda) = o_p(\delta_n^2)$. Thus, we obtain $\delta_n \|\hat{g}(\hat{\alpha}_n)\| - C\delta_n^2 \leq o_p(\delta_n^2)$, and hence $\|\hat{g}(\hat{\alpha}_n)\| = O_p(\delta_n)$. ■

Proof of Theorem 3.1 Let $W = \left(n^{-1} \sum_{i=1}^n I \otimes p_i p_i'\right)^{-1}$. We define

$$\begin{aligned} \hat{R}(\alpha) &= \hat{g}(\alpha)' W \hat{g}(\alpha), \\ R(\alpha) &= E\left[E[\rho(Z,\alpha)|X]' E[\rho(Z,\alpha)|X]\right]. \end{aligned}$$

By Assumption 3.3, we have

$$R(\alpha) = E\left[E[\rho(Z,\alpha)|X]' E[\rho(Z,\alpha)|X]\right] > 0 = R(\alpha_0)$$

for all $\alpha \neq \alpha_0$. Also, Corollary 4.2 of Newey (1991) implies that $R(\alpha)$ is continuous and $\sup_{\alpha \in \mathcal{A}} |\hat{R}(\alpha) - R(\alpha)| \xrightarrow{p} 0$. Thus, by Lemma A.1 of DIN (2003), it suffices to show that $\hat{R}(\hat{\alpha}_n) \xrightarrow{p} 0$. Similarly to Lemma A.1, we can obtain $\lambda_{\min}(W^{-1}) \geq C$ w.p.a.1. Thus it follows from Lemma A.4 that

$$\hat{g}(\hat{\alpha}_n)' W \hat{g}(\hat{\alpha}_n) \leq C \|\hat{g}(\hat{\alpha}_n)\|^2 = O_p(\delta_n^2)$$

and the desired result follows. ■

A.2 Rate of convergence

Let $Q_i = I \otimes p_i'$ and $Q = (Q_1', \ldots, Q_n')'$. Also, let $\rho(\alpha) = (\rho(Z_1,\alpha)', \ldots, \rho(Z_n,\alpha)')'$. Denote a sieve estimator of $m(X_i,\alpha) = E[\rho(Z_i,\alpha)|X_i]$ by

$$\hat{m}(X_i,\alpha) = Q_i (Q'Q)^{-1} Q' \rho(\alpha),$$

Also, let $\mathcal{A}_{0n} = \{\alpha \in \mathcal{A}_n : \|\alpha - \alpha_0\|_s = o(1)\}$ and $\eta_n = o(n^{-\tau})$ with $\tau \leq 1/4$.

Lemma A.5 *Suppose that Assumptions 3.1-3.4, 3.6-3.7, and 3.10-3.11 hold. Then we have (i)* $\|\hat{g}(\alpha)\| = o_p(1)$ *uniformly over* $\alpha \in \mathcal{A}_{0n}$*; (ii)* $\|\hat{g}(\alpha)\| = o_p(\eta_n)$ *uniformly over* $\alpha \in \mathcal{A}_{0n}$ *with* $\|\alpha - \alpha_0\|_w = o(\eta_n)$.

Proof. By using the similar argument as in the proof of Lemma A.1, we can show that $\lambda_{\min}(W) \geq C$ w.p.a.1. Hence

$$C\|\hat{g}(\alpha)\|^2 \leq \hat{g}(\alpha)' W \hat{g}(\alpha) = \frac{1}{n} \rho(\alpha)' Q (Q'Q)^{-1} Q' \rho(\alpha) = \frac{1}{n} \sum_{i=1}^n \|\hat{m}(X_i,\alpha)\|^2.$$

Also, we have

$$\frac{1}{n}\sum_{i=1}^{n}\|m(X_i,\alpha)\|^2 \leq \|\alpha-\alpha_0\|_s^{2\kappa}\frac{1}{n}\sum_{i=1}^{n}E[c(Z_i)|X_i]^2.$$

Then Corollary A.1 (i) of AC (2003) implies $\|\hat{g}(\alpha)\| = o_p(1)$ uniformly over $\alpha \in \mathcal{A}_{0n}$. Moreover, Assumption 3.11 implies that $E[\|m(X,\alpha)\|^2]$ and $\|\alpha-\alpha_0\|_w^2$ are equivalent. Then Corollary A.2 (i) of AC (2003) implies $\|\hat{g}(\alpha)\| = o_p(\eta_n)$ uniformly over $\alpha \in \mathcal{A}_{0n}$ with $\|\alpha-\alpha_0\|_w = o(\eta_n)$. ■

Lemma A.6 *Suppose that Assumptions 3.1-3.4, 3.6-3.7, 3.8 (ii), (iv), and 3.10-3.11 hold. Let* $t(\alpha) = -\left(n^{-1}\sum_{i=1}^{n} g_i(\alpha_0)g_i(\alpha_0)'\right)^{-1}\hat{g}(\alpha)$. *Then for any* $\eta_{0n} = o(n^{-1/4})$,

$$\max_{1\leq i\leq n}\sup_{\alpha\in\mathcal{A}_{0n}}|\eta_{0n}t(\alpha)'g_i(\alpha)| \xrightarrow{p} 0.$$

Proof. By Assumption 3.8 (ii) and Lemma A.1, we have $\lambda_{\min}(n^{-1}\sum_{i=1}^{n} g_i(\alpha_0)g_i(\alpha_0)') > C$ w.p.a.1. Thus, it follows from Lemma A.5 that

$$\|t(\alpha)\|^2 = \hat{g}(\alpha)'\left(\frac{1}{n}\sum_{i=1}^{n} g_i(\alpha_0)g_i(\alpha_0)'\right)^{-2}\hat{g}(\alpha) \leq C\|\hat{g}(\alpha)\|^2 = o_p(1)$$

uniformly over $\alpha \in \mathcal{A}_{0n}$. Also, we have $\max_{1\leq i\leq n}\sup_{\alpha\in\mathcal{A}}\|\rho(Z_i,\alpha)\| = O_p(n^{1/m})$ by Assumption 3.8 (ii) and the Markov inequality. Therefore, by Assumption 3.8 (iv), we obtain

$$\begin{aligned}\max_{1\leq i\leq n}\sup_{\alpha\in\mathcal{A}_{0n}}|\eta_{0n}t(\alpha)'g_i(\alpha)| &\leq \eta_{0n}\sup_{\alpha\in\mathcal{A}_{0n}}\|t(\alpha)\|\max_{1\leq i\leq n}\sup_{\alpha\in\mathcal{A}_{0n}}\|\rho(Z_i,\alpha)\|\zeta(k_n)\\ &= o(n^{-1/4})o_p(n^{1/m})\zeta(k_n) = o_p(1),\end{aligned}$$

and hence the desired result follows. ■

Let us define $\hat{\lambda}_n(\alpha_{n0})$ as $\hat{S}(\alpha_{n0},\hat{\lambda}_n(\alpha_{n0})) = \sup_{\lambda\in\hat{\Lambda}(\alpha_{n0})}\hat{S}(\alpha_{n0},\lambda)$.

Lemma A.7 *Suppose that Assumptions 3.1-3.7, 3.8 (i)-(iv), and 3.10-3.11 hold. Then we have* $\|\hat{\lambda}_n(\alpha_{n0})\| = o_p(n^{-1/4})$.

Proof. The proof is similar to that of Lemma A.3. By Assumption 3.5 (iii) and Lemma A.5, we have $\|\hat{g}(\alpha_{n0})\| = o_p(n^{-1/4})$. Let $\bar{\lambda} = \arg\max_{\lambda\in\Lambda_n}\hat{S}(\alpha_{n0},\lambda)$. Then by Lemma A.2, we have $\max_{1\leq i\leq n}\sup_{\lambda\in\Lambda_n}|\lambda' g_i(\alpha_{n0})| = o_p(1)$ and that $\bar{\lambda}$ exists w.p.a.1. A Taylor expansion yields

$$0 = \hat{S}(\alpha_{n0},0) \leq \hat{S}(\alpha_{n0},\bar{\lambda}) \leq \|\bar{\lambda}\|\|\hat{g}(\alpha_{n0})\| - C\|\bar{\lambda}\|^2.$$

Thus we have $\|\bar{\lambda}\| = o_p(n^{-1/4})$. Also, by Assumption 3.8 (iv), $\|\bar{\lambda}\| < \delta_n$ w.p.a.1. Hence we have $\bar{\lambda} = \hat{\lambda}_n(\alpha_{n0})$ and the result follows. ■

Define $\psi(X,\alpha) = \Sigma(X,\alpha)^{-1/2}m(X,\alpha)$. Let $Q(X_i,\alpha) = \Sigma(X_i,\alpha)^{1/2} \otimes p_i'$ and $Q(\alpha) = (Q(X_1,\alpha)', \ldots, Q(X_n,\alpha)')'$. We define the following sieve estimator for $\psi(X,\alpha)$:

$$\hat{\psi}(X_i,\alpha) = Q(X_i,\alpha)(Q(\alpha)'Q(\alpha))^{-1}\sum_{j=1}^{n} Q(X_j,\alpha)'\Sigma(X_j,\alpha)^{-1/2}\rho(Z_j,\alpha).$$

Lemma A.8 *Suppose that Assumptions 3.1-3.2, 3.4, 3.6-3.7, 3.8 (ii), and 3.9-3.10 hold. Then we have* $n^{-1}\sum_{i=1}^{n}\|\hat{\psi}(X_i,\alpha) - \psi(X_i,\alpha)\|^2 = o_p(n^{-1/2})$ *uniformly over* $\alpha \in \mathcal{A}_{0n}$.

