Econometric Theory, **27**, 2011, 8–46.
doi:10.1017/S0266466610000113

EMPIRICAL LIKELIHOOD ESTIMATION OF CONDITIONAL MOMENT RESTRICTION MODELS WITH UNKNOWN FUNCTIONS

TAISUKE OTSU
Yale University

This paper proposes an empirical likelihood-based estimation method for conditional moment restriction models with unknown functions, which include several semiparametric models. Our estimator is called the sieve conditional empirical likelihood (SCEL) estimator, which is based on the methods of conditional empirical likelihood and sieves. We derive (i) the consistency and a convergence rate of the SCEL estimator for the whole parameter, and (ii) the asymptotic normality and efficiency of the SCEL estimator for the parametric component. As an illustrating example, we consider a partially linear regression model with nonparametric endogeneity and heteroskedasticity.

1. INTRODUCTION

This paper studies a general estimation method for conditional moment restriction models with unknown functions:

$$\mathrm{E}[\rho(z, \theta_0, h_0)|x] = 0, \tag{1}$$

where $z \equiv (y', x_z')'$, x_z is a subset of x, ρ is a vector of known functions up to (θ_0, h_0), θ_0 is a vector of finite dimensional unknown parameters (the parametric component), and $h_0(\cdot) \equiv (h_{01}(\cdot), \ldots, h_{0q}(\cdot))$ is a vector of infinite dimensional unknown functions (the nonparametric component). The arguments of h_0 can be either a subset of y (endogenous) or a subset of x_z (exogenous). This model is semiparametric in two senses. First, the moment function ρ explicitly includes unknown functions h_0. Second, apart from the conditional moment restriction, the distribution form of z given x is unspecified. Indeed, the model (1) is general enough to include several semiparametric models, such as partially linear and single index models. Also, this model allows for h_0 to contain endogenous variables in its argument. See Ai and Chen (2003) for more examples.

The author is grateful to Xiaohong Chen, Bruce Hansen, John Kennan, Yuichi Kitamura, and Gautam Tripathi for helpful discussions and suggestions. Comments from a co-editor and referees substantially helped this revision. Address correspondence to Taisuke Otsu, Department of Economics, Yale University, Box 208281, New Haven, CT 06520-8281, USA; e-mail: taisuke.otsu@yale.edu.

 © Cambridge University Press 2010

https://doi.org/10.1017/S0266466610000113 Published online by Cambridge University Press

We propose an empirical likelihood-based estimation method for (1).[1] Our estimator is based on the method of conditional empirical likelihood by Zhang and Gijbels (2003) and Kitamura, Tripathi, and Ahn (2004).[2] Since we extend the conditional empirical likelihood approach by the method of sieves to include nonparametric components, our estimator is called the sieve conditional empirical likelihood (SCEL) estimator. The method of sieves overcomes the difficulty of estimating the nonparametric component h_0 by approximating the parameter space with some sieve space.[3] We derive (i) the consistency and a convergence rate of the SCEL estimator for the whole parameter (θ_0, h_0), and (ii) the asymptotic normality and efficiency of the SCEL estimator for the parametric component θ_0. Furthermore, as an illustrating example, we consider a partially linear regression model with nonparametric endogeneity and heteroskedasticity. This setup is not covered by Zhang and Gijbels (2003) or Kitamura et al. (2004).

For conditional moment restriction models *without* unknown functions (i.e., $\mathrm{E}[\rho(z, \theta_0)|x] = 0$), Donald, Imbens, and Newey (2003), Zhang and Gijbels (2003), and Kitamura et al. (2004) developed empirical likelihood-based estimators. Zhang and Gijbels (2003) and Zhang and Liu (2003) considered conditional moment restriction models with unknown functions but without finite dimensional parameters (i.e., $\mathrm{E}[\rho(z, h_0)|x] = 0$). Hjort, McKeague, and van Keilegom (2009) have extended the empirical likelihood approach to a semiparametric setup. They considered *unconditional* moment restriction models with unknown functions (i.e., $\mathrm{E}[\rho(z, \theta_0, h_0)] = 0$) and proposed the empirical likelihood ratio test for θ_0 with some plug-in estimator for h_0.

This paper extends the empirical likelihood approach to deal with the conditional moment restriction models (1), where parametric and nonparametric components coexist and both can contain endogenous variables. Although Otsu (2007) proposed an empirical likelihood-based estimator with the method of penalization to estimate conditional moment restriction models with unknown functions, the setup of Otsu (2007) does not allow nonparametric endogeneity. Since econometric models commonly contain endogenous regressors, this extension is important. Indeed, assumptions and proofs of this paper are very different from those of Otsu (2007). If h_0 have endogenous arguments, it is generally difficult to derive a convergence rate of the SCEL estimator under (strong) norms, such as the sup and L_2 norms. Therefore we employ a different approach developed by Ai and Chen (2003) to obtain a convergence rate under some weaker norm and to derive the asymptotic normality of the SCEL estimator. Ai and Chen (1999, 2003) proposed the sieve minimum distance (SMD) estimator to estimate the model (1). Our SCEL estimator is asymptotically (first-order) equivalent to their optimally weighted SMD estimator. This asymptotic equivalence between the SMD and SCEL estimators is analogous to the asymptotic equivalence between the generalized method of moments (GMM) estimator (Hansen, 1982) and the maximum empirical likelihood estimator (Qin and Lawless, 1994).

This paper is organized as follows. Section 2 introduces the SCEL estimator. Section 3 derives the asymptotic properties of the SCEL estimator. Section 4

https://doi.org/10.1017/S0266466610000113 Published online by Cambridge University Press

illustrates the main result by an example. Section 5 concludes. Appendixes contain proofs and lemmas.

2. ESTIMATOR

Our notation closely follows that of Ai and Chen (1999, 2003). Consider a random sample $\{y_i, x_i\}_{i=1}^n$ from the distribution of (y, x) with support $\mathcal{Y} \times \mathcal{X}$, where $\mathcal{Y}$ is a subset of $\mathbb{R}^{d_y}$ and $\mathcal{X}$ is a compact subset of $\mathbb{R}^s$. Let x_z be a subset of x with support $\mathcal{X}_z$, and $z \equiv (y', x_z')'$ with support $\mathcal{Z} \equiv \mathcal{Y} \times \mathcal{X}_z$. Define Θ and $\mathcal{H}$ as the parameter spaces of θ_0 and h_0, respectively. Assume that $\Theta \subset \mathbb{R}^{d_\theta}$ is compact with nonempty interior and that $\mathcal{H} \equiv \mathcal{H}^1 \times \cdots \times \mathcal{H}^{d_h}$ is a space of continuous functions. Let $\alpha_0 \equiv (\theta_0, h_0) \in \mathcal{A} \equiv \Theta \times \mathcal{H}$ be the whole parameters. In the model (1), the moment function $\rho : \mathcal{Z} \times \mathcal{A} \to \mathbb{R}^{d_\rho}$ is a vector of known functions up to α_0.

We now introduce conditional empirical likelihood by Zhang and Gijbels (2003) and Kitamura et al. (2004). To motivate conditional empirical likelihood, consider the following maximization problem:

$$\max_{\{p_{ji}\}_{i,j=1}^n} \sum_{i=1}^n \sum_{j=1}^n w_{ji} \log p_{ji} \quad \text{s.t.} \quad \sum_{j=1}^n p_{ji} = 1,$$

$$\sum_{j=1}^n p_{ji} g(z_j, \alpha) = 0, \qquad i, j = 1, \ldots, n, \tag{2}$$

where p_{ji} can be interpreted as the conditional probability $\Pr\{z = z_j | x = x_i\}$ and $w_{ji} \equiv K\left(\frac{x_j - x_i}{b_n}\right) / \sum_{j=1}^n K\left(\frac{x_j - x_i}{b_n}\right)$ is the kernel weight for x_j at x_i. Here $K : \mathbb{R}^s \to \mathbb{R}$ is a kernel function and b_n is a bandwidth parameter. We can regard the term $\sum_{j=1}^n w_{ji} \log p_{ji}$ as local likelihood at x_i obtained by borrowing information from nearby observations around x_i using kernel weights w_{ji}.[4]

Based on the dual representation of the maximization problem in (2), we define SCEL as follows:[5]

$$\hat{\ell}_n(\alpha) \equiv -\frac{1}{n} \sum_{i=1}^n \mathbb{I}_{in} \max_{\lambda_i \in \Lambda_n} \sum_{j=1}^n w_{ji} \log\left(1 + \lambda_i' \rho(z_j, \alpha)\right)$$

$$\equiv -\frac{1}{n} \sum_{i=1}^n \mathbb{I}_{in} \sum_{j=1}^n w_{ji} \log\left(1 + \hat{\lambda}(x_i, \alpha)' \rho\left(z_j, \alpha\right)\right), \tag{3}$$

where $\hat{\lambda}(x_i, \alpha)$ is implicitly defined, Λ_n is support of λ_i (defined in Assumption 3.5), $\mathbb{I}_{in} \equiv \mathbb{I}\{x_i \in \mathcal{X}_n\}$ is a trimming term to avoid the boundary bias of kernel estimators (see, e.g., Ai and Chen, 1999), $\mathbb{I}\{\cdot\}$ is the indicator function, and $\mathcal{X}_n$ is a closed subset of $\mathcal{X}$ (defined in Assumption 3.1(ii)). Note that $\hat{\lambda}(x_i, \alpha)$ can be interpreted as the Lagrange multiplier for the conditional moment restriction $\sum_{j=1}^n p_{ji} g(z_j, \alpha) = 0$ in (2).

https://doi.org/10.1017/S0266466610000113 Published online by Cambridge University Press

If α is finite dimensional (i.e., parametric), the maximizer of $\hat{\ell}_n(\alpha)$ with respect to α is asymptotically normal and efficient under certain regularity conditions (Zhang and Gijbels, 2003; Kitamura et al., 2004). However, if α includes infinite dimensional (i.e., nonparametric) components, the parameter space $\mathcal{A}$ is usually too large to obtain reasonable estimates. In order to overcome this difficulty, we approximate the infinite dimensional parameter space $\mathcal{H}$ by a sieve space $\mathcal{H}_n$, which is often finite dimensional and becomes dense in $\mathcal{H}$ as n increases. Let $\mathcal{A}_n \equiv \Theta \times \mathcal{H}_n$. Our SCEL estimator is defined as

$$\hat{\alpha} \equiv (\hat{\theta}, \hat{h}) \equiv \arg\max_{\alpha \in \mathcal{A}_n} \hat{\ell}_n(\alpha). \tag{4}$$

There are several choices for the sieve space $\mathcal{H}_n$, such as Fourier series, power series, splines, or wavelets. Once we choose a sieve space, the SCEL estimator $\hat{\alpha}$ can be computed as follows: (i) For each $\alpha \in \mathcal{A}_n$, compute $\hat{\lambda}(x_i, \alpha)$ by some Newton-type optimization and obtain $\hat{\ell}_n(\alpha)$ (inner loop), and then (ii) maximize $\hat{\ell}_n(\alpha)$ for $\alpha \in \mathcal{A}_n$ by some numerical optimization (outer loop). Zhang and Gijbels (2003) showed that if the parameter space $\mathcal{A}$ is compact, the maximizer of $\hat{\ell}_n(\alpha)$ with respect to α over the whole parameter space $\mathcal{A}$ attains the optimal convergence rate in a nonparametric regression setup. This paper approximates the compact space$\mathcal{A}$ by a sieve space $\mathcal{A}_n$ and focuses on the asymptotic properties of the estimator of the parametric component $\hat{\theta}$.

3. MAIN RESULTS

This section discusses the asymptotic properties of the SCEL estimator.

3.1. Consistency and Convergence Rate

First we derive the consistency of the SCEL estimator $\hat{\alpha}$ under some (strong) norm $\|\cdot\|_s$, such as the sup or L_2 norm. Let $||\cdot||_E$ be the euclidean norm and $\mathcal{N}(\epsilon, \mathcal{A}_n, \|\cdot\|_s)$ be a minimum bracket covering number by ϵ-brackets with respect to $||\cdot||_s$ to cover $\mathcal{A}_n$. Let $\underline{\gamma}$ be the largest integer satisfying $\underline{\gamma} < \gamma$ and $\nabla^a g(x) \equiv \frac{\partial^{a_1+\cdots+a_s}}{\partial x_1^{a_1} \cdots \partial x_s^{a_s}} g(x)$ for $a = (a_1, \ldots, a_s)$. We adopt the following definitions by Ai and Chen (2003).

DEFINITION 3.1 (Hölder space). *The Hölder space $\Lambda^\gamma(\mathcal{W})$ of order $\gamma > 0$ is a space of functions $g : \mathcal{W} \to \mathbb{R}$ such that the first $\underline{\gamma}$ derivatives are bounded, and the $\underline{\gamma}$th derivative satisfies* $\max_{\sum_{i=1}^s a_i = \underline{\gamma}} \left|\nabla^a g(w) - \nabla^a g(w')\right| \leq \text{const.} \times \left\|w - w'\right\|_E^{\gamma - \underline{\gamma}}$ *for all $w, w' \in \mathcal{W}$. The Hölder norm is defined as* $\|g\|_{\Lambda^\gamma} \equiv \sup_{w \in \mathcal{W}} |g(w)| + \max_{\sum_{i=1}^s a_i = \underline{\gamma}} \sup_{w \neq w'} \frac{|\nabla^a g(w) - \nabla^a g(w')|}{\|w - w'\|_E^{\gamma - \underline{\gamma}}}$. *The Hölder ball with radius c is defined as* $\Lambda_c^\gamma(\mathcal{W}) \equiv \left\{g \in \Lambda^\gamma(\mathcal{W}) : \|g\|_{\Lambda^\gamma} \leq c < \infty\right\}$.

https://doi.org/10.1017/S0266466610000113 Published online by Cambridge University Press

DEFINITION 3.2 (Envelope condition). *A real-valued function* $g(z,\alpha)$ *satisfies the* κ*th-order envelope condition in* $\alpha \in \mathcal{A}$ *if there exists a measurable function* $c_1(z)$ *with* $\mathrm{E}[c_1(z)^\kappa] < \infty$ *such that* $|g(z,\alpha)| \leq c_1(z)$ *for all* $z \in \mathcal{Z}$ *and* $\alpha \in \mathcal{A}$.

DEFINITION 3.3 (Hölder continuity). *A real-valued function* $g(z,\alpha)$ *is Hölder continuous in* $\alpha \in \mathcal{A}$ *if there exist a constant* $\kappa \in (0,1]$ *and a measurable function* $c_2(z)$ *with* $\mathrm{E}[c_2(z)^2|x]$ *bounded, such that* $|g(z,\alpha_1) - g(z,\alpha_2)| \leq c_2(z)\,\|\alpha_1 - \alpha_2\|_s^\kappa$ *for all* $z \in \mathcal{Z}$ *and* $\alpha_1, \alpha_2 \in \mathcal{A}$.

The Hölder space is a popular functional space for smooth functions. The envelope condition and Hölder continuity are used to control the boundedness and smoothness of functions, respectively. Based on these definitions, we impose the following assumptions to derive the consistency of the SCEL estimator under the norm $\|\cdot\|_s$.

Assumption 3.1. (i) $\{y_i, x_i\}_{i=1}^n$ is i.i.d.; (ii) $\mathcal{X} \equiv \prod_{k=1}^s [x_k^L, x_k^U] \subset \mathbb{R}^s$ for $x_k^L, x_k^U < \infty$ with $x_k^L \neq x_k^U$ $(k = 1, \ldots, s)$, and $\mathcal{X}_n \equiv \prod_{k=1}^s [x_k^L + b_n^{-\mu_1}, x_k^U - b_n^{-\mu_1}]$ for some $\mu_1 \in (0,1)$; (iii) the density function of x is bounded and bounded away from zero on $\mathcal{X}$ and has up to the Jth derivatives with $J \geq 2$.

Assumption 3.2. (i) α_0 is the only $\alpha \in \mathcal{A}$ satisfying $\mathrm{E}[\rho(z,\alpha)|x] = 0$; (ii) there exists a norm $\|\cdot\|_s$ such that $\mathcal{A}$ is compact under $\|\cdot\|_s$; (iii) $\{\mathcal{A}_n\}_{n=1}^\infty$ is a sequence of subsets of $\mathcal{A}$ and for all $\alpha \in \mathcal{A}$ there exists $\Pi_n \alpha \in \mathcal{A}_n$ satisfying $\|\Pi_n \alpha - \alpha\|_s = o(1)$.

Assumption 3.3. (i) $K(x)$ is a bounded and Lipschitz continuous function with bounded support and is symmetric around the origin; (ii) $K(x)$ is a Jth-order kernel function; (iii) as $n \to \infty$, $b_n \to 0$ and $n^{-1} b_n^{-s-2(s+1)/p} \log(b_n^{-s(s+1)}) \to 0$.

Assumption 3.4. (i) Each element of $\mathrm{E}[\rho(z,\alpha)|x] \in \Lambda_c^{J'}(\mathcal{X})$ with some $J' > J$ for all $\alpha \in \mathcal{A}_n$; (ii) $\rho(z,\alpha)$ satisfies the pth order envelope condition in $\alpha \in \mathcal{A}_n$ with $p > 8$; (iii) $\rho(z,\alpha)$ is Hölder continuous in $\alpha \in \mathcal{A}$.

Assumption 3.5. $\Lambda_n = \{\lambda \in \mathbb{R}^{d_\rho} : \|\lambda\|_E \leq \bar{c} n^{-1/p}\}$ for some $\bar{c} \in [1, \infty)$.

Several assumptions are similar to those of Ai and Chen (1999, 2003). Assumptions 3.1(i) and 3.1(ii) exclude dependent data and unbounded conditioning variables, respectively. We conjecture that the compactness assumption of $\mathcal{X}$ can be dropped by use of the trimming argument in Kitamura et al. (2004). Assumption 3.2(i) is an identification condition of α_0, which should be verified in each setup. Assumption 3.2(ii), which is adopted by, e.g., Ai and Chen (2003) and Newey and Powell (2003), is standard in this literature. This assumption is typically satisfied when $\mathcal{H}$ consists of bounded and smooth functions (see Section 4 for an example). Assumption 3.2(iii) is a weak condition on the sieve space $\mathcal{A}_n$. To derive the consistency of $\hat{\alpha}$, we do not need any specific convergence rate for

https://doi.org/10.1017/S0266466610000113 Published online by Cambridge University Press

$\|\Pi_n\alpha-\alpha\|_s$. Assumption 3.3 restricts the kernel function K and the bandwidth b_n. If $b_n=n^{-\zeta}$, Assumption 3.3(iii) is satisfied for $0<\zeta<1/\{s+2(s+1)/p\}$. Assumption 3.4 contains boundedness and smoothness conditions for $\rho(z,\alpha)$. In case of the SMD estimator by Ai and Chen (2003), Assumption 3.4(ii) is unnecessary to derive the consistency (although it is required to derive the convergence rate). Assumption 3.5 is a condition on support of λ_i. This assumption is also adopted by Kitamura et al. (2004, Assump. 3.6) to derive the asymptotic normality of their conditional empirical likelihood estimator. Since we need to control $\hat{\lambda}(x_i,\Pi_n\alpha_0)$ in the proof of consistency, this assumption is required even for deriving the consistency. This idea is also seen in the proof of Newey and Smith (2004, Thm. 3.1) to derive the consistency of the generalized empirical likelihood estimator. We conjecture that in order to drop Assumption 3.5 in our setup, we need to impose stronger assumptions on the derivative of $\rho(z,\alpha)$ with respect to α and the remainder of the linear approximation for $\rho(z,\alpha)$, which may not allow endogenous variables in the arguments of h_0. The consistency result is obtained below.

THEOREM 3.1. *Suppose that Assumptions 3.1, 3.2(i)–(iii), 3.3(i)–(iii), 3.4, and 3.5 hold. Then* $||\hat{\alpha}-\alpha_0||_s=o_p(1)$.

Theorem 3.1 provides the consistency of the SCEL estimator $\hat{\alpha}$ under the norm $\|\cdot\|_s$. To obtain the asymptotic normality of $\hat{\theta}$ (i.e., the SCEL estimator for the parametric component), we need a convergence rate for $\hat{\alpha}$ to α_0. However, it is generally difficult to derive the convergence rate of $\left\|\hat{\alpha}-\alpha_0\right\|_s$ particularly when the arguments of h_0 contain endogenous variables (see Ai and Chen, 2003, p. 1805, for the case of the SMD estimator). Therefore, following Ai and Chen (2003), we introduce another (weaker) norm $\|\cdot\|$ and derive a convergence rate under $\|\cdot\|$, which is actually sufficient to derive the asymptotic normality of $\hat{\theta}$.

