SUPPLEMENT TO "INSTRUMENTAL VARIABLE TREATMENT OF NONCLASSICAL MEASUREMENT ERROR MODELS"

BY YINGYAO HU AND SUSANNE M. SCHENNACH[1]

This supplementary material contains some of the more technical details omitted from the main paper. First, the asymptotic theory of the proposed sieve maximum likelihood estimator is fully developed, providing suitable regularity conditions, a nonparametric consistency result, and a semiparametric asymptotic normality and root n consistency result. Second, we provide an example that shows the necessity, for identification purposes, of our location constraint assumption regarding the measurement error. Third, a detailed example that illustrates the implementation of this location constraint with linear sieves is given. Finally, additional simulation results are reported.

S1. ASYMPTOTICS

LET US FIRST RECALL the assumptions needed for identification.

ASSUMPTION 1: *The joint density of y and x, x^*, z admits a bounded density with respect to the product measure of some dominating measure μ (defined on $\mathcal{Y}$) and the Lebesgue measure on $\mathcal{X} \times \mathcal{X}^* \times \mathcal{Z}$. All marginal and conditional densities are also bounded.*

ASSUMPTION 2: (i) $f_{y|xx^*z}(y|x, x^*, z) = f_{y|x^*}(y|x^*)$ *for all* $(y, x, x^*, z) \in \mathcal{Y} \times \mathcal{X} \times \mathcal{X}^* \times \mathcal{Z}$ *and* (ii) $f_{x|x^*z}(x|x^*, z) = f_{x|x^*}(x|x^*)$ *for all* $(x, x^*, z) \in \mathcal{X} \times \mathcal{X}^* \times \mathcal{Z}$.

ASSUMPTION 3: *The operators $L_{x|x^*}$ and $L_{z|x}$ are injective (for either $\mathcal{G} = \mathcal{L}^1$ or $\mathcal{G} = \mathcal{L}^1_{\mathrm{bnd}}$).*

ASSUMPTION 4: *For all $x_1^*, x_2^* \in \mathcal{X}^*$, the set $\{y : f_{y|x^*}(y|x_1^*) \neq f_{y|x^*}(y|x_2^*)\}$ has positive probability (under the marginal of y) whenever $x_1^* \neq x_2^*$.*

ASSUMPTION 5: *There exists a known functional M such that $M[f_{x|x^*}(\cdot|x^*)] = x^*$ for all $x^* \in \mathcal{X}^*$.*

Our sieve estimator is based on the following expression for the observed density (following Theorem 1 in the main text):

$$(\text{S1}) \qquad f_{yx|z}(y, x|z; \alpha_0) = \int_{\mathcal{X}^*} f_{y|x^*}(y|x^*; \theta_0) f_{x|x^*}(x|x^*) f_{x^*|z}(x^*|z)\, dx^*.$$

The unknown α_0 in the density function $f_{yx|z}$ includes θ_0 and density functions $f_{x|x^*}$ and $f_{x^*|z}$, that is, $\alpha_0 = (\theta_0, f_{x|x^*}, f_{x^*|z})^T$. The estimation procedure basically consists of replacing $f_{x|x^*}$ and $f_{x^*|z}$ (and $f_{y|x^*}$ if it contains an infinite-dimensional nuisance parameter η) by truncated series approximations and optimizing all parameters within a semiparametric maximum likelihood framework. The number of terms kept in the series approximations is allowed to grow with sample size at a controlled rate.

Our asymptotic analysis relies on standard smoothness restrictions (e.g., Ai and Chen (2003)) on the unknown functions η, $f_{x|x^*}$, and $f_{x^*|z}$. To describe them, let $\xi \in \mathcal{V} \subset \mathbb{R}^d$, $a = (a_1, \ldots, a_d)^T$, and

$$\nabla^a g(\xi) \equiv \frac{\partial^{a_1 + \cdots + a_d} g(\xi)}{\partial \xi_1^{a_1} \cdots \partial \xi_d^{a_d}}$$

denote the $(a_1 + \cdots + a_d)$th derivative. Let $\| \cdot \|_E$ denote the Euclidean norm. Let $\underline{\gamma}$ denote the largest integer satisfying $\gamma > \underline{\gamma}$. The Hölder space $\Lambda^\gamma(\mathcal{V})$ of order $\gamma > 0$ is a space of functions $g : \mathcal{V} \mapsto \mathbb{R}$ such that the first $\underline{\gamma}$ derivative is bounded, and the $\underline{\gamma}$th derivative are Hölder continuous with the exponent $\gamma - \underline{\gamma} \in (0, 1]$, that is,

$$\max_{a_1 + \cdots + a_d = \underline{\gamma}} |\nabla^a g(\xi) - \nabla^a g(\xi')| \leq c (\|\xi - \xi'\|_E)^{\gamma - \underline{\gamma}}$$

for all $\xi, \xi' \in \mathcal{V}$ and some constant c. The Hölder space becomes a Banach space with the Hölder norm as follows:

$$\|g\|_{\Lambda^\gamma} = \sup_{\xi \in \mathcal{V}} |g(\xi)| + \max_{a_1 + \cdots + a_d = \underline{\gamma}} \sup_{\xi \neq \xi' \in \mathcal{V}} \frac{|\nabla^a g(\xi) - \nabla^a g(\xi')|}{(\|\xi - \xi'\|_E)^{\gamma - \underline{\gamma}}}.$$

To facilitate the treatment of functions defined on noncompact domains, we follow the technique suggested in Chen, Hong, and Tamer (2005), introducing a weighting function of the form $\omega(\xi) = (1 + \|\xi\|_E^2)^{-\varsigma/2}$, $\varsigma > \gamma > 0$, and defining a weighted Hölder norm as $\|g\|_{\Lambda^{\gamma,\omega}} \equiv \|\tilde{g}\|_{\Lambda^\gamma}$ for $\tilde{g}(\xi) \equiv g(\xi)\omega(\xi)$. The corresponding weighted Hölder space is denoted by $\Lambda^{\gamma,\omega}(\mathcal{V})$, while a weighted Hölder ball can be defined as $\Lambda_c^{\gamma,\omega}(\mathcal{V}) \equiv \{g \in \Lambda^{\gamma,\omega}(\mathcal{V}) : \|g\|_{\Lambda^{\gamma,\omega}} \leq c < \infty\}$.

We assume the functions η, $f_{x|x^*}$, and $f_{x^*|z}$ belong to the sets $\mathcal{M}$, $\mathcal{F}_1$, and $\mathcal{F}_2$, respectively, defined below.

ASSUMPTION 6: $\eta \in \Lambda_c^{\gamma_1,\omega}(\mathcal{U})$ with $\gamma_1 > 1$.[2]

ASSUMPTION 7: $f_1 \in \Lambda_c^{\gamma_1,\omega}(\mathcal{X} \times \mathcal{X}^*)$ with $\gamma_1 > 1$ and $\int_{\mathcal{X}} f_1(x|x^*)\,dx = 1$ for all $x^* \in \mathcal{X}^*$.

---

[2]If η is a density function, certain restrictions should be added to Assumption 6 analogous to those in Assumptions 8 and 7.

ASSUMPTION 8: $f_2 \in \Lambda_c^{\gamma_1, \omega}(\mathcal{X}^* \times \mathcal{Z})$ *with* $\gamma_1 > 1$ *and* $\int_{\mathcal{X}^*} f_2(x^*|z) \, dx^* = 1$ *for all* $z \in \mathcal{Z}$.

$$\mathcal{M} = \{\eta(\cdot, \cdot) : \text{Assumption 6 holds}\},$$

$$\mathcal{F}_1 = \{f_1(\cdot|\cdot) : \text{Assumptions 3, 5, and 7 hold}\},$$

$$\mathcal{F}_2 = \{f_2(\cdot|\cdot) : \text{Assumptions 3 and 8 hold}\}.$$

The condition $\|f\|_{\Lambda^{\gamma_1, \omega}} \leq c < \infty$ is necessary for the method of sieves, which we will use in the next step. In principle, one can solve for the true value $\alpha_0 = (\theta_0, f_{x|x^*}, f_{x^*|z})^T$ as

$$\alpha_0 = \operatorname*{arg\,max}_{\alpha = (\theta, f_1, f_2)^T \in \mathcal{A}} E\left(\ln \int_{\mathcal{X}^*} f_{y|x^*}(y|x^*; \theta) f_1(x|x^*) f_2(x^*|z) \, dx^*\right),$$

where $\mathcal{A} = \Theta \times \mathcal{F}_1 \times \mathcal{F}_2$ with $\Theta = \mathcal{B} \times \mathcal{M}$. Let $p_j^{k_n}(\cdot)$ be a sequence of known univariate basis functions, such as power series, splines, Fourier series, and so forth. To approximate functions of two variables, we use a tensor–product linear sieve basis, denoted by $p^{k_n}(\cdot, \cdot) = (p_1^{k_n}(\cdot, \cdot), p_2^{k_n}(\cdot, \cdot), \ldots, p_{k_n}^{k_n}(\cdot, \cdot))^T$. In the sieve approximation, we consider η, f_1, and f_2 in finite-dimensional spaces $\mathcal{M}_n$, $\mathcal{F}_{1n}$, and $\mathcal{F}_{2n}$, where[3]

$$\mathcal{M}_n = \{\eta(\xi_1, \xi_2) = p^{k_n}(\xi_1, \xi_2)^T \delta \text{ for all } \delta$$

$$\text{s.t. Assumption 6 holds}\},$$

$$\mathcal{F}_{1n} = \{f(x|x^*) = p^{k_n}(x, x^*)^T \beta \text{ for all } \beta$$

$$\text{s.t. Assumptions 3, 5, and 7 hold}\},$$

$$\mathcal{F}_{2n} = \{f(x^*|z) = p^{k_n}(x^*, z)^T \gamma \text{ for all } \gamma$$

$$\text{s.t. Assumptions 3 and 8 hold}\}.$$

Therefore, we replace $\mathcal{M} \times \mathcal{F}_1 \times \mathcal{F}_2$ with $\mathcal{M}_n \times \mathcal{F}_{1n} \times \mathcal{F}_{2n}$ in the optimization problem and then estimate α_0 by $\widehat{\alpha}_n$ as

$$\widehat{\alpha}_n = (\widehat{\theta}_n, \widehat{f}_{1n}, \widehat{f}_{2n})^T$$

$$= \operatorname*{arg\,max}_{\alpha = (\theta, f_1, f_2)^T \in \mathcal{A}_n} \frac{1}{n} \sum_{i=1}^n \ln \int_{\mathcal{X}^*} f_{y|x^*}(y_i|x^*; \theta) f_1(x_i|x^*) f_2(x^*|z_i) \, dx^*,$$

where $\mathcal{A}_n = \Theta_n \times \mathcal{F}_{1n} \times \mathcal{F}_{2n}$ with $\Theta_n = \mathcal{B} \times \mathcal{M}_n$. In practice, the above integral can be conveniently carried out through either one of a number of numerical

---

[3]For simplicity, the notation $p^{k_n}(\cdot, \cdot)$ implicitly assumes that the sieves for η, $f(x|x^*)$, and $f(x^*|z)$ are the same, although this can be easily relaxed.

techniques, including Gaussian quadrature, Simpson's rules, importance sampling, or Markov chain Monte Carlo. In the sequel, we simply assume that this integral can be evaluated, for a given sample and a given truncated sieve, with a numerical accuracy that is far better than the statistical noise associated with the estimation procedure.

This setup is the same as in Shen (1997). We also use techniques described in Ai and Chen (2003) to state more primitive regularity conditions. In their paper, there are two sieve approximations: One is used to directly estimate the conditional mean as a function of the unknown parameter; the other is the sieve approximation of the infinite-dimensional parameter estimated through the maximization procedure. Our setup is, in some ways, simpler than in Ai and Chen (2003), because all the unknown parameters in α are estimated through a single-step semiparametric sieve MLE (maximum likelihood estimator). Since our estimator takes the form of a semiparametric sieve estimator, the very general treatment of Shen (1997) and Chen and Shen (1998) can be used to establish asymptotic normality and root n consistency under a very wide variety of conditions, including dependent and nonidentically distributed data. However, for the purposes of simplicity and conciseness, this section provides specific sufficient regularity conditions for the independent and identically distributed (i.i.d.) case.

The restrictions in the definitions of $\mathcal{F}_{1n}$ and $\mathcal{F}_{2n}$ are easy to impose on a sieve estimator. We have the sieve expressions of f_1 and f_2 as

$$f_1(x|x^*) = \sum_{i=0}^{i_n} \sum_{j=0}^{j_n} \beta_{ij} p_i(x - x^*) p_j(x^*),$$

$$f_2(x^*|z) = \sum_{i=0}^{i_n} \sum_{j=0}^{j_n} \gamma_{ij} p_i(x^* - z) p_j(z),$$

where $p_i(\cdot)$ are user-specified basis functions. Define $k_n = (i_n + 1)(j_n + 1)$ and assume that i_n/j_n is bounded and bounded away from zero for all n. We also define the projection of the true value α_0 onto the space $\mathcal{A}_n$ associated with k_n,

$$\Pi_n \alpha \equiv \alpha_n$$
$$\equiv \underset{\alpha_n = (\theta, f_1, f_2)^T \in \mathcal{A}_n}{\arg\max} E\left(\ln \int_{\mathcal{X}^*} f_{y|x^*}(y|x^*; \theta) f_1(x|x^*) f_2(x^*|z) \, dx^*\right),$$

and we let the smoothing parameter $k_n \to \infty$ as the sample size $n \to \infty$. The restriction $\int_{\mathcal{X}} f_1(x|x^*) \, dx = 1$ in the definition of $\mathcal{F}_{1n}$ implies $\sum_{j=0}^{j_n}(\sum_{i=0}^{i_n} \beta_{ij} \times \int_{\mathcal{E}} p_i(\varepsilon) \, d\varepsilon) p_j(x^*) = 1$ for all x^*, where $\varepsilon = x - x^*$. Suppose $p_0(\cdot)$ is the only constant in $p_j(\cdot)$. That equation implies that $\sum_{i=0}^{i_n} \beta_{i0} \int_{\mathcal{E}} p_i(\varepsilon) \, d\varepsilon = 1$ and $\sum_{i=0}^{i_n} \beta_{ij} \int_{\mathcal{E}} p_i(\varepsilon) \, d\varepsilon = 0$ for $j = 1, 2, \ldots, j_n$. Similar restrictions can be found

for $\int_{\mathcal{X}^*} f_2(x^*|z)\,dx^* = 1$. Moreover, the identification assumption, Assumption 5, also implies restrictions on the sieve coefficients. For example, consider the zero mode case. If the mode is unique and not at a boundary, we then have $\frac{\partial}{\partial x} f_{x|x^*}(x|x^*) = 0$ if and only if $x = x^*$. The restriction $\frac{\partial}{\partial x} f_{x|x^*}(x|x^*)|_{x=x^*} = 0$ in the definition of $\mathcal{F}_{1n}$ implies $\sum_{j=0}^{j_n}(\sum_{i=0}^{i_n} \beta_{ij}(\partial p_i(0))/\partial x)q_j(x^*) = 0$. Since it must hold for all x^*, we have additional j_n constraints $\sum_{i=0}^{i_n} \beta_{ij} \frac{\partial p_i(0)}{\partial x} = 0$ for $j = 1, 2, \ldots, j_n$. Similar restrictions can be found for the zero mean and the zero median cases. In all three cases, Assumption 5 can be expressed as linear restrictions on β, which are easy to implement. See Section S4 for an explicit expression for the restrictions in the case where Fourier series are used in the sieve approximation.