Proof. We modify Lemma A.1 of AC (2003) with $\delta_{1n} = \delta_{2n} = o(n^{-1/4})$. We replace $p^{k_n}(X)$ in their lemma with $Q(X,\alpha)$. The main difference is that $Q(X,\alpha)$ depends on α while $p^{k_n}(X)$ does not, which produces some extra terms that do not appear in the proof of AC (2003).

Let $N(\epsilon, \mathcal{A}_{0n}, \|\cdot\|_s)$ be the minimal number of ϵ-radius covering balls of $\mathcal{A}_{0n}$ under the metric $\|\cdot\|_s$. Also, let $\xi(k_n) = \sup_{X\in\mathcal{X}}\|\partial p^{k_n}(X)/\partial X\|$. Define $\epsilon_i(\alpha) = \Sigma(X_i,\alpha)^{-1/2}[\rho(Z_i,\alpha) - E[\rho(Z_i,\alpha)|X_i]]$ and $\epsilon(\alpha) = (\epsilon_1(\alpha)', \ldots, \epsilon_n(\alpha)')'$.

First we show that $\left\|Q(X,\alpha)(Q(\alpha)'Q(\alpha))^{-1}Q(\alpha)'\epsilon(\alpha)\right\| = o_p(n^{-1/4})$ uniformly over $(X,\alpha) \in \mathcal{X}\times\mathcal{A}_{0n}$. Let $\mathcal{W}_n = \mathcal{X}\times\mathcal{A}_{0n}$. For any pair $(X^1,\alpha^1) \in \mathcal{W}_n$ and $(X^2,\alpha^2) \in \mathcal{W}_n$,

$$\begin{aligned}
& \left\|Q(X^1,\alpha^1)(Q(\alpha^1)'Q(\alpha^1))^{-1}Q(\alpha^1)'\epsilon(\alpha^1) - Q(X^2,\alpha^2)(Q(\alpha^2)'Q(\alpha^2))^{-1}Q(\alpha^2)'\epsilon(\alpha^2)\right\| \\
\leq\ & \left\|\left(Q(X^1,\alpha^1) - Q(X^1,\alpha^2)\right)(Q(\alpha^1)'Q(\alpha^1))^{-1}Q(\alpha^1)'\epsilon(\alpha^1))\right\| \\
& + \left\|\left(Q(X^1,\alpha^2) - Q(X^2,\alpha^2)\right)(Q(\alpha^1)'Q(\alpha^1))^{-1}Q(\alpha^1)'\epsilon(\alpha^1))\right\| \\
& + \left\|Q(X^2,\alpha^2)(Q(\alpha^2)'Q(\alpha^2))^{-1}[Q(\alpha^2)'Q(\alpha^2) - Q(\alpha^1)'Q(\alpha^1)]\right. \\
& \quad \left.\times (Q(\alpha^1)'Q(\alpha^1))^{-1}Q(\alpha^1)'\epsilon(\alpha^1))\right\| \\
& + \left\|Q(X^2,\alpha^2)(Q(\alpha^2)'Q(\alpha^2))^{-1}\left[Q(\alpha^1)'(\epsilon(\alpha^1) - \epsilon(\alpha^2))\right]\right\| \\
& + \left\|Q(X^2,\alpha^2)(Q(\alpha^2)'Q(\alpha^2))^{-1}\left[(Q(\alpha^1) - Q(\alpha^2))'\epsilon(\alpha^2)\right]\right\|.
\end{aligned}$$

By Assumption 3.9 (i),

$$\begin{aligned}
\left\|Q(X^1,\alpha^1) - Q(X^1,\alpha^2)\right\|^2 &\leq \left\|\Sigma(X^1,\alpha^1)^{1/2} - \Sigma(X^1,\alpha^2)^{1/2}\right\|^2\zeta(k_n)^2 \\
&\leq C\|\alpha^1 - \alpha^2\|_s^{2\kappa}\zeta(k_n)^2.
\end{aligned}$$

Also by Assumption 3.4,

$$\begin{aligned}\left\|Q(X^1,\alpha^2)-Q(X^2,\alpha^2)\right\|^2 &\le \sup_{\alpha\in\mathcal{A}_{0n},X\in\mathcal{X}}\left\|\Sigma(X,\alpha)^{1/2}\right\|^2\|X^1-X^2\|^2\xi(k_n)^2\\ &\le C\|X^1-X^2\|^2\xi(k_n)^2.\end{aligned}$$

It follows from Assumption 3.6 (iii) and law of large numbers that $n^{-1}\left\|\epsilon(\alpha^1)'\epsilon(\alpha^1)\right\|^2=O_p(1)$. Also, by Assumption 3.8 (ii) and Lemma A.1, $\lambda_{\min}(Q(\alpha)'Q(\alpha)/n)>C$ w.p.a.1 for $\alpha\in\mathcal{A}_{0n}$. Therefore, we have

$$\begin{aligned}&\left\|(Q(\alpha^1)'Q(\alpha^1))^{-1}Q(\alpha^1)'\epsilon(\alpha^1)\right\|^2\\ &= \operatorname{tr}\left(\epsilon(\alpha^1)'Q(\alpha^1)(Q(\alpha^1)'Q(\alpha^1)/n)^{-1}(Q(\alpha^1)'Q(\alpha^1))^{-1}Q(\alpha^1)'\epsilon(\alpha^1)/n\right)\\ &\le C\operatorname{tr}(\epsilon(\alpha^1)'\epsilon(\alpha^1)/n)=O_p(1).\end{aligned}$$

Then we have

$$\begin{aligned}&P\left(\sup_{X\in\mathcal{X},\alpha^1,\alpha^2\in\mathcal{A}_{0n}}\frac{\left\|\left(Q(X,\alpha^1)-Q(X,\alpha^2)\right)(Q(\alpha^1)'Q(\alpha^1))^{-1}Q(\alpha^1)'\epsilon(\alpha^1))\right\|}{\|\alpha^1-\alpha^2\|_s^\kappa}>C\zeta(k_n)\right)<\eta,\\ &P\Bigg(\sup_{X^1,X^2\in\mathcal{X},\alpha^1,\alpha^2\in\mathcal{A}_{0n}}\left\|\left(Q(X^1,\alpha^2)-Q(X^2,\alpha^2)\right)\right.\\ &\qquad\left.\times(Q(\alpha^1)'Q(\alpha^1))^{-1}Q(\alpha^1)'\epsilon(\alpha^1))\right\|\Big/\|\alpha^1-\alpha^2\|_s^\kappa>C\xi(k_n)\Bigg)<\eta\end{aligned}$$

for any small $\eta>0$ and sufficiently large n. Also,

$$\begin{aligned}\left\|\frac{1}{n}Q(\alpha^2)'Q(\alpha^2)-\frac{1}{n}Q(\alpha^1)'Q(\alpha^1)\right\| &\le \frac{1}{n}\sum_{i=1}^n\left\|\Sigma(X_i,\alpha^1)-\Sigma(X_i,\alpha^2)\right\|\|p_i\|^2\\ &\le C\|\alpha^1-\alpha^2\|_s^\kappa k_n.\end{aligned}$$

Hence, for sufficiently large n

$$\begin{aligned}&P\Bigg(\sup_{X\in\mathcal{X},\alpha^1,\alpha^2\in\mathcal{A}_{0n}}\left\|Q(X,\alpha^2)(Q(\alpha^2)'Q(\alpha^2))^{-1}[Q(\alpha^2)'Q(\alpha^2)-Q(\alpha^1)'Q(\alpha^1)]\right.\\ &\qquad\left.\times(Q(\alpha^1)'Q(\alpha^1))^{-1}Q(\alpha^1)'\epsilon(\alpha^1))\right\|\Big/\|\alpha^1-\alpha^2\|_s^\kappa>C\zeta(k_n)k_n\Bigg)<\eta.\end{aligned}$$

Moreover, Assumption 3.6 (ii) implies that

$$\left\|\frac{1}{n}Q(\alpha^1)'(\epsilon(\alpha^1)-\epsilon(\alpha^2))\right\|\le C\|\alpha^1-\alpha^2\|_s^\kappa\zeta(k_n)\sqrt{\frac{1}{n}\sum_{i=1}^n c_2(Z_i)^2},$$

where $\sum_{i=1}^n c_2(Z_i)/n=O_p(1)$ by the weak law of large numbers. Also,

$$\left\|\frac{1}{n}(Q(\alpha^1)-Q(\alpha^2))'\epsilon(\alpha^2)\right\|\le C\|\alpha^1-\alpha^2\|_s^\kappa\zeta(k_n)\sqrt{\frac{1}{n}\operatorname{tr}\left(\epsilon(\alpha^2)'\epsilon(\alpha^2)\right)}.$$