Suppose that (i) for any $\alpha_1,\alpha_2\in\mathcal{A}$, there exists a continuous path $\alpha(\tau)\in\mathcal{A}$ for $\tau\in[0,1]$ such that $\alpha(0)=\alpha_1$ and $\alpha(1)=\alpha_2$; (ii) $\mathcal{A}$ is convex at α_0, i.e., for any $\alpha\in\mathcal{A}$, $(1-\tau)\alpha_0+\tau\alpha\in\mathcal{A}$ for small $\tau>0$; and (iii) for almost all z, $\rho(z,(1-\tau)\alpha_0+\tau\alpha)$ is continuously differentiable at $\tau=0$. Then we can define the pathwise first-order derivative of $\rho(z,\alpha_0)$ at the direction $[\alpha-\alpha_0]$ evaluated at α_0, that is

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

Also denote

$$m(x,\alpha)\equiv\mathrm{E}[\rho(z,\alpha)|x],\qquad V(x,\alpha)\equiv\mathrm{E}[\rho(z,\alpha)\rho(z,\alpha)'|x],$$

$$\frac{d\rho(z,\alpha_0)}{d\alpha}[\alpha_1-\alpha_2]\equiv\frac{d\rho(z,\alpha_0)}{d\alpha}[\alpha_1-\alpha_0]-\frac{d\rho(z,\alpha_0)}{d\alpha}[\alpha_2-\alpha_0],$$

$$\frac{dm(x,\alpha_0)}{d\alpha}[\alpha_1-\alpha_2]\equiv\mathrm{E}\left[\left.\frac{d\rho(z,\alpha_0)}{d\alpha}[\alpha_1-\alpha_2]\right|x\right].$$

https://doi.org/10.1017/S0266466610000113 Published online by Cambridge University Press

For all $\alpha_1, \alpha_2 \in \mathcal{A}$, the norm $||\cdot||$ is defined as

$$\|\alpha_1 - \alpha_2\| \equiv \sqrt{\mathrm{E}\left[\left\{\frac{dm(x,\alpha_0)}{d\alpha}[\alpha_1 - \alpha_2]\right\}' V(x,\alpha_0)^{-1}\left\{\frac{dm(x,\alpha_0)}{d\alpha}[\alpha_1 - \alpha_2]\right\}\right]}. \tag{5}$$

This norm is motivated by the second-order term of the GMM-type population objective function ($\mathrm{E}\left[m(x,\alpha)'V(x,\alpha_0)^{-1}m(x,\alpha)\right]$). In this sense, this norm is considered as an extension of the Fisher norm to the conditional moment restriction models. To derive the convergence rate of $\hat{\alpha}$ under the norm $\|\cdot\|$, we add the assumptions below.

Assumption 3.2 (Continued)**.** (iv) There exists a constant $\mu_2 \in (0, \infty)$ such that, for all $\alpha \in \mathcal{A}$, there is $\Pi_n \alpha \in \mathcal{A}_n$ satisfying $\|\Pi_n \alpha - \alpha\| = O(k_n^{-\mu_2})$ with $k_n^{-\mu_2} = o(n^{-1/4})$.

Assumption 3.3 (Continued)**.** (iv) As $n \to \infty$, $nb_n^{4J} \to 0$ and $n^{-1/2} b_n^{-s} k_n \max\{1, n^{-1/4+1/p} b_n^{-4(s+1)/p}\} \log n \to 0$.

Assumption 3.6. (i) Each element of $V(x,\alpha) \in \Lambda_c^{J'}(\mathcal{X})$ with some $J' > J$ for all $\alpha \in \mathcal{A}_n$; (ii) the smallest eigenvalue of $V(x,\alpha)$ is positive and the largest eigenvalue of $V(x,\alpha)$ is finite uniformly on $(x,\alpha) \in \mathcal{X} \times \mathcal{A}_n$.

Assumption 3.7. $\log \mathcal{N}(\epsilon^{1/\kappa}, \mathcal{A}_n, \|\cdot\|_s) \leq \text{const.} \times k_n \log(k_n/\epsilon)$.

Assumption 3.8. (i) $\mathcal{A}$ is convex in α_0, and $\rho(z,\alpha)$ is pathwise differentiable at α_0; (ii) for some constants $c_1, c_2 \in (0, \infty)$,

$$c_1 \mathrm{E}\left[m(x,\alpha)'V(x,\alpha_0)^{-1}m(x,\alpha)\right] \leq \|\alpha - \alpha_0\|^2 \leq c_2 \mathrm{E}\left[m(x,\alpha)'V(x,\alpha_0)^{-1}m(x,\alpha)\right]$$

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

Assumption 3.2(iv) is on the sieve approximation error by $\mathcal{A}_n$. This convergence rate, $o(n^{-1/4})$, is satisfied by commonly used sieve approximations (e.g., power series and Fourier series) to standard functional spaces (e.g., Sobolev and Hölder spaces). Assumption 3.3(iv) contains additional conditions on the bandwidth b_n, which are required to obtain the uniform convergence rates in Lemma B.2. If $b_n = n^{-\zeta}$ and $k_n = n^{-\psi}$, Assumption 3.3(iv) is satisfied for $\frac{1}{4J} < \zeta < \min\left\{\frac{1}{s}\left(\frac{1}{2} + \psi\right), \frac{1}{sp+4s+4}\left(\frac{3}{4}p - 1 + \psi p\right)\right\}$. Assumption 3.6 contains smoothness and boundedness conditions for $V(x,\alpha)$. Assumptions 3.7 and 3.8 are employed by Ai and Chen (2003). Assumption 3.7 is an entropy condition for the sieve space $\mathcal{A}_n$, which controls the size of $\mathcal{A}_n$ not to be too large. This assumption is typically satisfied by popular sieve spaces (see Section 4 of this paper and

https://doi.org/10.1017/S0266466610000113 Published online by Cambridge University Press

Chen and Shen, 1998, for specific examples). Assumption 3.8(i) guarantees that the norm $\|\cdot\|$ is well defined. Assumption 3.8(ii) is crucial to obtain a convergence rate for $\hat{\alpha}$. Intuitively, this assumption locally approximates (the quadratic approximation of) the SCEL population objective function with $\|\alpha - \alpha_0\|^2$. The convergence rate of $\hat{\alpha}$ under the norm $\|\cdot\|$ is obtained below.

THEOREM 3.2. *Suppose that Assumptions 3.1–3.8 hold. Then* $\|\hat{\alpha} - \alpha_0\| = o_p(n^{-1/4})$.

In the next subsection, we will see that this convergence rate is sufficient for deriving the asymptotic normality of $\hat{\theta}$.

3.2. Asymptotic Normality and Efficiency

To derive the asymptotic normality of the parametric component of the SCEL estimator $\hat{\theta}$, we follow the Riesz representation approach of Ai and Chen (1999, 2003), which is based on Shen (1997). Since the parameter $\alpha = (\theta, h)$ includes the infinite-dimensional component h, we generally cannot adopt the usual approach to derive the asymptotic normality, which is based on the first-order condition for $\hat{\alpha}$. For this problem, Shen (1997) proposed an alternative approach based on the Riesz representation theorem for the Hilbert space. Using the Riesz representor, Shen (1997) derived the asymptotic linear form for the finite-dimensional parameter of interest and then applied a central limit theorem. Ai and Chen (2003) extended this approach to the conditional moment restriction models with unknown functions.

We introduce the following notation based on Ai and Chen (2003): Let $\bar{\mathcal{V}}$ be the linear completion of $\mathcal{A} - \{\alpha_0\}$ under the norm $\|\cdot\|$, and

$$\frac{d\rho(z,\alpha)}{d\alpha}[v] \equiv \left.\frac{d\rho(z,\alpha+\tau v)}{d\tau}\right|_{\tau=0}, \qquad \frac{dm(x,\alpha)}{d\alpha}[v] \equiv \mathrm{E}\left[\left.\frac{d\rho(z,\alpha)}{d\alpha}[v]\right| x\right],$$

$$\langle v_1, v_2\rangle \equiv \mathrm{E}\left[\left\{\frac{dm(x,\alpha_0)}{d\alpha}[v_1]\right\}' V(x,\alpha_0)^{-1}\left\{\frac{dm(x,\alpha_0)}{d\alpha}[v_2]\right\}\right],$$

where $\langle v_1, v_2\rangle$ is the inner product for the Hilbert space $(\bar{\mathcal{V}}, \|\cdot\|)$. Write $\frac{dm(x,\alpha_0)}{d\alpha}[\alpha - \alpha_0] = \frac{dm(x,\alpha_0)}{d\theta'}(\theta - \theta_0) + \frac{dm(x,\alpha_0)}{dh}[h - h_0]$. For each θ_k $(k = 1, \ldots, d_\theta)$, let $\bar{\mathcal{W}}$ be the linear completion of $\mathcal{H} - \{h_0\}$,

$$w_k^* \equiv \arg\min_{w_k \in \bar{\mathcal{W}}} \mathrm{E}\left[\left\{\frac{dm(x,\alpha_0)}{d\theta_k} - \frac{dm(x,\alpha_0)}{dh}[w_k]\right\}' V(x,\alpha_0)^{-1} \times \left\{\frac{dm(x,\alpha_0)}{d\theta_k} - \frac{dm(x,\alpha_0)}{dh}[w_k]\right\}\right],$$

$w^* \equiv (w_1^*, \ldots, w_{d_\theta}^*)$, $\frac{dm(x,\alpha_0)}{dh}[w^*] \equiv \left(\frac{dm(x,\alpha_0)}{dh}[w_1^*], \ldots, \frac{dm(x,\alpha_0)}{dh}[w_{d_\theta}^*]\right)$, and $D_{w^*}(x) \equiv \frac{dm(x,\alpha_0)}{d\theta'} - \frac{dm(x,\alpha_0)}{dh}[w^*]$. Let $f(\alpha) \equiv \xi'\theta$ be the parameter of interest

https://doi.org/10.1017/S0266466610000113 Published online by Cambridge University Press

for each $\xi \in \mathbb{R}^{d_\theta}$ with $\xi \neq 0$. If $\mathrm{E}\left[D_{w^*}(x)'V(x,\alpha_0)^{-1}D_{w^*}(x)\right]$ is finite and positive definite, the linearity of f and the Riesz representation theorem imply that there exists the Riesz representor $v^* \in \bar{\mathcal{V}}$ satisfying

$$f(\alpha) - f(\alpha_0) \equiv \xi'(\theta - \theta_0) = \langle \alpha - \alpha_0, v^* \rangle, \tag{6}$$

for all $\alpha \in \mathcal{A}$, where $v^* \equiv (v_\theta^*, v_h^*)$, $v_\theta^* \equiv \mathrm{E}\left[D_{w^*}(x)'V(x,\alpha_0)^{-1}D_{w^*}(x)\right]^{-1}\xi$, and $v_h^* \equiv -w^* \times v_\theta^*$ (see Ai and Chen 2003, p. 1809). Thus, for the asymptotic normality of $\sqrt{n}\xi'(\hat{\theta} - \theta_0)$, it is sufficient to find an asymptotic linear form that approximates $\langle \hat{\alpha} - \alpha_0, v^* \rangle$. Let $\mathcal{N}_{0n} \equiv \left\{\alpha \in \mathcal{A}_n : \|\alpha - \alpha_0\|_s = o(1), \|\alpha - \alpha_0\| = o(n^{-1/4})\right\}$ and $\mathcal{N}_0 \equiv \left\{\alpha \in \mathcal{A} : \|\alpha - \alpha_0\|_s = o(1), \|\alpha - \alpha_0\| = o(n^{-1/4})\right\}$. To derive the asymptotic normality of $\hat{\theta}$, we add three assumptions.

Assumption 3.9. (i) $\theta_0 \in int(\Theta)$; (ii) $\mathrm{E}\left[D_{w^*}(x)'V(x,\alpha_0)^{-1}D_{w^*}(x)\right]$ is positive definite; (iii) there exists $v_n^* = (v_\theta^*, -\Pi_n w^* \times v_\theta^*) \in \mathcal{A}_n - \alpha_0$ such that $\left\|v_n^* - v^*\right\| = O(n^{-1/4})$.

Assumption 3.10. (i) $\frac{d\rho(z,\tilde{\alpha})}{d\alpha}[v]$ exists a.s. z for all $\tilde{\alpha} \in \mathcal{N}_0$; (ii) each element of $\frac{dm(x,\tilde{\alpha})}{d\alpha}[v_n^*] \in \Lambda_c^{J'}(\mathcal{X})$ with some $J' > J$ for all $\tilde{\alpha} \in \mathcal{N}_0$; (iii) $\frac{d\rho(z,\tilde{\alpha})}{d\alpha}[v_n^*]$ satisfies the second-order envelope condition in $\tilde{\alpha} \in \mathcal{N}_{0n}$ and is Hölder continuous in $\tilde{\alpha} \in \mathcal{N}_{0n}$; (iv) $\left.\frac{d^2\rho(z,\tilde{\alpha}+\tau v_n^*)}{d\tau^2}\right|_{\tau=0}$ exists a.s. z for all $\tilde{\alpha} \in \mathcal{N}_{0n}$ and satisfies the second-order envelope condition in $\tilde{\alpha} \in \mathcal{N}_{0n}$.

Assumption 3.11. (i) $\mathrm{E}\left[\left\|V(x,\alpha)^{-1}\frac{dm(x,\tilde{\alpha})}{d\alpha}[v_n^*] - V(x,\alpha_0)^{-1}\frac{dm(x,\alpha_0)}{d\alpha}[v_n^*]\right\|_E^2\right] = o(n^{-1/2})$ uniformly on $\tilde{\alpha} \in \mathcal{N}_{0n}$; (ii) $\mathrm{E}\left[\left\{\frac{dm(x,\alpha_0)}{d\alpha}[v^*]\right\}' V(x,\alpha_0)\left\{\mathbb{I}_{in}\frac{dm(x,\tilde{\alpha})}{d\alpha}[\tilde{\alpha} - \alpha_0] - \frac{dm(x,\alpha_0)}{d\alpha}[\tilde{\alpha} - \alpha_0]\right\}\right] = o(n^{-1/2})$ uniformly on $\bar{\alpha} \in \mathcal{N}_0$ and $\tilde{\alpha} \in \mathcal{N}_{0n}$.

These assumptions are also employed by Ai and Chen (2003). Assumption 3.9(i) is standard. Assumption 3.9(ii) is a local identification condition of θ_0. Assumption 3.9(iii) is on the approximation error of v_n^* to the Riesz representor v^*. Assumption 3.10 contains boundedness and smoothness conditions on the pathwise derivatives $\frac{dm(z,\tilde{\alpha})}{d\alpha}[v_n^*]$, $\frac{d\rho(z,\tilde{\alpha})}{d\alpha}[v_n^*]$, and $\left.\frac{d^2\rho(z,\tilde{\alpha}+\tau v_n^*)}{d\tau^2}\right|_{\tau=0}$. Assumption 3.10(i)–(iii) are required to obtain the uniform convergence rate in Lemma B.2(iii). Assumption 3.11 controls asymptotic bias terms when the moment function ρ is highly nonlinear in α. Assumption 3.11(ii) also restricts that the trimming term $\mathbb{I}_{in}$ should converge to 1 fast enough, which can be satisfied by choosing reasonably μ_1 in Assumption 3.1(ii). If ρ is linear in α, Assumptions 3.10 and 3.11 are trivially satisfied. The asymptotic normality of $\hat{\theta}$ is obtained below.

THEOREM 3.3. *Suppose that Assumptions 3.1–3.11 hold. Then* $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \Sigma^{-1})$ *with* $\Sigma \equiv \mathrm{E}\left[D_{w^*}(x)'V(x,\alpha_0)^{-1}D_{w^*}(x)\right]$.

https://doi.org/10.1017/S0266466610000113 Published online by Cambridge University Press

The limiting distribution of $\hat{\theta}$ is the same as that of the optimally weighted SMD estimator by Ai and Chen (2003). From Ai and Chen (2003, Thm. 6.1), we can see that under certain regularity conditions the asymptotic variance Σ^{-1} becomes the semiparametric efficiency bound of the model (1). Therefore the SCEL estimator $\hat{\theta}$ is asymptotically efficient as well. This first-order asymptotic equivalence between the SMD and SCEL estimators is analogous to the equivalence between the GMM and empirical likelihood estimators and is considered an extension of the asymptotic equivalence described in Zhang and Gijbels (2003, p. 7) to our semiparametric setup.

In contrast to the two-step SMD estimator by Ai and Chen (2003), our SCEL estimator does not need the preliminary weight estimation. Ai and Chen (2003) also suggested the continuously updated GMM-type estimator. For moment restriction models without unknown functions, Newey and Smith (2004) conducted higher-order comparisons for the GMM and empirical likelihood estimators. In our setup, we can also expect that higher-order properties of the SMD and SCEL estimators are different.

To estimate the covariance matrix Σ^{-1}, we can apply Ai and Chen's (1999) consistent estimator:

$$\hat{\Sigma} \equiv \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{in}\hat{D}_{\hat{w}}(x_i)'\hat{V}(x_i,\hat{\alpha})^{-1}\hat{D}_{\hat{w}}(x_i),$$

where $\hat{V}(x_i,\hat{\alpha}) \equiv \sum_{j=1}^{n} w_{ji}\rho\left(z_j,\hat{\alpha}\right)\rho\left(z_j,\hat{\alpha}\right)'$, $\hat{D}_{\hat{w}}(x_i) \equiv (\hat{D}_{\hat{w}_1}(x_i),\ldots,\hat{D}_{\hat{w}_{d_\theta}}(x_i))$, $\hat{w}_k \equiv \arg\min_{w_k\in\bar{\mathcal{W}}_n}\sum_{i=1}^{n}\mathbb{I}_{in}\hat{D}_{w_k}(x_i)'\hat{V}(x_i,\hat{\alpha})^{-1}\hat{D}_{w_k}(x_i)$, $\hat{D}_{w_k}(x_i) \equiv \sum_{j=1}^{n} w_{ji}\left(\frac{d\rho(z_j,\hat{\alpha})}{d\theta_k} - \frac{d\rho(z_j,\hat{\alpha})}{dh}[w_k]\right)$, and $\bar{\mathcal{W}}_n$ is a sieve space for $\bar{\mathcal{W}}$. We can show that under certain additional regularity conditions, $\hat{\Sigma}$ is consistent for Σ (see Ai and Chen, 1999, Thm. 5.1).

4. EXAMPLE

This section illustrates our main results by the partially linear regression model with nonparametric endogeneity and heteroskedasticity:

$$\rho(z,\alpha) = y_1 - x_1'\theta - h(y_2), \qquad \mathrm{E}[\rho(z,\alpha_0)|x] = 0, \qquad \mathrm{E}\left[\rho(z,\alpha_0)^2|x\right] = V(x,\alpha_0), \tag{7}$$

where $\alpha_0 = (\theta_0, h_0)$, $y_1 \in \mathbb{R}$, $z = \left(y_1, y_2', x_1'\right)'$, and $x = (x_1', x_2')'$. Denote $\dim(x_2) = \dim(y_2) = d$, $\dim(x_1) = d_\theta$, and $s = \dim(x) = d + d_\theta$. Assume that x_1 and y_2 do not contain the constant terms. This model is slightly different from Ai and Chen (2003, Ex. 2.1) because of the heteroskedastic errors. Due to the unknown function h, we cannot apply the result of Kitamura et al. (2004). In particular, it is difficult to obtain the convergence rate of $\|\hat{\alpha} - \alpha_0\|_s$ (see Ai and Chen, 2003, p. 1805).

https://doi.org/10.1017/S0266466610000113 Published online by Cambridge University Press

We consider the following setup. Let $\hat{\alpha} = (\hat{\theta}, \hat{h})$ be the SCEL estimator with $\mathcal{H}_n = \{h(y_2) = q^{k_n}(y_2)'\beta$ such that $\|h\|_{\Lambda^\gamma} \leq c\}$, where $q^{k_n}(y_2)$ is a $k_n \times 1$ vector of the tensor-product sieve basis by a univariate B-spline. This $\mathcal{H}_n$ is a sieve space to approximate $\mathcal{H} = \Lambda_c^\gamma(\mathbb{R}^d)$. Let $\|\alpha\|_s = \|\theta\|_E + \sup_{y_2 \in \mathbb{R}^d} \left| h(y_2) \left(1 + \|y_2\|_E^2\right)^{-a/2} \right|$ for some $a > \gamma > 0$, and $w^* = \left(w_1^*, \ldots, w_{d_\theta}^*\right)$ be the solution to $\min_{w_j \in \bar{\mathcal{W}}} \mathrm{E}\left[V(x, \alpha_0)\left\{x_{1j} - \mathrm{E}\left[w_j(y_2)|x\right]\right\}^2\right]$. The general results obtained in the preceding section are applied below.

THEOREM 4.1. *Suppose that Assumptions 3.1, 3.3, 3.5, 3.6(ii), and 3.9(i) hold. Furthermore, assume*

(i) *(a)* $y_2|x$ *is complete; (b)* $h_0 \in \mathcal{H} = \Lambda_c^\gamma(\mathbb{R}^d)$*; (c)* $\mathrm{E}[h_0(y_2)|x]$ *does not belong to the linear span of* x_1 *and* $\mathrm{E}\left[x_1 x_1'\right]$ *is finite and positive definite; (d)* $\mathrm{E}\left[\left\{x_1' - \mathrm{E}\left[w^*(y_2)|x\right]\right\}' V(x, \alpha_0)^{-1}\left\{x_1' - \mathrm{E}\left[w^*(y_2)|x\right]\right\}\right]$ *is positive definite.*

(ii) *(a)* $\mathrm{E}\left[|y_1|^p\right] < \infty$ *with* $p > 8$*; (b)* $\mathrm{E}\left[(1 + \|y_2\|_E^2)^a |x\right]$ *is bounded for* $a > \gamma > 0$.

(iii) $\mathrm{E}[y_1|x] \in \Lambda_c^{J'}(\mathcal{X})$ *and* $\mathrm{E}\left[q^{k_n}(y_2)|x\right] \in \Lambda_c^{J'}(\mathcal{X})$ *with some* $J' > J$.

(iv) $\mathrm{E}[y_1^2|x] \in \Lambda_c^{J'}(\mathcal{X})$, $\mathrm{E}\left[q^{k_n}(y_2) q^{k_n}(y_2)'|x\right] \in \Lambda_c^{J'}(\mathcal{X})$, *and* $\mathrm{E}[y_1 q^{k_n}(y_2)|x] \in \Lambda_c^{J'}(\mathcal{X})$.

(v) $n \geq k_n + d_\theta$ *and* $k_n^{-\gamma/d} = o(n^{-1/4})$.