S1.1. *Consistency*

We use the results in Newey and Powell (2003) to show consistency of the sieve estimator. Define $D \equiv (y, x, z)$ for $y \in \mathcal{Y}$, $x \in \mathcal{X}$, and $z \in \mathcal{Z}$. The random variables x, y, and z can have unbounded support $\mathbb{R}$. Following Ai and Chen (2003), we first show consistency under a strong norm $\|\cdot\|_s$ as a stepping stone to establishing a convergence rate under a suitably constructed weaker norm. Let

$$\|\alpha\|_s = \|b\|_E + \|\eta\|_{\infty,\omega} + \|f_1\|_{\infty,\omega} + \|f_2\|_{\infty,\omega},$$

where $\|g\|_{\infty,\omega} \equiv \sup_\xi |g(\xi)\omega(\xi)|$ with $\omega(\xi) = (1 + \|\xi\|_E^2)^{-\varsigma/2}$, $\varsigma > \gamma_1 > 0$. We make the following assumptions:

ASSUMPTION 9: (i) *The data* $\{(Y_i, X_i, Z_i)_{i=1}^n\}$ *are i.i.d.* (ii) *The density of* $D \equiv (y, x, z)$, f_D, *satisfies* $\int \omega(D)^{-2} f_D(D)\,dD < \infty$.

ASSUMPTION 10: (i) $b_0 \in \mathcal{B}$, *a compact subset of* $\mathbb{R}^b$. (ii) *Assumptions 6–8 hold for* (b, η, f_1, f_2) *in a neighborhood of* α_0 *(in the norm* $\|\cdot\|_s$*).*

ASSUMPTION 11: (i) $E[(\ln f_{yx|z}(D))^2]$ *is bounded.* (ii) *There exists a measurable function* $h_1(D)$ *with* $E\{(h_1(D))^2\} < \infty$ *such that, for any* $\overline{\alpha} = (\overline{\theta}, \overline{f}_1, \overline{f}_2)^T \in \mathcal{A}$,

$$\left| \frac{f_{yx|z}^{|1|}(D, \overline{\alpha}, \bar{\omega})}{f_{yx|z}(D, \overline{\alpha})} \right| \le h_1(D),$$

where $f_{yx|z}^{|1|}(D, \overline{\alpha}, \bar{\omega})$ *is defined as* $\frac{d}{dt} f_{yx|z}(D; \overline{\alpha} + t\bar{\omega})|_{t=0}$ *with each linear term, that is,* $\frac{d}{d\theta} f_{y|x^*}$, $\overline{f}_1$, *and* $\overline{f}_2$, *replaced by its absolute value, and* $\overline{\omega}(\xi, x, x^*, z) = [1, \omega^{-1}(\xi), \omega^{-1}((x, x^*)^T), \omega^{-1}((x^*, z)^T)]^T$ *with* $\xi \in \mathcal{U}$. *(The explicit expression of* $f_{yx|z}^{|1|}(D, \overline{\alpha}, \bar{\omega})$ *can be found in Equation* (S6) *in the proof of Lemma 2.)*

ASSUMPTION 12: $\|\Pi_n\alpha_0 - \alpha_0\|_s = o(1)$ (as $k_n \to \infty$) and $k_n/n \to 0$.

Assumption 9 is commonly used in cross-sectional analyses. Assumption 9(ii) is a typical condition on the tail behavior on the density, analogous to Assumption 3.2 in Chen, Hong, and Tamer (2005). Assumption 10 imposes restrictions on the parameter space. Detailed discussions on this assumption can be found in Gallant and Nychka (1987). Assumption 11 imposes an envelope condition on the first derivative of the log likelihood function and guarantees a Hölder continuity property for the log likelihood. Assumption 12 states that the sieve can approximate the true α_0 arbitrarily well, to control the bias, while ensuring that the number of terms in the sieve grows slower than the sample size, to control the variance. We show consistency in the following lemma.

LEMMA 2: *Under Assumptions 1–5 and 9–12, we have* $\|\widehat{\alpha}_n - \alpha_0\|_s = o_p(1)$.

See Section S2 for the proof.

Consistency under the norm $\|\cdot\|_s$ is the first step needed to obtain the asymptotic properties of the estimator. To proceed toward our main semiparametric asymptotic normality and root n consistency result, we then need to establish convergence at the rate $o_p(n^{-1/4})$ in a suitable norm. To achieve this convergence rate under relatively weak assumptions, we employ a device introduced by Ai and Chen (2003) and employ a weaker norm $\|\cdot\|$, under which $o_p(n^{-1/4})$ convergence is easier to establish.

We now recall the concept of pathwise derivative, which is central to the asymptotics of sieve estimators. Consider $\alpha_1, \alpha_2 \in \mathcal{A}$, and assume the existence of a continuous path $\{\alpha(\tau) : \tau \in [0,1]\}$ in $\mathcal{A}$ such that $\alpha(0) = \alpha_1$ and $\alpha(1) = \alpha_2$. If $\ln f_{yx|z}(D, (1-\tau)\alpha_0 + \tau\alpha)$ is continuously differentiable at $\tau = 0$ for almost all D and any $\alpha \in \mathcal{A}$, the pathwise derivative of $\ln f_{yx|z}(D, \alpha_0)$ at α_0 evaluated at $\alpha - \alpha_0$ can be defined as

$$\frac{d\ln f_{yx|z}(D, \alpha_0)}{d\alpha}[\alpha - \alpha_0] \equiv \frac{d\ln f_{yx|z}(D, (1-\tau)\alpha_0 + \tau\alpha)}{d\tau}\bigg|_{\tau=0}$$

almost everywhere (under the probability measure of D). The pathwise derivative is a linear functional that approximates $\ln f_{yx|z}(D, \alpha_0)$ in the neighborhood of α_0, that is, for small values of $\alpha - \alpha_0$. Note that this functional can also be evaluated for other values of the argument. For instance, by linearity,

$$\frac{d\ln f_{yx|z}(D, \alpha_0)}{d\alpha}[\alpha_1 - \alpha_2]$$
$$\equiv \frac{d\ln f_{yx|z}(D, \alpha_0)}{d\alpha}[\alpha_1 - \alpha_0] - \frac{d\ln f_{yx|z}(D, \alpha_0)}{d\alpha}[\alpha_2 - \alpha_0].$$

In our setting, the pathwise derivative at α_0 is (from Equation (S1))

$$\frac{d \ln f_{yx|z}(D, \alpha_0)}{d\alpha}[\alpha - \alpha_0]$$

$$= \frac{1}{f_{yx|z}(D, \alpha_0)} \left\{ \int_{\mathcal{X}^*} \frac{d}{d\theta} f_{y|x^*}(y|x^*; \theta_0) \right.$$

$$\times [\theta - \theta_0] f_{x|x^*}(x|x^*) f_{x^*|z}(x^*|z) \, dx^*$$

$$+ \int_{\mathcal{X}^*} f_{y|x^*}(y|x^*; \theta_0)[f_1(x|x^*) - f_{x|x^*}(x|x^*)] f_{x^*|z}(x^*|z) \, dx^*$$

$$\left. + \int_{\mathcal{X}^*} f_{y|x^*}(y|x^*; \theta_0) f_{x|x^*}(x|x^*)[f_2(x^*|z) - f_{x^*|z}(x^*|z)] \, dx^* \right\}.$$

Note that the denominator $f_{yx|z}(D, \alpha_0)$ is nonzero with probability 1. We use the Fisher norm $\| \cdot \|$ defined as

$$(S2) \qquad \|\alpha_1 - \alpha_2\| \equiv \sqrt{E\left\{ \left(\frac{d \ln f_{yx|z}(D, \alpha_0)}{d\alpha}[\alpha_1 - \alpha_2] \right)^2 \right\}}$$

for any $\alpha_1, \alpha_2 \in \mathcal{A}$. To establish the asymptotic normality of $\widehat{b}_n$, one typically needs $\widehat{\alpha}_n$ to converge to α_0 at a rate faster than $n^{-1/4}$. We need the following assumptions to obtain this rate of convergence:

ASSUMPTION 13: $\|\Pi_n \alpha_0 - \alpha_0\| = O(k_n^{-\gamma_1/d_1}) = o(n^{-1/4})$ *with* $d_1 = 2$ *and* $\gamma_1 > d_1$, *for* γ_1 *as in Assumptions 6–8.*[4]

ASSUMPTION 14: (i) *There exists a measurable function* $c(D)$ *with* $E\{c(D)^4\} < \infty$ *such that* $|\ln f_{yx|z}(D; \alpha)| \leq c(D)$ *for all* D *and* $\alpha \in \mathcal{A}_n$. (ii) $\ln f_{yx|z}(D; \alpha) \in \Lambda_c^{\gamma,\omega}(\mathcal{Y} \times \mathcal{X} \times \mathcal{Z})$ *for some constant* $c > 0$ *with* $\gamma > d_D/2$, *for all* $\alpha \in \mathcal{A}_n$, *where* d_D *is the dimension of* D.

ASSUMPTION 15: $\mathcal{A}$ *is convex in* α_0 *and* $f_{y|x^*}(y|x^*; \theta)$ *is pathwise differentiable at* θ_0.

ASSUMPTION 16: *For some* $c_1, c_2 > 0$,

$$c_1 E\left(\ln \frac{f_{yx|z}(D; \alpha_0)}{f_{yx|z}(D; \alpha)} \right) \leq \|\alpha - \alpha_0\|^2 \leq c_2 E\left(\ln \frac{f_{yx|z}(D; \alpha_0)}{f_{yx|z}(D; \alpha)} \right)$$

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

---

[4]In general, $d_1 = \max\{\dim(\mathcal{U}), \dim(\mathcal{X} \times \mathcal{X}^*), \dim(\mathcal{X}^* \times \mathcal{Z})\}$.

ASSUMPTION 17: $(k_n n^{-1/2} \ln n) \sup_{(\xi_1, \xi_2) \in (\mathcal{U} \cup (\mathcal{X} \times \mathcal{X}^*) \cup (\mathcal{X}^* \times \mathcal{Z}))} \| p^{k_n}(\xi_1, \xi_2) \|_E^2 = o(1)$.

ASSUMPTION 18: $\ln N(\varepsilon, \mathcal{A}_n) = O(k_n \ln(k_n/\varepsilon))$, *where* $N(\varepsilon, \mathcal{A}_n)$ *is the minimum number of balls* (*in the* $\| \cdot \|_s$ *norm*) *needed to cover the set* $\mathcal{A}_n$.

Assumption 13 controls the approximation error of $\Pi_n \alpha_0$ to α_0 and the selection of k_n. It is usually satisfied by using sieve functions such as power series, Fourier series, and so forth (see Newey (1995, 1997) for more discussion). Assumption 14 imposes an envelope condition and a smoothness condition on the log likelihood function. Assumption 15 implies that the norm $\| \cdot \|$ is well defined. Define $K(\alpha, \alpha_0) = E(\ln(f_{yx|z}(D; \alpha_0))/(f_{yx|z}(D; \alpha)))$, which is the Kullback–Leibler discrepancy. Assumption 16 implies that $\| \cdot \|$ is a norm equivalent to the $(K(\cdot, \cdot))^{1/2}$ discrepancy on $\mathcal{A}_n$. Under the norm $\| \cdot \|$, the sieve estimator can be shown to converge at the requisite rate $o_p(n^{-1/4})$.

THEOREM 2: *Under Assumptions 1–5 and 9–18, we have* $\| \widehat{\alpha}_n - \alpha_0 \| = o_p(n^{-1/4})$.

The proof is given in Section S2.

It may appear surprising at first that such a fast convergence rate could be obtained in a nonparametric estimation problem that includes, as a special case, models traditionally handled through deconvolution approaches and that are known to be prone to slow convergence issues (e.g., Fan (1991)). These issues can be circumvented, thanks to the fact that the Fisher norm downweighs each dimension of the estimation error $\hat{\alpha} - \alpha_0$ according to its own standard error. In other words, more error is tolerated along the dimensions that are more difficult to estimate. Assumption 16 does impose a limit on how weak the Fisher norm can be, however. In the limit where the Fisher norm becomes singular (i.e., completely insensitive to some dimensions of α), the local quadratic behavior of the objective function is lost and Assumption 16 no longer holds.

Thanks to the Fisher norm's downweighting property, as the number of terms in the sieve increases, each new degree of freedom that gets included in the estimation problem does not appear increasingly difficult to estimate. A relatively fast convergence in the Fisher norm is therefore possible and does not conflict with slower convergence obtained in some other norm. Naturally, for the same reason, convergence in the Fisher norm is not a very useful concept for the sole purpose of establishing a nonparametric convergence result. In nonparametric settings, convergence in some well-understood L_p norm would be a more useful result. However, our ultimate goal is to establish the asymptotics for some parametric component of our semiparametric model. In that context, the Fisher norm is a very useful device that was employed in Ai and Chen (2003) and that guarantees the important intermediate results of $o_p(n^{-1/4})$ convergence under rather weak conditions.

S1.2. *Asymptotic Normality*

We follow the semiparametric MLE framework of Shen (1997) to show the asymptotic normality of the estimator $\widehat{b}_n$. We define the inner product

$$(S3) \qquad \langle v_1, v_2 \rangle = E\left\{ \left(\frac{d \ln f_{yx|z}(D, \alpha_0)}{d\alpha}[v_1] \right) \left(\frac{d \ln f_{yx|z}(D, \alpha_0)}{d\alpha}[v_2] \right) \right\}.$$

Obviously, the weak norm $\| \cdot \|$ defined in Equation (S2) can be induced by this inner product. Let $\overline{\mathbf{V}}$ denote the closure of the linear span of $\mathcal{A} - \{\alpha_0\}$ under the norm $\| \cdot \|$ (i.e., $\overline{\mathbf{V}} = \mathbb{R}^{d_b} \times \overline{\mathcal{W}}$ with $\overline{\mathcal{W}} \equiv \overline{\mathcal{M} \times \mathcal{F}_1 \times \mathcal{F}_2} - \{(\eta_0, f_{x|x^*}, f_{x^*|z})^T\}$) and define the Hilbert space $(\overline{\mathbf{V}}, \langle \cdot, \cdot \rangle)$ with its inner product defined in Equation (S3).

