*Econometric Theory*, **33**, 2017, 1242–1258.
doi:10.1017/S0266466616000360

# A NOTE ON GENERALIZED EMPIRICAL LIKELIHOOD ESTIMATION OF SEMIPARAMETRIC CONDITIONAL MOMENT RESTRICTION MODELS

NAOYA SUEISHI
*Kobe University*

This paper proposes an empirical likelihood-based estimation method for semiparametric conditional moment restriction models, which contain finite dimensional unknown parameters and unknown functions. We extend the results of Donald, Imbens, and Newey (2003, *Journal of Econometrics* 117, 55–93) by allowing unknown functions to be included in the conditional moment restrictions. We approximate unknown functions by a sieve method and estimate the finite dimensional parameters and unknown functions jointly. We establish consistency and derive the convergence rate of the estimator. We also show that the estimator of the finite dimensional parameters is $\sqrt{n}$-consistent, asymptotically normally distributed, and asymptotically efficient.

## 1. INTRODUCTION

This paper is concerned with an empirical likelihood (EL)-based estimation method for semiparametric conditional moment restriction models given by the form:

$$E[\rho(Z,\theta_0,h_0)|X]=0, \tag{1}$$

where $Z=(Y,X_z)$, $X_z$ is a subset of $X$, and $\rho$ is a $d_\rho$ dimensional vector of known (residual) functions up to parameters. The conditional distribution of $Y$ given $X$ is not specified. The parameters of interest $\alpha_0=(\theta_0,h_0)$ contain a finite dimensional vector $\theta_0$ and a vector of unknown functions $h_0(\cdot)=(h_{01}(\cdot),\ldots,h_{0q}(\cdot))$. The model is semiparametric in the sense that it contains unknown functions $h_0$. Unknown functions $h_0$ may depend on the endogenous variables $Y$, conditioning

Early versions of this paper were titled "Generalized Empirical Likelihood Estimation via Conditional Moment Restrictions Containing Unknown Functions." This paper is based on part of my dissertation at the University of Wisconsin. I would like to thank Bruce Hansen and Jack Porter for their helpful comments and suggestions. The comments from a co-editor and two referees improved the paper. I also thank Yoshihiko Nishiyama, Kohtaro Hitomi, Qingfeng Liu, Ryo Okui, and seminar participants at Kobe University, Kyoto University, and University of Wisconsin. All remaining errors are mine. Address correspondence to Naoya Sueishi, Graduate School of Economics, Kobe University, Rokkodai-cho, Nada-ku, Kobe 657-8501, Japan; e-mail: sueishi@econ.kobe-u.ac.jp.

 © Cambridge University Press 2016

https://doi.org/10.1017/S0266466616000360 Published online by Cambridge University Press

(or instrumental) variables $X$, or known index functions $\delta(X;\beta_0)$ up to an unknown parameter $\beta_0$.

Ai and Chen (2003) (hereafter AC, 2003) and Newey and Powell (2003) independently propose the sieve minimum distance (SMD) estimator for the model (1) using sieves to approximate the unknown functions. Newey and Powell (2003) establish consistency of their estimator and discuss a sufficient condition for the identification of $h_0$. AC (2003) show that the SMD estimator of the parametric component $\theta_0$ is $\sqrt{n}$-consistent and asymptotically normally distributed. Furthermore, AC (2003) derive the semiparametric efficiency bound of the model (1) and show that the optimally weighted SMD estimator of $\theta_0$ attains the bound. Chen and Pouzo (2009) extend the results of AC (2003) and propose the penalized SMD estimator to deal with possibly nonsmooth residual functions. Blundell, Chen, and Kristensen (2007) apply the SMD estimator to the estimation of a system of Engel curves.

Without unknown functions $h_0$, (1) is reduced to

$$E[\rho(Z,\theta_0)|X]=0. \tag{2}$$

A large body of research has been conducted for analyzing the model (2), including Chamberlain (1987), Robinson (1987), Newey (1990), Carrasco and Florens (2000), and Dominguez and Lobato (2004). There are basically two approaches in the EL (Qin and Lawless, 1994) literature to deal with (2). Donald, Imbens, and Newey (2003) (hereafter DIN, 2003) utilize the fact that the conditional moment restriction is equivalent to a countable number of unconditional moment restrictions implied by (2) (Chamberlain, 1987). They show that their generalized empirical likelihood (GEL) estimator attains the semiparametric efficiency bound when the number of unconditional moment restrictions gets large with the sample size. Another approach is proposed by Zhang and Gijbels (2003) and Kitamura, Tripathi, and Ahn (2004), who use the kernel smoothing method to incorporate local restrictions implied by (2). Their estimator also attains the semiparametric efficiency bound. Smith (2006) extends the idea of Kitamura et al. (2004) by replacing the log likelihood criterion with the general Cressie-Read discrepancy family.

The purpose of this paper is to propose a GEL-based estimation method for (1) using a sieve method to approximate unknown functions $h_0$. Our GEL-based estimator is called the sieve generalized empirical likelihood (SGEL) estimator. We utilize an expanding set of unconditional moment restrictions based on approximating functions so that the optimal instrument is spanned asymptotically. The estimator was suggested by Nishiyama, Liu, and Sueishi (2005), but we did not provide the theoretical justification. We show that the SGEL estimator of $\theta_0$ is $\sqrt{n}$-consistent, asymptotically normally distributed, and asymptotically efficient.