(vi) $\mathrm{E}\left[w_j^*(y_2)|x\right] \in \Lambda_c^{J'}(\mathcal{X})$ *for* $j = 1, \ldots, d_\theta$.

Then $\|\hat{\alpha} - \alpha_0\|_s = o_p(1)$, $\|\hat{\theta} - \theta_0\|_E = o_p(n^{-1/4})$, $\mathrm{E}\left[\mathrm{E}[\hat{h}(y_2) - h_0(y_2)|x]^2\right] = o_p(n^{-1/2})$, *and* $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N\left(0, \mathrm{E}[\{x_1' - \mathrm{E}[w^*(y_2)|x]\}' V(x, \alpha_0)^{-1}\{x_1' - \mathrm{E}[w^*(y_2)|x]\}]^{-1}\right)$.

Conditions in Theorem 4.1 are similar to those in Ai and Chen (2003, Ex. 2.1). Condition (i) contains identification conditions for (θ_0, h_0). Conditions (ii) and (iii) are boundedness and smoothness conditions for $\rho(z, \alpha)$ and $\mathrm{E}[\rho(z, \alpha_0)|x]$. Condition (iv) includes assumptions to deal with heteroskedasticity. Condition (v) is on the sieve tuning parameter k_n. Condition (vi) is on the smoothness of w^*.

5. CONCLUSION

In this paper we propose an empirical likelihood-based estimator for conditional moment restriction models with unknown functions. We extend the method of conditional empirical likelihood (Zhang and Gijbels, 2003; Kitamura et al., 2004) by adding sieve approximations for unknown functions. For the whole parameter, our estimator is consistent and shows a sufficiently fast convergence rate under a suitably chosen norm. For the parametric component, our estimator is asymptotically normal and attains the semiparametric efficiency bound of Ai and Chen

https://doi.org/10.1017/S0266466610000113 Published online by Cambridge University Press

(2003). Therefore our estimator is asymptotically (first-order) equivalent to the sieve minimum distance estimator of Ai and Chen (2003). There are several directions for future research, such as the extension to nonsmooth moment restrictions (e.g., conditional quantile) or weakly dependent data, higher-order comparisons with Ai and Chen's (2003) sieve minimum distance estimator, and the optimal choice rule for the bandwidth. Since the results of Zhang and Gijbels (2003) allow nonsmooth moment functions, we expect that an application of their empirical process-based approach yields similar results in our semiparametric setup. It is also interesting to extend the results of Fan and Zhang (2004) and consider hypothesis test statistics based on conditional empirical likelihood, such as test statistics for shape restrictions on the unknown functions (i.e., $H_0 : h_0 \in \mathcal{H}_R$) and model specification (i.e., $H_0 : \mathrm{E}[\rho(z,\alpha)|x] = 0$ a.s. x for some $\alpha \in \mathcal{A}$).

NOTES

1. See Owen (2001) for a comprehensive review on empirical likelihood.
2. Zhang and Gijbels's (2003) "sieve" empirical likelihood and Kitamura et al.'s (2004) "smoothed" empirical likelihood are quite similar concepts. To avoid confusion, we introduce another terminology, conditional empirical likelihood, suggested by Kitamura (2003).
3. See Chen (2007) for a comprehensive review of the method of sieves.
4. Fan and Zhang (2004) applied the conditional likelihood approach to construct test statistics for nonparametric hypotheses.
5. SCEL can be extended to the generalized empirical likelihood analog (Smith, 1997); i.e.,

$$\hat{\ell}_n^{GEL}(\alpha) \equiv \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{in} \max_{\lambda_i \in \Lambda_n} \sum_{j=1}^{n} w_{ji}\, \psi\left(\lambda_i' \rho(z_j,\alpha)\right),$$

where ψ is a concave function. For conditional moment restriction models without unknown functions, this analog is considered by Smith (2007).

REFERENCES

Ai, C. & X. Chen (1999) Efficient estimation of models with conditional moment restrictions containing unknown functions. Working paper, New York University.

Ai, C. & 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 J.J. Heckman & E.E. Learner (eds.), *Handbook of Econometrics*, pp. 5549–5632. Elsevier.

Chen, X., O. Linton, & I. Van Keilegom (2003) Estimation of semiparametric models when the criterion function is not smooth. *Econometrica* 71, 1591–1608.

Chen, X. & X. Shen (1998) Sieve extremum estimates for weakly dependent data. *Econometrica* 66, 289–314.

Donald, S.G., G.W. Imbens, & W.K. Newey (2003) Empirical likelihood estimation and consistent tests with conditional moment restrictions. *Journal of Econometrics* 117, 55–93.

Fan, J. & J. Zhang (2004) Sieve empirical likelihood ratio tests for nonparametric functions. *Annals of Statistics* 32, 1858–1907.

Hansen, L.P. (1982) Large sample properties of generalized method of moments estimators. *Econometrica* 50, 1029–1054.

https://doi.org/10.1017/S0266466610000113 Published online by Cambridge University Press

Hjort, N.L., I.W. McKeague, & I. van Keilegom (2009) Extending the scope of empirical likelihood. *Annals of Statistics* 37, 1079–1111.
Kitamura, Y. (2003) A Likelihood-Based Approach to the Analysis of a Class of Nested and Non-Nested Models. Working paper, University of Pennsylvania.
Kitamura, Y., G. Tripathi, & H. Ahn (2004) Empirical likelihood-based inference in conditional moment restriction models. *Econometrica* 72, 1667–1714.
Newey, W.K. & J.L. Powell (2003) Instrumental variable estimation of nonparametric models. *Econometrica* 71, 1565–1578.
Newey, W.K. & R.J. Smith (2004) Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica* 72, 219–256.
Otsu, T (2007) Penalized empirical likelihood estimation of semiparametric models. *Journal of Multivariate Analysis* 98, 1923–1954.
Owen, A. (2001) *Empirical Likelihood*. Chapman & Hall/CRC.
Qin, J. & J. Lawless (1994) Empirical likelihood for linear models. *Annals of Statistics* 22, 300–325.
Shen, X. (1997) On methods of sieves and penalization. *Annals of Statistics* 25, 2555–2591.
Shen, X. & W.H. Wong (1994) Convergence rate of sieve estimates. *Annals of Statistics* 22, 580–615.
Smith, R.J. (1997) Alternative semi-parametric likelihood approaches to generalized method of moments estimation. *Economic Journal* 107, 503–519.
Smith, R.J. (2007) Local GEL estimation with conditional moment restrictions. In G.D.A. Phillips & E. Tzavalis (eds.), *The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis*, pp. 100–122. Cambridge University Press.
Van der Vaart, A. & J. Wellner (1996) *Weak Convergence and Empirical Process*. Springer-Verlag.
Zhang, J. & I. Gijbels (2003) Sieve empirical likelihood and extensions of the generalized least squares. *Scandinavian Journal of Statistics* 30, 1–24.
Zhang, J. & A. Liu (2003) Local polynomial fitting based on empirical likelihood. *Bernoulli* 9, 579–605.

APPENDIX A: Mathematical Proofs

Hereafter, we denote $\alpha_{n0} \equiv \Pi_n \alpha_0$, $\hat{m}(x_i, \alpha) \equiv \sum_{j=1}^n w_{ji} \rho(z_j, \alpha)$, and $\hat{V}(x_i, \alpha) \equiv \sum_{j=1}^n w_{ji} \rho(z_j, \alpha) \rho(z_j, \alpha)'$.

Proof of Theorem 3.1. The outline of the proof is similar to that of Kitamura et al. (2004, Thm. 3.1). Let $u(x, \alpha) \equiv m(x, \alpha)/(1 + ||m(x, \alpha)||_E)$. Note that

$$\sup_{(x,\alpha) \in \mathcal{X} \times \mathcal{A}_n} \|u(x, \alpha)\|_E \leq 1, \tag{A.1}$$

for all $n \in \mathbb{N}$. Since $\max_{1 \leq i, j \leq n} \sup_{\alpha \in \mathcal{A}_n} \left| n^{-1/p} u(x_i, \alpha)' \rho(z_j, \alpha) \right| = o_{as}(1)$ (by (A.1) and Lemma B.1(i)), $\log\left(1 + \left\{n^{-1/p} u(x_i, \alpha)\right\}' \rho(z_j, \alpha)\right)$ is well-defined a.s. for all n large enough. Since $n^{-1/p} u(x_i, \alpha) \in \Lambda_n = \{\lambda \in \mathbb{R}^{d_\rho} : \|\lambda\|_E \leq \bar{c} n^{-1/p}\}$ a.s. for all $\alpha \in \mathcal{A}_n$, we have

$$\begin{aligned} \hat{\ell}_n(\alpha) &= -\frac{1}{n} \sum_{i=1}^n \mathbb{I}_{in} \max_{\lambda_i \in \Lambda_n} \sum_{j=1}^n w_{ji} \log\left(1 + \lambda_i' \rho(z_j, \alpha)\right) \\ &\leq -\frac{1}{n} \sum_{i=1}^n \mathbb{I}_{in} \sum_{j=1}^n w_{ji} \log\left(1 + \left\{n^{-1/p} u(x_i, \alpha)\right\}' \rho(z_j, \alpha)\right) \equiv \hat{\ell}_n^u(\alpha), \end{aligned} \tag{A.2}$$

a.s. for all $\alpha \in \mathcal{A}_n$ and all n large enough.

https://doi.org/10.1017/S0266466610000113 Published online by Cambridge University Press

On the other hand, if we expand the logs in $\hat{\ell}_n(\alpha_{n0}) \equiv -\frac{1}{n}\sum_{i=1}^n \mathbb{I}_{in}\sum_{j=1}^n w_{ji}\log\left(1+\hat{\lambda}(x_i,\alpha_{n0})'\rho\left(z_j,\alpha_{n0}\right)\right)$, there exists $t_1 \in (0,1)$ such that

$$-\log\left(1+\hat{\lambda}(x_i,\alpha_{n0})'\rho\left(z_j,\alpha_{n0}\right)\right) = -\hat{\lambda}(x_i,\alpha_{n0})'\rho\left(z_j,\alpha_{n0}\right) + \frac{\left(\hat{\lambda}\left(x_i,\alpha_{n0}\right)'\rho\left(z_j,\alpha_{n0}\right)\right)^2}{2\left(1-t_1\hat{\lambda}\left(x_i,\alpha_{n0}\right)'\rho\left(z_j,\alpha_{n0}\right)\right)^2}$$

$$\geq -\hat{\lambda}(x_i,\alpha_{n0})'\rho\left(z_j,\alpha_{n0}\right), \tag{A.3}$$

for all $i, j = 1,\ldots,n$ (the second term in (A.3) is always positive). This implies

$$\hat{\ell}_n(\alpha_{n0}) \geq -\frac{1}{n}\sum_{i=1}^n \mathbb{I}_{in}\hat{\lambda}(x_i,\alpha_{n0})'\hat{m}\left(x_i,\alpha_{n0}\right) \equiv \hat{Q}_n\left(\alpha_{n0}\right). \tag{A.4}$$

Now let $\mathcal{B}_\delta^c \equiv \{\alpha \in \mathcal{A}_n : ||\alpha-\alpha_0||_s \geq \delta\}$ for each $\delta \in (0,\infty)$. If $\hat{\alpha} \in \mathcal{B}_\delta^c$, then $\sup_{\alpha\in\mathcal{B}_\delta^c}\hat{\ell}_n(\alpha) = \sup_{\alpha\in\mathcal{A}_n}\hat{\ell}_n(\alpha) \geq \hat{\ell}_n(\alpha_{n0})$ (because $\alpha_{n0} \in \mathcal{A}_n$). Thus we have

$$\Pr\left\{\hat{\alpha}\in\mathcal{B}_\delta^c\right\} \leq \Pr\left\{\sup_{\alpha\in\mathcal{B}_\delta^c}\hat{\ell}_n(\alpha) \geq \hat{\ell}_n(\alpha_{n0})\right\} \leq \Pr\left\{\sup_{\alpha\in\mathcal{B}_\delta^c}\hat{\ell}_n^u(\alpha) \geq \hat{Q}_n(\alpha_{n0})\right\}, \tag{A.5}$$

for all n large enough, where the second inequality follows from (A.2) and (A.4). Therefore, to derive the consistency of $\hat{\alpha}$ under $\|\cdot\|_s$, it is sufficient to show that

$$\Pr\left\{\sup_{\alpha\in\mathcal{B}_\delta^c}\hat{\ell}_n^u(\alpha) \geq \hat{Q}_n(\alpha_{n0})\right\} \to 0, \tag{A.6}$$

for arbitrary small $\delta \in (0,\infty)$.

An expansion of $\hat{\ell}_n^u(\alpha)$ yields that there exists $t_2 \in (0,1)$ such that

$$\hat{\ell}_n^u(\alpha) = -\frac{1}{n^{1+\frac{1}{p}}}\sum_{i=1}^n \mathbb{I}_{in}u(x_i,\alpha)'\hat{m}(x_i,\alpha) + \frac{1}{n^{1+\frac{2}{p}}}\sum_{i=1}^n \mathbb{I}_{in}\sum_{j=1}^n w_{ji}r_n(x_i,z_j,\alpha), \tag{A.7}$$

where $r_n(x_i,z_j,\alpha) \equiv \frac{\left(u(x_i,\alpha)'\rho(z_j,\alpha)\right)^2}{2\left(1-t_2n^{-1/p}u(x_i,\alpha)'\rho\left(z_j,\alpha\right)\right)^2}$. The first term of (A.7) (multiplied by $n^{1/p}$) satisfies

$$\begin{aligned}
&\sup_{\alpha\in\mathcal{A}_n}\left|-\frac{1}{n}\sum_{i=1}^n \mathbb{I}_{in}u(x_i,\alpha)'\hat{m}\left(x_i,\alpha\right) + \frac{1}{n}\sum_{i=1}^n \mathbb{I}_{in}u\left(x_i,\alpha\right)'m(x_i,\alpha)\right| \\
&\quad\leq \sup_{(x_i,\alpha)\in\mathcal{X}\times\mathcal{A}_n}\|u(x_i,\alpha)\|_E \sup_{(x_i,\alpha)\in\mathcal{X}_n\times\mathcal{A}_n}\left\|\hat{m}(x_i,\alpha)-m(x_i,\alpha)\right\|_E\left(\frac{1}{n}\sum_{i=1}^n \mathbb{I}_{in}\right) \\
&\quad= o_p(1),
\end{aligned} \tag{A.8}$$

https://doi.org/10.1017/S0266466610000113 Published online by Cambridge University Press

where the equality follows from (A.1), $\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{in} = O_p(1)$ (by a law of large numbers), and $\sup_{(x_i,\alpha)\in\mathcal{X}_n\times\mathcal{A}_n}\left\|\hat{m}(x_i,\alpha)-m(x_i,\alpha)\right\|_E = o_p(1)$ (by applying Ai and Chen (1999, Lem. A.1) under Assumptions 3.1, 3.2(ii)–(iii), 3.3, and 3.4). For the second term of (A.7), observe that

$$\begin{aligned}
&\sup_{(x_i,\alpha)\in\mathcal{X}_n\times\mathcal{A}_n}\left|r_n(x_i,z_j,\alpha)\right| \\
&\quad\le \left(\sup_{(x_i,\alpha)\in\mathcal{X}_n\times\mathcal{A}_n}\|u(x_i,\alpha)\|_E\right)^2\left(\sup_{\alpha\in\mathcal{A}_n}\left\|\rho(z_j,\alpha)\right\|_E\right)^2 \\
&\qquad\times \max_{1\le i,j\le n}\sup_{\alpha\in\mathcal{A}_n}\frac{1}{2\left(1-t_2 n^{-1/p}u(x_i,\alpha)'\rho(z_j,\alpha)\right)^2} \\
&\quad\le c_1(z_j)^2 O_{as}(1), \qquad \textbf{(A.9)}
\end{aligned}$$

where the second inequality follows from (A.1), Assumption 3.4(ii), and $\max_{1\le i,j\le n}\sup_{\alpha\in\mathcal{A}_n}\left|n^{-1/p}u(x_i,\alpha)'\rho\left(z_j,\alpha\right)\right| = o_{as}(1)$ (by (A.1) and Lemma B.1(i)). Thus the second term of (A.7) satisfies

$$\begin{aligned}
\sup_{\alpha\in\mathcal{A}_n}\frac{1}{n^{1+\frac{2}{p}}}\sum_{i=1}^{n}\mathbb{I}_{in}\sum_{j=1}^{n}w_{ji}r_n(x_i,z_j,\alpha) &\le O_{as}(n^{-2/p})\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{in}\sum_{j=1}^{n}w_{ji}c_1(z_j)^2\right) \\
&= O_p(n^{-2/p}), \qquad \textbf{(A.10)}
\end{aligned}$$

where the equality follows from the uniform convergence of kernel estimators and Kitamura et al. (2003, Lem. D.4).

Combining (A.7), (A.8), and (A.10),

$$\sup_{\alpha\in\mathcal{A}_n}\left|n^{1/p}\hat{\ell}_n^u(\alpha)+\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{in}u(x_i,\alpha)'m(x_i,\alpha)\right| = o_p(1). \qquad \textbf{(A.11)}$$

Also, observe that

$$\begin{aligned}
&\sup_{\alpha\in\mathcal{A}_n}\left|\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{in}u(x_i,\alpha)'m(x_i,\alpha)-\frac{1}{n}\sum_{i=1}^{n}u(x_i,\alpha)'m(x_i,\alpha)\right| \\
&\quad\le \sup_{\alpha\in\mathcal{A}_n}\left|\left(\frac{1}{n}\sum_{i=1}^{n}(\mathbb{I}_{in}-1)^2\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n}\|u(x_i,\alpha)\|_E^2\|m(x_i,\alpha)\|_E^2\right)^{1/2}\right| \\
&\quad\le \left(\frac{1}{n}\sum_{i=1}^{n}(\mathbb{I}_{in}-1)^2\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left[c_2(z_i)|x_i\right]^2\sup_{\alpha\in\mathcal{A}_n}\|\alpha-\alpha_0\|_s^{2\kappa}\right)^{1/2} \\
&\quad= o_p(1), \qquad \textbf{(A.12)}
\end{aligned}$$

where the first inequality follows from the Cauchy-Schwarz inequality, the second inequality follows from (A.1) and $\|m(x_i,\alpha)\| \le \mathrm{E}[c_2(z_i)|x_i]\,\|\alpha-\alpha_0\|_s^{\kappa}$ (by Assumptions 3.2(i)

https://doi.org/10.1017/S0266466610000113 Published online by Cambridge University Press

and 3.4(iii)), and the equality follows from $\frac{1}{n}\sum_{i=1}^{n}(\mathbb{I}_{in}-1)^2=\frac{1}{n}\sum_{i=1}^{n}(1-\mathbb{I}_{in})=o_p(1)$ (by a law of large numbers and $\mathrm{E}\left[\mathbb{I}_{in}\right]\to 1$), $\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left[c_2(z_i)|x_i\right]^2=O_p(1)$ (by a law of large numbers), and Assumption 3.2(ii).

We now show the following uniform convergence and continuity result:

$$\sup_{\alpha\in\mathcal{A}}\left|\frac{1}{n}\sum_{i=1}^{n}u\left(x_i,\alpha\right)'m(x_i,\alpha)-\mathrm{E}\left[u\left(x_i,\alpha\right)'m(x_i,\alpha)\right]\right|=o_p(1), \tag{A.13}$$

and $\mathrm{E}\left[u(x_i,\alpha)'m(x_i,\alpha)\right]$ is continuous in $\alpha\in\mathcal{A}$. To this end, it is sufficient to check the assumptions of Newey and Powell (2003, Lem. A.2). Observe that (i) $\mathcal{A}$ is compact under the norm $\|\cdot\|_s$ (Assumption 3.2(ii)); (ii) for each $\alpha\in\mathcal{A}$, a law of large numbers yields $\frac{1}{n}\sum_{i=1}^{n}u\left(x_i,\alpha\right)'m(x_i,\alpha)\xrightarrow{p}\mathrm{E}\left[u\left(x_i,\alpha\right)'m(x_i,\alpha)\right]$; and (iii) for each $\alpha,\tilde{\alpha}\in\mathcal{A}$,

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{i=1}^{n}u(x_i,\alpha)'m(x_i,\alpha)-\frac{1}{n}\sum_{i=1}^{n}u\left(x_i,\tilde{\alpha}\right)'m(x_i,\tilde{\alpha})\right|\\
&\quad=\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\|m(x_i,\alpha)\|_E+\|m(x_i,\tilde{\alpha})\|_E+\|m(x_i,\alpha)\|_E\,||m(x_i,\tilde{\alpha})||_E}{(1+||m(x_i,\alpha)||_E)\,(1+||m(x_i,\tilde{\alpha})||_E)}\right)\\
&\qquad\times\left(\|m(x_i,\alpha)\|_E-\|m(x_i,\tilde{\alpha})\|_E\right)\\
&\quad\le\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\|m(x_i,\alpha)\|_E+\|m(x_i,\tilde{\alpha})\|_E+\|m(x_i,\alpha)\|_E\,||m(x_i,\tilde{\alpha})||_E}{(1+||m(x_i,\alpha)||_E)\,(1+||m(x_i,\tilde{\alpha})||_E)}\right)\\
&\qquad\times\|m(x_i,\alpha)-m(x_i,\tilde{\alpha})\|_E\\
&\quad\le\left(\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[c_2(z_i)|x_i]\right)\|\alpha-\tilde{\alpha}\|_s^{\kappa},
\end{aligned}$$

where the equality follows from a straightforward calculation (recall that $u\left(x,\alpha\right)\equiv m\left(x,\alpha\right)/(1+||m(x,\alpha)||_E)$), the first inequality follows from the triangle inequality, and the second inequality follows from $\sup_{x_i\in\mathcal{X}}\left|\frac{\|m(x_i,\alpha)\|_E+\|m(x_i,\tilde{\alpha})\|_E+\|m(x_i,\alpha)\|_E\|m(x_i,\tilde{\alpha})||_E}{(1+||m(x_i,\alpha)||_E)(1+||m(x_i,\tilde{\alpha})||_E)}\right|\le 1$ and Assumption 3.4(iii). Therefore, by applying Newey and Powell (2003, Lem. A.2), we obtain (A.13) and the continuity of $\mathrm{E}\left[-u\left(x_i,\alpha\right)'m(x_i,\alpha)\right]$ in $\alpha\in\mathcal{A}$.