Therefore, for sufficiently large n, we have

$$P\left(\sup_{X\in\mathcal{X},\alpha^1,\alpha^2\in\mathcal{A}_{0n}}\frac{\left\|Q(X,\alpha^2)(Q(\alpha^2)'Q(\alpha^2))^{-1}\left[Q(\alpha^1)'(\epsilon(\alpha^1)-\epsilon(\alpha^2))\right]\right\|}{\|\alpha^1-\alpha^2\|_s^\kappa}>C\zeta(k_n)^2\right)<\eta$$

and

$$P\left(\sup_{X\in\mathcal{X},\alpha^1,\alpha^2\in\mathcal{A}_{0n}}\frac{\left\|Q(X,\alpha^2)(Q(\alpha^2)'Q(\alpha^2))^{-1}\left[(Q(\alpha^1)-Q(\alpha^2))'\epsilon(\alpha^2)\right]\right\|}{\|\alpha^1-\alpha^2\|_s^\kappa}>C\zeta(k_n)^2\right)<\eta.$$

Similarly to the proof of AC (2003), for any small ϵ, we divide $\mathcal{W}_n$ into b_n mutually exclusive subsets $\mathcal{W}_{nm}$, $m=1,2,\dots b_n$, where $(X^1,\alpha^1)\in\mathcal{W}_{nm}$ and $(X^2,\alpha^2)\in\mathcal{W}_{nm}$ imply $\left\|X^1-X^2\right\|\le\epsilon n^{-1/4}/(C\xi(k_n))$ and $\|\alpha^1-\alpha^2\|_s^\kappa\le\epsilon n^{-1/4}/(C\zeta(k_n)k_n)$. Then w.p.a.1, we have

$$\left\|Q(X^1,\alpha^1)(Q(\alpha^1)'Q(\alpha^1))^{-1}Q(\alpha^1)'\epsilon(\alpha^1)-Q(X^2,\alpha^2)(Q(\alpha^2)'Q(\alpha^2))^{-1}Q(\alpha^2)'\epsilon(\alpha^2)\right\|\le 2\epsilon n^{-1/4}.$$

For any (X,α), there exists an m such that $\|X-X^m\|\le\epsilon n^{-1/4}/(C\xi(k_n))$ and $\|\alpha-\alpha^m\|_s^\kappa\le\epsilon n^{-1/4}/(C\zeta(k_n)k_n)$. Thus, w.p.a.1,

$$\begin{aligned}&\sup_{(X,\alpha)\in\mathcal{X}\times\mathcal{A}_{0n}}\left\|Q(X,\alpha)(Q(\alpha)'Q(\alpha))^{-1}Q(\alpha)'\epsilon(\alpha))\right\|\\ &\le\ 2\epsilon n^{-1/4}+\max_m\left\|Q(X^m,\alpha^m)(Q(\alpha^m)'Q(\alpha^m))^{-1}Q(\alpha^m)'\epsilon(\alpha^m))\right\|.\end{aligned}$$

Hence we have

$$\begin{aligned}&P\left(\sup_{(X,\alpha)\in\mathcal{X}\times\mathcal{A}_{0n}}\left\|Q(X,\alpha)(Q(\alpha)'Q(\alpha))^{-1}Q(\alpha)'\epsilon(\alpha))\right\|>4\epsilon n^{-1/4}\right)\\ &\le\ 5\eta+P\left(\max_m\left\|Q(X,\alpha)(Q(\alpha)'Q(\alpha))^{-1}Q(\alpha)'\epsilon(\alpha))\right\|>2\epsilon n^{-1/4}\right).\end{aligned}$$

By a slight modification of the proof of AC (2003), we can show that the second term of the right hand side can be arbitrarily small if

$$\frac{n^{1/2}}{\zeta(k_n)^2}-\ln b_n\to\infty. \tag{A.3}$$

Since $\mathcal{X}$ is compact, we have

$$b_n=O\left(\left(\frac{n^{-1/4}}{\xi(k_n)}\right)^{-d_x}\times N\left(\left\{\frac{n^{-1/4}}{\zeta(k_n)k_n}\right\}^{1/\kappa},\mathcal{A}_{0n},\|\cdot\|_s\right)\right).$$

Therefore, (A.3) holds if

$$\left\{\ln(n^{1/4}\xi(k_n))^{d_x}+\ln\left[N\left((n^{1/4}\zeta(k_n)k_n)^{-1/\kappa},\mathcal{A}_n,\|\cdot\|_s\right)\right]\right\}\zeta(k_n)^2n^{-1/2}=o(1),$$

which is implied by Assumptions 3.7 (ii) and 3.10. Hence we have

$$\sup_{(X,\alpha)\in(\mathcal{X},\mathcal{A}_{0n})}\left\|Q(\alpha,X)(Q(\alpha)'Q(\alpha))^{-1}Q(\alpha)'\epsilon(\alpha))\right\|=o_p(n^{-1/4}). \tag{A.4}$$

Next, by Assumptions 3.2 (iii), 3.6 (iv), and 3.9 (ii), there exists $\Pi_{k_n}(\alpha)$ such that

$$\psi(X_i,\alpha) = E[\Sigma(X_i,\alpha)^{-1/2}\rho(Z_i,\alpha)|X_i] = \Pi_{k_n}(\alpha)p_i + o_p(n^{-1/4})$$

for all $X \in \mathcal{X}$ and $\alpha \in \mathcal{A}_n$. Thus it follows that

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^{n}\left\| Q(X_i,\alpha)(Q(\alpha)'Q(\alpha))^{-1}\sum_{j=1}^{n}Q(X_j,\alpha)'\psi(X_j,\alpha) - \psi(X_i,\alpha)\right\|^2 \\ = & \frac{1}{n}\sum_{i=1}^{n}\|\psi(X_i,\alpha) - \Pi_{k_n}(\alpha)p_i\|^2 + o_p(n^{-1/2}) = o_p(n^{-1/2}) \end{aligned} \tag{A.5}$$

uniformly over $\alpha \in \mathcal{A}_{0n}$.

The result follows from (A.4) and (A.5). ■

Lemma A.9 *Suppose that Assumptions 3.1-3.3, 3.4, 3.6 (iii)-(iv), 3.9 (ii), and 3.11 hold. Then* $n^{-1}\sum_{i=1}^{n}\|\psi(X_i,\alpha)\|^2 - E\left[\|\psi(X,\alpha)\|^2\right] = o_p(n^{-1/2})$ *uniformly over* $\alpha \in \mathcal{A}_{0n}$ *with* $\|\alpha - \alpha_0\|_w = o(1)$.

Proof. The result can be obtained by replacing $m(X,\alpha)$ with $\psi(X,\alpha)$ in Corollary A.2 (i) of AC (2003). ■

Lemma A.10 *Suppose that Assumptions 3.1-3.4, 3.6-3.7, 3.8 (ii), and 3.9-3.11 hold. Then we have* $n^{-1}\sum_{i=1}^{n}\|\hat{\psi}(X_i,\alpha)\|^2 = o_p(\eta_n^2)$ *uniformly over* $\alpha \in \mathcal{A}_{0n}$ *with* $\|\alpha - \alpha_0\|_w \le \eta_n$.

Proof. Assumptions 3.4 and 3.11 imply that $E[\|\psi(X,\alpha)\|^2]$ is equivalent to $\|\alpha-\alpha_0\|_w^2$. Thus the result follows from Lemmas A.8 and A.9. ■

Define $\psi_0(X,\alpha) \equiv \Sigma(X,\alpha_0)^{-1/2}m(X,\alpha)$ and denote

$$\tilde{\psi}_0(X_i,\alpha) \equiv Q(X_i,\alpha_0)(Q(\alpha_0)'Q(\alpha_0))^{-1}\sum_{j=1}^{n}Q(X_j,\alpha_0)'\Sigma(X_j,\alpha_0)^{-1/2}\rho(Z_j,\alpha).$$

Lemma A.11 *Suppose that Assumptions 3.1-3.4, 3.6-3.7, 3.8 (ii), and 3.9-3.11 hold. Then we have (i)* $n^{-1}\sum_{i=1}^{n}\|\tilde{\psi}_0(X_i,\alpha) - \psi_0(X_i,\alpha)\|^2 = o_p(n^{-1/2})$ *uniformly over* $\alpha \in \mathcal{A}_{0n}$*; (ii)* $n^{-1}\sum_{i=1}^{n}\|\tilde{\psi}_0(X_i,\alpha)\|^2 = o_p(\eta_n^2)$ *uniformly over* $\alpha \in \mathcal{A}_{0n}$ *with* $\|\alpha - \alpha_0\|_w \le \eta_n$.