Otsu (2011) proposes the sieve conditional empirical likelihood (SCEL) estimator, which is an extension of the conditional EL estimator of Kitamura et al. (2004). Moreover, the quadratic approximation of Zhang and Gijbels (2003) implies that the SCEL estimator is an information-theoretic alternative to the SMD estimator. In contrast, our method is viewed as an extension of DIN (2003).

https://doi.org/10.1017/S0266466616000360 Published online by Cambridge University Press

The rest of the paper is organized as follows. Section 2 introduces the SGEL estimator. Section 3 discusses consistency of the estimator and derives the rate of convergence under a certain metric. Section 4 shows that the estimator of the parametric component is asymptotically normally distributed. Section 5 concludes. The Appendix provides outlines of the proofs. Detailed proofs and some definitions and assumptions are given in the supplemental material that is available online.

## 2. SGEL ESTIMATOR

The environment we consider is the same as that of AC (2003). In what follows, we assume that the observations $\{(Y_i, X_i) : i = 1, 2, \ldots, n\}$ are drawn independently 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 $\mathbb{R}^{d_x}$. The vector function $\rho : \mathcal{Z} \times \mathcal{A} \to \mathbb{R}^{d_\rho}$ is known up to the unknown parameters $\alpha_0 = (\theta_0, h_0) \in \mathcal{A} \equiv \Theta \times \mathcal{H}$. Assume that $\Theta \subseteq \mathbb{R}^{d_\theta}$ is compact and $\mathcal{H} = \mathcal{H}^1 \times \cdots \times \mathcal{H}^q$ is a space of continuous functions. We allow dependence of $h_0$ on $\theta_0$. Assume that $Z = (Y, X_z)' \in \mathcal{Z} = \mathcal{Y} \times \mathcal{X}_z$ and $\mathcal{X}_z \subseteq \mathcal{X}$. Denote $\|A\| = (\mathrm{tr}(A'A))^{1/2}$ for any matrix $A$.

The conditional moment restriction (2) implies that

$$E[a(X)\rho(Z, \theta_0)] = 0 \tag{3}$$

for an arbitrary matrix-valued function $a(x)$. Then, $\theta_0$ can be estimated by using (3). However, this procedure has two deficiencies. First, an arbitrary choice of $a(x)$ leads to efficiency loss. Second, unconditional moment restrictions may fail to identify the parameter even when an optimal instrument is used. Chamberlain (1987) shows that the estimator which solves $\sum_{i=1}^{n} B(X_i)\rho(Z_i, \theta) = 0$ with

$$B(X) = E\left[\frac{\partial \rho(Z, \theta_0)}{\partial \theta} | X\right]' E[\rho(Z, \theta_0)\rho(Z, \theta_0)' | X]^{-1} \tag{4}$$

is asymptotically efficient. Although the optimal instrument is estimable, it must be used with caution. Dominguez and Lobato (2004) point out that there exists a case where (4) does not identify the parameter of interest even though (2) holds for a single value $\theta_0$. Carrasco and Florens (2000) address this issue by using a continuum of unconditional moment restrictions rather than using a finite number of moment restrictions.

Let $\{p_{0j}(X), j = 1, 2, \ldots\}$ be a sequence of basis functions. Motivated by DIN (2003), we introduce the GEL-based estimator based on the unconditional moment restrictions:

$$E[\rho(Z, \alpha_0) \otimes p^{k_n}(X)] = 0,$$

where $p^{k_n}(X) = (p_{01}(X), \ldots, p_{0k_n}(X))'$ is a $k_n \times 1$ vector. Let $s(v)$ be a concave function on its domain $\mathcal{V}$, which is an open interval containing 0. We normalize the function so that $s(0) = 0$ and $s_1(0) = s_2(0) = -1$, where $s_j(v) = \partial^j s(v)/\partial v^j$. The SGEL estimator is given by

https://doi.org/10.1017/S0266466616000360 Published online by Cambridge University Press

$$\hat{\alpha}_n = \arg \min_{\alpha=(\theta,h)\in\Theta\times\mathcal{H}_n} \max_{\lambda\in\hat{\Lambda}(\alpha)} \sum_{i=1}^{n} s(\lambda' g_i(\alpha)),$$

where $g_i(\alpha) = \rho(Z_i,\alpha) \otimes p^{k_n}(X_i)$ and $\hat{\Lambda}(\alpha) = \{\lambda : \lambda' g_i(\alpha) \in \mathcal{V}, i = 1,\dots,n\}$. The minimum is taken over a sieve space $\mathcal{H}_n = \mathcal{H}_n^1 \times \cdots \times \mathcal{H}_n^q$. Hence we utilize the sieve method in two ways: (i) approximation for the unknown best instrument and (ii) approximation for the unknown functions $h_0$.

## 3. CONSISTENCY AND CONVERGENCE RATE

We first prove consistency of the SGEL estimator. We establish consistency of the estimator under a metric $\|\cdot\|_s$ such as the sup or $L_2$ metric. Then we derive a convergence rate of the estimators under a certain metric, which is weaker than $\|\cdot\|_s$. Following AC (2003), we introduce a weaker metric $\|\cdot\|_w$ and establish that $\|\hat{\alpha}_n - \alpha_0\|_w = o_p(n^{-1/4})$. It turns out that it is enough to have the fast rate of convergence under $\|\cdot\|_w$ to derive the asymptotic normality of $\hat{\theta}_n$.

Because most of the definitions and assumptions duplicate those of AC (2003) and DIN (2003), we do not replicate them here. For complete definitions and assumptions, see the online supplement. Only conditions that are different from those of AC (2003) and DIN (2003) are discussed.

Let $k_{1n} = \dim(\mathcal{H}_n)$ denote the number of unknown sieve parameters in $h \in \mathcal{H}_n = \mathcal{H}_n^1 \times \cdots \times \mathcal{H}_n^q$. Also, let $\zeta(k_n)$ be a constant satisfying $\sup_{X\in\mathcal{X}} \|p^{k_n}(X)\| \le \zeta(k_n)$ and $\sqrt{k_n} \le \zeta(k_n)$. We impose the following conditions.