Combining (A.11), (A.12), and (A.13),

$$\sup_{\alpha\in\mathcal{A}_n}\left|n^{1/p}\hat{\ell}_n^{u}\left(\alpha\right)-\mathrm{E}\left[-u(x_i,\alpha)'m(x_i,\alpha)\right]\right|=o_p(1). \tag{A.14}$$

Observe that $\mathrm{E}\left[-u(x_i,\alpha)'m(x_i,\alpha)\right]=\mathrm{E}\left[-\frac{\|m(x_i,\alpha)\|_E^2}{1+||m(x_i,\alpha)||_E}\right]<0$ for all $\alpha\neq\alpha_0$ by the identification condition (Assumption 3.2(i)). The continuity of $\mathrm{E}\left[-u(x_i,\alpha)'m(x_i,\alpha)\right]$ in $\alpha\in\mathcal{A}$ and the compactness of $\mathcal{A}$ imply that for each $\delta\in(0,\infty)$ there exists a constant $H_\delta\in(0,\infty)$ such that $\sup_{\alpha\in\mathcal{B}_\delta^c}\mathrm{E}\left[-u(x_i,\alpha)'m(x_i,\alpha)\right]\le -H_\delta$. Therefore, from (A.14), we have

$$\Pr\left\{\sup_{\alpha\in\mathcal{B}_\delta^c}\hat{\ell}_n^{u}\left(\alpha\right)>-n^{-1/p}H_\delta\right\}<\frac{\delta}{3}, \tag{A.15}$$

for all n large enough.

https://doi.org/10.1017/S0266466610000113 Published online by Cambridge University Press

We now consider $\hat{Q}_n(\alpha_{n0}) \equiv -\frac{1}{n}\sum_{i=1}^n \mathbb{I}_{in}\hat{\lambda}(x_i,\alpha_{n0})'\hat{m}(x_i,\alpha_{n0})$ in (A.5). Observe that

$$\begin{aligned}
\left|\hat{Q}_n(\alpha_{n0})\right| &\le \max_{1\le i\le n}\left\|\hat{\lambda}(x_i,\alpha_{n0})\right\|_E \frac{1}{n}\sum_{i=1}^n \mathbb{I}_{in}\left\|\hat{m}(x_i,\alpha_{n0})\right\|_E \\
&\le \bar{c}n^{-1/p}\frac{1}{n}\sum_{i=1}^n \mathbb{I}_{in}\left\|\hat{m}(x_i,\alpha_{n0})-\hat{m}(x_i,\alpha_0)\right\|_E \\
&\quad + \bar{c}n^{-1/p}\frac{1}{n}\sum_{i=1}^n \mathbb{I}_{in}\left\|\hat{m}(x_i,\alpha_0)\right\|_E \\
&\le \bar{c}n^{-1/p}\left(\frac{1}{n}\sum_{i=1}^n \mathbb{I}_{in}\sum_{j=1}^n w_{ji}c_2(z_j)\right)\|\alpha_{n0}-\alpha_0\|_s^{\kappa} + \bar{c}n^{-1/p}o_p(n^{-1/4}) \\
&= o_p(d_n),
\end{aligned}$$

where $n^{1/p}d_n \to 0$, the second inequality follows from Assumption 3.5 and the triangle inequality, the third inequality follows from Assumption 3.4(iii) and $\sup_{x_i\in\mathcal{X}_n}\left\|\hat{m}(x_i,\alpha_0)-m(x_i,\alpha_0)\right\|_E = o_p(n^{-1/4})$ (obtained by applying Ai and Chen, 1999, Lem. A.1, at $\alpha=\alpha_0$), and the equality follows from Assumption 3.2(iii), the uniform convergence of kernel estimators, and Kitamura et al. (2003, Lem. D.4). Therefore, for each $\delta\in(0,\infty)$, we have

$$\Pr\left\{\hat{Q}_n(\alpha_{n0}) < -d_n H_\delta\right\} < \frac{\delta}{3}, \tag{A.16}$$

for all n large enough. From (A.15) and (A.16), set inclusion relations imply that, for each $\delta\in(0,\infty)$,

$$\begin{aligned}
&\Pr\left\{\sup_{\alpha\in\mathcal{B}_\delta^c}\hat{\ell}_n^u(\alpha) \ge \hat{Q}_n(\alpha_{n0})\right\} \\
&\le \Pr\left\{\sup_{\alpha\in\mathcal{B}_\delta^c}\hat{\ell}_n^u(\alpha) \ge \hat{Q}_n(\alpha_{n0}),\ \sup_{\alpha\in\mathcal{B}_\delta^c}\hat{\ell}_n^u(\alpha) \le -n^{-1/p}H_\delta\right\} \\
&\quad + \Pr\left\{\sup_{\alpha\in\mathcal{B}_\delta^c}\hat{\ell}_n^u(\alpha) > -n^{-1/p}H_\delta\right\} \\
&\le \Pr\left\{\sup_{\alpha\in\mathcal{B}_\delta^c}\hat{\ell}_n^u(\alpha) \ge \hat{Q}_n(\alpha_{n0}),\ \sup_{\alpha\in\mathcal{B}_\delta^c}\hat{\ell}_n^u(\alpha) \le -n^{-1/p}H_\delta,\ \hat{Q}_n(\alpha_{n0}) \ge -d_nH_\delta\right\} \\
&\quad + \Pr\left\{\hat{Q}_n(\alpha_{n0}) < -d_nH_\delta\right\} + \Pr\left\{\sup_{\alpha\in\mathcal{B}_\delta^c}\hat{\ell}_n^u(\alpha) > -n^{-1/p}H_\delta\right\} < \delta
\end{aligned}$$

holds for all n large enough. Therefore, (A.6) is obtained. ■

Proof of Theorem 3.2. We modify the proof of Ai and Chen (2003, Thm. 3.1) to deal with the Lagrange multiplier $\hat{\lambda}(x_i,\alpha)$ in SCEL $\hat{\ell}_n(\alpha)$. Let $0<\eta_{0n}=o(n^{-1/4})$ and $0<$

https://doi.org/10.1017/S0266466610000113 Published online by Cambridge University Press

$\delta_{0n} = 2\sqrt{\eta_{0n}} = o(n^{-1/8})$. First we show that $||\hat{\alpha} - \alpha_0|| = o_p(\delta_{0n}) = o_p(n^{-1/8})$. Define $t(x,\alpha) \equiv V(x,\alpha)^{-1} m(x,\alpha)$. From Assumption 3.6(ii), $V(x,\alpha)^{-1}$ exists uniformly on $(x,\alpha) \in \mathcal{X} \times \mathcal{A}_n$. Observe that

$$\sup_{(x,\alpha)\in\mathcal{X}\times\mathcal{A}_n} \|\eta_{0n} t(x,\alpha)\|_E \leq \eta_{0n} \sup_{(x,\alpha)\in\mathcal{X}\times\mathcal{A}_n} \left\|V(x,\alpha)^{-1}\right\|_E \sup_{x\in\mathcal{X}} \mathrm{E}[c_2(z)|x]$$

$$\times \sup_{\alpha\in\mathcal{A}_n} \|\alpha - \alpha_0\|_s^{\kappa} \leq \bar{c} n^{-1/p} \tag{A.17}$$

holds for all n large enough (note that $p > 8$), where the first inequality follows from Assumptions 3.2(i) and 3.4(iii), and the second inequality follows from Assumptions 3.2(ii)–(iii), 3.4(iii), and 3.6. Now (A.17) implies that for all $\alpha \in \mathcal{A}_n$, $\eta_{0n} t(x_i,\alpha) \in \Lambda_n$ a.s. for all n large enough. Thus we have

$$\hat{\ell}_n(\alpha) = -\frac{1}{n}\sum_{i=1}^n \mathbb{I}_{in} \max_{\lambda_i\in\Lambda_n} \sum_{j=1}^n w_{ji} \log\left(1 + \lambda_i' \rho(z_j,\alpha)\right)$$

$$\leq -\frac{1}{n}\sum_{i=1}^n \mathbb{I}_{in} \sum_{j=1}^n w_{ji} \log\left(1 + \{\eta_{0n} t(x_i,\alpha)\}' \rho(z_j,\alpha)\right) \equiv \hat{\ell}_{0n}(\alpha), \tag{A.18}$$

a.s. for all $\alpha \in \mathcal{A}_n$ and all n large enough. Moreover, from (A.17) and Lemma B.1(i),

$$\max_{1\leq i,j\leq n} \sup_{\alpha\in\mathcal{A}_n} \left\|\eta_{0n} t(x_i,\alpha)' \rho(z_j,\alpha)\right\|_E \leq \bar{c} n^{-1/p} \max_{1\leq j\leq n} \sup_{\alpha\in\mathcal{A}_n} \left\|\rho(z_j,\alpha)\right\|_E = o_{as}(1). \tag{A.19}$$

So, at each $\alpha \in \mathcal{A}_n$, the logs in $\hat{\ell}_{0n}(\alpha)$ are well-defined a.s. for all n large enough. Let $\mathcal{B}^c(k\delta_{0n}) \equiv \{\alpha \in \mathcal{A}_n : ||\alpha - \alpha_0|| \geq k\delta_{0n}, \|\alpha - \alpha_0\|_s = o(1)\}$. Observe that

$$\Pr\left\{\sup_{\alpha\in\mathcal{B}^c(k\delta_{0n})} \hat{\ell}_n(\alpha) \geq \hat{\ell}_n(\alpha_{n0})\right\} \leq \Pr\left\{\sup_{\alpha\in\mathcal{B}^c(k\delta_{0n})} \hat{\ell}_{0n}(\alpha) \geq \hat{Q}_n(\alpha_{n0})\right\},$$

for all $k \in [1,\infty)$ and all n large enough, where the inequality follows from (A.4) and (A.18). Thus, to obtain $||\hat{\alpha} - \alpha_0|| = o_p(n^{-1/8})$, it is sufficient to show that for all $k \in [1,\infty)$,

$$\Pr\left\{\sup_{\alpha\in\mathcal{B}^c(k\delta_{0n})} \hat{\ell}_{0n}(\alpha) \geq \hat{Q}_n(\alpha_{n0})\right\} \to 0. \tag{A.20}$$

Let $L_{0n}(\alpha) \equiv -\frac{1}{n}\sum_{i=1}^n \mathbb{I}_{in} m(x_i,\alpha)' V(x_i,\alpha)^{-1} m(x_i,\alpha)$ and $\mathcal{A}_{0n} \equiv \{\alpha \in \mathcal{A}_n : \|\alpha - \alpha_0\|_s = o(1)\}$. Several set inclusion relations imply

$$\Pr\left\{\sup_{\alpha\in\mathcal{B}^c(k\delta_{0n})} \hat{\ell}_{0n}(\alpha) \geq \hat{Q}_n(\alpha_{n0})\right\}$$

$$\leq \Pr\left\{\left|\hat{Q}_n(\alpha_{n0}) - \eta_{0n} L_{0n}(\alpha_{n0})\right| > \eta_{0n}^2\right\}$$

https://doi.org/10.1017/S0266466610000113 Published online by Cambridge University Press

$$+\Pr\left\{\left\{\left|\hat{Q}_n(\alpha_{n0})-\eta_{0n}L_{0n}(\alpha_{n0})\right|\le\eta_{0n}^2\right\}\cap\left\{\sup_{\alpha\in\mathcal{B}^c(k\delta_{0n})}\hat{\ell}_{0n}(\alpha)\ge\hat{Q}_n(\alpha_{n0})\right\}\right\}$$

$$\le\Pr\left\{\left|\hat{Q}_n(\alpha_{n0})-\eta_{0n}L_{0n}(\alpha_{n0})\right|>\eta_{0n}^2\right\}$$

$$+\Pr\left\{\sup_{\alpha\in\mathcal{B}^c(k\delta_{0n})}\hat{\ell}_{0n}(\alpha)\ge\eta_{0n}L_{0n}(\alpha_{n0})-\eta_{0n}^2\right\}$$

$$\le\Pr\left\{\left|\hat{Q}_n(\alpha_{n0})-\eta_{0n}L_{0n}(\alpha_{n0})\right|>\eta_{0n}^2\right\}+\Pr\left\{\sup_{\alpha\in\mathcal{A}_{0n}}\left|\hat{\ell}_{0n}(\alpha)-\eta_{0n}L_{0n}(\alpha)\right|>\eta_{0n}^2\right\}$$

$$+\Pr\left\{\left\{\sup_{\alpha\in\mathcal{A}_{0n}}\left|\hat{\ell}_{0n}(\alpha)-\eta_{0n}L_{0n}(\alpha)\right|\le\eta_{0n}^2\right\}\cap\left\{\sup_{\alpha\in\mathcal{B}^c(k\delta_{0n})}\hat{\ell}_{0n}(\alpha)\ge\eta_{0n}L_{0n}(\alpha_{n0})-\eta_{0n}^2\right\}\right\}$$

$$\le\Pr\left\{\left|\hat{Q}_n(\alpha_{n0})-\eta_{0n}L_{0n}(\alpha_{n0})\right|>\eta_{0n}^2\right\}+\Pr\left\{\sup_{\alpha\in\mathcal{A}_{0n}}\left|\hat{\ell}_{0n}(\alpha)-\eta_{0n}L_{0n}(\alpha)\right|>\eta_{0n}^2\right\}$$

$$+\Pr\left\{\sup_{\alpha\in\mathcal{B}^c(k\delta_{0n})}\eta_{0n}L_{0n}(\alpha)\ge\eta_{0n}L_{0n}(\alpha_{n0})-2\eta_{0n}^2\right\}\equiv P_1+P_2+P_3, \qquad \textbf{(A.21)}$$

for all n large enough.

For P_1, observe that

$$\left|\hat{Q}_n(\alpha_{n0})-\eta_{0n}L_{0n}(\alpha_{n0})\right|\le\frac{1}{n}\sum_{i=1}^n\mathbb{I}_{in}\left\|\hat{\lambda}(x_i,\alpha_{n0})\right\|_E\left\|\hat{m}(x_i,\alpha_{n0})\right\|_E$$

$$+\frac{\eta_{0n}}{n}\sum_{i=1}^n\mathbb{I}_{in}\left\|V(x_i,\alpha_{n0})^{-1}\right\|_E\left\|m(x_i,\alpha_{n0})\right\|_E^2$$

$$=o_p(n^{-1/2})=o_p(\eta_{0n}^2),$$

where the first equality follows from Lemma B.3(ii) and (viii) and Assumptions 3.2(iv) (i.e., $\|\alpha_{n0}-\alpha_0\|=o(n^{-1/4})$) and 3.6. This implies that $P_1\to 0$.

Now consider P_2. Similar to (A.7), an expansion of $\hat{\ell}_{0n}(\alpha)$ yields that there exists $r_0\in(0,1)$ such that

$$\hat{\ell}_{0n}(\alpha)=-\frac{\eta_{0n}}{n}\sum_{i=1}^n\mathbb{I}_{in}t(x_i,\alpha)'\hat{m}(x_i,\alpha)+\frac{\eta_{0n}^2}{n}\sum_{i=1}^n\mathbb{I}_{in}\sum_{j=1}^n w_{ji}r_{0n}(x_i,z_j,\alpha)\equiv T_1+T_2, \qquad \textbf{(A.22)}$$

https://doi.org/10.1017/S0266466610000113 Published online by Cambridge University Press

where $r_{0n}(x,z,\alpha) \equiv \frac{(t(x,\alpha)'\rho(z,\alpha))^2}{2(1-r_0\eta_{0n}t(x,\alpha)'\rho(z,\alpha))^2}$. Recall that $t(x,\alpha) \equiv V(x,\alpha)^{-1}m(x,\alpha)$. Here T_1 satisfies

$$\begin{aligned}
&\sup_{\alpha\in\mathcal{A}_{0n}} |T_1 - \eta_{0n}L_{0n}(\alpha)| \\
&= \eta_{0n} \sup_{\alpha\in\mathcal{A}_{0n}} \left| \frac{1}{n}\sum_{i=1}^n \mathbb{I}_{in} m(x_i,\alpha)' V(x_i,\alpha)^{-1} \left\{ -\hat{m}(x_i,\alpha) + m(x_i,\alpha) \right\} \right| \\
&\leq \eta_{0n} \sup_{(x_i,\alpha)\in\mathcal{X}_n\times\mathcal{A}_{0n}} \left\| \hat{m}(x_i,\alpha) - m(x_i,\alpha) \right\|_E \sup_{(x_i,\alpha)\in\mathcal{X}_n\times\mathcal{A}_{0n}} \left\| V(x_i,\alpha)^{-1} \right\|_E \\
&\quad \times \sup_{\alpha\in\mathcal{A}_{0n}} \frac{1}{n}\sum_{i=1}^n \mathbb{I}_{in} \| m(x_i,\alpha) \|_E \\
&\leq \eta_{0n} o_p(n^{-1/4}) \left(\frac{1}{n}\sum_{i=1}^n \mathbb{I}_{in} \| \mathrm{E}[c_2(z_i)|x_i] \|_E \right) \sup_{\alpha\in\mathcal{A}_{0n}} \|\alpha-\alpha_0\|_s^\kappa = o_p(\eta_{0n}^2), \quad \textbf{(A.23)}
\end{aligned}$$

where the second inequality follows from Lemma B.2(i) and Assumptions 3.2(i), 3.4(iii), and 3.6, and the last equality follows from $\frac{1}{n}\sum_{i=1}^n \mathbb{I}_{in} \| \mathrm{E}[c_2(z_i)|x_i] \|_E = O_p(1)$ (by a law of large numbers) and Assumption 3.2(ii)–(iii). For T_2, a similar argument to (A.10) yields

$$\begin{aligned}
|T_2| &\leq \eta_{0n}^2 \max_{1\leq i,j\leq n} \sup_{\alpha\in\mathcal{A}_{0n}} \left| \frac{1}{2\left(1-r_0\eta_{0n}t(x_i,\alpha)'\rho(z_j,\alpha)\right)^2} \right| \sup_{(x_i,\alpha)\in\mathcal{X}_n\times\mathcal{A}_{0n}} \|t(x_i,\alpha)\|_E^2 \\
&\quad \times \left(\frac{1}{n}\sum_{i=1}^n \mathbb{I}_{in} \sum_{j=1}^n w_{ji} \sup_{\alpha\in\mathcal{A}_{0n}} \left\| \rho(z_j,\alpha) \right\|_E^2 \right) \\
&\leq \eta_{0n}^2 O_{as}(1) \sup_{(x_i,\alpha)\in\mathcal{X}_n\times\mathcal{A}_{0n}} \left\| V(x_i,\alpha)^{-1} \right\|_E^2 \sup_{x_i\in\mathcal{X}_n} \mathrm{E}[c_2(z_i)|x_i]^2 \\
&\quad \times \sup_{\alpha\in\mathcal{A}_{0n}} \|\alpha-\alpha_0\|_s^{2\kappa} \left(\frac{1}{n}\sum_{i=1}^n \mathbb{I}_{in} \sum_{j=1}^n w_{ji} c_1(z_j)^2 \right) \\
&= o_p(\eta_{0n}^2), \quad \textbf{(A.24)}
\end{aligned}$$

where the second inequality follows from (A.19) and Assumption 3.4(ii)–(iii), and the equality follows from $\sup_{\alpha\in\mathcal{A}_{0n}} \|\alpha-\alpha_0\|_s^{2\kappa} = o(1)$, Assumptions 3.4(iii) and 3.6, the uniform convergence of kernel estimators and Kitamura et al. (2003, Lem. D.4).

Combining (A.22), (A.23), and (A.24), we have

$$\sup_{\alpha\in\mathcal{A}_{0n}} \left| \hat{\ell}_{0n}(\alpha) - \eta_{0n}L_{0n}(\alpha) \right| = o_p(\eta_{0n}^2),$$

which implies that $P_2 \to 0$.

https://doi.org/10.1017/S0266466610000113 Published online by Cambridge University Press

For P_3, our assumptions guarantee the conditions for Shen and Wong (1994, Thm. 1), and then we have $P_3 \to 0$. Therefore, we have (A.20), and $||\hat{\alpha} - \alpha_0|| = o_p(n^{-1/8})$ is obtained.