Proof. The results follow immediately from Lemmas A.8 and A.10. ■

Hereafter, denote

$$\hat{L}_n(\alpha) = -\sup_{\lambda \in \hat{\Lambda}(\alpha)} \hat{S}(\alpha, \lambda),$$
$$\bar{L}_n(\alpha) = -\hat{g}(\alpha)' \left(\frac{1}{n} \sum_{i=1}^{n} g_i(\alpha_0) g_i(\alpha_0)' \right)^{-1} \hat{g}(\alpha),$$
$$L_n(\alpha) = -\frac{1}{n} \sum_{i=1}^{n} \psi_0(X_i, \alpha)' \psi_0(X_i, \alpha).$$

Lemma A.12 *Suppose that Assumptions 3.1-3.4, 3.6-3.7, 3.8 (v), and 3.9-3.11 hold. Then we have (i) $\bar{L}_n(\alpha) - L_n(\alpha) = o_p(n^{-1/4})$ uniformly over $\alpha \in \mathcal{A}_{0n}$; (ii) $\bar{L}_n(\alpha) - L_n(\alpha) = o_p(n^{-1/4}\eta_n)$ uniformly over $\alpha \in \mathcal{A}_{0n}$ with $\|\alpha - \alpha_0\|_w = o(\eta_n)$.*

Proof. By Assumption 3.8 (v), we can choose $\zeta(k_n)\sqrt{k_n/n} = o_p(n^{-1/4})$. Then by Lemma A.1, we have

$$\left\| \frac{1}{n} \sum_{i=1}^{n} g_i(\alpha_0) g_i(\alpha_0)' - \frac{1}{n} \sum_{i=1}^{n} \Sigma(X_i, \alpha_0) \otimes p_i p_i' \right\| = o_p(n^{-1/4}).$$

Also, we have $\lambda_{\min}(n^{-1} \sum_{i=1}^{n} g_i(\alpha_0) g_i(\alpha_0)) > C$ and $\lambda_{\min}(n^{-1} \sum_{i=1}^{n} \Sigma(X_i, \alpha_0) \otimes p_i p_i') > C$ w.p.a.1. Thus we obtain

$$\begin{aligned} & \left| \hat{g}(\alpha)' \left(\frac{1}{n} \sum_{i=1}^{n} g_i(\alpha_0) g_i(\alpha_0)' \right)^{-1} \hat{g}(\alpha) - \hat{g}(\alpha)' \left(\frac{1}{n} \sum_{i=1}^{n} \Sigma(X_i, \alpha_0) \otimes p_i p_i' \right)^{-1} \hat{g}(\alpha) \right| \\ \leq & \ C \|\hat{g}(\alpha)\|^2 \left\| \frac{1}{n} \sum_{i=1}^{n} g_i(\alpha_0) g_i(\alpha_0)' - \frac{1}{n} \sum_{i=1}^{n} \Sigma(X_i, \alpha_0) \otimes p_i p_i' \right\| \\ = & \ O_p(\|\hat{g}(\alpha)\|^2) o_p(n^{-1/4}). \end{aligned}$$

Also, let $\hat{\rho}_i(\alpha) = \Sigma(X_i, \alpha_0)^{-1/2} \rho(Z_i, \alpha)$ and $\hat{\rho}(\alpha) = (\hat{\rho}_1(\alpha)', \ldots, \hat{\rho}_n(\alpha)')'$. Then we have

$$\hat{g}(\alpha)' \left(\frac{1}{n} \sum_{i=1}^{n} \Sigma(X_i, \alpha_0) \otimes p_i p_i' \right)^{-1} \hat{g}(\alpha) = \frac{1}{n} \sum_{i=1}^{n} \tilde{\psi}_0(X_i, \alpha)' \tilde{\psi}_0(X_i, \alpha).$$

Thus it follows that

$$
\begin{aligned}
&\left|\bar{L}_n(\alpha) - L_n(\alpha)\right| \\
&= \left|\frac{1}{n}\sum_{i=1}^n \tilde{\psi}_0(X_i,\alpha)'\tilde{\psi}_0(X_i,\alpha) - \frac{1}{n}\sum_{i=1}^n \psi_0(X_i,\alpha)'\psi_0(X_i,\alpha)\right| + o_p(n^{-1/4})O_p(\|\hat{g}(\alpha)\|^2) \\
&\le \left|\frac{1}{n}\sum_{i=1}^n \left(\tilde{\psi}_0(X_i,\alpha) - \psi_0(X_i,\alpha)\right)'\tilde{\psi}_0(X_i,\alpha)\right| \\
&\quad + \left|\frac{1}{n}\sum_{i=1}^n \psi_0(X_i,\alpha)'\left(\tilde{\psi}_0(X_i,\alpha) - \psi_0(X_i,\alpha)\right)\right| + o_p(n^{-1/4})O_p(\|\hat{g}(\alpha)\|^2) \\
&\le \left(\frac{1}{n}\sum_{i=1}^n \left\|\tilde{\psi}_0(X_i,\alpha) - \psi_0(X_i,\alpha)\right\|^2\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^n \|\tilde{\psi}_0(X_i,\alpha)\|^2\right)^{1/2} \\
&\quad + \left(\frac{1}{n}\sum_{i=1}^n \left\|\tilde{\psi}_0(X_i,\alpha) - \psi_0(X_i,\alpha)\right\|^2\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^n \|\psi_0(X_i,\alpha)\|^2\right)^{1/2} \\
&\quad + o_p(n^{-1/4})O_p(\|\hat{g}(\alpha)\|^2).
\end{aligned}
$$

Therefore the result follows from Lemmas A.5 and A.11. ■

Proof of Theorem 3.2 Let $0 < \eta_{0n} = o(n^{-1/4})$. Define $\hat{L}_{0n}(\alpha) = -n^{-1}\sum_{i=1}^n s(\eta_{0n}t(\alpha)'g_i(\alpha))$. By Lemma A.6, for $\alpha \in \mathcal{A}_{0n}$, we have

$$
\begin{aligned}
\hat{L}_{0n}(\alpha) &= \eta_{0n}t(\alpha)'\hat{g}(\alpha) - \frac{\eta_{0n}^2}{2}t(\alpha)'\left(\frac{1}{n}\sum_{i=1}^n s_2(\eta_{0n}t'g_i(\alpha))g_i(\alpha)g_i(\alpha)'\right)t(\alpha) \\
&= \eta_{0n}\bar{L}_n(\alpha) + o_p(n^{-1/2}).
\end{aligned} \tag{A.6}
$$

Also, by Lemma A.7, a Taylor expansion yields

$$
\begin{aligned}
\hat{L}_n(\alpha_{n0}) &= -\frac{1}{n}\sum_{i=1}^n s(\lambda(\alpha_{n0})'g_i(\alpha_{n0})) \\
&= \hat{\lambda}_n(\alpha_{n0})'\hat{g}(\alpha_{n0}) - \frac{1}{2}\hat{\lambda}_n(\alpha_{n0})'\left(\frac{1}{n}\sum_{i=1}^n s_2(\tilde{\lambda}'g_i(\alpha_{n0}))g_i(\alpha_{n0})g_i(\alpha_{n0})'\right)\hat{\lambda}_n(\alpha_{n0}),
\end{aligned}
$$

for some $\tilde{\lambda}$ between 0 and $\hat{\lambda}_n(\alpha_{n0})$. Hence we have

$$
\left|\hat{L}_n(\alpha_{n0})\right| \le \left\|\hat{\lambda}_n(\alpha_{n0})\right\|\|\hat{g}(\alpha_{n0})\| + C\left\|\hat{\lambda}_n(\alpha_{n0})\right\|^2 = o_p(n^{-1/2}). \tag{A.7}
$$

Now we show that $\|\hat{\alpha}_n - \alpha_0\|_w = o_p(n^{-1/8})$. Let $\delta_{0n} = 2\sqrt{\eta_{0n}} = o(n^{-1/8})$. By the definition of $\hat{L}_n(\alpha)$, $\hat{L}_{0n}(\alpha) \ge \hat{L}_n(\alpha)$ for all $\alpha \in \mathcal{A}_{0n}$. Therefore, by using similar set inclusion relations as in the proof of Theorem 3.2 of Otsu (2011), we have

$$\begin{aligned}
&P\left(\|\hat{\alpha}_n-\alpha_0\|_w \geq C\delta_{0n}\right) \\
&\quad \leq P\left(\sup_{\|\hat{\alpha}_n-\alpha_0\|_w \geq C\delta_{0n}, \alpha\in\mathcal{A}_{0n}} \hat{L}_n(\alpha) \geq \hat{L}_n(\alpha_{n0})\right) \\
&\quad \leq P\left(\sup_{\|\hat{\alpha}_n-\alpha_0\|_w \geq C\delta_{0n}, \alpha\in\mathcal{A}_{0n}} \hat{L}_{0n}(\alpha) \geq \hat{L}_n(\alpha_{n0})\right) \\
&\quad \leq P\left(\left|\hat{L}_n(\alpha_{n0})-\eta_{0n}L_n(\alpha_{n0})\right| > \eta_{0n}^2\right) + P\left(\sup_{\alpha\in\mathcal{A}_{0n}}\left|\hat{L}_{0n}(\alpha)-\eta_{0n}L_n(\alpha)\right| > \eta_{0n}^2\right) \\
&\qquad + P\left(\sup_{\|\alpha-\alpha_0\|_w \geq C\delta_{0n}, \alpha\in\mathcal{A}_{0n}} \eta_{0n}L_n(\alpha) \geq \eta_{0n}L_n(\alpha_{n0}) - 2\eta_{0n}^2\right) \\
&\quad \equiv P_1+P_2+P_3, \quad \text{say.}
\end{aligned}$$