As shown above, we have

$$\frac{d \ln f_{yx|z}(D, \alpha_0)}{d\alpha}[\alpha - \alpha_0]$$

$$= \frac{d \ln f_{yx|z}(D, \alpha_0)}{db}[b - b_0] + \frac{d \ln f_{yx|z}(D, \alpha_0)}{d\eta}[\eta - \eta_0]$$

$$+ \frac{d \ln f_{yx|z}(D, \alpha_0)}{df_1}[f_1 - f_{x|x^*}] + \frac{d \ln f_{yx|z}(D, \alpha_0)}{df_2}[f_2 - f_{x^*|z}].$$

For each component b_j of b, $j = 1, 2, \ldots, d_b$, we define $w_j^* \in \overline{\mathcal{W}}$ as

$$w_j^* \equiv (\eta_j^*, f_{1j}^*, f_{2j}^*)^T$$

$$= \underset{(\eta, f_1, f_2)^T \in \overline{\mathcal{W}}}{\arg \min} E\left\{ \left(\frac{d \ln f_{yx|z}(D, \alpha_0)}{db_j} - \frac{d \ln f_{yx|z}(D, \alpha_0)}{d\eta}[\eta] \right.\right.$$

$$\left.\left. - \frac{d \ln f_{yx|z}(D, \alpha_0)}{df_1}[f_1] - \frac{d \ln f_{yx|z}(D, \alpha_0)}{df_2}[f_2] \right)^2 \right\}.$$

Define $w^* = (w_1^*, w_2^*, \ldots, w_{d_b}^*)$,

$$\frac{d \ln f_{yx|z}(D, \alpha_0)}{df}[w_j^*] = \frac{d \ln f_{yx|z}(D, \alpha_0)}{d\eta}[\eta_j^*] + \frac{d \ln f_{yx|z}(D, \alpha_0)}{df_1}[f_{1j}^*]$$

$$+ \frac{d \ln f_{yx|z}(D, \alpha_0)}{df_2}[f_{2j}^*],$$

$$\frac{d \ln f_{yx|z}(D, \alpha_0)}{df}[w^*] = \left(\frac{d \ln f_{yx|z}(D, \alpha_0)}{df}[w_1^*], \ldots, \right.$$

$$\left. \frac{d \ln f_{yx|z}(D, \alpha_0)}{df}[w_{d_b}^*] \right),$$

and the row vector

$$(S4) \qquad G_{w^*}(D) = \frac{d \ln f_{yx|z}(D, \alpha_0)}{db^T} - \frac{d \ln f_{yx|z}(D, \alpha_0)}{df}[w^*].$$

We want to show that $\widehat{b}_n$ has a multivariate normal distribution asymptotically. It is well known that if $\lambda^T b$ has a normal distribution for all λ, then b has a multivariate normal distribution. Therefore, we consider $\lambda^T b$ as a functional of α. Define $s(\alpha) \equiv \lambda^T b$ for $\lambda \in \mathbb{R}^{d_b}$ and $\lambda \neq 0$. If $E[G_{w^*}(D)^T G_{w^*}(D)]$ is finite positive definite, then the function $s(\alpha)$ is bounded, and the Riesz representation theorem implies that there exists a representor v^* such that

$$(S5) \qquad s(\alpha) - s(\alpha_0) \equiv \lambda^T(b - b_0) = \langle v^*, \alpha - \alpha_0 \rangle$$

for all $\alpha \in \mathcal{A}$. Here $v^* \equiv \begin{pmatrix} v_b^* \\ v_f^* \end{pmatrix}$, $v_b^* = J^{-1}\lambda$, and $v_f^* = -w^* v_b^*$ with $J = E[G_{w^*}(D)^T \times G_{w^*}(D)]$. Under suitable assumptions made below, the Riesz representor v^* exists and is bounded.

As mentioned in Begun, Hall, Huang, and Wellner (1983), v_f^* corresponds to a worst possible direction of approach to $(\eta_0, f_{x|x^*}, f_{x^*|z})$ for the problem of estimating b_0. In the language of Stein (1956), v_f^* yields the most difficult one-dimensional subproblem. Equation (S5) implies that it is sufficient to find the asymptotic distribution of $\langle v^*, \widehat{\alpha}_n - \alpha_0 \rangle$ to obtain that of $\lambda^T(\widehat{b}_n - b_0)$ under suitable conditions. We denote

$$\frac{d \ln f_{yx|z}(D, \alpha)}{d\alpha}[v] \equiv \frac{d \ln f_{yx|z}(D, \alpha + \tau v)}{d\tau}\bigg|_{\tau=0} \quad \text{a.s. } D \quad \text{for any } v \in \overline{\mathbf{V}}.$$

For a sieve MLE, we have that

$$\langle v^*, \widehat{\alpha}_n - \alpha_0 \rangle = \frac{1}{n} \sum_{i=1}^n \frac{d \ln f_{yx|z}(D_i, \alpha_0)}{d\alpha}[v^*] + o_p(n^{-1/2}).$$

Note that $((d \ln f_{yx|z}(D, \alpha))/d\alpha[v^*]) = G_{w^*}(D)J^{-1}\lambda$. Thus, by the classical central limit theorem, the asymptotic distribution of $\sqrt{n}(\widehat{b}_n - b_0)$ is $N(0, J^{-1})$. In fact, the matrix J is the efficient information matrix in this semiparametric estimation, under suitable regularity conditions given in Shen (1997).

We now present the sufficient conditions for the $\sqrt{n}$-normality of $\widehat{b}_n$. Define

$$\mathcal{N}_{0n} \equiv \{\alpha \in \mathcal{A}_n : \|\alpha - \alpha_0\|_s \leq \upsilon_n, \|\alpha - \alpha_0\| \leq \upsilon_n n^{-1/4}\}$$

with $\upsilon_n = o(1)$ and define $\mathcal{N}_0$ the same way with $\mathcal{A}_n$ replaced by $\mathcal{A}$. Note that $\mathcal{N}_0$ still depends on n. For $\alpha \in \mathcal{N}_{0n}$ we define a local alternative $\alpha^*(\alpha, \varepsilon_n) = (1 - \varepsilon_n)\alpha + \varepsilon_n(v^* + \alpha_0)$ with $\varepsilon_n = o(n^{-1/2})$. Let $\Pi_n \alpha^*(\alpha, \varepsilon_n)$ be the projection of $\alpha^*(\alpha, \varepsilon_n)$ onto $\mathcal{A}_n$.

ASSUMPTION 19: (i) $E[G_{w^*}(D)^T G_{w^*}(D)]$ *exists, is bounded, and is positive-definite.* (ii) $b_0 \in \text{int}(\mathcal{B})$.

ASSUMPTION 20: *There exists a measurable function $h_2(D)$ with $E\{(h_2(D))^2\} < \infty$ such that for any $\overline{\alpha} = (\overline{\theta}, \overline{f}_1, \overline{f}_2)^T \in \mathcal{N}_0$,*

$$\left|\frac{f_{yx|z}^{|1|}(D, \overline{\alpha}, \bar{\omega})}{f_{yx|z}(D, \overline{\alpha})}\right|^2 + \left|\frac{f_{yx|z}^{|2|}(D, \overline{\alpha}, \bar{\omega})}{f_{yx|z}(D, \overline{\alpha})}\right| < h_2(D),$$

where $f_{yx|z}^{|2|}(D, \overline{\alpha}, \bar{\omega})$ is defined as $(d^2/dt^2)f_{yx|z}(D; \overline{\alpha} + t\bar{\omega})|_{t=0}$ with each linear term, that is, $\frac{d}{d\theta}f_{y|x^*}$, $\frac{d^2}{d\theta^2}f_{y|x^*}$, $\overline{f}_1$, and $\overline{f}_2$, replaced by its absolute value. (The explicit expression of $f_{yx|z}^{|2|}(D, \overline{\alpha}, \bar{\omega})$ can be found in Equation (S17) in the proof of Theorem 3.)

We introduce the following notations for the next assumption: for $\widetilde{f} = \eta$, f_1, or f_2,

$$\frac{d \ln f_{yx|z}(D, \alpha_0)}{d\widetilde{f}}[p^{k_n}]$$
$$= \left(\frac{d \ln f_{yx|z}(D, \alpha_0)}{d\widetilde{f}}[p_1^{k_n}], \frac{d \ln f_{yx|z}(D, \alpha_0)}{d\widetilde{f}}[p_2^{k_n}],\right.$$
$$\left. \ldots, \frac{d \ln f_{yx|z}(D, \alpha_0)}{d\widetilde{f}}[p_{k_n}^{k_n}]\right)^T,$$

$$\frac{d \ln f_{yx|z}(D, \alpha_0)}{db}$$
$$= \left(\frac{d \ln f_{yx|z}(D, \alpha_0)}{db_1}, \frac{d \ln f_{yx|z}(D, \alpha_0)}{db_2}, \ldots, \frac{d \ln f_{yx|z}(D, \alpha_0)}{db_{d_b}}\right)^T,$$

$$\frac{d \ln f_{yx|z}(D, \alpha_0)}{d\alpha}[p^{k_n}]$$
$$= \left(\left(\frac{d \ln f_{yx|z}(D, \alpha_0)}{db}\right)^T, \left(\frac{d \ln f_{yx|z}(D, \alpha_0)}{d\eta}[p^{k_n}]\right)^T,\right.$$
$$\left.\left(\frac{d \ln f_{yx|z}(D, \alpha_0)}{df_1}[p^{k_n}]\right)^T, \left(\frac{d \ln f_{yx|z}(D, \alpha_0)}{df_2}[p^{k_n}]\right)^T\right)^T,$$

and

$$\Omega_{k_n} = E\left\{\left(\frac{d \ln f_{yx|z}(D, \alpha_0)}{d\alpha}[p^{k_n}]\right)\left(\frac{d \ln f_{yx|z}(D, \alpha_0)}{d\alpha}[p^{k_n}]\right)^T\right\}.$$

ASSUMPTION 21: *The smallest eigenvalue of the matrix Ω_{k_n} is bounded away from zero, and $\|p_j^{k_n}\|_{\infty, \omega} < \infty$ for $j = 1, 2, \ldots, k_n$ uniformly in k_n.*

ASSUMPTION 22: *There is a* $v_n^* = \left(\begin{smallmatrix} v_b^* \\ -(\Pi_n w^*)v_b^* \end{smallmatrix}\right) \in \mathcal{A}_n - \{\Pi_n\alpha_0\}$ *such that* $\|v_n^* - v^*\| = o(n^{-1/4})$.

ASSUMPTION 23: *For all* $\alpha \in \mathcal{N}_{0n}$, *there exists a measurable function* $h_4(D)$ *with* $E|h_4(D)| < \infty$ *such that*

$$\left| \frac{d^4}{dt^4} \ln f_{yx|z}(D; \overline{\alpha} + t(\alpha - \alpha_0)) \right|_{t=0} \leq h_4(D)\|\alpha - \alpha_0\|_s^4.$$

Assumption 19 is essential to obtain root n consistency since it ensures that the asymptotic variance exists and that b_0 is an "interior" solution. Assumption 20 imposes an envelope condition on the second derivative of the log likelihood function. This condition is related to the stochastic equicontinuity condition, Condition A, in Shen (1997). The condition guarantees the linear approximation of the likelihood function by its derivative near α_0. That condition can be replaced by a stronger condition that $f_{yx|z}(D, \alpha)$ is differentiable in quadratic mean. Assumption 21 is similar to Assumption 2 in Newey (1997). Intuitively, Assumptions 21 and 23 are used to characterizes the local quadratic behavior of the criterion difference, that is, Condition B in Shen (1997), and can be simplified to: for all $\alpha \in \mathcal{N}_{0n}$,

$$E\left(\ln \frac{f_{yx|z}(D, \alpha_0)}{f_{yx|z}(D, \alpha)} \right) = \frac{1}{2}\|\alpha - \alpha_0\|^2(1 + o(1)).$$

Assumption 22 states that the representor can be approximated by the sieve with an asymptotically negligible error, which is an important necessary condition for the asymptotic bias of the sieve estimator itself to be asymptotically negligible. A detailed discussion of these assumptions can be found in Shen (1997) and Chen and Shen (1998). By Theorem 1 in Shen (1997), we show that the estimator for the parametric component b_0 is $\sqrt{n}$ consistent and asymptotically normally distributed.

THEOREM 3: *Under Assumptions* 1–5, 9–16, *and* 19–23, $\sqrt{n}(\widehat{b}_n - b_0) \xrightarrow{d} N(0, J^{-1})$, *where* $J = E[G_{w^*}(D)^T G_{w^*}(D)]$ *for* $G_{w^*}(D)$ *given in Equation* (S4).

See Section S2 for the proof.

Achieving the level of generality provided by Theorem 3 forces us to state some of our regularity conditions is a relatively high-level form, as is often done in the sieve estimation literature (e.g., Ai and Chen (2003), Shen (1997), Chen and Shen (1998)). However, once the type of sieve and the particular form of $f_{y|x^*}(y|x^*; \theta)$ are specified, more primitive assumptions can be formulated, using some of the techniques found in Blundell, Chen, and Kristensen (2007), for instance.

It is known that obtaining a root n consistency and asymptotic normality result for a semiparametric estimator in the context of classical errors-in-variables models demands a balance between the smoothness of the measurement error and of the densities (or regression functions) of interest (e.g., Taupin (1998), Schennach (2004)). Our treatment, when specialized to classical measurement errors, does not evade this requirement. When the measurement error densities are "too smooth" and the functions of interest are "not smooth enough" to guarantee root n consistency and asymptotic normality, this will manifests itself as a violation of one of our assumptions. If the failure is first order, that is, it is due to the inexistence of an influence function with bounded variance, then a bounded Riesz representor v^* will fail to exist and Assumptions 19 and 22 will not hold. If the failure is of a "higher-order" nature, that is, when nonlinear remainder terms in the estimator's stochastic expansion are not negligible, then any one of Assumption 20, 21, or 23 will not hold. Intuitively, this represents a case where the local quadratic behavior of the objective function is lost.