Next, following the argument by Shen and Wong (1994), we derive an improved convergence rate for $\hat{\alpha}$, that is, $||\hat{\alpha} - \alpha_0|| = o_p(n^{-1/8-1/16})$. Let $0 < \eta_{1n} = n^{-1/4}\delta_{0n} = o(n^{-1/4-1/8})$ and $0 < \delta_{1n} = 2\sqrt{\eta_{1n}} = o(n^{-1/8-1/16})$. Observe that for all $\alpha \in \left\{\alpha \in \mathcal{A}_n : \delta_{0n} \geq ||\alpha - \alpha_0||\right\}$, we have

$$\begin{aligned}
\Pr\{t(x_i,\alpha) \notin \operatorname{int}(\Lambda_n)\} &\leq \Pr\left\{\|t(x_i,\alpha)\|_E \geq \bar{c}n^{-1/p}\right\} \leq \frac{\mathrm{E}\left[\|t(x_i,\alpha)\|_E\right]}{\bar{c}n^{-1/p}} \\
&\leq \bar{c}^{-1} \sup_{(x_i,\alpha)\in\mathcal{X}\times\mathcal{A}_n} \left\|V(x_i,\alpha)^{-1}\right\|_E \times n^{1/p}\mathrm{E}\left[\|m(x_i,\alpha)\|_E\right] \\
&\leq \text{const.} \times n^{1/p}\sqrt{\mathrm{E}\left[\|m(x_i,\alpha)\|_E^2\right]} \\
&\leq \text{const.} \times n^{1/p}\,\|\alpha - \alpha_0\| \to 0,
\end{aligned}$$

where the second inequality follows from the Markov inequality, the fourth inequality follows from Assumption 3.6 and Jensen's inequality, the last inequality follows from Assumption 3.8(ii), and the convergence follows from $p > 8$ in Assumption 3.4(ii). Therefore, for all $\alpha \in \{\alpha \in \mathcal{A}_n : \delta_{0n} \geq ||\alpha - \alpha_0||\}$ and $i = 1, \ldots, n$, we have $t(x_i,\alpha) \in \operatorname{int}(\Lambda_n)$ w.p.a.1, which implies

$$\begin{aligned}
\hat{\ell}_n(\alpha) &= -\frac{1}{n}\sum_{i=1}^n \mathbb{I}_{in} \max_{\lambda_i\in\Lambda_n} \sum_{j=1}^n w_{ji} \log\left(1 + \lambda_i'\rho(z_j,\alpha)\right) \\
&\leq -\frac{1}{n}\sum_{i=1}^n \mathbb{I}_{in} \sum_{j=1}^n w_{ji} \log\left(1 + t(x_i,\alpha)'\rho\left(z_j,\alpha\right)\right) \equiv \hat{\ell}_{1n}(\alpha), \qquad \textbf{(A.25)}
\end{aligned}$$

w.p.a.1. Observe that

$$\begin{aligned}
\Pr\left\{\delta_{0n} \geq ||\hat{\alpha} - \alpha_0|| \geq k\delta_{1n}\right\} &\leq \Pr\left\{\sup_{\{\alpha\in\mathcal{A}_n : \delta_{0n}\geq||\alpha-\alpha_0||\geq k\delta_{1n}\}} \hat{\ell}_n(\alpha) \geq \hat{\ell}_n(\alpha_{n0})\right\} \\
&\leq \Pr\left\{\sup_{\{\alpha\in\mathcal{A}_n : \delta_{0n}\geq||\alpha-\alpha_0||\geq k\delta_{1n}\}} \hat{\ell}_{1n}(\alpha) \geq \hat{Q}_n(\alpha_{n0})\right\},
\end{aligned}$$

where the second inequality follows from (A.4) and (A.25). Thus, to obtain $||\hat{\alpha} - \alpha_0|| = o_p(n^{-1/8-1/16})$, it is sufficient to show that

$$\Pr\left\{\sup_{\{\alpha\in\mathcal{A}_n : \delta_{0n}\geq||\alpha-\alpha_0||\geq k\delta_{1n}\}} \hat{\ell}_{1n}(\alpha) \geq \hat{Q}_n(\alpha_{n0})\right\} \to 0. \qquad \textbf{(A.26)}$$

An expansion of $\hat{\ell}_{1n}(\alpha)$ yields that there exists $r_1 \in (0,1)$ such that

$$\hat{\ell}_{1n}(\alpha) = -\frac{1}{n}\sum_{i=1}^n \mathbb{I}_{in} t(x_i,\alpha)'\hat{m}(x_i,\alpha) + \frac{1}{2n}\sum_{i=1}^n \mathbb{I}_{in} t(x_i,\alpha)'\hat{V}_1(x_i,\alpha)t(x_i,\alpha), \qquad \textbf{(A.27)}$$

https://doi.org/10.1017/S0266466610000113 Published online by Cambridge University Press

where $\hat{V}_1(x_i,\alpha)\equiv\sum_{j=1}^n \frac{w_{ji}\rho(z_j,\alpha)\rho(z_j,\alpha)'}{\left(1-r_1 t(x_i,\alpha)'\rho(z_j,\alpha)\right)^2}$, and r_1 depends on x_i and α. Let $V_1(x_i,\alpha)\equiv \mathrm{E}\left[\left.\frac{\rho(z_j,\alpha)\rho(z_j,\alpha)'}{\left(1-r_1 t(x_i,\alpha)'\rho(z_j,\alpha)\right)^2}\right| x_i\right]$ and

$$L_{1n}(\alpha)\equiv -\frac{1}{n}\sum_{i=1}^n \mathbb{I}_{in} t(x_i,\alpha)' m(x_i,\alpha)+\frac{1}{2n}\sum_{i=1}^n \mathbb{I}_{in} t(x_i,\alpha)' V_1(x_i,\alpha) t(x_i,\alpha).$$

For each $k\in[1,\infty)$, several set inclusion relations imply

$$\begin{aligned}
&\Pr\left\{\sup_{\{\alpha\in\mathcal{A}_n:\delta_{0n}\geq||\alpha-\alpha_0||\geq k\delta_{1n}\}} \hat{\ell}_{1n}(\alpha)\geq \hat{Q}_n(\alpha_{n0})\right\}\\
&\quad\leq \Pr\left\{\left|\hat{Q}_n(\alpha_{n0})-L_{1n}(\alpha_{n0})\right|>\eta_{1n}\right\}\\
&\qquad+\Pr\left\{\left\{\left|\hat{Q}_n(\alpha_{n0})-L_{1n}(\alpha_{n0})\right|\leq\eta_{1n}\right\}\right.\\
&\qquad\qquad\left.\cap\left\{\sup_{\{\alpha\in\mathcal{A}_n:\delta_{0n}\geq||\alpha-\alpha_0||\geq k\delta_{1n}\}} \hat{\ell}_{1n}(\alpha)\geq \hat{Q}_n(\alpha_{n0})\right\}\right\}\\
&\quad\leq \Pr\left\{\left|\hat{Q}_n(\alpha_{n0})-L_{1n}(\alpha_{n0})\right|>\eta_{1n}\right\}\\
&\qquad+\Pr\left\{\sup_{\{\alpha\in\mathcal{A}_n:\delta_{0n}\geq||\alpha-\alpha_0||\geq k\delta_{1n}\}} \hat{\ell}_{1n}(\alpha)\geq L_{1n}(\alpha_{n0})-\eta_{1n}\right\}\\
&\quad\leq \Pr\left\{\left|\hat{Q}_n(\alpha_{n0})-L_{1n}(\alpha_{n0})\right|>\eta_{1n}\right\}\\
&\qquad+\Pr\left\{\sup_{\{\alpha\in\mathcal{A}_n:\delta_{0n}\geq||\alpha-\alpha_0||\}}\left|\hat{\ell}_{1n}(\alpha)-L_{1n}(\alpha)\right|>\eta_{1n}\right\}\\
&\qquad+\Pr\left\{\begin{array}{c}\left\{\sup_{\{\alpha\in\mathcal{A}_n:\delta_{0n}\geq||\alpha-\alpha_0||\}}\left|\hat{\ell}_{1n}(\alpha)-L_{1n}(\alpha)\right|\leq\eta_{1n}\right\}\\ \cap\left\{\sup_{\{\alpha\in\mathcal{A}_n:\delta_{0n}\geq||\alpha-\alpha_0||\geq k\delta_{1n}\}}\hat{\ell}_{1n}(\alpha)\geq L_{1n}(\alpha_{n0})-\eta_{1n}\right\}\end{array}\right\}\\
&\quad\leq \Pr\left\{\left|\hat{Q}_n(\alpha_{n0})-L_{1n}(\alpha_{n0})\right|>\eta_{1n}\right\}\\
&\qquad+\Pr\left\{\sup_{\{\alpha\in\mathcal{A}_n:\delta_{0n}\geq||\alpha-\alpha_0||\}}\left|\hat{\ell}_{1n}(\alpha)-L_{1n}(\alpha)\right|>\eta_{1n}\right\}\\
&\qquad+\Pr\left\{\sup_{\{\alpha\in\mathcal{A}_n:\delta_{0n}\geq||\alpha-\alpha_0||\geq k\delta_{1n}\}} L_{1n}(\alpha)\geq L_{1n}(\alpha_{n0})-2\eta_{1n}\right\}\\
&\quad\equiv P_4+P_5+P_6,
\end{aligned}$$

for all n large enough.

https://doi.org/10.1017/S0266466610000113 Published online by Cambridge University Press

For P_4, observe that

$$\begin{aligned}\left|\hat{Q}_n(\alpha_{n0}) - L_{1n}(\alpha_{n0})\right| &\leq \frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{in}\left\|\hat{\lambda}(x_i,\alpha_{n0})\right\|_E \left\|\hat{m}(x_i,\alpha_{n0})\right\|_E \\ &\quad + \frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{in}\left\|t(x_i,\alpha_{n0})\right\|_E \left\|m(x_i,\alpha_{n0})\right\|_E \\ &\quad + \frac{1}{2n}\sum_{i=1}^{n}\mathbb{I}_{in}\left\|V_1(x_i,\alpha_{n0})\right\|_E \left\|t(x_i,\alpha_{n0})\right\|_E^2 \\ &= o_p(n^{-1/2}),\end{aligned}$$

where the equality follows from Lemmas B.1(ii) and B.3(ii) and (viii), and Assumption 3.6. This implies that $P_4 \to 0$.

For P_5, observe that

$$\begin{aligned}&\sup_{\{\alpha\in\mathcal{A}_n:\delta_{0n}\geq||\alpha-\alpha_0||\}}\left|\hat{\ell}_{1n}(\alpha) - L_{1n}(\alpha)\right| \\ &\leq \sup_{\{\alpha\in\mathcal{A}_n:\delta_{0n}\geq||\alpha-\alpha_0||\}}\left|-\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{in}t(x_i,\alpha)'\left\{\hat{m}(x_i,\alpha) - m(x_i,\alpha)\right\}\right| \\ &\quad + \sup_{\{\alpha\in\mathcal{A}_n:\delta_{0n}\geq||\alpha-\alpha_0||\}}\left|\frac{1}{2n}\sum_{i=1}^{n}\mathbb{I}_{in}t(x_i,\alpha)'\left\{\hat{V}_1(x_i,\alpha) - V_1(x_i,\alpha)\right\}t(x_i,\alpha)\right| \\ &\leq \sup_{(x_i,\alpha)\in\mathcal{X}_n\times\{\alpha\in\mathcal{A}_n:\delta_{0n}\geq||\alpha-\alpha_0||\}}\left\|\hat{m}(x_i,\alpha) - m(x_i,\alpha)\right\|_E \sup_{\{\alpha\in\mathcal{A}_n:\delta_{0n}\geq||\alpha-\alpha_0||\}} \\ &\quad \times\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{in}\left\|t(x_i,\alpha)\right\|_E + \sup_{(x_i,\alpha)\in\mathcal{X}_n\times\{\alpha\in\mathcal{A}_n:\delta_{0n}\geq||\alpha-\alpha_0||\}}\left\|\hat{V}_1(x_i,\alpha) - V_1(x_i,\alpha)\right\|_E \\ &\quad \times\sup_{\{\alpha\in\mathcal{A}_n:\delta_{0n}\geq||\alpha-\alpha_0||\}}\frac{1}{2n}\sum_{i=1}^{n}\mathbb{I}_{in}\left\|t(x_i,\alpha)\right\|_E^2 \\ &= o_p(n^{-1/4}\delta_{0n}) + o_p(n^{-1/4}\delta_{0n}^2) = o_p(\eta_{1n}),\end{aligned}$$

where the first equality follows from Assumption 3.6, and Lemmas B.2(i) and B.3(i)–(ii), and $\sup_{(x_i,\alpha)\in\mathcal{X}_n\times\{\alpha\in\mathcal{A}_n:\delta_{0n}\geq||\alpha-\alpha_0||\}}\left\|\hat{V}_1(x_i,\alpha) - V_1(x_i,\alpha)\right\|_E = o_p(n^{-1/4})$ (obtained by applying Ai and Chen (1999, Lem. A.1)). This implies that $P_5 \to 0$.

For P_6, observe that

$$P_6 = \Pr\left\{\sup_{\{\alpha\in\mathcal{A}_n:\delta_{0n}\geq||\alpha-\alpha_0||\geq k\delta_{1n}\}} L_{1n}(\alpha) - L_{1n}(\alpha_0) \geq L_{1n}(\alpha_{n0}) - L_{1n}(\alpha_0) - 2\eta_{1n}\right\},$$

and our assumptions guarantee that the conditions of Shen and Wong (1994, Thm. 1) hold, which yields $P_6 \to 0$.

https://doi.org/10.1017/S0266466610000113 Published online by Cambridge University Press

Therefore, (A.26) is obtained, i.e., $||\hat{\alpha} - \alpha_0|| = o_p(n^{-1/8-1/16})$. By repeating this refinement infinitely many times, we have $||\hat{\alpha} - \alpha_0|| = o_p(n^{-1/8(1+1/2+1/4+\cdots)}) = o_p(n^{-1/4})$. ■

Proof of Theorem 3.3. We first show that $\hat{\lambda}(x_i, \alpha)$ satisfies a first-order condition w.p.a.1 for all $\alpha \in \mathcal{N}_{0n} \equiv \left\{\alpha \in \mathcal{A}_n : \|\alpha - \alpha_0\|_s = o(1), \|\alpha - \alpha_0\| = o(n^{-1/4})\right\}$. Observe that, for all $\alpha \in \mathcal{N}_{0n}$,

$$\begin{aligned}\mathbb{I}_{in}\left\|\hat{\lambda}(x_i,\alpha)\right\|_E &\le C^{-1}\mathbb{I}_{in}\left\|\hat{m}(x_i,\alpha)\right\|_E \\ &\le C^{-1}\mathbb{I}_{in}\left\|m(x_i,\alpha)\right\|_E + C^{-1}\mathbb{I}_{in}\left\|\hat{m}(x_i,\alpha) - m(x_i,\alpha)\right\|_E,\end{aligned}$$

w.p.a.1, where the first inequality follows from (B.3) in the proof of Lemma B.3(vi) and the second inequality follows from the triangle inequality. This implies that, for all $\alpha \in \mathcal{N}_{0n}$,

$$\begin{aligned}\Pr\left\{\mathbb{I}_{in}\hat{\lambda}(x_i,\alpha) \notin \operatorname{int}(\Lambda_n)\right\} &\le \Pr\left\{\mathbb{I}_{in}\left\|\hat{\lambda}(x_i,\alpha)\right\|_E \ge \bar{c}n^{-1/p}\right\} \\ &\le \frac{C^{-1}\mathrm{E}\left[\mathbb{I}_{in}\left\|m(x_i,\alpha)\right\|_E\right] + o(n^{-1/4})}{\bar{c}n^{-1/p}} \to 0,\end{aligned}$$

i.e., $\mathbb{I}_{in}\hat{\lambda}(x_i, \alpha) \in \operatorname{int}(\Lambda_n)$, w.p.a.1, where the second inequality follows from the Markov inequality and the proof of Ai and Chen (1999, Lem. A.1), and the convergence follows from Assumption 3.8(ii) and $p > 8$. Therefore, $\hat{\lambda}(x_i, \alpha)$ satisfies the following first-order condition for all $\alpha \in \mathcal{N}_{0n}$ w.p.a.1:

$$0 = \mathbb{I}_{in}\sum_{j=1}^{n}\frac{w_{ji}\rho(z_j,\alpha)}{1+\hat{\lambda}(x_i,\alpha)'\rho(z_j,\alpha)}. \tag{A.28}$$

Now let $0 < \epsilon_n = o(n^{-1/2})$, $u^* \equiv \pm v^*$, and $u_n^* \equiv \Pi_n u^*$. Consider a continuous path $\{\alpha(\tau) : \tau \in [0,1]\}$ in $\mathcal{N}_{0n}$ such that $\alpha(0) = \hat{\alpha}$ and $\alpha(1) = \hat{\alpha} + \epsilon_n u_n^* \in \mathcal{N}_{0n}$, and denote

$$\frac{d\hat{m}(x_i,\alpha(\tau))}{d\alpha}\left[\epsilon_n u_n^*\right] \equiv \left.\frac{d\hat{m}(x_i,\alpha(t))}{dt}\right|_{t=\tau}, \qquad \frac{d\hat{\lambda}(x_i,\alpha(\tau))}{d\alpha}[\epsilon_n u_n^*] \equiv \left.\frac{d\hat{\lambda}(x_i,\alpha(t))}{dt}\right|_{t=\tau}.$$

An expansion of $\hat{\ell}_n(\alpha(t))$ around $t = 0$ yields

$$0 \le \hat{\ell}_n(\hat{\alpha}) - \hat{\ell}_n(\hat{\alpha} + \epsilon_n u_n^*) = \hat{\ell}_n(\alpha(0)) - \hat{\ell}_n(\alpha(1)) = -\left.\frac{d\hat{\ell}_n(\alpha(t))}{dt}\right|_{t=0} - \frac{1}{2}\left.\frac{d^2\hat{\ell}_n(\alpha(t))}{dt^2}\right|_{t=s}, \tag{A.29}$$

https://doi.org/10.1017/S0266466610000113 Published online by Cambridge University Press

for some $s \in (0,1)$, where the inequality follows from the fact that $\hat{\alpha}$ maximizes $\hat{\ell}_n(\alpha)$. First, consider the first term on the right-hand side of (A.29). Observe that

$$
\begin{aligned}
-\left.\frac{d\hat{\ell}_n(\alpha(t))}{dt}\right|_{t=0} &= \frac{1}{n}\sum_{i=1}^n \mathbb{I}_{in} \sum_{j=1}^n \frac{w_{ji}\hat{\lambda}(x_i,\hat{\alpha})' \frac{d\rho(z_j,\hat{\alpha})}{d\alpha}[\epsilon_n u_n^*]}{1+\hat{\lambda}(x_i,\hat{\alpha})'\rho(z_j,\hat{\alpha})} \\
&= \frac{1}{n}\sum_{i=1}^n \mathbb{I}_{in}\hat{\lambda}(x_i,\hat{\alpha})' \frac{d\hat{m}(x_i,\hat{\alpha})}{d\alpha}\left[\epsilon_n u_n^*\right] + \frac{1}{n}\sum_{i=1}^n \mathbb{I}_{in}\hat{\lambda}(x_i,\hat{\alpha})' \\
&\quad \times \left(\sum_{j=1}^n \frac{w_{ji}\frac{d\rho(z_j,\hat{\alpha})}{d\alpha}[\epsilon_n u_n^*]}{1+\hat{\lambda}(x_i,\hat{\alpha})'\rho(z_j,\hat{\alpha})} - \sum_{j=1}^n w_{ji}\frac{d\rho(z_j,\hat{\alpha})}{d\alpha}[\epsilon_n u_n^*] \right) \\
&= \frac{\epsilon_n}{n}\sum_{i=1}^n \mathbb{I}_{in}\hat{\lambda}(x_i,\hat{\alpha})' \frac{d\hat{m}(x_i,\hat{\alpha})}{d\alpha}\left[u_n^*\right] \\
&\quad - \frac{\epsilon_n}{n}\sum_{i=1}^n \mathbb{I}_{in}\hat{\lambda}(x_i,\hat{\alpha})' \left(\sum_{j=1}^n \frac{w_{ji}\frac{d\rho(z_j,\hat{\alpha})}{d\alpha}[u_n^*]\rho(z_j,\hat{\alpha})'}{1+\hat{\lambda}(x_i,\hat{\alpha})'\rho(z_j,\hat{\alpha})} \right) \hat{\lambda}(x_i,\hat{\alpha}) \\
&= \frac{\epsilon_n}{n}\sum_{i=1}^n \mathbb{I}_{in}\hat{m}(x_i,\hat{\alpha})'\hat{V}(x_i,\hat{\alpha})^{-1}\frac{d\hat{m}(x_i,\hat{\alpha})}{d\alpha}\left[u_n^*\right] + o_p(\epsilon_n n^{-1/2}) \\
&= \frac{\epsilon_n}{n}\sum_{i=1}^n \mathbb{I}_{in}\hat{m}(x_i,\hat{\alpha})'V(x_i,\alpha_0)^{-1}\frac{dm(x_i,\alpha_0)}{d\alpha}\left[u^*\right] + o_p(\epsilon_n n^{-1/2}) \\
&= \frac{\epsilon_n}{n}\sum_{i=1}^n \mathbb{I}_{in}\hat{m}(x_i,\alpha_0)'V(x_i,\alpha_0)^{-1}\frac{dm(x_i,\alpha_0)}{d\alpha}\left[u^*\right] \\
&\quad + \left\langle u^*, \hat{\alpha}-\alpha_0 \right\rangle + o_p(\epsilon_n n^{-1/2}) \\
&= \frac{\epsilon_n}{n}\sum_{i=1}^n \rho(z_i,\alpha_0)'V(x_i,\alpha_0)^{-1}\frac{dm(x_i,\alpha_0)}{d\alpha}\left[u^*\right] \\
&\quad + \left\langle u^*, \hat{\alpha}-\alpha_0 \right\rangle + o_p(\epsilon_n n^{-1/2}), \qquad \textbf{(A.30)}
\end{aligned}
$$

where the first equality follows from the chain rule and (A.28), the third equality follows from a straightforward calculation, the fourth equality follows from Lemma B.4(ii) and (iii), the fifth equality follows from Lemma B.4(iv), the sixth equality follows from Lemma B.4(v), and the last equality follows from Lemma B.4(vi).

https://doi.org/10.1017/S0266466610000113 Published online by Cambridge University Press

Next we consider the second term on the right-hand side of (A.29). Let $\frac{d\phi(\alpha(s))}{d\alpha}[\epsilon_n u_n^*] \equiv \frac{d\phi(\alpha(t))}{dt}\Big|_{t=s}$ and $\frac{d^2\phi(\alpha(s))}{d\alpha d\alpha}[\epsilon_n u_n^*, \epsilon_n u_n^*] \equiv \frac{d^2\phi(\alpha(t))}{dt^2}\Big|_{t=s}$ be the pathwise derivatives for some function ϕ. Similar to Kitamura et al. (2004, p. 1706), the chain rule yields

$$\begin{aligned}
&\frac{d^2\ell_n(\alpha(t))}{dt^2}\Big|_{t=s} \\
&= \frac{\epsilon_n^2}{n}\sum_{i=1}^n \mathbb{I}_{in}\sum_{j=1}^n \frac{w_{ji}\left\{\frac{d\hat{\lambda}(x_i,\alpha(s))'\rho(z_j,\alpha(s))}{d\alpha}[u_n^*]\right\}\hat{\lambda}(x_i,\alpha(s))'\frac{d\rho(z_j,\alpha(s))}{d\alpha}[u_n^*]}{\left(1+\hat{\lambda}(x_i,\alpha(s))'\rho(z_j,\alpha(s))\right)^2} \\
&\quad -\frac{\epsilon_n^2}{n}\sum_{i=1}^n \mathbb{I}_{in}\sum_{j=1}^n \frac{w_{ji}\frac{d\hat{\lambda}(x_i,\alpha(s))}{d\alpha}[u_n^*]'\frac{d\rho(z_j,\alpha(s))}{d\alpha}[u_n^*]}{1+\hat{\lambda}(x_i,\alpha(s))'\rho(z_j,\alpha(s))} \\
&\quad -\frac{\epsilon_n^2}{n}\sum_{i=1}^n \mathbb{I}_{in}\sum_{j=1}^n \frac{w_{ji}}{1+\hat{\lambda}(x_i,\alpha(s))'\rho(z_j,\alpha(s))}\sum_{k=1}^{d_\rho}\left\{\frac{d^2\rho_k(z_j,\alpha(s))}{d\alpha d\alpha}[u_n^*,u_n^*]\right\} \\
&\quad \times\hat{\lambda}_k(x_i,\alpha(s)) \\
&= o_p(\epsilon_n^2), \qquad \textbf{(A.31)}
\end{aligned}$$

where the second equality follows from Lemma B.5.