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

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

Assumption 3.5(ii) is more restrictive than Assumption 3.5(ii) of AC (2003) and Assumption 3.2 of Otsu (2011), which only require that $\|\Pi_n\alpha - \alpha\|_s = o(1)$. We need Assumption 3.5(ii) and Assumption 3.8(iii) to control the order of $\|\hat{\lambda}_n(\Pi_n\alpha_0)\|$, where $\hat{\lambda}_n(\alpha) = \arg\max_{\lambda\in\hat{\Lambda}(\alpha)} \sum_{i=1}^n s(\lambda' g_i(\alpha))$. Instead of specifying a convergence rate for $\|\Pi_n\alpha - \alpha\|_s$, Otsu (2011) controls the order of the Lagrange multiplier by restricting its support (Assumption 3.5 in Otsu, 2011).

Assumption 3.8(iii) requires that the functions in $\mathcal{H}$ must be sufficiently smooth. Suppose that $h_0$ is a $p$-smooth real valued function on $[0,1]$. It is known that if the sieve space $\mathcal{H}_n$ is the space of splines of order $r \ge [p]+1$ with $J_n$ knots, then $\|\Pi_n h_0 - h_0\|_s = O(J_n^{-p}) = O(k_{1n}^{-p})$ with respect to the sup norm $\|h\|_s \equiv \sup_{x\in[0,1]}|h(x)|$. Thus we have $\|\Pi_n\alpha - \alpha\|_s^\kappa = O(k_{1n}^{-\kappa p})$. Also, Assumption 3.8(ii) requires that $k_n^2/n^{1-2/m} \to 0$. We now suppose that $m = 8$.

https://doi.org/10.1017/S0266466616000360 Published online by Cambridge University Press

Then Assumption 3.8(ii) is satisfied with $k_n = o(n^{3/8})$. On the other hand, for identification, $k_{1n}$ cannot grow faster than $k_n$. Thus the rate of $k_{1n}$ is at most $O(k_n) = o(n^{3/8})$. Under these conditions, Assumption 3.8(iii) is satisfied if $n^{1/8}k_n k_n^{-\kappa p} = n^{1/8}n^{3/8-3\kappa p/8} = o(1)$, which requires $\kappa p > 4/3$. Therefore if $\kappa$ is small then we need the existence of high order derivatives of $h_0$. If $X$ is multivariate, then it might be difficult to satisfy Assumption 3.8(iii). In that case, Assumption 3.8 is rather a "high-level" assumption. Note that Assumptions 3.5(ii) and 3.8(iii) are sufficient but not necessary conditions. We conjecture that consistency will hold under more mild regularity conditions.

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

Next, we obtain the rate of convergence of the estimator under a weak metric $\|\cdot\|_w$ (see equation (14) in AC, 2003 or the online supplement).

Let $m(X,\alpha) = E[\rho(Z,\alpha)|X]$ and $\Sigma(X,\alpha) = E[\rho(Z,\alpha)\rho(Z,\alpha)'|X]$. Let $\Lambda_c^p(\mathcal{X})$ be the Hölder ball with radius $c$ and smoothness $p = m + \gamma$:

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

Moreover, let $\mathcal{A}_{0n} = \{\alpha \in \mathcal{A}_n : \|\alpha - \alpha_0\|_s = o(1)\}$.

The following assumption is imposed.

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

Assumption 3.9 is a smoothness condition for $\Sigma(X,\alpha)^{1/2}$, that is, the square root of $\Sigma(X,\alpha)$. We impose Assumption 3.9 instead of Assumption 3.4(iii) in AC (2003) because the SGEL estimator implicitly estimates the optimal weight. Assumption 3.9 is not necessary to obtain $\|\hat{\alpha}_n - \alpha_0\|_w = o_p(n^{-1/4})$, but is used to obtain the asymptotic normality of $\hat{\theta}_n$. An intuition is the following. Notice that the objective function of the SMD estimator is given by $n^{-1}\sum_{i=1}^n \hat{m}(X_i,\alpha)'\widehat{\Sigma}(X_i)^{-1}\hat{m}(X_i,\alpha)$, where $\hat{m}(X,\alpha)$ is a sieve estimator of $m(X,\alpha)$ and $\widehat{\Sigma}(X)$ is a consistent estimator of $\Sigma(X,\alpha_0)$. To get the convergence rate of the estimator, AC (2003) utilize the result that

$$\sup_{\alpha\in\mathcal{A}_n} \left| \frac{1}{n}\sum_{i=1}^n \hat{m}(X_i,\alpha)'\widehat{\Sigma}(X_i)^{-1}\hat{m}(X_i,\alpha) - \frac{1}{n}\sum_{i=1}^n m(X_i,\alpha)'\Sigma(X_i,\alpha_0)^{-1}m(X_i,\alpha) \right|$$
$$= o_p(n^{-1/4})$$

under the assumption that $\hat{\Sigma}(X) = \Sigma(X,\alpha_0) + o_p(n^{-1/4})$ (Assumption 3.4(ii) of AC, 2003). Thus they show that $\hat{m}(X,\alpha)$ converges to $m(X,\alpha)$ at a certain rate.

https://doi.org/10.1017/S0266466616000360 Published online by Cambridge University Press

On the other hand, we cannot estimate the weight matrix separately. Thus we show that a sieve estimator of $\Sigma(X,\alpha)^{-1/2}m(X,\alpha)$ converges to the true function (see Lemma A.8 in the online supplement). Assumption 3.9 is helpful to show the uniform convergence result.