S2. PROOFS

PROOF OF LEMMA 2: First note that Assumptions 1–5 imply that the model is identified so that α_0 is uniquely defined. We prove the results by checking the conditions in Theorem 4.1 in Newey and Powell (2003). Their Assumption 1 on identification of the unknown parameter is assumed directly. We assume $k_n \to \infty$ and $k_n/n \to 0$ in Assumption 12 so that the relevant part of their Assumption 2 is satisfied. Note that we do not have any "plug-in" nonparametric part in the likelihood function. The first part of their Condition 3 is assumed in our Assumption 11(i). For the rest of their Condition 3, we consider pathwise derivative

$$\ln f_{yx|z}(D; \alpha_1) - \ln f_{yx|z}(D; \alpha_2)$$
$$= \frac{d \ln f_{yx|z}(D, \overline{\alpha}_0)}{d\alpha}[\alpha_1 - \alpha_2]$$
$$= \frac{d}{dt} \ln f_{yx|z}(D; \overline{\alpha}_0 + t(\alpha_1 - \alpha_2)) \bigg|_{t=0},$$

where $\overline{\alpha}_0 = (\overline{\theta}, \overline{f}_1, \overline{f}_2)^T$ is a mean value between α_1 and α_2. Letting $\alpha_1 = (\theta_1, f_{11}, f_{21})^T$ and $\alpha_2 = (\theta_2, f_{12}, f_{22})^T$, we have

$$\frac{d}{dt} \ln f_{yx|z}(D; \overline{\alpha}_0 + t(\alpha_1 - \alpha_2)) \bigg|_{t=0}$$
$$= \frac{1}{f_{yx|z}(D, \overline{\alpha}_0)} \left\{ \int \frac{d}{d\theta} f_{y|x^*}(y|x^*; \overline{\theta})(\theta_1 - \theta_2)\overline{f}_1(x|x^*)\overline{f}_2(x^*|z) \, dx^* \right.$$

$$+ \int f_{y|x^*}(y|x^*; \overline{\theta})[f_{11} - f_{12}]\overline{f}_2(x^*|z) \, dx^*$$

$$+ \int f_{y|x^*}(y|x^*; \overline{\theta})\overline{f}_1(x|x^*)[f_{21} - f_{22}] \, dx^* \Big\}.$$

The bounds can be found as

$$(S6) \qquad \left| \frac{d}{dt} \ln f_{yx|z}(D; \overline{\alpha}_0 + t(\alpha_1 - \alpha_2)) \right|_{t=0}$$

$$\leq \frac{1}{|f_{yx|z}(D, \overline{\alpha}_0)|} \Big\{ \int \left| \frac{d}{d\theta} f_{y|x^*}(y|x^*; \overline{\theta})\omega^{-1}(\xi)\overline{f}_1(x|x^*)\overline{f}_2(x^*|z) \right| dx^*$$

$$\times \|\theta_1 - \theta_2\|_s$$

$$+ \int |f_{y|x^*}(y|x^*; \overline{\theta})\omega^{-1}(x, x^*)\overline{f}_2(x^*|z)| \, dx^* \, \|f_{11} - f_{12}\|_s$$

$$+ \int |f_{y|x^*}(y|x^*; \overline{\theta})\overline{f}_1(x|x^*)\omega^{-1}(x^*, z)| \, dx^* \, \|f_{21} - f_{22}\|_s \Big\}$$

$$\leq \frac{1}{|f_{yx|z}(D, \overline{\alpha}_0)|} \Big\{ \int \left| \frac{d}{d\theta} f_{y|x^*}(y|x^*; \overline{\theta})\omega^{-1}(\xi)\overline{f}_1(x|x^*)\overline{f}_2(x^*|z) \right| dx^*$$

$$+ \int |f_{y|x^*}(y|x^*; \overline{\theta})\omega^{-1}(x, x^*)\overline{f}_2(x^*|z)| \, dx^*$$

$$+ \int |f_{y|x^*}(y|x^*; \overline{\theta})\overline{f}_1(x|x^*)\omega^{-1}(x^*, z)| \, dx^* \Big\} \|\alpha - \alpha_0\|_s$$

$$\equiv \left| \frac{f_{yx|z}^{|1|}(D, \overline{\alpha}_0, \bar{\omega})}{f_{yx|z}(D, \overline{\alpha}_0)} \right| \|\alpha - \alpha_0\|_s,$$

where $f_{yx|z}^{|1|}(D, \overline{\alpha}_0, \bar{\omega})$ is defined as $\frac{d}{dt} f_{yx|z}(D; \overline{\alpha}_0 + t\bar{\omega})|_{t=0}$ with each linear term, that is, $\frac{d}{d\theta} f_{y|x^*}$, $\overline{f}_1$, and $\overline{f}_2$, replaced by its absolute value. The function $\bar{\omega}$ is defined as

$$\overline{\omega}(\xi, x, x^*, z) = \left[1, \omega^{-1}(\xi), \omega^{-1}((x, x^*)^T), \omega^{-1}((x^*, z)^T) \right]^T$$

with $\xi \in \mathcal{U}$. Therefore, our Assumption 11(ii), that is, $E((f_{yx|z}^{|1|}(D, \overline{\alpha}_0, \bar{\omega}))/(f_{yx|z}(D, \overline{\alpha}_0)))^2 \leq E(h_1(D))^2 < \infty$, implies that $\ln f_{yx|z}(D, \alpha)$ is Hölder continuous in α. Therefore, their Condition 3 holds. Assumption 10 guarantees that $\mathcal{A}$ is compact under the norm $\|\cdot\|_s$, which is their Condition 4. From Chen, Hansen, and Scheinkman (1997), for any $\alpha \in \mathcal{A}$

$$(S7) \qquad \|\alpha - \Pi_n \alpha\|_s \leq \|\eta - \Pi_n \eta\|_s + \|f_1 - \Pi_n f_1\|_s + \|f_2 - \Pi_n f_2\|_s$$

$$= O(k_n^{-\gamma_1/d_1})$$

with $d_1 = 2$. Therefore, their Condition 5 is satisfied with our Assumption 12. A similar proof can also be found in that of Lemma 3.1 and Proposition 3.1 in Ai and Chen (2003).   *Q.E.D.*

PROOF OF THEOREM 2: First note that Assumptions 2–5 imply that the model is identified so that α_0 is uniquely defined. We prove the results by checking the conditions in Theorem 3.1 in Ai and Chen (2003). Note that there are two different estimated criterion functions, that is, $L_n(\alpha)$ and $\widehat{L}_n(\alpha)$ in their Appendix B (Ai and Chen (2003, p. 1825)). In our setup, we do not have that distinction and their proof still applies with $L_n(\alpha) = \frac{1}{n}\sum_{i=1}^{n} \ln f_{yx|z}(D_i, \alpha)$. From the proof of Lemma 2, Assumptions 11 and 13 imply their Condition 3.5(iii), that is, $\|\alpha - \Pi_n\alpha\| = o(n^{-1/4})$. Assumptions 3.6(iii), 3.7, and 3.8 in Chen and Shen (1998) are assumed directly in our Assumptions 14, 17, and 18, respectively. According to its expression, $f_{yx|z}(D; \alpha)$ is pathwise differentiable at α_0 if $f_{y|x^*}(y|x^*; \theta)$ is pathwise differentiable at θ_0. Therefore, Assumption 15 implies their Condition 3.9(i). Condition 3.9(ii) in Ai and Chen (2003) is assumed directly in Assumption 16. Thus, the results of consistency follow.   *Q.E.D.*

PROOF OF THEOREM 3: First note that Assumptions 1–5 imply that the model is identified so that α_0 is uniquely defined. We prove the results by checking the conditions in Theorem 1 in Shen (1997). We define the remainder term as

$$r[\alpha - \alpha_0, D] \equiv \ln f_{yx|z}(D, \alpha) - \ln f_{yx|z}(D, \alpha_0)$$
$$- \frac{d\ln f_{yx|z}(D, \alpha_0)}{d\alpha}[\alpha - \alpha_0].$$

We also define $\mu_n(g) = \frac{1}{n}\sum_{i=1}^{n}[g(D, \alpha) - Eg(D, \alpha)]$ as the empirical process induced by g. We have the sieve estimator $\widehat{\alpha}_n$ for α_0 and a local alternative $\alpha^*(\widehat{\alpha}_n, \varepsilon_n) = (1 - \varepsilon_n)\widehat{\alpha}_n + \varepsilon_n(v^* + \alpha_0)$ with $\varepsilon_n = o(n^{-1/2})$. Let $\Pi_n\alpha^*(\alpha, \varepsilon_n)$ be the projection of $\alpha^*(\alpha, \varepsilon_n)$ to $\mathcal{A}_n$.

First of all, the Riesz representor v^* is finite because the matrix J is invertible and w^* is bounded. Second, Equation (4.2) in Shen (1997), that is,

$$\left| s(\alpha) - s(\alpha_0) - \frac{ds(\alpha)}{d\alpha}[\alpha - \alpha_0] \right| \leq c\|\alpha - \alpha_0\|^\omega$$

as $\|\alpha - \alpha_0\| \to 0$, is required by Theorem 1 in that paper and holds trivially in our paper with $\omega = \infty$ because we have $s(\alpha) \equiv \lambda^T b$.

Third, Condition A in Shen (1997) requires

$$\sup_{\alpha \in \mathcal{N}_{0n}} \mu_n \big(r[\alpha - \alpha_0, D] - r[\Pi_n \alpha^*(\alpha, \varepsilon_n) - \alpha_0, D] \big) = O_p(\varepsilon_n^2).$$

By the definition of $r[\alpha - \alpha_0, D]$, we have

$$\mu_n \big(r[\alpha - \alpha_0, D] - r[\Pi_n \alpha^*(\alpha, \varepsilon_n) - \alpha_0, D] \big)$$

$$= \mu_n \left\{ \left(\ln f_{yx|z}(D, \alpha) - \ln f_{yx|z}(D, \alpha_0) - \frac{d \ln f_{yx|z}(D, \alpha_0)}{d\alpha}[\alpha - \alpha_0] \right) \right.$$

$$- \left(\ln f_{yx|z}(D, \Pi_n \alpha^*(\alpha, \varepsilon_n)) - \ln f_{yx|z}(D, \alpha_0) \right.$$

$$\left. \left. - \frac{d \ln f_{yx|z}(D, \alpha_0)}{d\alpha}[\Pi_n \alpha^*(\alpha, \varepsilon_n) - \alpha_0] \right) \right\}$$

$$= \mu_n \left(\ln f_{yx|z}(D, \alpha) - \ln f_{yx|z}(D, \Pi_n \alpha^*(\alpha, \varepsilon_n)) \right.$$

$$\left. - \frac{d \ln f_{yx|z}(D, \alpha_0)}{d\alpha}[\alpha - \Pi_n \alpha^*(\alpha, \varepsilon_n)] \right).$$

The Taylor expansion gives

$$\ln f_{yx|z}(D, \alpha) - \ln f_{yx|z}(D, \Pi_n \alpha^*(\alpha, \varepsilon_n))$$

$$= \frac{d \ln f_{yx|z}(D, \Pi_n \alpha^*(\alpha, \varepsilon_n))}{d\alpha}[\alpha - \Pi_n \alpha^*(\alpha, \varepsilon_n)]$$

$$+ \frac{1}{2} \frac{d^2 \ln f_{yx|z}(D, \widetilde{\alpha}_1)}{d\alpha \, d\alpha^T}[\alpha - \Pi_n \alpha^*(\alpha, \varepsilon_n), \alpha - \Pi_n \alpha^*(\alpha, \varepsilon_n)],$$

where $\widetilde{\alpha}_1$ is a mean value between α and $\Pi_n \alpha^*(\alpha, \varepsilon_n)$. Therefore, we have

(S8) $\quad \mu_n \big(r[\alpha - \alpha_0, D] - r[\Pi_n \alpha^*(\alpha, \varepsilon_n) - \alpha_0, D] \big)$

$$= \mu_n \left(\frac{d \ln f_{yx|z}(D, \Pi_n \alpha^*(\alpha, \varepsilon_n))}{d\alpha}[\alpha - \Pi_n \alpha^*(\alpha, \varepsilon_n)] \right.$$

$$\left. - \frac{d \ln f_{yx|z}(D, \alpha_0)}{d\alpha}[\alpha - \Pi_n \alpha^*(\alpha, \varepsilon_n)] \right)$$

$$+ \mu_n \left(\frac{1}{2} \frac{d^2 \ln f_{yx|z}(D, \widetilde{\alpha}_1)}{d\alpha \, d\alpha^T}[\alpha - \Pi_n \alpha^*(\alpha, \varepsilon_n), \alpha - \Pi_n \alpha^*(\alpha, \varepsilon_n)] \right).$$

Since

$$\alpha - \Pi_n \alpha^*(\alpha, \varepsilon_n) = \varepsilon_n \Pi_n(\alpha - \alpha_0 - v^*),$$

the right-hand side of Equation (S8) equals

$$
\begin{aligned}
\text{(S9)} \quad &= \mu_n\left(\frac{d^2\ln f_{yx|z}(D,\overline{\alpha}_1)}{d\alpha\, d\alpha^T}[\alpha - \Pi_n\alpha^*(\alpha,\varepsilon_n), \Pi_n\alpha^*(\alpha,\varepsilon_n) - \alpha_0]\right) \\
&\quad + \mu_n\left(\frac{1}{2}\frac{d^2\ln f_{yx|z}(D,\widetilde{\alpha}_1)}{d\alpha\, d\alpha^T}[\varepsilon_n\Pi_n(\alpha - \alpha_0 - v^*), \varepsilon_n\Pi_n(\alpha - \alpha_0 - v^*)]\right) \\
&= \mu_n\left(\frac{d^2\ln f_{yx|z}(D,\overline{\alpha}_1)}{d\alpha\, d\alpha^T}[\varepsilon_n\Pi_n(\alpha - \alpha_0 - v^*), \Pi_n\alpha^*(\alpha,\varepsilon_n) - \alpha_0]\right) \\
&\quad + \mu_n\left(\frac{1}{2}\frac{d^2\ln f_{yx|z}(D,\widetilde{\alpha}_1)}{d\alpha\, d\alpha^T}[\varepsilon_n\Pi_n(\alpha - \alpha_0 - v^*), \varepsilon_n\Pi_n(\alpha - \alpha_0 - v^*)]\right) \\
&= \varepsilon_n\mu_n\left(\frac{d^2\ln f_{yx|z}(D,\overline{\alpha}_1)}{d\alpha\, d\alpha^T}\right. \\
&\qquad \times [\Pi_n(\alpha - \alpha_0 - v^*), \varepsilon_n\Pi_n(v^* + \alpha_0 - \alpha) + (\alpha - \alpha_0)]\Big) \\
&\quad + \varepsilon_n^2\mu_n\left(\frac{1}{2}\frac{d^2\ln f_{yx|z}(D,\widetilde{\alpha}_1)}{d\alpha\, d\alpha^T}[\Pi_n(\alpha - \alpha_0 - v^*), \Pi_n(\alpha - \alpha_0 - v^*)]\right) \\
&= \varepsilon_n\mu_n\left(\frac{d^2\ln f_{yx|z}(D,\overline{\alpha}_1)}{d\alpha\, d\alpha^T}[\Pi_n(\alpha - \alpha_0 - v^*), \alpha - \alpha_0]\right) \\
&\quad - \varepsilon_n^2\mu_n\left(\frac{1}{2}\frac{d^2\ln f_{yx|z}(D,\overline{\alpha}_1)}{d\alpha\, d\alpha^T}[\Pi_n(\alpha - \alpha_0 - v^*), \Pi_n(\alpha - \alpha_0 - v^*)]\right) \\
&\quad + \varepsilon_n^2\mu_n\left(\frac{1}{2}\frac{d^2\ln f_{yx|z}(D,\widetilde{\alpha}_1)}{d\alpha\, d\alpha^T}[\Pi_n(\alpha - \alpha_0 - v^*), \Pi_n(\alpha - \alpha_0 - v^*)]\right) \\
&= A_1 + A_2 + A_3,
\end{aligned}
$$

where $\overline{\alpha}_1$ is a mean value between α_0 and $\Pi_n\alpha^*(\alpha,\varepsilon_n)$. We consider the term A_1 as

$$
\text{(S10)} \quad \sup_{\alpha\in\mathcal{N}_{0n}} A_1 = \varepsilon_n \sup_{\alpha\in\mathcal{N}_{0n}} \mu_n\left(\frac{d^2\ln f_{yx|z}(D,\overline{\alpha}_1)}{d\alpha\, d\alpha^T}[\Pi_n(\alpha - \alpha_0 - v^*), \alpha - \alpha_0]\right).
$$