Since $n^{-1}\sum_{i=1}^n \|\psi_0(X_i,\alpha_{n0})\|^2 = o_p(n^{-1/2})$, it follows from (A.7) that

$$\begin{aligned}
\left|\hat{L}_n(\alpha_{n0})-\eta_{0n}L_n(\alpha_{n0})\right| &\leq \left\|\hat{\lambda}_n(\alpha_{n0})\right\| \|\hat{g}(\alpha_{n0})\| + C\left\|\hat{\lambda}_n(\alpha_{n0})\right\|^2 + \frac{\eta_{0n}}{n}\sum_{i=1}^n \|\psi_0(X_i,\alpha_{n0})\|^2 \\
&= o_p(n^{-1/2}) = o_p(\eta_{0n}^2),
\end{aligned}$$

which implies $P_1 \to 0$. Also, it follows from Lemma A.12 and (A.6) that

$$\begin{aligned}
\sup_{\alpha\in\mathcal{A}_{0n}}\left|\hat{L}_{0n}(\alpha)-\eta_{0n}L_n(\alpha)\right| &\leq \sup_{\alpha\in\mathcal{A}_{0n}}\left|\eta_{0n}\bar{L}_n(\alpha)-\eta_{0n}L_n(\alpha)\right| + o_p(n^{-1/2}) \\
&= o_p(n^{-1/2}) = o_p(\eta_{0n}^2).
\end{aligned}$$

Therefore, we obtain $P_2 \to 0$. Finally, using Theorem 1 of Shen and Wong (1994), we have $P_3 \to 0$. Therefore we obtain $\|\hat{\alpha}_n-\alpha_0\|_w = o_p(n^{-1/8})$.

We can refine the convergence rate by using the logic that is introduced by AC (2003) and adopted in Otsu (2011). Then we obtain $\|\hat{\alpha}_n-\alpha_0\|_w = o_p(n^{-1/8(1+1/2+1/4+\cdots)}) = o_p(n^{-1/4})$. ■

A.3 Asymptotic normality

Denote

$$\begin{aligned}
&\frac{d\psi(X_i,\alpha)}{d\alpha}[v_n^*] = \Sigma(X_i,\alpha)^{-1/2}\frac{dm(X_i,\alpha)}{d\alpha}[v_n^*] \\
&\frac{d\hat{\psi}(X_i,\alpha)}{d\alpha}[v_n^*] = Q(X_i,\alpha)(Q(\alpha)'Q(\alpha))^{-1}\sum_{j=1}^n Q(X_j,\alpha)'\Sigma(X_j,\alpha)^{-1/2}\frac{d\rho(Z_j,\alpha)}{d\alpha}[v_n^*].
\end{aligned}$$

Lemma A.13 *Suppose that Assumptions 3.1-3.2, 3.4, 3.7, 3.8 (ii), 3.9-3.10, and 4.1-4.4 hold.*

Then

$$\sup_{\alpha \in \mathcal{N}_{0n}} \frac{1}{n} \sum_{i=1}^{n} \left\| \frac{d\hat{\psi}(X_i, \alpha)}{d\alpha}[v_n^*] - \frac{d\psi(X_i, \alpha)}{d\alpha}[v_n^*] \right\|^2 = o_p(n^{-1/2}),$$

$$\sup_{\alpha \in \mathcal{N}_{0n}} \frac{1}{n} \sum_{i=1}^{n} \left\| \frac{d\psi(X_i, \alpha)}{d\alpha}[v_n^*] - \frac{d\psi(X_i, \alpha_0)}{d\alpha}[v_n^*] \right\|^2 = o_p(n^{-1/2}).$$

Proof. The first equation can be proved by replacing $\rho(Z_i, \alpha)$ with $(d\rho(Z_i, \alpha)/d\alpha)[v_n^*]$ in Lemma A.8. The proof of the second equality is almost the same as that of Corollary C.1 of AC (2003). ■

Denote

$$\frac{d^2\rho(Z_i, \alpha)}{d\alpha d\alpha}[v_n^*, v_n^*] = \left. \frac{d^2\rho(Z_i, \alpha + \tau v_n^*)}{d\tau^2} \right|_{\tau=0}$$

$$\frac{d^2\hat{\psi}(X_i, \alpha)}{d\alpha d\alpha}[v_n^*, v_n^*] = Q(X_i, \alpha)(Q(\alpha)'Q(\alpha))^{-1} \sum_{j=1}^{n} Q(X_j, \alpha)\Sigma(X_j, \alpha)^{-1/2} \frac{d^2\rho(Z_j, \alpha)}{d\alpha d\alpha}[v_n^*, v_n^*].$$

Lemma A.14 *Suppose that Assumptions 3.1-3.2, 3.4, 3.7, 3.8 (ii), 3.9-3.10, and 4.1-4.4 hold. Then*

$$\frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{d\hat{\psi}(X_i, \alpha)}{d\alpha}[v_n^*] \right\}' \hat{\psi}(X_i, \alpha) = \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{d\psi(X_i, \alpha)}{d\alpha}[v^*] \right\}' \hat{\psi}(X_i, \alpha) + o_p(n^{-1/2})$$

uniformly over $\alpha \in \mathcal{N}_{0n}$.

Proof. Observe that

$$\begin{aligned}
&\frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{d\hat{\psi}(X_i, \alpha)}{d\alpha}[v_n^*] \right\}' \hat{\psi}(X_i, \alpha) - \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{d\psi(X_i, \alpha)}{d\alpha}[v^*] \right\}' \hat{\psi}(X_i, \alpha) \\
&= \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{d\hat{\psi}(X_i, \alpha)}{d\alpha}[v_n^*] - \frac{d\psi(X_i, \alpha)}{d\alpha}[v_n^*] \right\}' \hat{\psi}(X_i, \alpha) \\
&\quad + \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{d\psi(X_i, \alpha)}{d\alpha}[v_n^* - v^*] \right\}' \hat{\psi}(X_i, \alpha).
\end{aligned}$$

Thus the result follows from Lemmas A.10, A.13, and Assumption 4.2. ■

Lemma A.15 *Suppose that Assumptions 3.1-3.3, 3.6 (iv), 3.9 (ii), 4.1 (i), and 4.2-4.5 hold. Then*

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{d\psi(X_i, \alpha)}{d\alpha}[v^*] \right\}' \hat{\psi}(X_i, \alpha) &= \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{d\psi(X_i, \alpha_0)}{d\alpha}[v^*] \right\}' \hat{\psi}(X_i, \alpha_0) \\
&\quad + \langle v^*, \alpha - \alpha_0 \rangle + o_p(n^{-1/2})
\end{aligned}$$

uniformly over $\alpha \in \mathcal{N}_{0n}$.

Proof. We modify the proof of Corollary C.3 (ii) in AC (2003). The main difference is that $\frac{d\psi(X,\alpha)}{d\alpha}[v^*]$ depend on α while $g(X, v^*)$ in AC (2003) does not. Define the following set of functions:

$$\mathcal{F} = \left\{ \left\{ \frac{d\psi(X,\alpha)}{d\alpha}[v^*] \right\}' \left(\hat{\psi}(X,\alpha) - \psi(X,\alpha) \right) : \alpha \in \mathcal{N}_{0n} \right\}.$$

Assumptions 3.6 (iv), 3.9 (ii), and 4.3 (ii) imply that $\mathcal{F}$ is a Donsker class. Also,

$$E\left[\left| \left\{ \frac{d\psi(X_i,\alpha)}{d\alpha}[v^*] \right\}' \left(\hat{\psi}(X_i,\alpha) - \psi(X_i,\alpha) \right) \right|^2 \right] = o_p(1)$$

uniformly over $\alpha \in \mathcal{N}_{0n}$. Thus, as in AC (2003), we have

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^{n} \left\{ \frac{d\psi(X_i,\alpha)}{d\alpha}[v^*] \right\}' \left(\hat{\psi}(X_i,\alpha) - \psi(X_i,\alpha) \right) \\ &= E\left[\left\{ \frac{d\psi(X_i,\alpha)}{d\alpha}[v^*] \right\}' \left(\hat{\psi}(X_i,\alpha) - \psi(X_i,\alpha) \right) \right] + o_p(n^{-1/2}) \end{aligned}$$

uniformly over $\alpha \in \mathcal{N}_{0n}$. Also,

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^{n} \left\{ \frac{d\psi(X_i,\alpha_0)}{d\alpha}[v^*] \right\}' \left(\hat{\psi}(X_i,\alpha_0) - \psi(X_i,\alpha_0) \right) \\ &= E\left[\left\{ \frac{d\psi(X_i,\alpha_0)}{d\alpha}[v^*] \right\}' \left(\hat{\psi}(X_i,\alpha_0) - \psi(X_i,\alpha_0) \right) \right] + o_p(n^{-1/2}). \end{aligned}$$