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

## 4. ASYMPTOTIC NORMALITY AND EFFICIENCY

Now we derive the asymptotic distribution of $\hat{\theta}_n$. Following AC (2003), we introduce some notations. Let $\overline{V}$ be the closure of the linear span of $\mathcal{A}-\alpha_0$ under $\|\cdot\|_w$. For $v \in \overline{V}$, let

$$\frac{d\rho(Z,\alpha)}{d\alpha}[v] = \left.\frac{d\rho(Z,\alpha+\tau v)}{d\tau}\right|_{\tau=0} \text{ a.s. } Z,$$
$$\frac{dm(Z,\alpha)}{d\alpha}[v] = E\left[\frac{d\rho(Z,\alpha)}{d\alpha}[v]|X\right] \text{ a.s. } X.$$

Then $(\overline{V}, \|\cdot\|_w)$ is a Hilbert space with the inner product:

$$\langle v_1, v_2\rangle = E\left[\left\{\frac{dm(X,\alpha_0)}{d\alpha}[v_1]\right\}' \Sigma(X,\alpha_0)^{-1}\left\{\frac{dm(X,\alpha_0)}{d\alpha}[v_2]\right\}\right].$$

With $\overline{V} = \mathbb{R}^{d_\theta} \times \overline{\mathcal{W}}$ and $\overline{\mathcal{W}} \equiv \overline{\mathcal{H}} - h_0$, we 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 $\theta_j$, $j=1,\ldots,d_\theta$, let $w_j^* \in \overline{\mathcal{W}}$ be the solution to

$$\min_{w_j\in\overline{\mathcal{W}}} E\left[\left(\frac{dm(X,\alpha_0)}{d\theta_j} - \frac{dm(X,\alpha_0)}{dh}[w_j]\right)' \Sigma(X,\alpha_0)^{-1}\left(\frac{dm(X,\alpha_0)}{d\theta_j} - \frac{dm(X,\alpha_0)}{dh}[w_j]\right)\right].$$

Also, we define

$$w^* = (w_1^*,\ldots,w_{d_\theta}^*),$$
$$\frac{dm(X,\alpha_0)}{dh}[w^*] = \left(\frac{dm(X,\alpha_0)}{dh}[w_1^*],\ldots,\frac{dm(X,\alpha_0)}{dh}[w_{d_\theta}^*]\right),$$
$$D_{w^*}(X) \equiv \frac{dm(X,\alpha_0)}{d\theta'} - \frac{dm(X,\alpha_0)}{dh}[w^*].$$

Let $f(\alpha) = \xi'\theta$ for some fixed $\xi \in \mathbb{R}^{d_\theta}$. Because $f(\alpha)$ is a linear functional, by the Riesz representation theorem, we have

$$f(\alpha) - f(\alpha_0) = \langle v^*, \alpha-\alpha_0\rangle \quad \text{for all } \alpha \in \mathcal{A},$$

https://doi.org/10.1017/S0266466616000360 Published online by Cambridge University Press

where $v^* = (v_\theta^*, v_h^*) \in \overline{V}$ with $v_\theta^* = (E[D_{w^*}(X)'\Sigma(X,\alpha_0)^{-1}D_{w^*}(X)])^{-1}\xi$ and $v_h^* = -w \times v_\theta^*$.

Additional assumptions are imposed.

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

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

Assumption 4.7 is satisfied for commonly used GEL estimators. Assumptions 4.7 and 4.8 are used to prove that the directional derivative of the objective function of the SGEL estimator converges to the directional derivative of the objective function of the SMD estimator at the order $o_p(n^{-1/2})$. Under these assumptions, we have the following results.

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

The asymptotic variance is the same as that of the optimally weighted SMD estimator of AC (2003). Therefore, the SGEL estimator is asymptotically efficient.

Now, we discuss an estimation method for the asymptotic variance of the SGEL estimator, which is used for inference. The estimator can be obtained in the following way. For each $\theta_j$, we can estimate $w_j^*$ by solving the minimization problem:

$$\min_{w_j \in \mathcal{H}_n^j} \left(\frac{1}{n}\sum_{i=1}^n \left\{\frac{d\rho(Z_i,\hat{\alpha}_n)}{d\theta_j} - \frac{d\rho(Z_i,\hat{\alpha}_n)}{dh}[w_j]\right\} \otimes p^{k_n}(X_i)\right)' \left(\frac{1}{n}\sum_{i=1}^n g_i(\hat{\alpha}_n)g_i(\hat{\alpha}_n)'\right)^{-1} \times \left(\frac{1}{n}\sum_{i=1}^n \left\{\frac{d\rho(Z_i,\hat{\alpha}_n)}{d\theta_j} - \frac{d\rho(Z_i,\hat{\alpha}_n)}{dh}[w_j]\right\} \otimes p^{k_n}(X_i)\right).$$

Let $\widehat{w^*} = (\widehat{w_1^*}, \ldots, \widehat{w_{d_\theta}^*})$ be the estimator of $w^*$. The estimator of $V^{-1}$ is obtained by

$$\widehat{V^{-1}} = \left[\left(\frac{1}{n}\sum_{i=1}^n \left\{\frac{d\rho(Z_i,\hat{\alpha}_n)}{d\theta'} - \frac{d\rho(Z_i,\hat{\alpha}_n)}{dh}[\widehat{w^*}]\right\} \otimes p^{k_n}(X_i)\right)' \left(\frac{1}{n}\sum_{i=1}^n g_i(\hat{\alpha}_n)g_i(\hat{\alpha}_n)'\right)^{-1} \times \left(\frac{1}{n}\sum_{i=1}^n \left\{\frac{d\rho(Z_i,\hat{\alpha}_n)}{d\theta'} - \frac{d\rho(Z_i,\hat{\alpha}_n)}{dh}[\widehat{w^*}]\right\} \otimes p^{k_n}(X_i)\right)\right]^{-1}. \quad (5)$$

In contrast to the covariance estimator of AC (2003), we do not need to estimate $\Sigma(X,\alpha_0)$ directly. AC (2003) recommend to estimate $\Sigma(X,\alpha_0)$ by the least squares estimator. Although the sieve estimator is consistent for $\Sigma(X,\alpha_0)$, the fitted value can be negative or very close to zero in a finite sample.