Let $\overline{\alpha}_1 = (\overline{\theta}, \overline{f}_1, \overline{f}_2)$ and $v_n = \Pi_n(\alpha - \alpha_0 - v^*) = ([v_n]_\theta, [v_n]_{f_1}, [v_n]_{f_2})$. We consider the term

$$
\text{(S11)} \quad \left|\sup_{\alpha\in\mathcal{N}_{0n}} \frac{d^2\ln f_{yx|z}(D,\overline{\alpha}_1)}{d\alpha\, d\alpha^T}[v_n, \alpha - \alpha_0]\right|
$$

$$
\leq \sup_{\alpha\in\mathcal{N}_{0n}} \left|\frac{1}{f_{yx|z}(D,\overline{\alpha}_1)}\frac{d^2 f_{yx|z}(D,\overline{\alpha}_1)}{d\alpha\, d\alpha^T}[v_n, (\alpha - \alpha_0)]\right.
$$

$$- \frac{d\ln f_{yx|z}(D,\overline{\alpha}_1)}{d\alpha}[v_n] \frac{d\ln f_{yx|z}(D,\overline{\alpha}_1)}{d\alpha}[\alpha - \alpha_0] \Bigg|$$

$$\leq \sup_{\alpha \in \mathcal{N}_{0n}} \left(\left| \frac{1}{f_{yx|z}(D,\overline{\alpha}_1)} \frac{d^2 f_{yx|z}(D,\overline{\alpha}_1)}{d\alpha\, d\alpha^T}[v_n, (\alpha - \alpha_0)] \right| \right.$$

$$\left. + \left| \frac{d\ln f_{yx|z}(D,\overline{\alpha}_1)}{d\alpha}[v_n] \right| \left| \frac{d\ln f_{yx|z}(D,\overline{\alpha}_1)}{d\alpha}[\alpha - \alpha_0] \right| \right).$$

We need to find the bounds on three terms in the absolute value. We have

$$(S12) \qquad \frac{d\ln f_{yx|z}(D,\overline{\alpha}_1)}{d\alpha}[\alpha - \alpha_0]$$

$$= \frac{1}{f_{yx|z}(D,\overline{\alpha}_1)} \left\{ \int \frac{d}{d\theta} f_{y|x^*}(y|x^*;\overline{\theta})(\theta - \theta_0)\overline{f}_1(x|x^*)\overline{f}_2(x^*|z)\, dx^* \right.$$

$$+ \int f_{y|x^*}(y|x^*;\overline{\theta})[f_1 - f_{x|x^*}]\overline{f}_2(x^*|z)\, dx^*$$

$$\left. + \int f_{y|x^*}(y|x^*;\overline{\theta})\overline{f}_1(x|x^*)[f_2 - f_{x^*|z}]\, dx^* \right\}.$$

Therefore, the term $|(d\ln f_{yx|z}(D,\overline{\alpha}_1))/d\alpha[\alpha - \alpha_0]|$ can be bounded through

$$(S13) \qquad \left| \frac{d\ln f_{yx|z}(D,\overline{\alpha}_1)}{d\alpha}[\alpha - \alpha_0] \right|$$

$$\leq \frac{1}{|f_{yx|z}(D,\overline{\alpha}_1)|} \left\{ \int \left| \frac{d}{d\theta} f_{y|x^*}(y|x^*;\overline{\theta})\omega^{-1}(\xi)\overline{f}_1(x|x^*)\overline{f}_2(x^*|z) \right| dx^* \right.$$

$$\times \|\theta - \theta_0\|_s$$

$$+ \int |f_{y|x^*}(y|x^*;\overline{\theta})\omega^{-1}(x,x^*)\overline{f}_2(x^*|z)|\, dx^* \|f_1 - f_{x|x^*}\|_s$$

$$\left. + \int |f_{y|x^*}(y|x^*;\overline{\theta})\overline{f}_1(x|x^*)\omega^{-1}(x^*,z)|\, dx^* \|f_2 - f_{x^*|z}\|_s \right\}$$

$$\leq \left| \frac{f_{yx|z}^{|1|}(D,\overline{\alpha}_1,\overline{\omega})}{f_{yx|z}(D,\overline{\alpha}_1)} \right| \|\alpha - \alpha_0\|_s,$$

where $f_{yx|z}^{|1|}(D,\overline{\alpha}_1,\overline{\omega})$ is defined in Assumption 11 and Equation (S6). Similarly, we also have

$$(S14) \qquad \left| \frac{d\ln f_{yx|z}(D,\overline{\alpha}_1)}{d\alpha}[v_n] \right| \leq \left| \frac{f_{yx|z}^{|1|}(D,\overline{\alpha}_1,\overline{\omega})}{f_{yx|z}(D,\overline{\alpha}_1)} \right| \|v_n\|_s$$

with

(S15) $\qquad \|v_n\|_s = \|\Pi_n(\alpha - \alpha_0 - v^*)\|_s \leq \|v_n^*\|_s + \|\Pi_n(\alpha - \alpha_0)\|_s < \infty.$

We then consider the term $1/(f_{yx|z}(D, \overline{\alpha}_1))(d^2 f_{yx|z}(D, \overline{\alpha}_1))/(d\alpha\, d\alpha^T)[v_n, (\alpha - \alpha_0)]$ as

(S16) $\qquad \dfrac{1}{f_{yx|z}(D, \overline{\alpha}_1)} \dfrac{d^2 f_{yx|z}(D, \overline{\alpha}_1)}{d\alpha\, d\alpha^T}[v_n, (\alpha - \alpha_0)]$

$$= \frac{1}{f_{yx|z}(D, \overline{\alpha}_1)} \left\{ \int \frac{d^2}{d\theta^2} f_{y|x^*}(y|x^*; \overline{\theta})[v_n]_\theta (\theta - \theta_0) \right.$$

$$\times \overline{f}_1(x|x^*)\overline{f}_2(x^*|z)\, dx^*$$

$$+ \int \frac{d}{d\theta} f_{y|x^*}(y|x^*; \overline{\theta})[v_n]_\theta [f_1 - f_{x|x^*}]\overline{f}_2(x^*|z)\, dx^*$$

$$+ \int \frac{d}{d\theta} f_{y|x^*}(y|x^*; \overline{\theta})[v_n]_\theta \overline{f}_1(x|x^*)[f_2 - f_{x^*|z}]\, dx^*$$

$$+ \int \frac{d}{d\theta} f_{y|x^*}(y|x^*; \overline{\theta})(\theta - \theta_0)[v_n]_{f_1}\overline{f}_2(x^*|z)\, dx^*$$

$$+ \int f_{y|x^*}(y|x^*; \overline{\theta})[v_n]_{f_1}[f_2 - f_{x^*|z}]\, dx^*$$

$$+ \int \frac{d}{d\theta} f_{y|x^*}(y|x^*; \overline{\theta})(\theta - \theta_0)\overline{f}_1(x|x^*)[v_n]_{f_2}\, dx^*$$

$$\left. + \int f_{y|x^*}(y|x^*; \overline{\theta})[f_1 - f_{x|x^*}][v_n]_{f_2}\, dx^* \right\}.$$

Therefore, the term $|1/(f_{yx|z}(D, \overline{\alpha}_1))(d^2 f_{yx|z}(D, \overline{\alpha}_1))/(d\alpha\, d\alpha^T)[v_n, (\alpha - \alpha_0)]|$ can be bounded through

(S17) $\qquad \left| \dfrac{1}{f_{yx|z}(D, \overline{\alpha}_1)} \dfrac{d^2 f_{yx|z}(D, \overline{\alpha}_1)}{d\alpha\, d\alpha^T}[v_n, (\alpha - \alpha_0)] \right|$

$$\leq \frac{1}{|f_{yx|z}(D, \overline{\alpha}_1)|} \left\{ \int \left| \frac{d^2}{d\theta^2} f_{y|x^*}(y|x^*; \overline{\theta})\omega^{-1}(\xi)\omega^{-1}(\xi) \right. \right.$$

$$\left. \times \overline{f}_1(x|x^*)\overline{f}_2(x^*|z) \right| dx^* \, \|[v_n]_\theta\|_s \|\theta - \theta_0\|_s$$

$$+ \int \left| \frac{d}{d\theta} f_{y|x^*}(y|x^*; \overline{\theta})\omega^{-1}(\xi)\omega^{-1}(x, x^*)\overline{f}_2(x^*|z) \right| dx^*$$

$$\times \|[v_n]_\theta\|_s \|f_1 - f_{x|x^*}\|_s$$

$$+ \int \left| \frac{d}{d\theta} f_{y|x^*}(y|x^*; \overline{\theta}) \omega^{-1}(\xi) \overline{f}_1(x|x^*) \omega^{-1}(x^*, z) \right| dx^*$$

$$\times \| [v_n]_\theta \|_s \| f_2 - f_{x^*|z} \|_s$$

$$+ \int \left| \frac{d}{d\theta} f_{y|x^*}(y|x^*; \overline{\theta}) \omega^{-1}(\xi) \omega^{-1}(x, x^*) \overline{f}_2(x^*|z) \right| dx^*$$

$$\times \| \theta - \theta_0 \|_s \| [v_n]_{f_1} \|_s$$

$$+ \int |f_{y|x^*}(y|x^*; \overline{\theta}) \omega^{-1}(x, x^*) \omega^{-1}(x^*, z)| \, dx^* \| [v_n]_{f_1} \|_s \| f_2 - f_{x^*|z} \|_s$$

$$+ \int \left| \frac{d}{d\theta} f_{y|x^*}(y|x^*; \overline{\theta}) \omega^{-1}(\xi) \overline{f}_1(x|x^*) \omega^{-1}(x^*, z) \right| dx^*$$

$$\times \| \theta - \theta_0 \|_s \| [v_n]_{f_2} \|_s$$

$$+ \int |f_{y|x^*}(y|x^*; \overline{\theta}) \omega^{-1}(x, x^*) \omega^{-1}(x^*, z)| \, dx^*$$

$$\times \| f_1 - f_{x|x^*} \|_s \| [v_n]_{f_2} \|_s \Big\}$$

$$\leq \frac{1}{|f_{yx|z}(D, \overline{\alpha}_1)|} \left\{ \int \left| \frac{d^2}{d\theta^2} f_{y|x^*}(y|x^*; \overline{\theta}) \omega^{-1}(\xi) \omega^{-1}(\xi) \right. \right.$$

$$\times \overline{f}_1(x|x^*) \overline{f}_2(x^*|z) \Big| dx^*$$

$$+ \int \left| \frac{d}{d\theta} f_{y|x^*}(y|x^*; \overline{\theta}) \omega^{-1}(\xi) \omega^{-1}(x, x^*) \overline{f}_2(x^*|z) \right| dx^*$$

$$+ \int \left| \frac{d}{d\theta} f_{y|x^*}(y|x^*; \overline{\theta}) \omega^{-1}(\xi) \overline{f}_1(x|x^*) \omega^{-1}(x^*, z) \right| dx^*$$

$$+ \int \left| \frac{d}{d\theta} f_{y|x^*}(y|x^*; \overline{\theta}) \omega^{-1}(\xi) \omega^{-1}(x, x^*) \overline{f}_2(x^*|z) \right| dx^*$$

$$+ \int |f_{y|x^*}(y|x^*; \overline{\theta}) \omega^{-1}(x, x^*) \omega^{-1}(x^*, z)| \, dx^*$$

$$+ \int \left| \frac{d}{d\theta} f_{y|x^*}(y|x^*; \overline{\theta}) \omega^{-1}(\xi) \overline{f}_1(x|x^*) \omega^{-1}(x^*, z) \right| dx^*$$

$$+ \int |f_{y|x^*}(y|x^*; \overline{\theta}) \omega^{-1}(x, x^*) \omega^{-1}(x^*, z)| \, dx^* \Big\} \| \alpha - \alpha_0 \|_s \| v_n \|_s$$

$$\equiv \left| \frac{f_{yx|z}^{|2|}(D, \overline{\alpha}_1, \bar{\omega})}{f_{yx|z}(D, \overline{\alpha}_1)} \right| \| \alpha - \alpha_0 \|_s \| v_n \|_s,$$

where $f_{yx|z}^{|2|}(D, \overline{\alpha}_1, \bar{\omega})$ is defined in Assumption 20. Plugging the bounds in Equations (S13), (S14), and (S17) back in to Equation (S11), we have

$$\left| \sup_{\alpha \in \mathcal{N}_{0n}} \frac{d^2 \ln f_{yx|z}(D, \overline{\alpha}_1)}{d\alpha \, d\alpha^T}[v_n, (\alpha - \alpha_0)] \right|$$

$$\leq \sup_{\overline{\alpha}_1 \in \mathcal{N}_{0n}} \left[\left| \frac{f_{yx|z}^{|1|}(D, \overline{\alpha}_1, \bar{\omega})}{f_{yx|z}(D, \overline{\alpha}_1)} \right|^2 + \left| \frac{f_{yx|z}^{|2|}(D, \overline{\alpha}_1, \bar{\omega})}{f_{yx|z}(D, \overline{\alpha}_1)} \right| \right] \|\alpha - \alpha_0\|_s \|v_n\|_s$$

$$\leq h_2(D)\|\alpha - \alpha_0\|_s \|v_n\|_s.$$

By the envelope condition in Assumption 20, Equation (S10) becomes

$$\sup_{\alpha \in \mathcal{N}_{0n}} A_1$$

$$= \varepsilon_n O_p(n^{-1/2})$$

$$\times \sqrt{E \left(\sup_{\alpha \in \mathcal{N}_{0n}} \frac{d^2 \ln f_{yx|z}(D, \overline{\alpha}_1)}{d\alpha \, d\alpha^T}[\Pi_n(\alpha - \alpha_0 - v^*), (\alpha - \alpha_0)] \right)^2}$$

$$\leq \varepsilon_n O_p(n^{-1/2}) \sqrt{E(h_2(D))^2} \|\alpha - \alpha_0\|_s \|v_n\|_s$$

$$= O_p(\varepsilon_n^2)$$

with $\|\alpha - \alpha_0\|_s = o(1)$. The last two terms, A_2 and A_3 in Equation (S9), are bounded as

$$\left| \sup_{\alpha \in \mathcal{N}_{0n}} A_2 \right| = \varepsilon_n^2 \left| \sup_{\alpha \in \mathcal{N}_{0n}} \mu_n \left(\frac{1}{2} \frac{d^2 \ln f_{yx|z}(D, \overline{\alpha}_1)}{d\alpha \, d\alpha^T} \right. \right.$$

$$\left. \left. \times [\Pi_n(\alpha - \alpha_0 - v^*), \Pi_n(\alpha - \alpha_0 - v^*)] \right) \right|$$

$$\leq \varepsilon_n^2 \frac{1}{2} \mu_n \left(\left| \frac{f_{yx|z}^{|1|}(D, \overline{\alpha}_1, \bar{\omega})}{f_{yx|z}(D, \overline{\alpha}_1)} \right|^2 + \left| \frac{f_{yx|z}^{|2|}(D, \overline{\alpha}_1, \bar{\omega})}{f_{yx|z}(D, \overline{\alpha}_1)} \right| \right)$$

$$\times \|\Pi_n(\alpha - \alpha_0 - v^*)\|_s^2$$

$$\leq \varepsilon_n^2 \frac{1}{2} O_p\big(E|h_2(D)|\big) \|\Pi_n(\alpha - \alpha_0 - v^*)\|_s^2$$

$$= O_p(\varepsilon_n^2).$$

The same result holds for $|\sup_{\alpha \in \mathcal{N}_{0n}} A_3|$ and, therefore, Condition A in Shen (1997) holds.