Finally, combining (A.29)–(A.39) and $u^* \equiv \pm v^*$, we have

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

for any fixed nonzero $\xi \in R^{d_\theta}$. Therefore, a central limit theorem implies the conclusion. ■

Proof of Theorem 4.1. To obtain the conclusion, it is sufficient to verify Assumptions 3.1–3.11. Assumption 3.1 is directly assumed. From Newey and Powell (2003), condition (i) guarantees Assumption 3.2(i). Since $\mathcal{H} = \Lambda_c^\gamma\left(\mathbb{R}^d\right)$ is compact under $\|\alpha\|_s = \|\theta\|_E + \sup_{y_2\in\mathbb{R}^d}\left|h(y_2)\left(1+\|y_2\|_E^2\right)^{-a/2}\right|$, Assumption 3.2(ii) is satisfied. From the proof of Ai and Chen (2003, Prop. 3.1), we have $\|\alpha-\Pi_n\alpha\|_s \leq \text{const.}\times k_n^{-\gamma/d}$. Thus condition (v) guarantees Assumption 3.2(iii) and (iv). Assumption 3.3 is directly assumed. For each $\alpha\in\mathcal{A}_n$, we have $m(x,\alpha) = \mathrm{E}[y_1|x] - x_1'\theta - \mathrm{E}[q^{k_n}(y_2)|x]'\beta \in \Lambda_c^{J'}(\mathcal{X})$ by condition (iii). Thus Assumption 3.4(i) is satisfied. Observe that $\sup_{\alpha\in\mathcal{A}_n}|\rho(z,\alpha)| \leq c_1(z) \equiv |y_1| + \sup_{\theta\in\Theta}\left|x_1'\theta\right| + \sup_{h\in\mathcal{H}_n}\|h\|_{\Lambda^\gamma}$. Since $\mathrm{E}[|y_1|^p] < \infty$ (by condition (ii)(a)), $\mathrm{E}[\sup_{\theta\in\Theta}\left|x_1'\theta\right|^p] < \infty$ (by Assumption 3.1 and the compactness of Θ), and $\mathrm{E}[\left(\sup_{h\in\mathcal{H}_n}\right.$

https://doi.org/10.1017/S0266466610000113 Published online by Cambridge University Press

$\left\|h\right\|_{\Lambda^\gamma}\Big)^p] < c$ (by $\mathcal{H}_n \subset \Lambda_c^\gamma\left(\mathbb{R}^d\right)$), Assumption 3.4(ii) is satisfied. Similar to the proof of Ai and Chen (2003, Prop. 3.1),

$$\begin{aligned}|\rho(z,\alpha_1)-\rho(z,\alpha_2)| &\le \|x_1\|_E \|\theta_1-\theta_2\|_E \\ &\quad + \left(1+\|y_2\|_E^2\right)^{a/2} \sup_{y_2\in\mathbb{R}^d}\left|\{h_1(y_2)-h_2(y_2)\}\left(1+\|y_2\|_E^2\right)^{-a/2}\right| \\ &\le c_2(z)\,\|\alpha-\alpha_0\|_s,\end{aligned}$$

where $c_2(z) = \|x_1\|_E + \left(1+\|y_2\|_E^2\right)^{a/2}$, and $\mathrm{E}\left[c_2(z)^2|x\right] < \infty$ (by Assumption 3.1 and condition (ii)(b)). Thus, Assumption 3.4(iii) is satisfied with $\kappa = 1$. Assumption 3.5 is directly assumed. For each $\alpha \in \mathcal{A}_n$, we have

$$\begin{aligned}V(x,\alpha) &= \mathrm{E}[y_1^2|x] + \left|x_1'\theta\right|^2 + \beta'\mathrm{E}\left[q^{k_n}(y_2)q^{k_n}(y_2)'|x\right]\beta - 2\left(x_1'\theta\right)\mathrm{E}[y_1|x] \\ &\quad -2\mathrm{E}\left[y_1 q^{k_n}(y_2)|x\right]'\beta + 2\left(x_1'\theta\right)\mathrm{E}\left[q^{k_n}(y_2)|x\right]'\beta. \qquad \textbf{(A.32)}\end{aligned}$$

Since each term in (A.32) belongs to $\Lambda_c^\gamma\left(\mathbb{R}^d\right)$ by conditions (iii) and (iv), Assumption 3.6(i) is satisfied. Assumption 3.6(ii) is directly assumed. From Chen and Shen (1998), we can see that Assumption 3.7 is satisfied. Observe that $\frac{d\rho(z,\alpha_0)}{d\alpha}[\alpha-\alpha_0] = -x_1'(\theta-\theta_0) - (h(y_2)-h_0(y_2)) = \rho(z,\alpha)-\rho(z,\alpha_0)$, which implies $\frac{dm(x,\alpha_0)}{d\alpha}[\alpha-\alpha_0] = m(x,\alpha)$. Thus, in this case, the norm $\|\cdot\|$ is written as $\|\alpha-\alpha_0\|^2 = \mathrm{E}\left[m(x,\alpha)'V(x,\alpha_0)^{-1}m(x,\alpha)\right]$, and Assumption 3.8 is satisfied. Assumption 3.9(i) is directly assumed. Condition (i)(d) guarantees Assumption 3.9(ii). Under conditions (iii) and (vi) with our specification of $\mathcal{H}_n$, Assumption 3.9(iii) is satisfied. Observe that $\frac{d\rho(z,\tilde{\alpha})}{d\alpha}[v] = -x_1'v_\theta - v_h(y_2)$. Thus, Assumption 3.10(i) is satisfied. Since $\frac{dm(x,\tilde{\alpha})}{d\alpha}[v] = -x_1'v_\theta - v_h(y_2)$, conditions (iii) and (vi) guarantee Assumption 3.10(ii). From $\frac{d^2\rho(z,\tilde{\alpha})}{d\alpha d\alpha}[v,v] = 0$ for all α and v, Assumption 3.10(iii) is trivially satisfied. Since $\frac{dm(x,\tilde{\alpha})}{d\alpha}[v] - \frac{dm(x,\alpha_0)}{d\alpha}[v] = 0$ for all α and v, Assumption 3.11 always holds. Therefore, all the assumptions are verified. ■

APPENDIX B: Lemmas

LEMMA B.1 (Kitamura et al., 2004, Lem. D.2).

(i) Suppose that Assumptions 3.1(i) and 3.4(ii) hold. Then

$$\max_{1\le j\le n}\sup_{\alpha\in\mathcal{A}_n}\left\|\rho\left(z_j,\alpha\right)\right\|_E = o_{as}(n^{1/p}).$$

(ii) Suppose that Assumptions 3.1(i), 3.4(ii), and 3.5 hold. Then

$$\max_{1\le j\le n}\sup_{(x_i,\alpha)\in\mathcal{X}\times\mathcal{A}_n}\left|\hat{\lambda}(x_i,\alpha)'\rho\left(z_j,\alpha\right)\right| = o_{as}(1).$$

https://doi.org/10.1017/S0266466610000113 Published online by Cambridge University Press

LEMMA B.2 (Applications of Ai and Chen, 1999, Lem. A.1).

(i) Suppose that Assumptions 3.1, 3.3, 3.4, and 3.7 hold. Then

$$\sup_{(x_i,\alpha)\in\mathcal{X}_n\times\mathcal{A}_n} \left\|\hat{m}(x_i,\alpha)-m(x_i,\alpha)\right\|_E = o_p(n^{-1/4}).$$

(ii) Suppose that Assumptions 3.1, 3.3, 3.4, 3.6, and 3.7 hold. Then

$$\sup_{(x_i,\alpha)\in\mathcal{X}_n\times\mathcal{A}_n} \left\|\hat{V}(x_i,\alpha)-V(x_i,\alpha)\right\|_E = o_p(n^{-1/4}),$$

$$\sup_{(x_i,\alpha)\in\mathcal{X}_n\times\mathcal{A}_n} \left\|\hat{V}(x_i,\alpha)^{-1}-V(x_i,\alpha)^{-1}\right\|_E = o_p(n^{-1/4}).$$

(iii) Suppose that Assumptions 3.1, 3.3, 3.7, and 3.10(i)–(iii) hold. Then

$$\sup_{(x_i,\tilde{\alpha})\in\mathcal{X}_n\times\mathcal{N}_{0n}} \left\|\frac{d\hat{m}(x_i,\tilde{\alpha})}{d\alpha}\left[v_n^*\right]-\frac{dm(x_i,\tilde{\alpha})}{d\alpha}\left[v_n^*\right]\right\|_E = o_p(n^{-1/4}).$$

LEMMA B.3. *Let $\eta_n = n^{-\tau}$ with $\tau\in(0,1/4]$ and $\mathcal{B}(\eta_n)\equiv\{\alpha\in\mathcal{A}_n\colon\|\alpha-\alpha_0\|\le o(\eta_n)\}$. Suppose that Assumptions 3.1, 3.2(i), 3.3, 3.4, 3.7, and 3.8 hold. Then*

(i) $$\sup_{\alpha\in\mathcal{B}(\eta_n)}\frac{1}{n}\sum_{i=1}^n \mathbb{I}_{in}\|m(x_i,\alpha)\|_E = o_p(\eta_n),$$

(ii) $$\sup_{\alpha\in\mathcal{B}(\eta_n)}\frac{1}{n}\sum_{i=1}^n \mathbb{I}_{in}\|m(x_i,\alpha)\|_E^2 = o_p(\eta_n^2),$$

(iii) $$\sup_{\alpha\in\mathcal{B}(\eta_n)}\frac{1}{n}\sum_{i=1}^n \mathbb{I}_{in}\left\|\hat{m}(x_i,\alpha)\right\|_E\|m(x_i,\alpha)\|_E = o_p(\eta_n^2),$$

(iv) $$\sup_{\alpha\in\mathcal{B}(\eta_n)}\frac{1}{n}\sum_{i=1}^n \mathbb{I}_{in}\left\|\hat{m}(x_i,\alpha)\right\|_E = o_p(\eta_n),$$

(v) $$\sup_{\alpha\in\mathcal{B}(\eta_n)}\frac{1}{n}\sum_{i=1}^n \mathbb{I}_{in}\left\|\hat{m}(x_i,\alpha)\right\|_E^2 = o_p(\eta_n^2).$$

Suppose that Assumptions 3.1, 3.2(i), and 3.3–3.8 hold. Then

(vi) $$\sup_{\alpha\in\mathcal{B}(\eta_n)}\frac{1}{n}\sum_{i=1}^n \mathbb{I}_{in}\left\|\hat{\lambda}(x_i,\alpha)\right\|_E = o_p(\eta_n),$$

(vii) $$\sup_{\alpha\in\mathcal{B}(\eta_n)}\frac{1}{n}\sum_{i=1}^n \mathbb{I}_{in}\left\|\hat{\lambda}(x_i,\alpha)\right\|_E^2 = o_p(\eta_n^2),$$

(viii) $$\sup_{\alpha\in\mathcal{B}(\eta_n)}\frac{1}{n}\sum_{i=1}^n \mathbb{I}_{in}\left\|\hat{\lambda}(x_i,\alpha)\right\|_E\left\|\hat{m}(x_i,\alpha)\right\|_E = o_p(\eta_n^2).$$

https://doi.org/10.1017/S0266466610000113 Published online by Cambridge University Press

Proof of (i). The proof is similar to that of Ai and Chen (2003, Cor. A.2(ii)). From $\mathbb{I}_{in} \leq 1$ and the triangle inequality,

$$\begin{aligned}\sup_{\alpha\in\mathcal{B}(\eta_n)} \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{in}\|m(x_i,\alpha)\|_E &\leq \sup_{\alpha\in\mathcal{B}(\eta_n)} \frac{1}{n}\sum_{i=1}^{n} \|m(x_i,\alpha)\|_E \\ &\leq \sup_{\alpha\in\mathcal{B}(\eta_n)} \left|\frac{1}{n}\sum_{i=1}^{n} \|m(x_i,\alpha)\|_E - \mathrm{E}\left[\|m(x_i,\alpha)\|_E\right]\right| \\ &\quad + \sup_{\alpha\in\mathcal{B}(\eta_n)} \mathrm{E}\left[\|m(x_i,\alpha)\|_E\right] \equiv T_1 + T_2.\end{aligned}$$

To obtain $T_1 = o_p(n^{-1/2})$, it is sufficient to check the conditions of Chen, Linton, and van Keilegom (2003, Lem. 1). Observe that (i) $m(x_i,\alpha_0) = 0$ (by Assumption 3.2(i)); (ii) $\{\|m(x_i,\alpha)\|_E : \alpha \in \mathcal{A}_n\}$ is a Donsker class (by Assumptions 3.4 and 3.7 and van der Vaart and Wellner (1996, Thm. 2.5.6)); and (iii) Assumption 3.4(iii) implies that as $\|\alpha-\alpha_0\| \to 0$,

$$\mathrm{E}\left[\left\{\|m(x_i,\alpha)\|_E - \|m(x_i,\alpha_0)\|_E\right\}^2\right] = \mathrm{E}\left[\|m(x_i,\alpha)\|_E^2\right] \leq \text{const.} \times \|\alpha-\alpha_0\|^2 \to 0, \tag{B.1}$$

where the equality follows from $m(x_i,\alpha_0) = 0$, and the inequality follows from Assumptions 3.6 and 3.8(ii). Thus, Chen et al. (2003, Lem. 1) yields $T_1 = o_p(n^{-1/2})$.

For T_2, we have

$$\begin{aligned}\sup_{\alpha\in\mathcal{B}(\eta_n)} \mathrm{E}\left[\|m(x_i,\alpha)\|_E\right] &\leq \sup_{\alpha\in\mathcal{B}(\eta_n)} \sqrt{\mathrm{E}\left[\|m(x_i,\alpha)\|_E^2\right]} \\ &\leq \sup_{\alpha\in\mathcal{B}(\eta_n)} \sqrt{\text{const.} \times \|\alpha-\alpha_0\|^2} = o(\eta_n),\end{aligned}$$

where the first inequality follows from Jensen's inequality and the second inequality follows from (B.1). ■

Proof of (ii). The proof is similar to that of part (i) (replace $\|m(x_i,\alpha)\|_E$ with $\|m(x_i,\alpha)\|_E^2$). ■

Proof of (iii). Observe that

$$\begin{aligned}&\sup_{\alpha\in\mathcal{B}(\eta_n)} \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{in}\left\|\hat{m}(x_i,\alpha)\right\|_E \|m(x_i,\alpha)\|_E \\ &\leq \sup_{\alpha\in\mathcal{B}(\eta_n)} \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{in}\left\|\hat{m}(x_i,\alpha) - m(x_i,\alpha)\right\|_E \|m(x_i,\alpha)\|_E \\ &\quad + \sup_{\alpha\in\mathcal{B}(\eta_n)} \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{in}\|m(x_i,\alpha)\|_E^2 \\ &\leq \sup_{(x_i,\alpha)\in\mathcal{X}_n\times\mathcal{A}_n} \left\|\hat{m}(x_i,\alpha) - m(x_i,\alpha)\right\|_E \sup_{\alpha\in\mathcal{B}(\eta_n)} \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{in}\|m(x_i,\alpha)\|_E\end{aligned}$$

https://doi.org/10.1017/S0266466610000113 Published online by Cambridge University Press

$$+ \sup_{\alpha \in \mathcal{B}(\eta_n)} \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{in} \|m(x_i, \alpha)\|_E^2$$

$$= o_p(n^{-1/4}\eta_n) + o_p(\eta_n^2) = o_p(\eta_n^2),$$

where the first inequality follows from the triangle inequality, and the equality follows from Lemma B.2 (i) and parts (i)–(ii) of this lemma. ■

Proof of (iv). The proof is similar to that of part (iii). ■

Proof of (v). The proof is similar to that of part (iii). ■

Proof of (vi). Since $0 \in \Lambda_n$, we have

$$0 = \sum_{j=1}^{n} w_{ji} \log\left(1 + 0'\rho\left(z_j, \alpha\right)\right) \leq \sum_{j=1}^{n} w_{ji} \log\left(1 + \hat{\lambda}(x_i, \alpha)'\rho(z_j, \alpha)\right), \tag{B.2}$$

for all $\alpha \in \mathcal{B}(\eta_n)$ and $i = 1, \ldots, n$. A second-order expansion of $\sum_{j=1}^{n} w_{ji} \log\left(1 + \hat{\lambda}(x_i, \alpha)'\right.$ $\left.\rho(z_j, \alpha)\right)$ yields that there exists $\bar{t} \in (0, 1)$ satisfying

$$0 \leq \hat{\lambda}(x_i, \alpha)'\hat{m}(x_i, \alpha) - \hat{\lambda}(x_i, \alpha)' \left\{ \sum_{j=1}^{n} w_{ji} \frac{\rho(z_j, \alpha)\rho(z_j, \alpha)'}{2\left(1 - \bar{t}\hat{\lambda}(x_i, \alpha)'\rho\left(z_j, \alpha\right)\right)^2} \right\} \hat{\lambda}(x_i, \alpha)$$

$$\leq \left\|\hat{\lambda}(x_i, \alpha)\right\|_E \left\|\hat{m}(x_i, \alpha)\right\|_E - C\left\|\hat{\lambda}(x_i, \alpha)\right\|_E^2,$$

w.p.a.1, for some constant $C > 0$ and for all $\alpha \in \mathcal{B}(\eta_n)$ and $i = 1, \ldots, n$, where the first inequality follows from (B.2), and the second inequality follows from the fact that $\sum_{j=1}^{n} w_{ji} \frac{\rho(z_j, \alpha)\rho\left(z_j, \alpha\right)'}{2\left(1 - \bar{t}\hat{\lambda}(x_i, \alpha)'\rho(z, \alpha)\right)^2}$ converges to a positive definite matrix uniformly on $(x_i, \alpha) \in \mathcal{X}_n \times \mathcal{B}(\eta_n)$ (by Lemmas B.1(ii) and B.2(ii) and Assumption 3.6(ii)). This implies

$$\left\|\hat{\lambda}(x_i, \alpha)\right\|_E \leq C^{-1} \left\|\hat{m}(x_i, \alpha)\right\|_E, \tag{B.3}$$

w.p.a.1, for all $\alpha \in \mathcal{B}(\eta_n)$ and $i = 1, \ldots, n$. Therefore, part (iv) of this lemma implies

$$\sup_{\alpha \in \mathcal{B}(\eta_n)} \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{in} \left\|\hat{\lambda}(x_i, \alpha)\right\|_E \leq \sup_{\alpha \in \mathcal{B}(\eta_n)} \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{in} C^{-1} \left\|\hat{m}(x_i, \alpha)\right\|_E = o_p(\eta_n).$$ ■

Proof of (vii). The proof is similar to that of part (vi). ■

Proof of (viii). The proof is similar to that of part (vi). ■

https://doi.org/10.1017/S0266466610000113 Published online by Cambridge University Press