To prove consistency of the estimator, we need the following additional condition.

https://doi.org/10.1017/S0266466616000360 Published online by Cambridge University Press

**Assumption 4.9.** For $j = 1, \ldots, d_\theta$,

$$\frac{d\rho(Z,\alpha)}{d\theta_j} - \frac{d\rho(Z,\alpha)}{dh}[w_j]$$

satisfies an envelope condition and is Hölder continuous in $\alpha \in \mathcal{N}_0$ and $w_j \in \{v \in \overline{W} : \|v\|_s \leq c < \infty\}$.

THEOREM 4.2. *Under Assumptions 3.1–3.11, 4.1–4.2, and 4.9,* $\widehat{V^{-1}} = V^{-1} + o_p(1)$.

Finally, we make a brief remark on another inference method. Chen and Pouzo (2009) propose an alternative method for constructing a confidence region for $\theta_0$ that avoids estimating the asymptotic variance of the estimator. Their confidence region is obtained by inverting the objective function of the SMD estimator, which is asymptotically chi-square distributed. Although we do not give a rigorous proof, it will also be possible to obtain a confidence region by inverting the objecting function of the SGEL estimator.

## 5. CONCLUSION

In this paper, we propose a GEL-based estimation method for semiparametric conditional moment restriction models. We extend the GEL estimator of DIN (2003) by allowing unknown functions to be included in the conditional moment restriction. Our SGEL estimator can be viewed as complementary to the methods of Otsu (2011) and AC (2003). We show that the SGEL estimator of the parametric component is asymptotically efficient.

Newey and Smith (2004) show that the GEL estimator outperforms the two-step GMM estimator in terms of the asymptotic bias for unconditional moment restriction models without unknown functions. We conjecture that a similar result also holds for semiparametric conditional moment restriction models. We can extend our theoretic result in this direction in future research.

Detailed proofs and some definitions and other assumptions are provided online at Cambridge Journals Online in supplementary material to this article. Readers may refer to the supplementary material associated with this article, available at Cambridge Journals Online (`journals.cambridge.org/ect`).

## *REFERENCES*

Ai, C. & X. Chen (2003) Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica* 71, 1795–1843.

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

Carrasco, M. & J.-P. Florens (2000) Generalization of GMM to a continuum of moment conditions. *Econometric Theory* 16, 797–834.

Chamberlain, G. (1987) Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics* 34, 305–334.

https://doi.org/10.1017/S0266466616000360 Published online by Cambridge University Press

Chen, X. & D. Pouzo (2009) Efficient estimation of semiparametric conditional moment models with possibly nonsmooth residuals. *Journal of Econometrics* 152, 46–60.
Dominguez, M.A. & I.N. Lobato (2004) Consistent estimation of models defined by conditional moment restrictions. *Econometrica* 2004, 1601–1615.
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.
Kitamura, Y., G. Tripathi, & H. Ahn (2004) Empirical likelihood-based inference in conditional moment restriction models. *Econometrica* 72, 1667–1714.
Newey, W.K. (1990) Efficient instrumental variable estimation of nonlinear models. *Econometrica* 58, 809–837.
Newey, W.K. (1991) Uniform convergence in probability and stochastic equicontinuity. *Econometrica* 59, 1161–1167.
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, 219–255.
Nishiyama, Y., Q. Liu, & N. Sueishi (2005) Semiparametric estimators for conditional moment restrictions containing nonparametric functions: Comparison of GMM and empirical likelihood procedures. In A. Zerger & R.M. Argent (eds.), *Proceedings of International Congress on Modelling and Simulation*, pp. 926–932. Modelling and Simulation Society of Australia and New Zealand.
Otsu, T. (2011) Empirical likelihood estimation of conditional moment restriction models with unknown functions. *Econometric Theory* 27, 8–46.
Qin, J. & J. Lawless (1994) Empirical likelihood and general estimating equations. *Annals of Statistics* 22, 300–325.
Robinson, P.M. (1987) Asymptotically efficient estimation in the presence of heteroskedasticity of unknown form. *Econometrica* 55, 875–891.
Shen, X. & W.H. Wong (1994) Convergence rate of sieve estimates. *Annals of Statistics* 22, 580–615.
Smith, R.J. (2006) Efficient information theoretic inference for conditional moment restrictions. *Journal of Econometrics* 138, 430–460.
Zhang, J. & I. Gijbels (2003) Sieve empirical likelihood and extensions of the generalized least squares. *Scandinavian Journal of Statistics* 30, 1–24.

# APPENDIX A

Throughout the Appendix, $C$ denotes a generic positive constant which may be different in different uses. For conciseness, only key lemmas are presented and their proofs are omitted. See the online supplement for the detailed proofs.