Hence we obtain

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^{n} \left\{ \left\{ \frac{d\psi(X_i,\alpha)}{d\alpha}[v^*] \right\}' \hat{\psi}(X_i,\alpha) - \left\{ \frac{d\psi(X_i,\alpha_0)}{d\alpha}[v^*] \right\}' \hat{\psi}(X_i,\alpha_0) \right\} \\ &= \frac{1}{n}\sum_{i=1}^{n} \left\{ \left\{ \frac{d\psi(X_i,\alpha)}{d\alpha}[v^*] \right\}' \psi(X_i,\alpha) - \left\{ \frac{d\psi(X_i,\alpha_0)}{d\alpha}[v^*] \right\}' \psi(X_i,\alpha_0) \right\} \\ &\quad + E\left[\left\{ \frac{d\psi(X_i,\alpha)}{d\alpha}[v^*] \right\}' \hat{\psi}(X_i,\alpha) - \left\{ \frac{d\psi(X_i,\alpha_0)}{d\alpha}[v^*] \right\}' \hat{\psi}(X_i,\alpha_0) \right] \\ &\quad - E\left[\left\{ \frac{d\psi(X_i,\alpha)}{d\alpha}[v^*] \right\}' \psi(X_i,\alpha) - \left\{ \frac{d\psi(X_i,\alpha_0)}{d\alpha}[v^*] \right\}' \psi(X_i,\alpha_0) \right] + o_p(n^{-1/2}). \end{aligned}$$

Note that

$$\begin{aligned} & E\left[\left\{ \frac{d\psi(X_i,\alpha)}{d\alpha}[v^*] \right\}' \hat{\psi}(X_i,\alpha) \right] \\ &= E\left[\left\{ Q(X_i,\alpha)(Q(\alpha)'Q(\alpha))^{-1} \sum_{j=1}^{n} Q(X_j,\alpha) \frac{d\psi(X_j,\alpha)}{d\alpha}[v^*] \right\}' \psi(X_i,\alpha) \right]. \end{aligned}$$

Also, by Assumptions 3.2 (iii), 3.6 (iv), and 3.9 (ii),

$$\left\| Q(X,\alpha)(Q(\alpha)'Q(\alpha))^{-1} \sum_{j=1}^{n} Q(X_j,\alpha)' \frac{d\psi(X_j,\alpha)}{d\alpha}[v^*] - \frac{d\psi(X,\alpha)}{d\alpha}[v^*] \right\| = o_p(n^{-1/4})$$

uniformly over $X \in \mathcal{X}$ and $\alpha \in \mathcal{N}_{0n}$. Hence, by the Cauchy-Schwarz inequality,

$$
\begin{aligned}
& E\left[\left\{Q(X_i,\alpha)(Q(\alpha)'Q(\alpha))^{-1}\sum_{j=1}^{n}Q(X_j,\alpha)'\frac{d\psi(X_j,\alpha)}{d\alpha}[v^*]\right\}'\psi(X_i,\alpha)\right] \\
& \quad -E\left[\left\{Q(X_i,\alpha_0)(Q(\alpha_0)'Q(\alpha_0))^{-1}\sum_{j=1}^{n}Q(X_j,\alpha_0)'\frac{d\psi(X_j,\alpha_0)}{d\alpha}[v^*]\right\}'\psi(X_i,\alpha_0)\right] \\
& \quad -E\left[\left\{\frac{d\psi(X_i,\alpha)}{d\alpha}[v^*]\right\}'\psi(X_i,\alpha)-\left\{\frac{d\psi(X_i,\alpha_0)}{d\alpha}[v^*]\right\}'\psi(X_i,\alpha_0)\right] \\
= {} & E\left[\left\{Q(X_i,\alpha)(Q(\alpha)'Q(\alpha))^{-1}\sum_{j=1}^{n}Q(X_j,\alpha)'\frac{d\psi(X_j,\alpha)}{d\alpha}[v^*]-\frac{d\psi(X_i,\alpha)}{d\alpha}[v^*]\right\}'\psi(X_i,\alpha)\right] \\
= {} & o_p(n^{-1/2})
\end{aligned}
$$

uniformly over $\alpha \in \mathcal{N}_{0n}$. Therefore, we obtain

$$
\begin{aligned}
& \frac{1}{n}\sum_{i=1}^{n}\left\{\left\{\frac{d\psi(X_i,\alpha)}{d\alpha}[v^*]\right\}'\hat{\psi}(X_i,\alpha)-\left\{\frac{d\psi(X_i,\alpha_0)}{d\alpha}[v^*]\right\}'\hat{\psi}(X_i,\alpha_0)\right\} \\
& \quad = \frac{1}{n}\sum_{i=1}^{n}\left\{\left\{\frac{d\psi(X_i,\alpha)}{d\alpha}[v^*]\right\}'\psi(X_i,\alpha)-\left\{\frac{d\psi(X_i,\alpha_0)}{d\alpha}[v^*]\right\}'\psi(X_i,\alpha_0)\right\}+o_p(n^{-1/2}).
\end{aligned}
$$

Now we consider the following class of functions:

$$
\mathcal{G}=\left\{\left\{\frac{d\psi(X,\alpha)}{d\alpha}[v^*]\right\}'\psi(X,\alpha):\alpha\in\mathcal{N}_{0n}\right\}.
$$

Again $\mathcal{G}$ is a Donsker class. Hence, we obtain

$$
\sup_{\alpha\in\mathcal{N}_{0n}}\left|\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{d\psi(X_i,\alpha)}{d\alpha}[v^*]\right\}'\psi(X_i,\alpha)-E\left[\left\{\frac{d\psi(X_i,\alpha)}{d\alpha}[v^*]\right\}'\psi(X_i,\alpha)\right]\right|=o_p(n^{-1/2}).
$$

Therefore by Assumptions 4.1 (ii), 4.4, and 4.5, we obtain

$$
\begin{aligned}
& \frac{1}{n}\sum_{i=1}^{n}\left\{\left\{\frac{d\psi(X_i,\alpha)}{d\alpha}[v^*]\right\}'\hat{\psi}(X_i,\alpha)-\left\{\frac{d\psi(X_i,\alpha_0)}{d\alpha}[v^*]\right\}'\hat{\psi}(X_i,\alpha_0)\right\} \\
& \quad = E\left[\left\{\frac{d\psi(X_i,\alpha)}{d\alpha}[v^*]\right\}'\psi(X_i,\alpha)-\left\{\frac{d\psi(X_i,\alpha_0)}{d\alpha}[v^*]\right\}'\psi(X_i,\alpha_0)\right]+o_p(n^{-1/2}) \\
& \quad = E\left[\left\{\frac{dm(X_i,\alpha_0)}{d\alpha}[v^*]\right\}'\Sigma(X_i,\alpha_0)^{-1}\left\{m(X_i,\alpha)-m(X_i,\alpha_0)\right\}\right]+o_p(n^{-1/2}) \\
& \quad = E\left[\left\{\frac{dm(X_i,\alpha_0)}{d\alpha}[v^*]\right\}'\Sigma(X_i,\alpha_0)^{-1}\left\{\frac{dm(X_i,\bar{\alpha})}{d\alpha}[\alpha-\alpha_0]-\frac{dm(X_i,\alpha_0)}{d\alpha}[\alpha-\alpha_0]\right\}\right] \\
& \qquad +\langle v^*,\alpha-\alpha_0\rangle+o_p(n^{-1/2}) \\
& \quad = \langle v^*,\alpha-\alpha_0\rangle+o_p(n^{-1/2})
\end{aligned}
$$

for some $\bar{\alpha} \in \mathcal{N}_0$ between α and α_0. ∎

Lemma A.16 *Suppose that Assumptions 3.1-3.4, 3.7, 3.9 (ii), 3.10, and 4.3 hold. Then*

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{d\psi(X_i,\alpha_0)}{d\alpha}[v^*]\right\}'\hat{\psi}(X_i,\alpha_0)=\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{dm(X_i,\alpha_0)}{d\alpha}[v^*]\right\}'\Sigma(X_i,\alpha_0)^{-1}\rho(Z_i,\alpha_0)+o_p(n^{-1/2}).$$

Proof. Notice that

$$\left\|Q(X,\alpha_0)(Q(\alpha_0)'Q(\alpha_0))^{-1}\sum_{j=1}^{n}Q(X_j,\alpha_0)\frac{d\psi(X_j,\alpha_0)}{d\alpha}[v^*]-\frac{d\psi(X,\alpha_0)}{d\alpha}[v^*]\right\|=o_p(n^{-1/4})$$

uniformly over $X\in\mathcal{X}$. Then we can prove the result by replacing $g(X,v^*)$ and $\hat{m}(X,\alpha_0)$ in Corollary C.3 (iii) of AC (2003) with $\frac{d\psi(X_i,\alpha_0)}{d\alpha}[v^*]$ and $\hat{\psi}(X_i,\alpha_0)$, respectively. ■