Fourth, Condition B requires

$$\sup_{\alpha \in \mathcal{N}_{0n}} \left[E\left(\ln \frac{f_{yx|z}(D, \alpha_0)}{f_{yx|z}(D, \Pi_n \alpha^*(\alpha, \varepsilon_n))} \right) - E\left(\ln \frac{f_{yx|z}(D, \alpha_0)}{f_{yx|z}(D, \alpha)} \right) \right.$$
$$\left. - \frac{1}{2} \left(\|\alpha^*(\alpha, \varepsilon_n) - \alpha_0\|^2 - \|\alpha - \alpha_0\|^2 \right) \right] = O(\varepsilon_n^2).$$

As Corollary 2 in Shen (1997) points out, Condition B can be replaced by Condition B′ as

$$E\left(\ln \frac{f_{yx|z}(D, \alpha_0)}{f_{yx|z}(D, \alpha)} \right) = \frac{1}{2} \|\alpha - \alpha_0\|^2 (1 + o(h_n))$$

with some positive sequence $\{h_n\} \to 0$ as $n \to \infty$. We consider the Taylor expansion

$$(S18) \quad E[\ln f_{yx|z}(D, \alpha) - \ln f_{yx|z}(D, \alpha_0)]$$
$$= E\left(\frac{d \ln f_{yx|z}(D, \alpha_0)}{d\alpha} [\alpha - \alpha_0] \right)$$
$$+ \frac{1}{2} E\left(\frac{d^2 \ln f_{yx|z}(D, \alpha_0)}{d\alpha \, d\alpha^T} [\alpha - \alpha_0, \alpha - \alpha_0] \right)$$
$$+ \frac{1}{6} E \frac{d^3}{dt^3} \ln f_{yx|z}(D; \alpha_0 + t(\alpha - \alpha_0)) \Big|_{t=0}$$
$$+ \frac{1}{24} E \frac{d^4}{dt^4} \ln f_{yx|z}(D; \overline{\alpha} + t(\alpha - \alpha_0)) \Big|_{t=0},$$

where $\overline{\alpha}$ is a mean value between α and α_0.

As for the leading terms on the right-hand side, we have η satisfying $\int_{\mathcal{Y}} (\partial/\partial\eta) f_{y|x^*}(y|x^*; \theta) \, dy = 0$, $\int_{\mathcal{Y}} (\partial^2/\partial\eta^2) f_{y|x^*}(y|x^*; \theta) \, dy = 0$, and $\int_{\mathcal{Y}} (\partial^3/\partial\eta^3) f_{y|x^*}(y|x^*; \theta) \, dy = 0$ for all $\theta \in \Theta$, and we have f_1 and f_2 satisfying $\int_{\mathcal{X}} f_1(x|x^*) \, dx = 1$ and $\int_{\mathcal{X}^*} f_2(x^*|z) \, dx = 1$. It is then tedious but straightforward to show [5]

$$E\left(\frac{d \ln f_{yx|z}(D, \alpha_0)}{d\alpha} [\alpha - \alpha_0] \right) = 0,$$
$$E\left(\frac{1}{f_{yx|z}(D, \alpha_0)} \frac{d^2 f_{yx|z}(D, \alpha_0)}{d\alpha \, d\alpha^T} [\alpha - \alpha_0, \alpha - \alpha_0] \right) = 0,$$

---

[5] We abuse the notation $(d^3 \ln f_{yx|z})/d\alpha^3$ to stand for the third order derivative with respect to a vector α.

$$E\left[\frac{1}{f_{yx|z}(D,\alpha_0)}\frac{d^3 f_{yx|z}(D,\alpha_0)}{d\alpha^3}[\alpha-\alpha_0,\alpha-\alpha_0,\alpha-\alpha_0]\right]=0.$$

Therefore,

$$E\left(\frac{d^2\ln f_{yx|z}(D,\alpha_0)}{d\alpha\,d\alpha^T}[\alpha-\alpha_0,\alpha-\alpha_0]\right)$$

$$=E\left[\frac{1}{f_{yx|z}(D,\alpha_0)}\frac{d^2 f_{yx|z}(D,\alpha_0)}{d\alpha\,d\alpha^T}[\alpha,\alpha]-\left(\frac{d\ln f_{yx|z}(D,\alpha_0)}{d\alpha}[\alpha-\alpha_0]\right)\right.$$

$$\left.\times\left(\frac{d\ln f_{yx|z}(D,\alpha_0)}{d\alpha}[\alpha-\alpha_0]\right)\right]$$

$$=-E\left[\left(\frac{d\ln f_{yx|z}(D,\alpha_0)}{d\alpha}[\alpha-\alpha_0]\right)\left(\frac{d\ln f_{yx|z}(D,\alpha_0)}{d\alpha}[\alpha-\alpha_0]\right)\right]$$

$$=-\|\alpha-\alpha_0\|^2.$$

Therefore, Equation (S18) becomes

$$
\begin{aligned}
\text{(S19)}\quad & E[\ln f_{yx|z}(D,\alpha)-\ln f_{yx|z}(D,\alpha_0)] \\
&= -\frac{1}{2}\|\alpha-\alpha_0\|^2+\frac{1}{6}E\frac{d^3}{dt^3}\ln f_{yx|z}(D;\alpha_0+t(\alpha-\alpha_0))\Big|_{t=0} \\
&\quad +\frac{1}{24}E\frac{d^4}{dt^4}\ln f_{yx|z}(D;\overline{\alpha}+t(\alpha-\alpha_0))\Big|_{t=0}.
\end{aligned}
$$

For the second term on the right-hand side, we have

$$\frac{d^3}{dt^3}\ln f_{yx|z}(D;\alpha_0+t(\alpha-\alpha_0))\Big|_{t=0}$$

$$=E\left[\frac{1}{f_{yx|z}(D,\alpha_0)}\frac{d^3 f_{yx|z}(D,\alpha_0)}{d\alpha^3}[\alpha-\alpha_0,\alpha-\alpha_0,\alpha-\alpha_0]\right]$$

$$-3E\left[\frac{d\ln f_{yx|z}(D,\alpha_0)}{d\alpha}[\alpha-\alpha_0]\frac{1}{f_{yx|z}(D,\alpha_0)}\frac{d^2 f_{yx|z}(D,\alpha_0)}{d\alpha\,d\alpha^T}\right.$$

$$\left.\times[\alpha-\alpha_0,\alpha-\alpha_0]\right]$$

$$+2E\left(\frac{d\ln f_{yx|z}(D,\alpha_0)}{d\alpha}[\alpha-\alpha_0]\right)^3$$

$$=B_1+B_2+B_3.$$

Again, it is straightforward to show $B_1 = 0$. The term B_2 is bounded as

$$E\left[\frac{d\ln f_{yx|z}(D,\alpha_0)}{d\alpha}[\alpha - \alpha_0]\right.$$

$$\left. \times \frac{1}{f_{yx|z}(D,\alpha_0)}\frac{d^2 f_{yx|z}(D,\alpha_0)}{d\alpha\,d\alpha^T}[\alpha - \alpha_0, \alpha - \alpha_0]\right]$$

$$\leq E\left[\left|\frac{d\ln f_{yx|z}(D,\alpha_0)}{d\alpha}[\alpha - \alpha_0]\right|\right.$$

$$\left. \times \left|\frac{1}{f_{yx|z}(D,\alpha_0)}\frac{d^2 f_{yx|z}(D,\alpha_0)}{d\alpha\,d\alpha^T}[\alpha - \alpha_0, \alpha - \alpha_0]\right|\right]$$

$$\leq \left[E\left|\frac{1}{f_{yx|z}(D,\alpha_0)}\frac{d^2 f_{yx|z}(D,\alpha_0)}{d\alpha\,d\alpha^T}[\alpha - \alpha_0, \alpha - \alpha_0]\right|^2\right]^{1/2}$$

$$\times \left[E\left|\frac{d\ln f_{yx|z}(D,\alpha_0)}{d\alpha}[\alpha - \alpha_0]\right|^2\right]^{1/2}$$

$$= \left[E\left|\frac{1}{f_{yx|z}(D,\alpha_0)}\frac{d^2 f_{yx|z}(D,\alpha_0)}{d\alpha\,d\alpha^T}[\alpha - \alpha_0, \alpha - \alpha_0]\right|^2\right]^{1/2}\|\alpha - \alpha_0\|$$

$$\leq \left[E\left|\frac{f_{yx|z}^{|2|}(D,\alpha_0,\bar{\omega})}{f_{yx|z}(D,\alpha_0)}\right|^2\right]^{1/2}\|\alpha - \alpha_0\|_s^2\|\alpha - \alpha_0\|$$

$$\leq \left[E|h_2(D)|^2\right]^{1/2}\|\alpha - \alpha_0\|_s^2\|\alpha - \alpha_0\|.$$

For the term B_3, we have

$$B_3 \leq E\left|\frac{d\ln f_{yx|z}(D,\alpha_0)}{d\alpha}[\alpha - \alpha_0]\right|^3$$

$$\leq \left[E\left|\frac{d\ln f_{yx|z}(D,\alpha_0)}{d\alpha}[\alpha - \alpha_0]\right|^4\right]^{1/2}$$

$$\times \left[E\left|\frac{d\ln f_{yx|z}(D,\alpha_0)}{d\alpha}[\alpha - \alpha_0]\right|^2\right]^{1/2}$$

$$= \left[E\left(\frac{d\ln f_{yx|z}(D,\alpha_0)}{d\alpha}[\alpha - \alpha_0]\right)^4\right]^{1/2}\|\alpha - \alpha_0\|$$

$$\leq \left[E\left|\frac{f_{yx|z}^{|1|}(D,\alpha_0,\bar{\omega})}{f_{yx|z}(D,\alpha_0)}\right|^4\right]^{1/2}\|\alpha - \alpha_0\|_s^2\|\alpha - \alpha_0\|$$

$$\leq \left[E|h_1(D)|^4\right]^{1/2}\|\alpha - \alpha_0\|_s^2\|\alpha - \alpha_0\|.$$

Note that $E|h_2(D)|^2 < \infty$ implies $E|h_1(D)|^4 < \infty$. Therefore, Equation (S19) becomes

$$E[\ln f_{yx|z}(D, \alpha) - \ln f_{yx|z}(D, \alpha_0)]$$

$$= -\frac{1}{2}\|\alpha - \alpha_0\|^2 + O(\|\alpha - \alpha_0\|_s^2 \|\alpha - \alpha_0\|)$$

$$+ \frac{1}{24}E\frac{d^4}{dt^4}\ln f_{yx|z}(D; \overline{\alpha} + t(\alpha - \alpha_0))\Big|_{t=0}.$$

By Assumption 23, we have

$$E\frac{d^4}{dt^4}\ln f_{yx|z}(D; \overline{\alpha} + t(\alpha - \alpha_0))\Big|_{t=0}$$

$$\leq E\left|\frac{d^4}{dt^4}\ln f_{yx|z}(D; \overline{\alpha} + t(\alpha - \alpha_0))\Big|_{t=0}\right|$$

$$\leq E|h_4(D)|\|\alpha - \alpha_0\|_s^4$$

$$= O(\|\alpha - \alpha_0\|_s^4)$$

and, therefore,

$$(S20) \qquad E[\ln f_{yx|z}(D, \alpha_0) - \ln f_{yx|z}(D, \alpha)] = \frac{1}{2}\|\alpha - \alpha_0\|^2(1 + O(h_n))$$

with

$$h_n = \frac{\|\alpha - \alpha_0\|_s^2}{\|\alpha - \alpha_0\|} + \frac{\|\alpha - \alpha_0\|_s^4}{\|\alpha - \alpha_0\|^2}.$$

Next, we show that $\|\alpha - \alpha_0\|_s^2/\|\alpha - \alpha_0\| \to 0$ as $n \to \infty$. We will need the convergence rate of the sieve coefficients. Therefore, we define for $\alpha \in \mathcal{N}_{0n}$,

$$\alpha = (b^T, \Pi_n\eta, \Pi_n f_1, \Pi_n f_2)^T$$

$$= (b^T, p^{k_n}(\xi_1, \xi_2)^T\delta, p^{k_n}(x, x^*)^T\beta, p^{k_n}(x^*, z)^T\gamma)^T,$$

$$\Pi_n\alpha_0 = (b_0^T, \Pi_n\eta_0, \Pi_n f_{x|x^*}, \Pi_n f_{x^*|z})^T$$

$$= (b_0^T, p^{k_n}(\xi_1, \xi_2)^T\delta_0, p^{k_n}(x, x^*)^T\beta_0, p^{k_n}(x^*, z)^T\gamma_0)^T,$$

where p^{k_n}'s are k_n-by-1 vectors, that is, $p^{k_n}(\cdot, \cdot) = (p_1^{k_n}(\cdot, \cdot), p_2^{k_n}(\cdot, \cdot), \ldots, p_{k_n}^{k_n}(\cdot, \cdot))^T$. Note that all the vectors are column vectors. We also define the vector of the sieve coefficients as

$$\alpha^c = (b^T, \delta^T, \beta^T, \gamma^T)^T,$$

$$\alpha_0^c = (b_0^T, \delta_0^T, \beta_0^T, \gamma_0^T)^T.$$

We then have

$$
\begin{aligned}
\alpha - \alpha_0 &= \alpha - \Pi_n \alpha_0 + \Pi_n \alpha_0 - \alpha_0 \\
&= ((b^T - b_0^T),\ p^{k_n}(\xi_1, \xi_2)^T (\delta - \delta_0), \\
&\quad\ p^{k_n}(x, x^*)^T (\beta - \beta_0),\ p^{k_n}(x^*, z)^T (\gamma - \gamma_0)) \\
&\quad\ + \Pi_n \alpha_0 - \alpha_0.
\end{aligned}
$$

Suppose that

$$
\|\alpha - \alpha_0\| = O(n^{-1/4 - \varsigma_0})
$$

with some small $\varsigma_0 > 0$. By Assumption 13 and Equation (S7), we let

$$
\|\Pi_n \alpha_0 - \alpha_0\|_s = O(k_n^{-\gamma_1/d_1}) = O(n^{-1/4 - \varsigma})
$$

for some small $\varsigma > \varsigma_0$.