## A.1. Consistency

**Proof of Theorem 3.1.** Let $p_i = p^{k_n}(X_i)$ and $W = \left(n^{-1}\sum_{i=1}^n I \otimes p_i p_i'\right)^{-1}$. Also, let $\hat{g}(\alpha) = n^{-1}\sum_{i=1}^n g_i(\alpha)$. We define

$$\hat{R}(\alpha) = \hat{g}(\alpha)' W \hat{g}(\alpha),$$
$$R(\alpha) = E[E[\rho(Z,\alpha)|X]' E[\rho(Z,\alpha)|X]].$$

By Assumption 3.3 in the online supplement, which is an assumption on the identification of $\alpha_0$, we have

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

https://doi.org/10.1017/S0266466616000360 Published online by Cambridge University Press

for all $\alpha \neq \alpha_0$. Also, Corollary 4.2 of Newey (1991) implies that $R(\alpha)$ is continuous and $\sup_{\alpha\in\mathcal{A}}|\hat{R}(\alpha) - R(\alpha)| \xrightarrow{p} 0$. Thus, by Lemma A.1 of DIN (2003), it suffices to show that $\hat{R}(\hat{\alpha}_n) \xrightarrow{p} 0$. Because the minimum eigenvalue of $W^{-1}$ is bound from below with probability approaching one, it follows from Lemma A.4 in the online supplement that

$$\hat{g}(\hat{\alpha}_n)'W\hat{g}(\hat{\alpha}_n) \leq C\|\hat{g}(\hat{\alpha}_n)\|^2 = o_p(1)$$

and the desired result follows. ■

## A.2. Rate of Convergence

Let $Q(X_i,\alpha) = \Sigma(X_i,\alpha)^{1/2} \otimes p_i'$ and $Q(\alpha) = (Q(X_1,\alpha)', \ldots, Q(X_n,\alpha)')'$. Also, let $\psi(X,\alpha) = \Sigma(X,\alpha)^{-1/2}m(X,\alpha)$ and $\psi_0(X,\alpha) = \Sigma(X,\alpha_0)^{-1/2}m(X,\alpha)$. We define

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

Moreover, we denote

$$\hat{L}_n(\alpha) = -\sup_{\lambda\in\hat{\Lambda}(\alpha)}\frac{1}{n}\sum_{i=1}^{n} s(\lambda' g_i(\alpha)),$$

$$\bar{L}_n(\alpha) = -\hat{g}(\alpha)'\left(\frac{1}{n}\sum_{i=1}^{n} g_i(\alpha_0)g_i(\alpha_0)'\right)^{-1}\hat{g}(\alpha),$$

$$L_n(\alpha) = -\frac{1}{n}\sum_{i=1}^{n}\psi_0(X_i,\alpha)'\psi_0(X_i,\alpha).$$

Let $\mathcal{A}_{0n} = \{\alpha \in \mathcal{A}_n : \|\alpha - \alpha_0\|_s = o(1)\}$ and $\alpha_{n0} = \Pi_n\alpha_0$. Let $\hat{\lambda}_n(\alpha_{n0})$ satisfy

$$\hat{\lambda}_n(\alpha_{n0}) = \arg\max_{\lambda\in\hat{\Lambda}(\alpha_{n0})}\sum_{i=1}^{n} s(\lambda' g_i(\alpha_{n0})).$$

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

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

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

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

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

https://doi.org/10.1017/S0266466616000360 Published online by Cambridge University Press

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

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

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

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

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

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

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

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

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

which implies $P_1 \to 0$. Also, it follows from Lemma A.12 and (A.1) that

$$\sup_{\alpha \in \mathcal{A}_{0n}} |\hat{L}_{0n}(\alpha) - \eta_{0n} L_n(\alpha)| \le \sup_{\alpha \in \mathcal{A}_{0n}} |\eta_{0n} \bar{L}_n(\alpha) - \eta_{0n} L_n(\alpha)| + o_p(n^{-1/2})$$
$$= o_p(n^{-1/2}) = o_p(\eta_{0n}^2).$$

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

https://doi.org/10.1017/S0266466616000360 Published online by Cambridge University Press

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

## A.3. Asymptotic Normality

Let $\mathcal{N}_{0n} = \{\alpha \in \mathcal{A}_n : \|\alpha - \alpha_0\|_s = o(1), \|\alpha - \alpha_0\|_w = o(n^{-1/4})\}$. Let $v_n^* = (v_\theta^*, -\Pi_n w^* \times v_\theta^*)$ (see Assumption 4.2 in the online supplement). Denote

$$\frac{d\psi(X_i,\alpha)}{d\alpha}[v_n^*] = \Sigma(X_i,\alpha)^{-1/2}\frac{dm(X_i,\alpha)}{d\alpha}[v_n^*]$$

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

Moreover, we define

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

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

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

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

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

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

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

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

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

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

https://doi.org/10.1017/S0266466616000360 Published online by Cambridge University Press

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

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

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

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

$$0=\frac{1}{n}\sum_{i=1}^{n}s_1(\hat{\lambda}_n(\alpha)' g_i(\alpha))g_i(\alpha) \tag{A.3}$$

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

By Assumption 4.7, expanding (A.3) yields

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

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

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

Hence it follows that $\hat{\lambda}_n(\alpha)=-(n^{-1}\sum_{i=1}^{n}g_i(\alpha)g_i(\alpha)')^{-1}\hat{g}(\alpha)+o_p(n^{-1/2})$. Also, it can be shown that

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

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

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

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

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

https://doi.org/10.1017/S0266466616000360 Published online by Cambridge University Press

$$0 \le \hat{L}_n(\alpha(0)) - \hat{L}_n(\alpha(1)) = -\frac{d\hat{L}_n(\alpha(t))}{dt}\Big|_{t=0} - \frac{1}{2}\frac{d^2\hat{L}_n(\alpha(t))}{dt^2}\Big|_{t=s} \tag{A.4}$$