LEMMA B.4. *Suppose that Assumptions 3.1–3.11 hold. Then*

$$(i)\ \sup_{\tilde{\alpha}\in\mathcal{N}_{0n}} \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{in} \left\| \hat{\lambda}(x_i,\tilde{\alpha})' \frac{d\hat{\lambda}(x_i,\tilde{\alpha})}{d\alpha}[v_n^*] \right\|_E = o_p(n^{-1/4}),$$

$$(ii)\ \sup_{\tilde{\alpha}\in\mathcal{N}_{0n}} \left| \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{in}\hat{\lambda}(x_i,\tilde{\alpha})' \frac{d\hat{m}(x_i,\tilde{\alpha})}{d\alpha}[v_n^*] - \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{in}\hat{m}(x_i,\tilde{\alpha})'\hat{V}(x_i,\tilde{\alpha})^{-1}\frac{d\hat{m}(x_i,\tilde{\alpha})}{d\alpha}[v_n^*] \right| = o_p(n^{-1/2}),$$

$$(iii)\ \sup_{\tilde{\alpha}\in\mathcal{N}_{0n}} \left| \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{in}\hat{\lambda}(x_i,\tilde{\alpha})' \left(\sum_{j=1}^{n} \frac{w_{ji}\frac{d\rho(z_j,\tilde{\alpha})}{d\alpha}[v_n^*]\rho(z_j,\tilde{\alpha})'}{1+\hat{\lambda}(x_i,\tilde{\alpha})'\rho(z_j,\tilde{\alpha})} \right) \hat{\lambda}(x_i,\tilde{\alpha}) \right| = o_p(n^{-1/2}),$$

$$(iv)\ \sup_{\tilde{\alpha}\in\mathcal{N}_{0n}} \left| \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{in}\hat{m}(x_i,\tilde{\alpha})'\hat{V}(x_i,\tilde{\alpha})^{-1}\frac{d\hat{m}(x_i,\tilde{\alpha})}{d\alpha}[v_n^*] - \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{in}\hat{m}(x_i,\tilde{\alpha})'V(x_i,\alpha_0)^{-1}\frac{dm(x_i,\alpha_0)}{d\alpha}[v^*] \right| = o_p(n^{-1/2}),$$

$$(v)\ \sup_{\tilde{\alpha}\in\mathcal{N}_{0n}} \left| \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{in}\{\hat{m}(x_i,\tilde{\alpha})-\hat{m}(x_i,\alpha_0)\}'V(x_i,\alpha_0)^{-1}\frac{dm(x_i,\alpha_0)}{d\alpha}[v^*] - \langle v^*, \tilde{\alpha}-\alpha_0\rangle \right| = o_p(n^{-1/2}),$$

(vi) Ai and Chen (1999, Cor. A.7):

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{in}\hat{m}(x_i,\alpha_0)'V(x_i,\alpha_0)^{-1}\frac{dm(x_i,\alpha_0)}{d\alpha}[v^*] = \frac{1}{n}\sum_{i=1}^{n} \rho(z_i,\alpha_0)'V(x_i,\alpha_0)^{-1}\frac{dm(x_i,\alpha_0)}{d\alpha}[v^*] + o_p(n^{-1/2}),$$

Proof of (i). By taking the directional derivative for (A.28), we have

$$\begin{aligned} 0 &= \mathbb{I}_{in}\frac{d\left(\sum_{j=1}^{n}\frac{w_{ji}\rho(z_j,\tilde{\alpha})}{1+\hat{\lambda}(x_i,\tilde{\alpha})'\rho(z_j,\tilde{\alpha})}\right)}{d\alpha}[v_n^*] \\ &= \mathbb{I}_{in}\sum_{j=1}^{n}\frac{w_{ji}\frac{d\rho(z_j,\tilde{\alpha})}{d\alpha}[v_n^*]}{\left(1+\hat{\lambda}(x_i,\tilde{\alpha})'\rho(z_j,\tilde{\alpha})\right)} - \mathbb{I}_{in}\left(\sum_{j=1}^{n}\frac{w_{ji}\rho(z_j,\tilde{\alpha})\rho(z_j,\tilde{\alpha})'}{\left(1+\hat{\lambda}(x_i,\tilde{\alpha})'\rho(z_j,\tilde{\alpha})\right)^2}\right)\frac{d\hat{\lambda}(x_i,\tilde{\alpha})}{d\alpha}[v_n^*] \\ &\quad - \mathbb{I}_{in}\left(\sum_{j=1}^{n}\frac{w_{ji}\rho(z_j,\tilde{\alpha})\frac{d\rho(z_j,\tilde{\alpha})}{d\alpha}[v_n^*]'}{\left(1+\hat{\lambda}(x_i,\tilde{\alpha})'\rho(z_j,\tilde{\alpha})\right)^2}\right)\hat{\lambda}(x_i,\tilde{\alpha}), \end{aligned} \tag{B.4}$$

https://doi.org/10.1017/S0266466610000113 Published online by Cambridge University Press

w.p.a.1 for all $\tilde{\alpha} \in \mathcal{N}_{0n}$. Note that $\left(\Sigma_{j=1}^{n} \frac{w_{ji}\rho(z_j,\tilde{\alpha})\rho(z_j,\tilde{\alpha})'}{(1+\hat{\lambda}(x_i,\tilde{\alpha})'\rho(z_j,\tilde{\alpha}))^2}\right)$ is invertible uniformly on $(x_i,\tilde{\alpha}) \in \mathcal{X}_n \times \mathcal{N}_{0n}$ w.p.a.1 (by Lemmas B.1(ii) and B.2(ii)). Thus, solving (B.4) for $\frac{d\hat{\lambda}(x_i,\tilde{\alpha})}{d\alpha}\left[v_n^*\right]$ and the triangle inequality imply that

$$
\begin{aligned}
&\sup_{(x_i,\tilde{\alpha})\in\mathcal{X}_n\times\mathcal{N}_{0n}} \left\| \frac{d\hat{\lambda}(x_i,\tilde{\alpha})}{d\alpha}\left[v_n^*\right] \right\|_E \\
&\leq \sup_{(x_i,\tilde{\alpha})\in\mathcal{X}_n\times\mathcal{N}_{0n}} \left\| \left(\sum_{j=1}^{n} \frac{w_{ji}\rho\left(z_j,\tilde{\alpha}\right)\rho\left(z_j,\tilde{\alpha}\right)'}{\left(1+\hat{\lambda}(x_i,\tilde{\alpha})'\rho\left(z_j,\tilde{\alpha}\right)\right)^2} \right)^{-1} \right\|_E \\
&\quad \times \left\{ \sup_{(x_i,\tilde{\alpha})\in\mathcal{X}_n\times\mathcal{N}_{0n}} \left\| \sum_{j=1}^{n} \frac{w_{ji}\frac{d\rho(z_j,\tilde{\alpha})}{d\alpha}\left[v_n^*\right]}{\left(1+\hat{\lambda}(x_i,\tilde{\alpha})'\rho\left(z_j,\tilde{\alpha}\right)\right)} \right\|_E \right. \\
&\qquad \left. + \sup_{(x_i,\tilde{\alpha})\in\mathcal{X}_n\times\mathcal{N}_{0n}} \left\| \sum_{j=1}^{n} \frac{w_{ji}\rho\left(z_j,\tilde{\alpha}\right)\frac{d\rho(z_j,\tilde{\alpha})}{d\alpha}\left[v_n^*\right]'}{\left(1+\hat{\lambda}(x_i,\tilde{\alpha})'\rho\left(z_j,\tilde{\alpha}\right)\right)^2}\hat{\lambda}(x_i,\tilde{\alpha}) \right\|_E \right\} \\
&\leq O_p(1) \sup_{x_i\in\mathcal{X}_n} \left\| \sum_{j=1}^{n} w_{ji}\bar{c}_1(z_j) \right\|_E \\
&\quad + O_p(1) \sup_{x_i\in\mathcal{X}_n} \left\| \sum_{j=1}^{n} w_{ji}c_1(z_j)\bar{c}_1(z_j) \right\|_E \sup_{(x_i,\tilde{\alpha})\in\mathcal{X}_n\times\mathcal{N}_{0n}} \left\| \hat{\lambda}(x_i,\tilde{\alpha}) \right\|_E \\
&= O_p(1), \qquad \textbf{(B.5)}
\end{aligned}
$$

where the second inequality follows from Lemmas B.1(ii) and B.2(ii) and Assumptions 3.4(ii) and 3.10(iii) ($\bar{c}_1(z_j)$ and $c_1(z_j)$ are envelope functions for $\frac{d\rho(z_j,\tilde{\alpha})}{d\alpha}\left[v_n^*\right]_E$ and $\rho(z_j,\tilde{\alpha})$, respectively), and the equality follows from Assumption 3.5, the uniform convergence of kernel estimators, and Kitamura et al. (2003, Lem. D.4). Therefore, from (B.5) and Lemma B.3(vi),

$$
\begin{aligned}
&\sup_{\tilde{\alpha}\in\mathcal{N}_{0n}} \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{in} \left\| \hat{\lambda}\left(x_i,\tilde{\alpha}\right)' \frac{d\hat{\lambda}\left(x_i,\tilde{\alpha}\right)}{d\alpha}\left[v_n^*\right] \right\|_E \\
&\leq \sup_{(x_i,\tilde{\alpha})\in\mathcal{X}_n\times\mathcal{N}_{0n}} \left\| \frac{d\hat{\lambda}(x_i,\tilde{\alpha})}{d\alpha}\left[v_n^*\right] \right\|_E \sup_{\tilde{\alpha}\in\mathcal{N}_{0n}} \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{in} \left\| \hat{\lambda}\left(x_i,\tilde{\alpha}\right) \right\|_E . \\
&= o_p(n^{-1/4}).
\end{aligned}
$$

■

https://doi.org/10.1017/S0266466610000113 Published online by Cambridge University Press

Proof of (ii). From (A.28), we have

$$0 = \mathbb{I}_{in} \sum_{j=1}^{n} \frac{w_{ji}\rho\left(z_j,\alpha\right)}{1+\hat{\lambda}(x_i,\alpha)'\rho\left(z_j,\alpha\right)}$$

$$= \mathbb{I}_{in} \sum_{j=1}^{n} w_{ji}\rho(z_j,\alpha)\left\{1-\hat{\lambda}(x_i,\alpha)'\rho\left(z_j,\alpha\right)+\frac{\left(\hat{\lambda}(x_i,\alpha)'\rho\left(z_j,\alpha\right)\right)^2}{1+\hat{\lambda}(x_i,\alpha)'\rho(z_j,\alpha)}\right\}$$

$$= \mathbb{I}_{in}\hat{m}(x_i,\alpha)-\mathbb{I}_{in}\hat{V}(x_i,\alpha)\hat{\lambda}(x_i,\alpha)+\mathbb{I}_{in}\sum_{j=1}^{n}\frac{w_{ji}\rho(z_j,\alpha)\left(\hat{\lambda}(x_i,\alpha)'\rho(z_j,\alpha)\right)^2}{1+\hat{\lambda}(x_i,\alpha)'\rho(z_j,\alpha)} \quad \textbf{(B.6)}$$

w.p.a.1, where the second equality follows from a direct calculation. From Lemma B.2(ii) and Assumption 3.6, $\hat{V}(x,\alpha)^{-1}$ exists w.p.a.1 uniformly on $(x,\tilde{\alpha}) \in \mathcal{X}_n \times \mathcal{N}_{0n}$. Solving (B.6) for $\hat{\lambda}(x_i,\alpha)$ yields

$$\mathbb{I}_{in}\hat{\lambda}(x_i,\alpha) = \mathbb{I}_{in}\hat{V}(x_i,\alpha)^{-1}\hat{m}(x_i,\alpha)+\mathbb{I}_{in}\hat{V}(x_i,\alpha)^{-1}\hat{r}(x_i,\alpha), \quad \textbf{(B.7)}$$

w.p.a.1 for all $\tilde{\alpha} \in \mathcal{N}_{0n}$, where $\hat{r}(x_i,\alpha) \equiv \sum_{j=1}^{n}\frac{w_{ji}\rho(z_j,\alpha)\left(\hat{\lambda}(x_i,\alpha)'\rho(z_j,\alpha)\right)^2}{1+\hat{\lambda}(x_i,\alpha)'\rho(z_j,\alpha)}$. Thus, we have

$$\sup_{\tilde{\alpha}\in\mathcal{N}_{0n}}\left|\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{in}\hat{\lambda}(x_i,\tilde{\alpha})'\frac{d\hat{m}(x_i,\tilde{\alpha})}{d\alpha}\left[v_n^*\right]-\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{in}\hat{m}\left(x_i,\tilde{\alpha}\right)'\hat{V}(x_i,\tilde{\alpha})^{-1}\frac{d\hat{m}(x_i,\tilde{\alpha})}{d\alpha}\left[v_n^*\right]\right|$$

$$= \sup_{\tilde{\alpha}\in\mathcal{N}_{0n}}\left|\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}'_{in}\hat{r}\left(x_i,\tilde{\alpha}\right)'\hat{V}(x_i,\tilde{\alpha})^{-1}\frac{d\hat{m}(x_i,\tilde{\alpha})}{d\alpha}\left[v_n^*\right]\right|$$

$$\leq \max_{1\leq j\leq n}\sup_{(x_i,\tilde{\alpha})\in\mathcal{X}_n\times\mathcal{N}_{0n}}\left|\frac{1}{1+\hat{\lambda}(x_i,\tilde{\alpha})'\rho\left(z_j,\tilde{\alpha}\right)}\right|\sup_{(x_i,\tilde{\alpha})\in\mathcal{X}_n\times\mathcal{N}_{0n}}\left\|\hat{V}(x_i,\tilde{\alpha})^{-1}\right\|_E$$

$$\times\sup_{(x_i,\tilde{\alpha})\in\mathcal{X}_n\times\mathcal{N}_{0n}}\left\|\frac{d\hat{m}(x_i,\tilde{\alpha})}{d\alpha}\left[v_n^*\right]\right\|_E\sup_{(x_i,\tilde{\alpha})\in\mathcal{X}_n\times\mathcal{N}_{0n}}\left|\sum_{j=1}^{n}w_{ji}\left\|\rho\left(z_j,\tilde{\alpha}\right)\right\|_E^3\right|$$

$$\times\sup_{\tilde{\alpha}\in\mathcal{N}_{0n}}\left|\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{in}\left\|\hat{\lambda}(x_i,\tilde{\alpha})\right\|_E^2\right|$$

$$\leq O_p(1)\sup_{x_i\in\mathcal{X}_n}\left|\sum_{j=1}^{n}w_{ji}\bar{c}_1(z_j)\right|\sup_{x_i\in\mathcal{X}_n}\left|\sum_{j=1}^{n}w_{ji}c_1(z_j)^3\right|\sup_{\tilde{\alpha}\in\mathcal{N}_{0n}}\left|\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{in}\left\|\hat{\lambda}(x_i,\tilde{\alpha})\right\|_E^2\right|$$

$$= o_p(n^{-1/2}),$$

where the second inequality follows from Lemmas B.1(ii) and B.2(ii) and Assumptions 3.4(ii) and 3.10(iii) ($\bar{c}_1(z_j)$ and $c_1(z_j)$ are envelope functions for $\frac{d\rho(z_j,\tilde{\alpha})}{d\alpha}[v_n^*]_E$ and $\rho(z_j,\tilde{\alpha})$,

https://doi.org/10.1017/S0266466610000113 Published online by Cambridge University Press

respectively), and the last equality follows from Lemma B.3(vii), the uniform convergence of kernel estimators, and Kitamura et al. (2003, Lem. D.4). ■

Proof of (iii). Observe that

$$\begin{aligned}
&\sup_{\tilde{\alpha}\in\mathcal{N}_{0n}}\left|\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{in}\hat{\lambda}(x_i,\tilde{\alpha})'\left(\sum_{j=1}^{n}\frac{w_{ji}\frac{d\rho(z_j,\tilde{\alpha})}{d\alpha}[v_n^*]\rho(z_j,\tilde{\alpha})'}{1+\hat{\lambda}(x_i,\tilde{\alpha})'\rho(z_j,\tilde{\alpha})}\right)\hat{\lambda}(x_i,\tilde{\alpha})\right| \\
&\quad\le \max_{1\le j\le n}\sup_{(x_i,\tilde{\alpha})\in\mathcal{X}_n\times\mathcal{N}_{0n}}\left|\frac{1}{1+\hat{\lambda}(x_i,\tilde{\alpha})'\rho\left(z_j,\tilde{\alpha}\right)}\right| \\
&\qquad\times\sup_{(x_i,\tilde{\alpha})\in\mathcal{X}_n\times\mathcal{N}_{0n}}\left\|\sum_{j=1}^{n}w_{ji}\frac{d\rho(z_j,\tilde{\alpha})}{d\alpha}[v_n^*]\rho(z_j,\tilde{\alpha})'\right\|_E\sup_{\tilde{\alpha}\in\mathcal{N}_{0n}}\left|\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{in}\left\|\hat{\lambda}(x_i,\tilde{\alpha})\right\|_E^2\right| \\
&\quad\le O_p(1)\sup_{x_i\in\mathcal{X}_n}\left|\sum_{j=1}^{n}w_{ji}\bar{c}_1(z_j)c_1(z_j)\right|\sup_{\tilde{\alpha}\in\mathcal{N}_{0n}}\left|\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{in}\left\|\hat{\lambda}(x_i,\tilde{\alpha})\right\|_E^2\right| = o_p(n^{-1/2}),
\end{aligned}$$

where the second inequality follows from Lemma B.1(ii) and Assumptions 3.4(ii) and 3.10(iii), and the last equality follows from Lemma B.3(vii), the uniform convergence of kernel estimators, and Kitamura et al. (2003, Lem. D.4). ■

Proof of (iv). Observe that

$$\begin{aligned}
&\sup_{\tilde{\alpha}\in\mathcal{N}_{0n}}\left|\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{in}\hat{m}(x_i,\tilde{\alpha})'\hat{V}(x_i,\tilde{\alpha})^{-1}\frac{d\hat{m}(x_i,\tilde{\alpha})}{d\alpha}\left[v_n^*\right]\right. \\
&\qquad\left.-\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{in}\hat{m}(x_i,\tilde{\alpha})'V(x_i,\tilde{\alpha})^{-1}\frac{dm(x_i,\tilde{\alpha})}{d\alpha}\left[v_n^*\right]\right| \\
&\quad\le\sup_{\tilde{\alpha}\in\mathcal{N}_{0n}}\left|\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{in}\hat{m}(x_i,\tilde{\alpha})'\hat{V}(x_i,\tilde{\alpha})^{-1}\left\{\frac{d\hat{m}(x_i,\tilde{\alpha})}{d\alpha}\left[v_n^*\right]-\frac{dm(x_i,\tilde{\alpha})}{d\alpha}\left[v_n^*\right]\right\}\right| \\
&\qquad+\sup_{\tilde{\alpha}\in\mathcal{N}_{0n}}\left|\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{in}\hat{m}(x_i,\tilde{\alpha})'\left\{\hat{V}(x_i,\tilde{\alpha})^{-1}-V(x_i,\tilde{\alpha})^{-1}\right\}\frac{dm(x_i,\tilde{\alpha})}{d\alpha}\left[v_n^*\right]\right| \\
&\quad\le\sup_{(x_i,\tilde{\alpha})\in\mathcal{X}_n\times\mathcal{N}_{0n}}\left\|\hat{V}(x_i,\tilde{\alpha})^{-1}\right\|_E\sup_{(x_i,\tilde{\alpha})\in\mathcal{X}_n\times\mathcal{N}_{0n}}\left\|\frac{d\hat{m}(x_i,\tilde{\alpha})}{d\alpha}\left[v_n^*\right]-\frac{dm(x_i,\tilde{\alpha})}{d\alpha}\left[v_n^*\right]\right\|_E \\
&\qquad\times\sup_{\tilde{\alpha}\in\mathcal{N}_{0n}}\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{in}\left\|\hat{m}(x_i,\tilde{\alpha})\right\|_E+\sup_{(x_i,\tilde{\alpha})\in\mathcal{X}_n\times\mathcal{N}_{0n}}\left\|\hat{V}(x_i,\tilde{\alpha})^{-1}-V(x_i,\tilde{\alpha})^{-1}\right\|_E \\
&\qquad\times\sup_{(x_i,\tilde{\alpha})\in\mathcal{X}_n\times\mathcal{N}_{0n}}\left\|\frac{dm(x_i,\tilde{\alpha})}{d\alpha}\left[v_n^*\right]\right\|_E\sup_{\tilde{\alpha}\in\mathcal{N}_{0n}}\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{in}\left\|\hat{m}(x_i,\tilde{\alpha})\right\|_E \\
&\quad= o_p(n^{-1/2}), \qquad \textbf{(B.8)}
\end{aligned}$$

where the equality follows from Lemmas B.2(ii)–(iii) and B.3(iv).

https://doi.org/10.1017/S0266466610000113 Published online by Cambridge University Press

The Cauchy-Schwarz inequality yields

$$\begin{aligned}
&\sup_{\tilde{\alpha}\in\mathcal{N}_{0n}}\left|\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{in}\hat{m}(x_i,\tilde{\alpha})'V(x_i,\tilde{\alpha})^{-1}\frac{dm(x_i,\tilde{\alpha})}{d\alpha}\left[v_n^*\right]\right.\\
&\qquad\left.-\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{in}\hat{m}(x_i,\tilde{\alpha})'V(x_i,\alpha_0)^{-1}\frac{dm(x_i,\alpha_0)}{d\alpha}\left[v_n^*\right]\right|\\
&\le\left(\sup_{\tilde{\alpha}\in\mathcal{N}_{0n}}\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{in}\left\|\hat{m}(x_i,\tilde{\alpha})\right\|_E^2\right)^{1/2}\\
&\qquad\times\left(\sup_{\tilde{\alpha}\in\mathcal{N}_{0n}}\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{in}\left\|V(x_i,\tilde{\alpha})^{-1}\frac{dm(x_i,\tilde{\alpha})}{d\alpha}\left[v_n^*\right]-V(x_i,\alpha_0)^{-1}\frac{dm(x_i,\alpha_0)}{d\alpha}\left[v_n^*\right]\right\|_E^2\right)^{1/2}\\
&=o_P(n^{-1/2}),
\end{aligned}\tag{B.9}$$

where the equality follows from Lemma B.3(v) and an application of Ai and Chen (2003, Cor. C.1(ii)) (replace $\frac{dm(x_i,\tilde{\alpha})}{d\alpha}[v_n^*]$ in Ai and Chen (2003, Cor. C.1(ii)) with $V(x_i,\tilde{\alpha})^{-1}\frac{dm(x_i,\tilde{\alpha})}{d\alpha}[v_n^*]$ and use Assumption 3.11(i)).