Lemma A.17 *Suppose that Assumptions 3.1-3.2, 3.4, 3.7, 3.8 (ii), 3.9-3.10, and 4.1-4.4 hold. Then*

$$\sup_{\alpha\in\mathcal{N}_{0n}}\left\|\frac{1}{n}\sum_{i=1}^{n}\frac{d\rho(Z_i,\alpha)}{d\alpha}[v_n^*]\otimes p_i\right\|=O_p(1),$$
$$\sup_{\alpha\in\mathcal{N}_{0n}}\left\|\frac{1}{n}\sum_{i=1}^{n}\frac{d^2\rho(Z_i,\alpha)}{d\alpha d\alpha}[v_n^*,v_n^*]\otimes p_i\right\|=O_p(1).$$

Proof. Some calculation yields

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\left\|\frac{d\hat{\psi}(X_i,\alpha)}{d\alpha}[v_n^*]\right\|^2 &= \left(\frac{1}{n}\sum_{i=1}^{n}\frac{d\rho(Z_i,\alpha)}{d\alpha}[v_n^*]\otimes p_i\right)'\left(\frac{1}{n}\sum_{i=1}^{n}\Sigma(X_i,\alpha)^{-1}\otimes p_ip_i'\right)^{-1}\\ &\quad\times\left(\frac{1}{n}\sum_{i=1}^{n}\frac{d\rho(Z_i,\alpha)}{d\alpha}[v_n^*]\otimes p_i\right).\end{aligned}$$

By Lemma A.1, $\lambda_{\min}((\sum_{i=1}^{n}\Sigma(X_i,\alpha)^{-1}\otimes p_ip_i'/n)^{-1})>C$ w.p.a.1. Thus we have

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\frac{d\rho(Z_i,\alpha)}{d\alpha}[v_n^*]\otimes p_i\right\|^2\leq\frac{C}{n}\sum_{i=1}^{n}\left\|\frac{d\hat{\psi}(X_i,\alpha)}{d\alpha}[v_n^*]\right\|^2=O_p(1)$$

by Lemma A.13 and Assumption 4.3. Similarly

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\frac{d^2\rho(Z_i,\alpha)}{d\alpha d\alpha}[v_n^*,v_n^*]\otimes p_i\right\|^2\leq\frac{C}{n}\sum_{i=1}^{n}\left\|\frac{d^2\hat{\psi}(X_i,\alpha)}{d\alpha d\alpha}[v_n^*,v_n^*]\right\|^2=O_p(1)$$

by Assumption 4.6. ■

Proof of Theorem 4.1 Let $\hat{\lambda}_n(\alpha)=\arg\max_{\lambda\in\hat{\Lambda}(\alpha)}\hat{S}(\alpha,\lambda)$. Similarly to the proof of Lemma A.7, we can show that $\hat{\lambda}_n(\alpha)\in\Lambda_n$ and $\max_{1\leq i\leq n}|\hat{\lambda}_n(\alpha)'g_i(\alpha)|\xrightarrow{p}0$ for $\alpha\in\mathcal{N}_{0n}$. Then $\hat{\lambda}_n(\alpha)$ satisfies the following first order condition

$$0=\frac{1}{n}\sum_{i=1}^{n}s_1(\hat{\lambda}_n(\alpha)'g_i(\alpha))g_i(\alpha) \tag{A.8}$$

for all $\alpha \in \mathcal{N}_{0n}$.

By Assumption 4.7, expanding (A.8) yields

$$\begin{aligned} 0 &= \frac{1}{n}\sum_{i=1}^{n} s_1(\hat{\lambda}_n(\alpha)'g_i(\alpha))g_i(\alpha) \\ &= -\hat{g}(\alpha) - \left(\frac{1}{n}\sum_{i=1}^{n} g_i(\alpha)g_i(\alpha)'\right)\hat{\lambda}_n(\alpha) + \frac{1}{2n}\sum_{i=1}^{n} s_3(\tilde{\lambda}'g_i(\alpha))(\hat{\lambda}_n(\alpha)'g_i(\alpha))^2 g_i(\alpha) \end{aligned}$$

for some $\tilde{\lambda}$ and for all $\alpha \in \mathcal{N}_{0n}$. Assumption 4.8 implies that $\max_{1\le i\le n}|\hat{\lambda}_n(\alpha)'g_i(\alpha)| = o_p(n^{-1/8})$ for $\alpha \in \mathcal{N}_{0n}$. Thus we obtain

$$\left\|\frac{1}{n}\sum_{i=1}^{n} s_3(\tilde{\lambda}'g_i(\alpha))(\hat{\lambda}_n(\alpha)'g_i(\alpha))^2 g_i(\alpha)\right\| \le C\left(\max_{1\le i\le n}\left|\hat{\lambda}_n(\alpha)'g_i(\alpha)\right|\right)^2 \|\hat{g}(\alpha)\| = o_p(n^{-1/2}).$$

Hence it follows that $\hat{\lambda}_n(\alpha) = -(n^{-1}\sum_{i=1}^{n} g_i(\alpha)g_i(\alpha)')^{-1}\hat{g}(\alpha) + o_p(n^{-1/2})$. Also, by Lemma A.1, we obtain

$$\left\|\frac{1}{n}\sum_{i=1}^{n} g_i(\alpha)g_i(\alpha)' - \frac{1}{n}\sum_{i=1}^{n}\Sigma(X_i,\alpha)\otimes p_i p_i'\right\| = o_p(n^{-1/4})$$

uniformly over $\alpha \in \mathcal{N}_{0n}$. Moreover, by envelope conditions,

$$\left|\hat{\lambda}_n(\alpha)'\left(\frac{1}{n}\sum_{i=1}^{n} s_2(\tilde{\lambda}'g_i(\alpha))\frac{d\rho(Z_i,\alpha)}{d\alpha}[u_n^*]\rho(Z_i,\alpha)'\otimes p_i p_i'\right)\hat{\lambda}_n(\alpha)\right| \le C\left\|\hat{\lambda}_n(\alpha)\right\|^2 = o_p(n^{-1/2})$$

uniformly over $\alpha \in \mathcal{N}_{0n}$.

Let $0 < \epsilon_n = o(n^{-1/2})$ and $u^* \equiv \pm v^*$. Denote $u_n^* = \Pi_n u^*$. By assumption, we can take a continuous path $\{\alpha(t) : t \in [0,1]\}$ in $\mathcal{N}_{0n}$ such that $\alpha(0) = \hat{\alpha}_n$ and $\alpha(1) = \hat{\alpha}_n + \epsilon_n u_n^* \in \mathcal{N}_{0n}$. By the definition of the SGEL estimator, a Taylor expansion yields

$$0 \le \hat{L}_n(\alpha(0)) - \hat{L}_n(\alpha(1)) = -\left.\frac{d\hat{L}_n(\alpha(t))}{dt}\right|_{t=0} - \frac{1}{2}\left.\frac{d^2\hat{L}_n(\alpha(t))}{dt^2}\right|_{t=s} \tag{A.9}$$

for some $s \in [0,1]$.

Let $\hat{\lambda}_n = \hat{\lambda}_n(\hat{\alpha}_n)$. By the envelope theorem and Lemmas A.14-A.16, we obtain

$$
\begin{aligned}
& -\left.\frac{d\hat{L}_n(\alpha(t))}{dt}\right|_{t=0} \\
= \; & \frac{1}{n}\sum_{i=1}^{n} s_1(\hat{\lambda}_n' g_i(\hat{\alpha}_n))\hat{\lambda}_n' \frac{d\rho(Z_i,\hat{\alpha}_n)}{d\alpha}[\epsilon_n u_n^*] \otimes p_i \\
= \; & -\hat{\lambda}_n' \frac{1}{n}\sum_{i=1}^{n} \frac{d\rho(Z_i,\hat{\alpha}_n)}{d\alpha}[\epsilon_n u_n^*] \otimes p_i \\
& + \hat{\lambda}_n' \left(\frac{\epsilon_n}{n}\sum_{i=1}^{n} s_2(\tilde{\lambda}' g_i(\hat{\alpha}_n)) \frac{d\rho(Z_i,\hat{\alpha}_n)}{d\alpha}[u_n^*]\rho(Z_i,\hat{\alpha}_n)' \otimes p_i p_i' \right) \hat{\lambda}_n + o_p(\epsilon_n n^{-1/2}) \\
= \; & \hat{g}(\hat{\alpha}_n)' \left(\frac{1}{n}\sum_{i=1}^{n} g_i(\hat{\alpha}_n) g_i(\hat{\alpha}_n)' \right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} \frac{d\rho(Z_i,\hat{\alpha}_n)}{d\alpha}[\epsilon_n u_n^*] \otimes p_i \right) + o_p(\epsilon_n n^{-1/2}) \\
= \; & \hat{g}(\hat{\alpha}_n)' \left(\frac{1}{n}\sum_{i=1}^{n} \Sigma(X_i,\hat{\alpha}_n) \otimes p_i p_i' \right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} \frac{d\rho(Z_i,\hat{\alpha}_n)}{d\alpha}[\epsilon_n u_n^*] \otimes p_i \right) + o_p(\epsilon_n n^{-1/2}) \\
= \; & \frac{1}{n}\sum_{i=1}^{n} \hat{\psi}(X_i,\hat{\alpha}_n)' \left\{ \frac{d\hat{\psi}(X_i,\hat{\alpha}_n)}{d\alpha}[\epsilon_n u_n^*] \right\} + o_p(\epsilon_n n^{-1/2}) \\
= \; & \frac{\epsilon_n}{n}\sum_{i=1}^{n} \left\{ \frac{dm(X_i,\alpha_0)}{d\alpha}[u^*] \right\}' \Sigma(X_i,\alpha_0)^{-1}\rho(Z_i,\alpha_0) \\
& + \epsilon_n \langle u^*, \hat{\alpha}_n - \alpha_0 \rangle + o_p(\epsilon_n n^{-1/2}).
\end{aligned}
\tag{A.10}
$$