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

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

$$\begin{aligned}
&-\frac{d\hat{L}_n(\alpha(t))}{dt}\Big|_{t=0} \\
&= \frac{1}{n}\sum_{i=1}^n s_1(\hat{\lambda}_n' g_i(\hat{\alpha}_n))\hat{\lambda}_n' \frac{d\rho(Z_i,\hat{\alpha}_n)}{d\alpha}[\epsilon_n u_n^*] \otimes p_i \\
&= -\hat{\lambda}_n' \frac{1}{n}\sum_{i=1}^n \frac{d\rho(Z_i,\hat{\alpha}_n)}{d\alpha}[\epsilon_n u_n^*] \otimes p_i \\
&\quad + \hat{\lambda}_n' \left( \frac{\epsilon_n}{n}\sum_{i=1}^n s_2(\tilde{\lambda}' g_i(\hat{\alpha}_n)) \frac{d\rho(Z_i,\hat{\alpha}_n)}{d\alpha}[u_n^*]\rho(Z_i,\hat{\alpha}_n)' \otimes p_i p_i' \right) \hat{\lambda}_n + o_p(\epsilon_n n^{-1/2}) \\
&= \hat{g}(\hat{\alpha}_n)' \left( \frac{1}{n}\sum_{i=1}^n g_i(\hat{\alpha}_n) g_i(\hat{\alpha}_n)' \right)^{-1} \left( \frac{1}{n}\sum_{i=1}^n \frac{d\rho(Z_i,\hat{\alpha}_n)}{d\alpha}[\epsilon_n u_n^*] \otimes p_i \right) + o_p(\epsilon_n n^{-1/2}) \\
&= \hat{g}(\hat{\alpha}_n)' \left( \frac{1}{n}\sum_{i=1}^n \Sigma(X_i,\hat{\alpha}_n) \otimes p_i p_i' \right)^{-1} \left( \frac{1}{n}\sum_{i=1}^n \frac{d\rho(Z_i,\hat{\alpha}_n)}{d\alpha}[\epsilon_n u_n^*] \otimes p_i \right) + o_p(\epsilon_n n^{-1/2}) \\
&= \frac{1}{n}\sum_{i=1}^n \hat{\psi}(X_i,\hat{\alpha}_n)' \left\{ \frac{d\hat{\psi}(X_i,\hat{\alpha}_n)}{d\alpha}[\epsilon_n u_n^*] \right\} + o_p(\epsilon_n n^{-1/2}) \\
&= \frac{\epsilon_n}{n}\sum_{i=1}^n \left\{ \frac{dm(X_i,\alpha_0)}{d\alpha}[u^*] \right\}' \Sigma(X_i,\alpha_0)^{-1}\rho(Z_i,\alpha_0) \\
&\quad + \epsilon_n \langle u^*, \hat{\alpha}_n - \alpha_0 \rangle + o_p(\epsilon_n n^{-1/2}).
\end{aligned} \tag{A.5}$$

Next we denote $\frac{d\hat{\lambda}_n(\alpha(\tau))}{d\alpha}[\epsilon_n u_n^*] = \frac{d\hat{\lambda}_n(\alpha(t))}{dt}\big|_{t=\tau}$. By (A.3), we obtain

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

Since the minimum eigenvalue of $-n^{-1}\sum_{i=1}^n s_2(\hat{\lambda}_n(\alpha)' g_i(\alpha)) g_i(\alpha) g_i(\alpha)'$ is greater than $C$ with probability approaching one, we have

https://doi.org/10.1017/S0266466616000360 Published online by Cambridge University Press

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

Here we have

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

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

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

Denote $\hat{\lambda}' g_i(s) = \hat{\lambda}_n(\alpha(s))' g_i(\alpha(s))$. Then by Lemma A.17, we have

$$\frac{d^2 \hat{L}_n(\alpha(t))}{dt^2}\Big|_{t=s}$$
$$= \frac{1}{n}\sum_{i=1}^n s_2(\hat{\lambda}' g_i(s)) \frac{d\hat{\lambda}' g_i(s)}{d\alpha}[\epsilon_n u_n^*] \hat{\lambda}_n(\alpha(s))' \frac{d\rho(Z_i,\alpha(s))}{d\alpha}[\epsilon_n u_n^*] \otimes p_i$$
$$+ \frac{1}{n}\sum_{i=1}^n s_1(\hat{\lambda}' g_i(s)) \left\{ \frac{d\hat{\lambda}_n(\alpha(s))}{d\alpha}[\epsilon_n u_n^*] \right\}' \frac{d\rho(Z_i,\alpha(s))}{d\alpha}[\epsilon_n u_n^*] \otimes p_i$$
$$+ \frac{1}{n}\sum_{i=1}^n s_1(\hat{\lambda}' g_i(s)) \hat{\lambda}_n(\alpha(s))' \frac{d^2\rho(Z_i,\alpha(s))}{d\alpha d\alpha}[\epsilon_n u_n^*, \epsilon_n u_n^*] \otimes p_i$$
$$= \left\{ \frac{d\hat{\lambda}_n(\alpha(s))}{d\alpha}[u_n^*] \right\}' \left( \frac{\epsilon_n^2}{n}\sum_{i=1}^n s_2(\hat{\lambda}' g_i(s)) \rho(Z_i,\alpha) \left\{ \frac{d\rho(Z_i,\alpha(s))}{d\alpha}[u_n^*] \right\}' \otimes p_i p_i' \right) \hat{\lambda}_n(\alpha(s))$$
$$+ \hat{\lambda}_n(\alpha(s))' \left( \frac{\epsilon_n^2}{n}\sum_{i=1}^n s_2(\hat{\lambda}' g_i(s)) \left\{ \frac{d\rho(Z_i,\alpha(s))}{d\alpha}[u_n^*] \right\} \left\{ \frac{d\rho(Z_i,\alpha(s))}{d\alpha}[u_n^*] \right\}' \otimes p_i p_i' \right) \hat{\lambda}_n(\alpha(s))$$
$$+ \frac{\epsilon_n^2}{n}\sum_{i=1}^n s_1(\hat{\lambda}' g_i(s)) \left\{ \frac{d\hat{\lambda}_n(\alpha(s))}{d\alpha}[u_n^*] \right\}' \frac{d\rho(Z_i,\alpha(s))}{d\alpha}[u_n^*] \otimes p_i$$
$$+ \frac{\epsilon_n^2}{n}\sum_{i=1}^n s_1(\hat{\lambda}' g_i(s)) \hat{\lambda}_n(\alpha(s))' \frac{d^2\rho(Z_i,\alpha(s))}{d\alpha d\alpha}[u_n^*, u_n^*] \otimes p_i$$
$$= o_p(\epsilon_n^2). \qquad \textbf{(A.6)}$$