   We then show $\|\alpha^c - \alpha_0^c\|_E = O(n^{-1/4 - \varsigma_0})$ from $\|\alpha - \alpha_0\| = O(n^{-1/4 - \varsigma_0})$. For any $\alpha \in \mathcal{N}_{0n}$, we have

$$
\begin{aligned}
\big| \|\alpha - \alpha_0\| &- \|\Pi_n \alpha_0 - \alpha_0\| \big| \\
&\leq \|\alpha - \Pi_n \alpha_0\| \leq \|\alpha - \alpha_0\| + \|\Pi_n \alpha_0 - \alpha_0\|.
\end{aligned}
$$

We have shown that Assumption 11 implies $E|(f_{yx|z}^{|1|}(D, \overline{\alpha}_1, \bar{\omega}))/(f_{yx|z}(D, \overline{\alpha}_1))|^2 \leq E|h_1(D)|^2 < \infty$. We then have

$$
\begin{aligned}
\|\Pi_n \alpha_0 - \alpha_0\| &\leq \sqrt{E\left(\frac{f_{yx|z}^{|1|}(D, \overline{\alpha}_1, \bar{\omega})}{f_{yx|z}(D, \overline{\alpha}_1)}\right)^2} \|\Pi_n \alpha_0 - \alpha_0\|_s \\
&= O(\|\Pi_n \alpha_0 - \alpha_0\|_s) \\
&\leq O(k_n^{-\gamma_1/d_1}) \\
&= O(n^{-1/4 - \varsigma})
\end{aligned}
$$

and, therefore, for some constants $0 < C_1, C_2 < \infty$,

$$
\text{(S21)} \qquad C_1 \|\alpha - \alpha_0\| \leq \|\alpha - \Pi_n \alpha_0\| \leq C_2 \|\alpha - \alpha_0\|.
$$

Moreover, we define

$$
\begin{aligned}
\frac{d \ln f_{yx|z}(D, \alpha_0)}{db} \\
= \left(\frac{d \ln f_{yx|z}(D, \alpha_0)}{db_1}, \frac{d \ln f_{yx|z}(D, \alpha_0)}{db_2}, \ldots, \frac{d \ln f_{yx|z}(D, \alpha_0)}{db_{d_b}} \right)^T,
\end{aligned}
$$

$$\frac{d \ln f_{yx|z}(D, \alpha_0)}{d\eta}[p^{k_n}]$$

$$= \left(\frac{d \ln f_{yx|z}(D, \alpha_0)}{d\eta}[p_1^{k_n}], \right.$$

$$\left. \frac{d \ln f_{yx|z}(D, \alpha_0)}{d\eta}[p_2^{k_n}], \ldots, \frac{d \ln f_{yx|z}(D, \alpha_0)}{d\eta}[p_{k_n}^{k_n}] \right)^T,$$

$$\frac{d \ln f_{yx|z}(D, \alpha_0)}{df_1}[p^{k_n}]$$

$$= \left(\frac{d \ln f_{yx|z}(D, \alpha_0)}{df_1}[p_1^{k_n}], \right.$$

$$\left. \frac{d \ln f_{yx|z}(D, \alpha_0)}{df_1}[p_2^{k_n}], \ldots, \frac{d \ln f_{yx|z}(D, \alpha_0)}{df_1}[p_{k_n}^{k_n}] \right)^T,$$

$$\frac{d \ln f_{yx|z}(D, \alpha_0)}{df_2}[p^{k_n}]$$

$$= \left(\frac{d \ln f_{yx|z}(D, \alpha_0)}{df_2}[p_1^{k_n}], \right.$$

$$\left. \frac{d \ln f_{yx|z}(D, \alpha_0)}{df_2}[p_2^{k_n}], \ldots, \frac{d \ln f_{yx|z}(D, \alpha_0)}{df_2}[p_{k_n}^{k_n}] \right)^T,$$

$$\frac{d \ln f_{yx|z}(D, \alpha_0)}{d\alpha}[p^{k_n}]$$

$$= \left[\left(\frac{d \ln f_{yx|z}(D, \alpha_0)}{db} \right)^T, \left(\frac{d \ln f_{yx|z}(D, \alpha_0)}{d\eta}[p^{k_n}] \right)^T, \right.$$

$$\left. \left(\frac{d \ln f_{yx|z}(D, \alpha_0)}{df_1}[p^{k_n}] \right)^T, \left(\frac{d \ln f_{yx|z}(D, \alpha_0)}{df_2}[p^{k_n}] \right)^T \right]^T.$$

With the notations above, we have

$$\frac{d \ln f_{yx|z}(D, \alpha_0)}{d\alpha}[\alpha - \Pi_n \alpha_0]$$

$$= \left(\frac{d \ln f_{yx|z}(D, \alpha_0)}{db} \right)^T (b - b_0) + \left(\frac{d \ln f_{yx|z}(D, \alpha_0)}{d\eta}[p^{k_n}] \right)^T (\delta - \delta_0)$$

$$+ \left(\frac{d \ln f_{yx|z}(D, \alpha_0)}{df_1}[p^{k_n}] \right)^T (\beta - \beta_0)$$

$$+ \left(\frac{d \ln f_{yx|z}(D, \alpha_0)}{df_2} [p^{k_n}] \right)^T (\gamma - \gamma_0)$$

$$= \left(\frac{d \ln f_{yx|z}(D, \alpha_0)}{d\alpha} [p^{k_n}] \right)^T (\alpha^c - \alpha_0^c)$$

and

$$\| \alpha - \Pi_n \alpha_0 \|^2$$

$$= E \left\{ \left(\frac{d \ln f_{yx|z}(D, \alpha_0)}{d\alpha} [\alpha - \Pi_n \alpha_0] \right)^2 \right\}$$

$$= (\alpha^c - \alpha_0^c)^T E \left\{ \left(\frac{d \ln f_{yx|z}(D, \alpha_0)}{d\alpha} [p^{k_n}] \right) \left(\frac{d \ln f_{yx|z}(D, \alpha_0)}{d\alpha} [p^{k_n}] \right)^T \right\}$$

$$\times (\alpha^c - \alpha_0^c)$$

$$\equiv (\alpha^c - \alpha_0^c)^T \Omega_{k_n} (\alpha^c - \alpha_0^c).$$

The matrix Ω_{k_n} is positive definite with its smallest eigenvalue bounded away from zero uniformly in k_n according to Assumption 21. Since $\| \alpha - \Pi_n \alpha_0 \|$ is always finite, the largest eigenvalue of Ω_{k_n} is finite. Thus, we have for some constants $0 < C_1, C_2 < \infty$,

$$(S22) \qquad C_1 \| \alpha^c - \alpha_0^c \|_E \leq \| \alpha - \Pi_n \alpha_0 \| \leq C_2 \| \alpha^c - \alpha_0^c \|_E.$$

Note that C_1 and C_2 are general constants that may take different values at each appearance.

We then consider the ratio $\| \alpha - \alpha_0 \|_s^2 / \| \alpha - \alpha_0 \|$. From Equations (S21) and (S22), we have

$$(S23) \qquad \| \alpha - \alpha_0 \| \geq C_1 \| \alpha^c - \alpha_0^c \|_E$$

and $\| \alpha^c - \alpha_0^c \|_E = O(n^{-1/4 - \varsigma_0})$. Assumption 21 implies $\| \alpha - \Pi_n \alpha_0 \|_s^2 \leq C_2 \| \alpha^c - \alpha_0^c \|_1^2$, where $\| \cdot \|_1$ is the L_1 vector norm. Thus, we have

$$\| \alpha - \alpha_0 \|_s^2 \leq \| \alpha - \Pi_n \alpha_0 \|_s^2 + \| \Pi_n \alpha_0 - \alpha_0 \|_s^2$$

$$\leq C_2 \| \alpha^c - \alpha_0^c \|_1^2 + O(k_n^{-2\gamma_1/d_1})$$

$$\leq C_2 k_n \| \alpha^c - \alpha_0^c \|_E^2 + O(n^{2(-1/4 - \varsigma)}).$$

Since $\| \alpha^c - \alpha_0^c \|_E = O(n^{-1/4 - \varsigma_0})$ and $\varsigma > \varsigma_0$, we have

$$(S24) \qquad \| \alpha - \alpha_0 \|_s^2 \leq C_2 k_n \| \alpha^c - \alpha_0^c \|_E^2.$$

By Equations (S23) and (S24), we have

$$\frac{\| \alpha - \alpha_0 \|_s^2}{\| \alpha - \alpha_0 \|} \leq \frac{C_2 k_n \| \alpha^c - \alpha_0^c \|_E^2}{C_1 \| \alpha^c - \alpha_0^c \|_E} \leq O(k_n \| \alpha^c - \alpha_0^c \|_E).$$

Assumption 13 requires $k_n^{-\gamma_1/d_1} = O(n^{-1/4-\varsigma})$, that is, $k_n = n^{(1/4+\varsigma)1/(\gamma_1/d_1)}$. We then have

$$k_n\|\alpha^c - \alpha_0^c\|_E = O\big(n^{-1/4(1-1/(\gamma_1/d_1))+\varsigma(1/(\gamma_1/d_1))-\varsigma_0}\big) = o(1)$$

for $\varsigma < \frac{1}{4}(\gamma_1/d_1 - 1) + (\gamma_1/d_1)\varsigma_0$ with $\gamma_1/d_1 > 1$ in Assumption 13. Therefore, Equation (S20) holds with the positive sequence $\{h_n\} \to 0$ as $n \to \infty$. That means that Condition B$'$ in Shen (1997) holds.

Fifth, Condition C in Shen (1997) requires

$$\sup_{\alpha \in \mathcal{N}_{0n}} \|\alpha^*(\alpha, \varepsilon_n) - \Pi_n\alpha^*(\alpha, \varepsilon_n)\| = O(n^{-1/4}\varepsilon_n).$$

By definition, we have $\alpha^*(\alpha, \varepsilon_n) = (1 - \varepsilon_n)\alpha + \varepsilon_n(v^* + \alpha_0)$ with $\alpha \in \mathcal{N}_{0n}$. Therefore,

$$\begin{aligned}
&\|\alpha^*(\alpha, \varepsilon_n) - \Pi_n\alpha^*(\alpha, \varepsilon_n)\| \\
&\quad = \varepsilon_n\|v^* + \alpha_0 - \Pi_n(v^* + \alpha_0)\| \\
&\quad \leq \varepsilon_n\|v^* - \Pi_nv^*\| + \varepsilon_n\|\alpha_0 - \Pi_n\alpha_0\| \\
&\quad = O(n^{-1/4}\varepsilon_n).
\end{aligned}$$

The last step is due to Assumption 22. Condition C also requires

$$(S25) \qquad \sup_{\alpha \in \mathcal{N}_{0n}} \mu_n\left(\frac{d \ln f_{yx|z}(D, \alpha_0)}{d\alpha}[\alpha^*(\alpha, \varepsilon_n) - \Pi_n\alpha^*(\alpha, \varepsilon_n)]\right) = O_p(\varepsilon_n^2).$$

The left-hand side equals

$$\begin{aligned}
&\varepsilon_n\mu_n\left(\frac{d \ln f_{yx|z}(D, \alpha_0)}{d\alpha}[v^* - v_n^*]\right) \\
&\quad + \varepsilon_n\mu_n\left(\frac{d \ln f_{yx|z}(D, \alpha_0)}{d\alpha}[\alpha_0 - \Pi_n\alpha_0]\right).
\end{aligned}$$

By the envelope condition in Assumption 11, the first term (corresponding to v^*) is

$$\begin{aligned}
&\left|\mu_n\left(\frac{d \ln f_{yx|z}(D, \alpha_0)}{d\alpha}[v^* - v_n^*]\right)\right| \\
&\quad = \sqrt{E\left(\frac{d \ln f_{yx|z}(D, \alpha_0)}{d\alpha}[v^* - v_n^*]\right)^2} O_p(n^{-1/2}) \\
&\quad = \|v^* - v_n^*\| O_p(n^{-1/2}) \\
&\quad = o_p(n^{-1/2}),
\end{aligned}$$

and the second term (corresponding to α_0) is

$$
\left| \mu_n \left(\frac{d \ln f_{yx|z}(D, \alpha_0)}{d\alpha} [\alpha_0 - \Pi_n \alpha_0] \right) \right|
$$

$$
= \sqrt{ E \left(\frac{d \ln f_{yx|z}(D, \alpha_0)}{d\alpha} [\alpha_0 - \Pi_n \alpha_0] \right)^2 } O_p(n^{-1/2})
$$

$$
= \| \alpha_0 - \Pi_n \alpha_0 \| O_p(n^{-1/2})
$$

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

The last step is due to $\| \alpha_0 - \Pi_n \alpha_0 \| = o(n^{-1/4})$. Therefore, Condition C in Theorem 1 in Shen (1997) holds. Note that Condition C' in Corollary 2 is also satisfied, that is, $\| v_n^* - v^* \| = o(n^{-1/4})$ and $o(h_n) \| \alpha_0 - \Pi_n \alpha_0 \|^2 = o_p(n^{-1/2})$.