Next we denote $\frac{d\hat{\lambda}_n(\alpha(\tau))}{d\alpha}[\epsilon_n u_n^*] = \left.\frac{d\hat{\lambda}_n(\alpha(t))}{dt}\right|_{t=\tau}$. By (A.8), we obtain

$$
\begin{aligned}
0 \; = \; & \frac{1}{n}\sum_{i=1}^{n} s_2(\hat{\lambda}_n(\alpha)' g_i(\alpha)) g_i(\alpha) g_i(\alpha)' \frac{d\hat{\lambda}_n(\alpha)}{d\alpha}[v_n^*] \\
& + \frac{1}{n}\sum_{i=1}^{n} s_2(\hat{\lambda}_n(\alpha)' g_i(\alpha)) g_i(\alpha) \hat{\lambda}_n(\alpha)' \frac{d\rho(Z_i,\alpha)}{d\alpha}[v_n^*] \otimes p_i \\
& + \frac{1}{n}\sum_{i=1}^{n} s_1(\hat{\lambda}_n(\alpha)' g_i(\alpha)) \frac{d\rho(Z_i,\alpha)}{d\alpha}[v_n^*] \otimes p_i .
\end{aligned}
$$

Since $\lambda_{\min}(-n^{-1}\sum_{i=1}^{n} s_2(\hat{\lambda}_n(\alpha)' g_i(\alpha)) g_i(\alpha) g_i(\alpha)') > C$ w.p.a.1, we have

$$
\begin{aligned}
\left\| \frac{d\hat{\lambda}_n(\alpha)}{d\alpha}[v_n^*] \right\| \; \leq \; & C \left\| \frac{1}{n}\sum_{i=1}^{n} s_2(\hat{\lambda}_n(\alpha)' g_i(\alpha)) g_i(\alpha) \hat{\lambda}_n(\alpha)' \frac{d\rho(Z_i,\alpha)}{d\alpha}[v_n^*] \otimes p_i \right\| \\
& + C \left\| \frac{1}{n}\sum_{i=1}^{n} s_1(\hat{\lambda}_n(\alpha)' g_i(\alpha)) \frac{d\rho(Z_i,\alpha)}{d\alpha}[v_n^*] \otimes p_i \right\| .
\end{aligned}
$$

Here we have

$$
\begin{aligned}
& \left\| \frac{1}{n}\sum_{i=1}^{n} s_2(\hat{\lambda}_n(\alpha)' g_i(\alpha)) g_i(\alpha) \hat{\lambda}_n(\alpha)' \frac{d\rho(Z_i,\alpha)}{d\alpha}[v_n^*] \otimes p_i \right\| \\
= \; & \left\{ \hat{\lambda}_n(\alpha)' \left(\frac{1}{n}\sum_{i=1}^{n} s_2(\hat{\lambda}_n(\alpha)' g_i(\alpha)) \left\{ \frac{d\rho(Z_i,\alpha)}{d\alpha}[v_n^*] \right\} \rho(Z_i,\alpha)' \otimes p_i p_i' \right)^2 \hat{\lambda}_n(\alpha) \right\}^{1/2} \\
\leq \; & C \left\| \hat{\lambda}_n(\alpha) \right\| = o_p(n^{-1/4})
\end{aligned}
$$

uniformly over $\alpha \in \mathcal{N}_{0n}$. Thus by Lemma A.17, $\sup_{\alpha\in\mathcal{N}_{0n}} \left\| \frac{d\hat{\lambda}_n(\alpha)}{d\alpha}[v_n^*] \right\| = O_p(1)$. Also, by the envelope condition,

$$\left| \hat{\lambda}_n(\alpha)' \left(\frac{1}{n} \sum_{i=1}^{n} s_2(\hat{\lambda}_n(\alpha)' g_i(\alpha)) \left\{ \frac{d\rho(Z_i,\alpha)}{d\alpha}[v_n^*] \right\} \left\{ \frac{d\rho(Z_i,\alpha)}{d\alpha}[v_n^*] \right\}' \otimes p_i p_i' \right) \hat{\lambda}_n(\alpha) \right| = o_p(n^{-1/2}).$$

Denote $\hat{\lambda}' g_i(s) = \hat{\lambda}_n(\alpha(s))' g_i(\alpha(s))$. Then we have

$$\begin{aligned}
&\left.\frac{d^2 \hat{L}_n(\alpha(t))}{dt^2}\right|_{t=s} \\
=\ & \frac{1}{n}\sum_{i=1}^{n} s_2(\hat{\lambda}' g_i(s)) \frac{d\hat{\lambda}' g_i(s)}{d\alpha}[\epsilon_n u_n^*] \hat{\lambda}_n(\alpha(s))' \frac{d\rho(Z_i,\alpha(s))}{d\alpha}[\epsilon_n u_n^*] \otimes p_i \\
&+ \frac{1}{n}\sum_{i=1}^{n} s_1(\hat{\lambda}' g_i(s)) \left\{ \frac{d\hat{\lambda}_n(\alpha(s))}{d\alpha}[\epsilon_n u_n^*] \right\}' \frac{d\rho(Z_i,\alpha(s))}{d\alpha}[\epsilon_n u_n^*] \otimes p_i \\
&+ \frac{1}{n}\sum_{i=1}^{n} s_1(\hat{\lambda}' g_i(s)) \hat{\lambda}_n(\alpha(s))' \frac{d^2\rho(Z_i,\alpha(s))}{d\alpha d\alpha}[\epsilon_n u_n^*, \epsilon_n u_n^*] \otimes p_i \\
=\ & \left\{ \frac{d\hat{\lambda}_n(\alpha(s))}{d\alpha}[u_n^*] \right\}' \left(\frac{\epsilon_n^2}{n}\sum_{i=1}^{n} s_2(\hat{\lambda}' g_i(s)) \rho(Z_i,\alpha) \left\{ \frac{d\rho(Z_i,\alpha(s))}{d\alpha}[u_n^*] \right\}' \otimes p_i p_i' \right) \hat{\lambda}_n(\alpha(s)) \\
&+ \hat{\lambda}_n(\alpha(s))' \left(\frac{\epsilon_n^2}{n}\sum_{i=1}^{n} s_2(\hat{\lambda}' g_i(s)) \left\{ \frac{d\rho(Z_i,\alpha(s))}{d\alpha}[u_n^*] \right\} \left\{ \frac{d\rho(Z_i,\alpha(s))}{d\alpha}[u_n^*] \right\}' \otimes p_i p_i' \right) \hat{\lambda}_n(\alpha(s)) \\
&+ \frac{\epsilon_n^2}{n}\sum_{i=1}^{n} s_1(\hat{\lambda}' g_i(s)) \left\{ \frac{d\hat{\lambda}_n(\alpha(s))}{d\alpha}[u_n^*] \right\}' \frac{d\rho(Z_i,\alpha(s))}{d\alpha}[u_n^*] \otimes p_i \\
&+ \frac{\epsilon_n^2}{n}\sum_{i=1}^{n} s_1(\hat{\lambda}' g_i(s)) \hat{\lambda}_n(\alpha(s))' \frac{d^2\rho(Z_i,\alpha(s))}{d\alpha d\alpha}[u_n^*, u_n^*] \otimes p_i \\
=\ & o_p(\epsilon_n^2). \qquad \text{(A.11)}
\end{aligned}$$

Therefore, it follows from (A.9), (A.10) and (A.11) that

$$\begin{aligned}
\sqrt{n}\xi'(\hat{\theta}_n - \theta_0) &= \sqrt{n}\langle \hat{\alpha}_n - \alpha_0, v^* \rangle \\
&= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \left\{ \frac{dm(X_i,\alpha_0)}{d\alpha}[v^*] \right\}' \Sigma(X_i,\alpha_0)^{-1} \rho(Z_i,\alpha_0) + o_p(1)
\end{aligned}$$

for all $\xi \neq 0$. The result follows from a central limit theorem. ■

References

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

Chen, X. (2007): “Large Sample Sieve Estimation of Semi-Nonparametric Models,” in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 6, pp. 5549–5632. Elsevier.

Donald, S. G., G. W. Imbens, and W. K. Newey (2003): "Empirical Likelihood Estimation and Consistent Tests with Conditional Moment Restrictions," *Journal of Econometrics*, 117, 55–93.

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

Otsu, T. (2011): "Empirical Likelihood Estimation of Conditional Moment Restriction Models with Unknown Functions," *Econometric Theory*, 27, 8–46.

Shen, X., and W. H. Wong (1994): "Convergence Rate of Sieve Estimates," *Annals of Statistics*, 22, 580–615.