Furthermore, observe that

$$\begin{aligned}
&\sup_{\tilde{\alpha}\in\mathcal{N}_{0n}}\left|\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{in}\hat{m}(x_i,\tilde{\alpha})'V(x_i,\alpha_0)^{-1}\frac{dm(x_i,\alpha_0)}{d\alpha}\left[v_n^*-v^*\right]\right|\\
&\le\sup_{x_i\in\mathcal{X}_n}\left\|V(x_i,\alpha_0)^{-1}\right\|_E\sup_{x_i\in\mathcal{X}_n}\left\|\frac{dm(x_i,\alpha_0)}{d\alpha}\left[v_n^*-v^*\right]\right\|_E\sup_{\tilde{\alpha}\in\mathcal{N}_{0n}}\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{in}\left\|\hat{m}(x_i,\tilde{\alpha})\right\|_E\\
&=o_p(n^{-1/2}),
\end{aligned}\tag{B.10}$$

where the equality follows from Assumptions 3.6, 3.9(iii), and 3.10(iii), and Lemma B.3(iv). Combining (B.8)–(B.10), the conclusion is obtained. ■

Proof of (v). The outline of the proof is similar to that of Ai and Chen (2003, Cor. C.3(ii)). Let $g(x_i,v^*)\equiv V(x_i,\alpha_0)^{-1}\frac{dm(x_i,\alpha_0)}{d\alpha}\left[v^*\right]$. First, by applying Chen et al. (2003, Lem. 1), we show that

$$\begin{aligned}
&\sup_{\tilde{\alpha}\in\mathcal{N}_{0n}}\left|\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{in}\left\{\hat{m}(x_i,\tilde{\alpha})-m(x_i,\tilde{\alpha})\right\}'g(x_i,v^*)-\mathrm{E}\left[\mathbb{I}_{in}\left\{\hat{m}(x_i,\tilde{\alpha})-m(x_i,\tilde{\alpha})\right\}'g(x_i,v^*)\right]\right|\\
&\quad=o_p(n^{-1/2}).
\end{aligned}\tag{B.11}$$

Observe that (i) $\mathrm{E}\left[\mathbb{I}_{in}\left\{\hat{m}(x_i,\alpha_0)-m(x_i,\alpha_0)\right\}'g(x_i,v^*)|x_i\right]=0$ (by Assumption 3.2(i)); (ii) $\left\{\left\{\hat{m}(x_i,\tilde{\alpha})-m(x_i,\tilde{\alpha})\right\}'g(x_i,v^*):\alpha\in\mathcal{N}_{0n}\right\}$ is a Donsker class (by Assumptions 3.4, 3.6, 3.7, and 3.10(i)–(ii) and van der Vaart and Wellner (1996, Thm. 2.5.6)); and (iii)

https://doi.org/10.1017/S0266466610000113 Published online by Cambridge University Press

Assumptions 3.6 and 3.10(i)–(ii) and Lemma B.2(i) imply that

$$\mathrm{E}\left[\left\|\mathbb{I}_{in}\left\{\hat{m}(x_i,\tilde{\alpha})-m(x_i,\tilde{\alpha})\right\}'g(x_i,v^*)\right\|_E^2\right] \le \sup_{x_i\in\mathcal{X}_n}\left\|g(x_i,v^*)\right\|_E^2 \times \mathrm{E}\left[\left\|\mathbb{I}_{in}\left\{\hat{m}(x_i,\tilde{\alpha})-m(x_i,\tilde{\alpha})\right\}\right\|_E^2\right] \xrightarrow{p} 0.$$

Thus, (B.11) is obtained. Note that (B.11) also holds at $\tilde{\alpha}=\alpha_0$. Thus, we have

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^n \mathbb{I}_{in}\left\{\hat{m}(x_i,\tilde{\alpha})-\hat{m}(x_i,\alpha_0)\right\}'g(x_i,v^*)-\mathrm{E}\left[\mathbb{I}_{in}\left\{\hat{m}(x_i,\tilde{\alpha})-\hat{m}(x_i,\alpha_0)\right\}'g(x_i,v^*)\right]\\ &\quad=\frac{1}{n}\sum_{i=1}^n \mathbb{I}_{in}\left\{m(x_i,\tilde{\alpha})-m(x_i,\alpha_0)\right\}'g(x_i,v^*)-\mathrm{E}\left[\mathbb{I}_{in}\{m(x_i,\tilde{\alpha})-m(x_i,\alpha_0)\}'g(x_i,v^*)\right]\\ &\qquad+o_p(n^{-1/2}),\end{aligned} \tag{B.12}$$

uniformly on $\tilde{\alpha}\in\mathcal{N}_{0n}$. Let $\hat{g}(x_i,v^*)\equiv\sum_{j=1}^n w_{ji}g(x_j,v^*)$. From exchanging summation and Assumption 3.1(i),

$$\begin{aligned}&\mathrm{E}\left[\mathbb{I}_{in}\left\{\hat{m}(x_i,\tilde{\alpha})-\hat{m}(x_i,\alpha_0)\right\}'g(x_i,v^*)\right]\\ &\quad=\mathrm{E}\left[\mathbb{I}_{in}\left\{m(x_i,\tilde{\alpha})-m(x_i,\alpha_0)\right\}'\hat{g}(x_i,v^*)\right]\\ &\quad=\mathrm{E}\left[\mathbb{I}_{in}\left\{m(x_i,\tilde{\alpha})-m(x_i,\alpha_0)\right\}'g(x_i,v^*)\right]\\ &\qquad-\left\{\mathrm{E}\left[\mathbb{I}_{in}\left\{m(x_i,\tilde{\alpha})-m(x_i,\alpha_0)\right\}'\left\{g(x_i,v^*)-\hat{g}(x_i,v^*)\right\}\right]\right\}\\ &\quad=\mathrm{E}\left[\mathbb{I}_{in}\left\{m(x_i,\tilde{\alpha})-m(x_i,\alpha_0)\right\}'g(x_i,v^*)\right]+o_p(n^{-1/2}),\end{aligned} \tag{B.13}$$

uniformly on $\tilde{\alpha}\in\mathcal{N}_{0n}$, where the last equality follows from $\mathrm{E}[\mathbb{I}_{in}\|m(x_i,\tilde{\alpha})-m(x_i,\alpha_0)\|_E]\le\text{const.}\|\tilde{\alpha}-\alpha_0\|$ (by Assumptions 3.2(i) and 3.8(ii)) and $\sup_{x_i\in\mathcal{X}_n}\left\|\hat{g}(x_i,v^*)-g(x_i,v^*)\right\|_E=o_p(n^{-1/4})$ (obtained by applying Ai and Chen (1999, Lem. A.1)). Combining (B.12) and (B.13),

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^n \mathbb{I}_{in}\left\{\hat{m}(x_i,\tilde{\alpha})-\hat{m}(x_i,\alpha_0)\right\}'g(x_i,v^*)\\ &\quad=\frac{1}{n}\sum_{i=1}^n \mathbb{I}_{in}\left\{m(x_i,\tilde{\alpha})-m(x_i,\alpha_0)\right\}'g(x_i,v^*)+o_p(n^{-1/2})\\ &\quad=\left\{\frac{1}{n}\sum_{i=1}^n \mathbb{I}_{in}\{m(x_i,\tilde{\alpha})-m(x_i,\alpha_0)\}'g(x_i,v^*)-\mathrm{E}\left[\mathbb{I}_{in}\{m(x_i,\tilde{\alpha})-m(x_i,\alpha_0)\}'g(x_i,v^*)\right]\right\}\\ &\qquad+\mathrm{E}\left[\mathbb{I}_{in}\left\{m(x_i,\tilde{\alpha})-m(x_i,\alpha_0)\right\}'g(x_i,v^*)\right]+o_p(n^{-1/2}),\end{aligned} \tag{B.14}$$

https://doi.org/10.1017/S0266466610000113 Published online by Cambridge University Press

uniformly on $\tilde{\alpha} \in \mathcal{N}_{0n}$. For the first term on the right-hand side of (B.14), we can apply the same argument as (B.11) based on Chen et al. (2003, Lem. 1), which yields

$$\sup_{\tilde{\alpha}\in\mathcal{N}_{0n}}\left|\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{in}\{m(x_i,\tilde{\alpha})-m(x_i,\alpha_0)\}'g(x_i,v^*)-\mathrm{E}\left[\mathbb{I}_{in}\{m(x_i,\tilde{\alpha})-m(x_i,\alpha_0)\}'g(x_i,v^*)\right]\right|$$
$$=o_p(n^{-1/2}). \tag{B.15}$$

Also note that, for each $\tilde{\alpha} \in \mathcal{N}_{0n}$, there exists a convex combination of $\tilde{\alpha}$ and α_0 (denote $\bar{\alpha} \in \mathcal{N}_0$) such that

$$\mathrm{E}\left[\mathbb{I}_{in}\{m(x_i,\tilde{\alpha})-m(x_i,\alpha_0)\}'g(x_i,v^*)\right]=\mathrm{E}\left[\mathbb{I}_{in}\left\{\frac{dm(x_i,\bar{\alpha})}{d\alpha}\left[\tilde{\alpha}-\alpha_0\right]\right\}'g(x_i,v^*)\right]. \tag{B.16}$$

Thus, from (B.14)–(B.16), we have

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{in}\left\{\hat{m}(x_i,\tilde{\alpha})-\hat{m}(x_i,\alpha_0)\right\}'g(x_i,v^*)\\
&\quad=\mathrm{E}\left[\mathbb{I}_{in}\{m(x_i,\tilde{\alpha})-m(x_i,\alpha_0)\}'g(x_i,v^*)\right]+o_p(n^{-1/2})\\
&\quad=\left\langle v^*,\tilde{\alpha}-\alpha_0\right\rangle+\mathrm{E}\left[\left\{\mathbb{I}_{in}\frac{dm(x_i,\bar{\alpha})}{d\alpha}\left[\tilde{\alpha}-\alpha_0\right]\right\}'g(x_i,v^*)\right.\\
&\qquad\qquad\left.-\left\{\frac{dm(x_i,\alpha_0)}{d\alpha}\left[\tilde{\alpha}-\alpha_0\right]\right\}'g(x_i,v^*)\right]+o_p(n^{-1/2})\\
&\quad=\left\langle v^*,\tilde{\alpha}-\alpha_0\right\rangle+o_p(n^{-1/2}),
\end{aligned}$$

uniformly on $\tilde{\alpha} \in \mathcal{N}_{0n}$, where the last equality follows from Assumption 3.11(ii). The conclusion is obtained. ■

Proof of (vi). See Ai and Chen (1999, Cor. A.7). ■

LEMMA B.5. *Suppose that Assumptions 3.1–3.10 hold. Then*

(i)
$$\sup_{\tilde{\alpha}\in\mathcal{N}_{0n}}\left|\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{in}\sum_{j=1}^{n}\frac{w_{ji}\left\{\frac{d\hat{\lambda}(x_i,\tilde{\alpha})'\rho(z_j,\tilde{\alpha})}{d\alpha}\left[v_n^*\right]\right\}\hat{\lambda}(x_i,\tilde{\alpha})'\frac{d\rho(z_j,\tilde{\alpha})}{d\alpha}[v_n^*]}{\left(1+\hat{\lambda}(x_i,\tilde{\alpha})'\rho(z_j,\tilde{\alpha})\right)^2}\right|=o_p(n^{-1/4}),$$

(ii)
$$\sup_{\tilde{\alpha}\in\mathcal{N}_{0n}}\left|\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{in}\hat{\lambda}(x_i,\tilde{\alpha})'\left(\sum_{j=1}^{n}\frac{w_{ji}\frac{d\rho(z_j,\tilde{\alpha})}{d\alpha}[v_n^*]\rho(z_j,\tilde{\alpha})'}{1+\hat{\lambda}(x_i,\tilde{\alpha})'\rho(z_j,\tilde{\alpha})}\right)\hat{\lambda}(x_i,\tilde{\alpha})\right|=o_p(n^{-1/2}),$$

(iii)
$$\sup_{\tilde{\alpha}\in\mathcal{N}_{0n}}\left|\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{in}\sum_{j=1}^{n}\frac{w_{ji}}{1+\hat{\lambda}(x_i,\tilde{\alpha})'\rho(z_j,\tilde{\alpha})}\sum_{k=1}^{d_\rho}\left\{\frac{d^2\rho^{(k)}(z_j,\tilde{\alpha})}{d\alpha d\alpha}\left[v_n^*,v_n^*\right]\right\}\hat{\lambda}^{(k)}(x_i,\tilde{\alpha})\right|$$
$$=o_p(n^{-1/4}).$$

https://doi.org/10.1017/S0266466610000113 Published online by Cambridge University Press

Proof of (i). Observe that

$$\sup_{\tilde{\alpha}\in\mathcal{N}_{0n}}\left|\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{in}\sum_{j=1}^{n}\frac{w_{ji}\left\{\frac{d\hat{\lambda}(x_i,\tilde{\alpha})'\rho(z_j,\tilde{\alpha})}{d\alpha}[v_n^*]\right\}\hat{\lambda}(x_i,\tilde{\alpha})'\frac{d\rho(z_j,\tilde{\alpha})}{d\alpha}[v_n^*]}{\left(1+\hat{\lambda}(x_i,\tilde{\alpha})'\rho(z_j,\tilde{\alpha})\right)^2}\right|$$

$$\leq O_{as}(1)\sup_{\tilde{\alpha}\in\mathcal{N}_{0n}}\left|\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{in}\sum_{j=1}^{n}w_{ji}\left\{\rho(z_j,\tilde{\alpha})'\frac{d\hat{\lambda}(x_i,\tilde{\alpha})}{d\alpha}[v_n^*]+\hat{\lambda}(x_i,\tilde{\alpha})'\frac{d\rho(z_j,\tilde{\alpha})}{d\alpha}[v_n^*]\right\}\right.$$

$$\left.\times\,\hat{\lambda}(x_i,\tilde{\alpha})'\frac{d\rho(z_j,\tilde{\alpha})}{d\alpha}[v_n^*]\right|$$

$$\leq O_{as}(1)\sup_{(x_i,\tilde{\alpha})\in\mathcal{X}_n\times\mathcal{N}_{0n}}\left\|\sum_{j=1}^{n}w_{ji}\rho(z_j,\tilde{\alpha})'\frac{d\rho(z_j,\tilde{\alpha})}{d\alpha}[v_n^*]\right\|_E$$

$$\times\sup_{\tilde{\alpha}\in\mathcal{N}_{0n}}\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{in}\left\|\hat{\lambda}(x_i,\tilde{\alpha})'\frac{d\hat{\lambda}(x_i,\tilde{\alpha})}{d\alpha}[v_n^*]\right\|_E$$

$$+O_{as}(1)\sup_{(x_i,\tilde{\alpha})\in\mathcal{X}_n\times\mathcal{N}_{0n}}\left\|\sum_{j=1}^{n}w_{ji}\left\{\frac{d\rho(z_j,\tilde{\alpha})}{d\alpha}[v_n^*]\right\}'\frac{d\rho(z_j,\tilde{\alpha})}{d\alpha}[v_n^*]\right\|_E$$

$$\times\sup_{\tilde{\alpha}\in\mathcal{N}_{0n}}\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{in}\left\|\hat{\lambda}(x_i,\tilde{\alpha})\right\|_E^2$$

$$\leq o_p(n^{-1/4})\sup_{x_i\in\mathcal{X}_n}\left\|\sum_{j=1}^{n}w_{ji}c_1(z_j)\bar{c}_1(z_j)\right\|_E+o_p(n^{-1/2})\sup_{x_i\in\mathcal{X}_n}\left\|\sum_{j=1}^{n}w_{ji}\bar{c}_1(z_j)^2\right\|_E$$

$$=o_p(n^{-1/4})+o_p(n^{-1/2}),$$

where the first inequality follows from Lemma B.1(ii), the third inequality follows from Assumptions 3.4(ii) and 3.10(iii) ($c_1(z_j)$ and $\bar{c}_1(z_j)$ are envelope functions for $\rho(z_j,\tilde{\alpha})$ and $\frac{d\rho(z_j,\tilde{\alpha})}{d\alpha}[v_n^*]$, respectively) and Lemmas B.3(vii) and B.4(i), and the equality follows from the uniform convergence of kernel estimators. Thus, the conclusion is obtained. ■

Proof of (ii). Observe that

$$\sup_{\tilde{\alpha}\in\mathcal{N}_{0n}}\left|\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{in}\hat{\lambda}(x_i,\tilde{\alpha})'\left(\sum_{j=1}^{n}\frac{w_{ji}\frac{d\rho(z_j,\tilde{\alpha})}{d\alpha}[v_n^*]\rho(z_j,\tilde{\alpha})'}{1+\hat{\lambda}(x_i,\tilde{\alpha})'\rho(z_j,\tilde{\alpha})}\right)\hat{\lambda}(x_i,\tilde{\alpha})\right|$$

$$\leq\max_{1\leq j\leq n}\sup_{(x_i,\tilde{\alpha})\in\mathcal{X}_n\times\mathcal{N}_{0n}}\left|\frac{1}{1+\hat{\lambda}(x_i,\tilde{\alpha})'\rho(z_j,\tilde{\alpha})}\right|$$

https://doi.org/10.1017/S0266466610000113 Published online by Cambridge University Press

$$\times \sup_{(x_i,\tilde{\alpha})\in\mathcal{X}_n\times\mathcal{N}_{0n}} \left\| \sum_{j=1}^n w_{ji} \frac{d\rho(z_j,\tilde{\alpha})}{d\alpha}[v_n^*]\rho(z_j,\tilde{\alpha})' \right\|_E \sup_{\tilde{\alpha}\in\mathcal{N}_{0n}} \frac{1}{n}\sum_{i=1}^n \mathbb{I}_{in} \left\| \hat{\lambda}(x_i,\tilde{\alpha}) \right\|_E^2$$

$$\leq O_{as}(1) \sup_{x_i\in\mathcal{X}_n} \left\| \sum_{j=1}^n w_{ji}\bar{c}_1(z_j)c_1(z_j) \right\|_E \sup_{\tilde{\alpha}\in\mathcal{N}_{0n}} \frac{1}{n}\sum_{i=1}^n \mathbb{I}_{in} \left\| \hat{\lambda}(x_i,\tilde{\alpha}) \right\|_E^2 = o_p(n^{-1/2}),$$

where the second inequality follows from Lemma B.1(ii) and Assumptions 3.4(ii) and 3.10(iii), and the equality follows from Lemma B.3(vii) and the uniform convergence of kernel estimators. ■

Proof of (iii). Observe that

$$\sup_{\tilde{\alpha}\in\mathcal{N}_{0n}} \left| \frac{1}{n}\sum_{i=1}^n \mathbb{I}_{in} \sum_{j=1}^n \frac{w_{ji}}{1+\hat{\lambda}(x_i,\tilde{\alpha})'\rho(z_j,\tilde{\alpha})} \sum_{k=1}^{d_\rho} \left\{ \frac{d^2\rho_k(z_j,\tilde{\alpha})}{d\alpha d\alpha}[v_n^*,v_n^*] \right\} \hat{\lambda}_k(x_i,\tilde{\alpha}) \right|$$

$$\leq \max_{1\leq j\leq n} \sup_{(x_i,\alpha)\in\mathcal{X}_n\times\mathcal{N}_{0n}} \left| \frac{1}{1+\hat{\lambda}(x_i,\tilde{\alpha})'\rho(z_j,\tilde{\alpha})} \right|$$

$$\times \sup_{(x_i,\tilde{\alpha})\in\mathcal{X}_n\times\mathcal{N}_{0n}} \left\| \sum_{j=1}^n w_{ji} \sum_{k=1}^{d_\rho} \frac{d^2\rho_k(z_j,\tilde{\alpha})}{d\alpha d\alpha}[v_n^*,v_n^*] \right\|_E$$

$$\times \sup_{\tilde{\alpha}\in\mathcal{N}_{0n}} \frac{1}{n}\sum_{i=1}^n \mathbb{I}_{in} \left\| \max_{1\leq k\leq d_\rho} \hat{\lambda}_k(x_i,\tilde{\alpha}) \right\|_E$$

$$\leq O_{as}(1) \sup_{x_i\in\mathcal{X}_n} \left\| \sum_{j=1}^n w_{ji}c_1(z_j) \right\|_E \sup_{\tilde{\alpha}\in\mathcal{N}_{0n}} \frac{1}{n}\sum_{i=1}^n \mathbb{I}_{in} \left\| \max_{1\leq k\leq d_\rho} \hat{\lambda}_k(x_i,\tilde{\alpha}) \right\|_E = o_p(n^{-1/4}),$$

where the second inequality follows from Lemma B.1(ii) and Assumption 3.10(iv), and the equality follows from Lemma B.3(vi) and the uniform convergence of kernel estimators. ■

https://doi.org/10.1017/S0266466610000113 Published online by Cambridge University Press