https://doi.org/10.1017/S0266466616000360 Published online by Cambridge University Press

Therefore, it follows from (A.4), (A.5), and (A.6) that

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

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

**Proof of Theorem 4.2.** Let $D_w(X, \alpha) = E\left[ \frac{d\rho(Z,\alpha)}{d\theta'} - \frac{d\rho(Z,\alpha)}{dh}[w] | X \right]$. Define

$$\hat{D}_w(X, \alpha) = Q(X_i, \alpha)(Q(\alpha)'Q(\alpha))^{-1}$$
$$\times \sum_{j=1}^{n} Q(X_j, \alpha)\Sigma(X_j, \alpha)^{-1/2} \left\{ \frac{d\rho(Z_j, \alpha)}{d\theta_j} - \frac{d\rho(Z_j, \alpha)}{dh}[w_j] \right\}.$$

Then, similarly to the proof of Lemma A.8, which is given in the online supplement, we can show that

$$\left\| \hat{D}_w(X, \alpha) - \Sigma(X, \alpha)^{-1/2} D_w(X, \alpha) \right\| = o_p(1),$$
$$\left\| \hat{D}_w(X, \alpha) - \Sigma(X, \alpha_0)^{-1/2} D_w(X, \alpha_0) \right\| = o_p(1)$$

uniformly over $X \in \mathcal{X}$, $\alpha \in \mathcal{N}_{0n}$ and $w_j \in \mathcal{H}_n$, $j = 1, \ldots, d_\theta$. As is noted by AC (2003), we can also show that $\|\widehat{w_j^*}(\cdot)\|_s < c$ by some constant $c$. Thus we only need to consider the subset $\{v \in \overline{\mathcal{W}} : \|v\|_s \leq c\}$. We use the sieve space $\mathcal{H}_n$ to approximate this subset. Observe that

$$\left( \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{d\rho(Z_i, \alpha)}{d\theta'} - \frac{d\rho(Z_i, \alpha)}{dh}[w] \right\} \otimes p_i \right)' \left( \frac{1}{n} \sum_{i=1}^{n} g_i(\alpha) g_i(\alpha)' \right)^{-1}$$
$$\times \left( \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{d\rho(Z_i, \alpha)}{d\theta'} - \frac{d\rho(Z_i, \alpha)}{dh}[w] \right\} \otimes p_i \right)$$
$$= \left( \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{d\rho(Z_i, \alpha)}{d\theta'} - \frac{d\rho(Z_i, \alpha)}{dh}[w] \right\} \otimes p_i \right)' \left( \frac{1}{n} \sum_{i=1}^{n} \Sigma(X_i, \alpha) \otimes p_i p_i' \right)^{-1}$$
$$\times \left( \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{d\rho(Z_i, \alpha)}{d\theta'} - \frac{d\rho(Z_i, \alpha)}{dh}[w] \right\} \otimes p_i \right) + o_p(1)$$
$$= \frac{1}{n} \sum_{i=1}^{n} \hat{D}_w(X_i, \alpha)' \hat{D}_w(X_i, \alpha) + o_p(1)$$

uniformly over $\alpha \in \mathcal{N}_{0n}$ and $w \in \mathcal{H}_n$ with $\|w\|_s \leq c$. Since $\hat{\alpha}_n \in \mathcal{N}_{0n}$ with probability approaching one, we have

https://doi.org/10.1017/S0266466616000360 Published online by Cambridge University Press

$$\left(\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{d\rho(Z_i,\hat{\alpha}_n)}{d\theta'}-\frac{d\rho(Z_i,\hat{\alpha}_n)}{dh}[w]\right\}\otimes p_i\right)'\left(\frac{1}{n}\sum_{i=1}^{n}g_i(\hat{\alpha}_n)g_i(\hat{\alpha}_n)'\right)^{-1}$$
$$\times\left(\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{d\rho(Z_i,\hat{\alpha}_n)}{d\theta'}-\frac{d\rho(Z_i,\hat{\alpha}_n)}{dh}[w]\right\}\otimes p_i\right)$$
$$=\frac{1}{n}\sum_{i=1}^{n}D_w(X_i,\alpha_0)'\Sigma(X_i,\alpha_0)^{-1}D_w(X_i,\alpha_0)+o_p(1)$$

uniformly over $w\in\mathcal{H}_n$ with $\|w\|_s\le c$. Finally, by Lemma A.1 of Newey and Powell (2003), we have $\|\hat{w}-w^*\|_s=o_p(1)$ and hence the result follows. ■

https://doi.org/10.1017/S0266466616000360 Published online by Cambridge University Press