Finally, Condition D in Shen (1997), that is,

$$
\sup_{\alpha \in \mathcal{N}_{0n}} \mu_n \left(\frac{d \ln f_{yx|z}(D, \alpha_0)}{d\alpha} [\alpha - \alpha_0] \right) = o_p(n^{-1/2}),
$$

can be verified as follows: We first have

$$
\sup_{\alpha \in \mathcal{N}_{0n}} \left| \frac{d \ln f_{yx|z}(D, \alpha_0)}{d\alpha} [\alpha - \alpha_0] \right|
$$

$$
\leq \left| \frac{1}{f_{yx|z}(D, \alpha_0)} \int \frac{d}{d\theta} f_{y|x^*}(y|x^*; \theta_0) \omega^{-1}(\xi) f_{x|x^*}(x|x^*) f_{x^*|z}(x^*|z) \, dx^* \right|
$$

$$
\times \| \theta - \theta_0 \|_s
$$

$$
+ \left| \frac{1}{f_{yx|z}(D, \alpha_0)} \int f_{y|x^*}(y|x^*; \theta_0) \omega^{-1}(x, x^*) f_{x^*|z}(x^*|z) \, dx^* \right|
$$

$$
\times \| f_1 - f_{x|x^*} \|_s
$$

$$
+ \left| \frac{1}{f_{yx|z}(D, \alpha_0)} \int f_{y|x^*}(y|x^*; \theta_0) f_{x|x^*}(x|x^*) \omega^{-1}(x^*, z) \, dx^* \right|
$$

$$
\times \| f_2 - f_{x^*|z} \|_s
$$

$$
\leq \left| \frac{f_{yx|z}^{|1|}(D, \alpha_0, \bar{\omega})}{f_{yx|z}(D, \alpha_0)} \right| \| \alpha - \alpha_0 \|_s
$$

$$
\leq |h_1(D)| \| \alpha - \alpha_0 \|_s
$$

with $E|h_1(D)|^2 < \infty$ by the envelope condition in Assumption 11. We then have

$$\sup_{\alpha \in \mathcal{N}_{0n}} \mu_n \left(\frac{d \ln f_{yx|z}(D, \alpha_0)}{d\alpha} [\alpha - \alpha_0] \right)$$

$$= \sqrt{E \left(\sup_{\alpha \in \mathcal{N}_{0n}} \frac{d \ln f_{yx|z}(D, \alpha_0)}{d\alpha} [\alpha - \alpha_0] \right)^2} O_p(n^{-1/2})$$

$$\leq \sqrt{E|h_1(D)|^2} \|\alpha - \alpha_0\|_s O_p(n^{-1/2})$$

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

Thus, Condition D in Theorem 1 in Shen (1997) holds. Since all the conditions in Theorem 1 in Shen (1997) hold, the results of asymptotic normality follow.   *Q.E.D.*

S3. NONUNIQUENESS OF THE INDEXING OF THE EIGENVALUES

Let x^* and $\tilde{x}^*$ be related through $x^* = R(\tilde{x}^*)$, where $R(\tilde{x}^*)$ is a given piecewise differentiable function. We now show that, without Assumption 5, models in which x^* or $\tilde{x}^*$ is the unobserved true regressor are observationally equivalent, because

$$L_{x|\tilde{x}^*} \Delta_{y;\tilde{x}^*} L_{x|\tilde{x}^*}^{-1} = L_{x|x^*} \Delta_{y;x^*} L_{x|x^*}^{-1},$$

where the operators $\Delta_{y;\tilde{x}^*}$ and $L_{x|\tilde{x}^*}$ are defined as

$$[\Delta_{y;\tilde{x}^*} g](\tilde{x}^*) = f_{y|\tilde{x}^*}(y|\tilde{x}^*) g(\tilde{x}^*),$$

$$[L_{x|\tilde{x}^*} g](x) = \int f_{x|\tilde{x}^*}(x|\tilde{x}^*) g(\tilde{x}^*) \, d\tilde{x}^*.$$

We first note that the operators $\Delta_{y;\tilde{x}^*}$ and $L_{x|\tilde{x}^*}$ can also be written in terms of $f_{y|x^*}$ and $f_{x|x^*}$ as

$$[\Delta_{y;\tilde{x}^*} g](\tilde{x}^*) = f_{y|x^*}(y|R(\tilde{x}^*)) g(\tilde{x}^*),$$

$$[L_{x|\tilde{x}^*} g](x) = \int f_{x|x^*}(x|R(\tilde{x}^*)) g(\tilde{x}^*) \, d\tilde{x}^*.$$

It can be verified (by calculating $L_{x|\tilde{x}^*} L_{x|\tilde{x}^*}^{-1} g$) that $L_{x|\tilde{x}^*}^{-1}$ is given by

$$[L_{x|\tilde{x}^*}^{-1} g](\tilde{x}^*) = r(\tilde{x}^*)[L_{x|x^*}^{-1} g](R(\tilde{x}^*)),$$

where $r(\tilde{x}^*) = dR(\tilde{x}^*)/d\tilde{x}^*$ whenever this differential exists and $r(\tilde{x}^*) = 0$ otherwise.[6] We can then calculate

$$[L_{x|\tilde{x}^*}\Delta_{y;\tilde{x}^*}L^{-1}_{x|\tilde{x}^*}g](x)$$

$$= \int f_{x|x^*}(x|R(\tilde{x}^*))f_{y|x^*}(y|R(\tilde{x}^*))r(\tilde{x}^*)[L^{-1}_{x|x^*}g](R(\tilde{x}^*))\,d\tilde{x}^*$$

$$= \int f_{x|x^*}(x|R(\tilde{x}^*))f_{y|x^*}(y|R(\tilde{x}^*))[L^{-1}_{x|x^*}g](R(\tilde{x}^*))\,dR(\tilde{x}^*)$$

$$= \int f_{x|x^*}(x|x^*)f_{y|x^*}(y|x^*)[L^{-1}_{x|x^*}g](x^*)\,dx^*$$

$$\text{(substituting } x^* = R(\tilde{x}^*))$$

$$= [L_{x|x^*}\Delta_{y;x^*}L^{-1}_{x|x^*}g](x).$$

It follows that indexing the eigenfunctions by $\tilde{x}^*$ or x^* produces observationally equivalent models, but implies different joint densities of x and of the true regressor (x^* or $\tilde{x}^*$).

S4. RESTRICTIONS WITH FOURIER SERIES

As shown above, the sieve estimators are

$$f_1(x|x^*) = \sum_{i=0}^{i_n}\sum_{j=0}^{j_n}\beta_{ij}p_i(x - x^*)q_j(x^*),$$

$$f_2(x^*|z) = \sum_{i=0}^{i_n}\sum_{j=0}^{j_n}\gamma_{ij}p_i(x^* - z)q_j(z).$$

Let $z, x^* \in [0, l_x]$ and $(x - x^*) \in [-l_e, l_e]$. We use the Fourier series

$$p_k(x - x^*) = \cos\frac{k\pi}{l_e}(x - x^*) \quad\text{or}\quad \sin\frac{k\pi}{l_e}(x - x^*),$$

$$p_k(x^* - z) = \cos\frac{k\pi}{l_x}(x^* - z) \quad\text{or}\quad \sin\frac{k\pi}{l_x}(x^* - z),$$

and $q_k(x) = \cos(k\pi/l_x)x$. For simplicity, we consider the case where $i_n = 3$ and $j_n = 2$. Longer series can be handled similarly. We have

$$f_1(x|x^*) = \left(a_{00} + a_{01}\cos\frac{\pi}{l_x}x^* + a_{02}\cos\frac{2\pi}{l_x}x^*\right)$$

[6]Since $R(\tilde{x}^*)$ is piecewise differentiable, $dR(\tilde{x}^*)/d\tilde{x}^*$ exists almost everywhere and the points where it does not will not affect the value of the integral.

$$+ \sum_{k=1}^{3} \left(a_{k0} + a_{k1} \cos \frac{\pi}{l_x} x^* + a_{k2} \cos \frac{2\pi}{l_x} x^* \right) \cos \frac{k\pi}{l_e} (x - x^*)$$

$$+ \sum_{k=1}^{3} \left(b_{k0} + b_{k1} \cos \frac{\pi}{l_x} x^* + b_{k2} \cos \frac{2\pi}{l_x} x^* \right) \sin \frac{k\pi}{l_e} (x - x^*).$$

Consider the restriction $\int_{\mathcal{X}} f_1(x|x^*) \, dx = 1$. We can show that

$$\int_{\mathcal{X}} f_1(x|x^*) \, dx = 2l_e \left(a_{00} + a_{01} \cos \frac{\pi}{l_x} x^* + a_{02} \cos \frac{2\pi}{l_x} x^* \right)$$

for all x^*. Therefore, $a_{00} = 1/2l_e$ and $a_{01} = a_{02} = 0$. We can similarly find the sieve expression of the function $f_2(x^*|z)$ that satisfies $\int_{\mathcal{X}^*} f_2(x^*|z) \, dx^* = 1$.

Next, we consider the identification restrictions on $f_1(x|x^*)$. First, in the zero mode case, we have $\frac{\partial}{\partial x} f_1(x|x^*)|_{x=x^*} = 0$ for all x^* with

$$\frac{\partial}{\partial x} f_1(x|x^*) \bigg|_{x=x^*} = \sum_{k=1}^{3} \frac{k\pi}{l_e} \left(b_{k0} + b_{k1} \cos \frac{\pi}{l_x} x^* + b_{k2} \cos \frac{2\pi}{l_x} x^* \right).$$

Thus, the restrictions on the coefficients are

$$\sum_{k=1}^{3} k b_{k0} = \sum_{k=1}^{3} k b_{k1} = \sum_{k=1}^{3} k b_{k2} = 0.$$

Second, if we make the zero mean assumption instead of the zero mode one, we have $\int_{\mathcal{X}} (x - x^*) f_1(x|x^*) \, dx = 0$ for all x^* with

$$\int_{\mathcal{X}} (x - x^*) f_1(x|x^*) \, dx$$

$$= \sum_{k=1}^{3} \left(b_{k0} + b_{k1} \cos \frac{\pi}{l_x} x^* + b_{k2} \cos \frac{2\pi}{l_x} x^* \right) \left(-\frac{2l_e^2}{k\pi} (-1)^k \right).$$

We have

$$\sum_{k=1}^{3} \frac{(-1)^k}{k} b_{k0} = \sum_{k=1}^{3} \frac{(-1)^k}{k} b_{k1} = \sum_{k=1}^{3} \frac{(-1)^k}{k} b_{k2} = 0.$$

Third, if we make the zero median assumption, we have $\int_{\mathcal{X} \cap \{x < x^*\}} f_{x|x^*}(x|x^*) \, dx = \frac{1}{2}$ for all x^* with

$$\int_{\mathcal{X} \cap \{x < x^*\}} f_1(x|x^*) \, dx$$

$$= \frac{1}{2} + \sum_{k=1}^{3} \left(b_{k0} + b_{k1} \cos \frac{\pi}{l_x} x^* + b_{k2} \cos \frac{2\pi}{l_x} x^* \right) l_e \frac{(-1)^k - 1}{k\pi}.$$

Therefore,

$$\sum_{k=1}^{3} \frac{(-1)^k - 1}{k} b_{k0} = \sum_{k=1}^{3} \frac{(-1)^k - 1}{k} b_{k1} = \sum_{k=1}^{3} \frac{(-1)^k - 1}{k} b_{k2} = 0.$$

Fourth, if x^* is the 100th percentile of $f_{x|x^*}$, we assume $(x - x^*) \in [-l_e, 0]$. The sieve estimator of $f_1(x|x^*)$ is

$$f_1(x|x^*) = \left(a_{00} + a_{01} \cos \frac{\pi}{l_x} x^* + a_{02} \cos \frac{2\pi}{l_x} x^* \right)$$

$$+ \sum_{k=1}^{3} \left(a_{k0} + a_{k1} \cos \frac{\pi}{l_x} x^* + a_{k2} \cos \frac{2\pi}{l_x} x^* \right) \cos \frac{k\pi}{l_e}$$

$$\times (x - x^*).$$

The restriction $\int_{\mathcal{X} \cap \{x < x^*\}} f_{x|x^*}(x|x^*) \, dx = 1$ for all x^* is equivalent to the restrictions $a_{00} = 1/l_e$ and $a_{01} = a_{02} = 0$.

S5. ADDITIONAL SIMULATIONS

EXAMPLE IV—Heteroskedastic Error with Zero Mean: Consider a measurement error

$$(S26) \qquad x = x^* + \sigma(x^*)\nu$$

with $x^* \perp \nu$, $E(\nu) = 0$, and $\sigma(\cdot) > 0$ being an unknown nonstochastic function. These assumptions can also be written as $E(x - x^*|x^*) = 0$, that is, the measurement error is the conditional mean independent of the true value. The identification condition is also satisfied because it can be verified that $x^* = \int x f_{x|x^*}(x|x^*) \, dx$. The error structure in the simulation is $F_\nu(\nu) = \Phi(\nu)$ with $\sigma(x^*) = 0.5 \exp(-x^*)$. The simulation results are in Table SI.

EXAMPLE V—Nonadditive Error with Zero Mode: An error equation like (S26) is usually set up for convenience. The additive structure of (S26) with $x^* \perp \nu$ may not always be appropriate in applications. Therefore, we now consider a nonseparable example, where it is more natural to specify $f_{x|x^*}(x|x^*)$ directly for the purpose of generating the simulated data. Let

$$f_{x|x^*}(x|x^*) = \frac{g(x, x^*)}{\int_{-\infty}^{\infty} g(x, x^*) \, dx},$$

TABLE SI

SIMULATION RESULTS

	a = −1			b = 1		
	Mean	Std. Dev.	RMSE	Mean	Std. Dev.	RMSE
Example I						
Ignoring meas. error	−0.7601	0.0759	0.2516	0.7601	0.0686	0.2495
Accurate data	−0.9974	0.0823	0.0824	0.9989	0.0766	0.0766
Sieve MLE	−0.9556	0.1831	0.1884	0.9087	0.1315	0.1601
Smoothing parameters: $i_n = 6$, $j_n = 6$ in f_1; $i_n = 3$, $j_n = 2$ in f_2						
Example II						
Ignoring meas. error	−0.5167	0.0611	0.4871	0.5834	0.0590	0.4208
Accurate data	−1.0010	0.0813	0.0813	1.0030	0.0761	0.0761
Sieve MLE	−0.9232	0.2010	0.2152	0.9430	0.1440	0.1549
Smoothing parameters: $i_n = 7$, $j_n = 3$ in f_1; $i_n = 3$, $j_n = 2$ in f_2						
Example III						
Ignoring meas. error	−0.6351	0.0734	0.3722	0.6219	0.0647	0.3836
Accurate data	−1.0010	0.0802	0.0802	1.0020	0.0752	0.0753
Sieve MLE	−0.9741	0.2803	0.2815	0.9342	0.2567	0.2650
Smoothing parameters: $i_n = 8$, $j_n = 8$ in f_1; $i_n = 3$, $j_n = 2$ in f_2						

$$g(x, x^*) = \exp\left\{h(x^*)\left[\left(\frac{x - x^*}{\sigma(x^*)}\right) - \exp\left(\frac{x - x^*}{\sigma(x^*)}\right)\right]\right\}.$$

It is easy to show that $f_{x|x^*}$ has the unique mode at x^* for any $h(x^*) > 0$. Thus the model is identified with this error structure. When $h(x^*) = 1$, this density becomes the density generated by Equation (S26) with ν having an extreme value distribution. Furthermore, the fact that identification holds for a general $h(x^*)$ means the independence assumption $x^* \perp \nu$ in (S26) is not necessary. We can deal with more general measurement error using the estimator in this paper. In the simulation, we use $\sigma(x^*) = 0.5\exp(-x^*)$ and $h(x^*) = \exp(-0.1x^*)$. The simulation results are in Table SI.

EXAMPLE VI—Nonadditive Error with Zero Median: We let the cumulative distribution function that corresponds to $f_{x|x^*}$ be

$$F_{x|x^*}(x|x^*)$$
$$= \frac{1}{\pi}\arctan\left\{h(x^*)\left[\frac{1}{2} + \frac{1}{2}\exp\left(\frac{x - x^*}{\sigma(x^*)}\right) - \exp\left(-\frac{x - x^*}{\sigma(x^*)}\right)\right]\right\} + \frac{1}{2}$$

with $h(x^*) > 0$. Note that $F_{x|x^*}(x^*|x^*) = \frac{1}{2}$ for any $h(x^*)$. Moreover, this distribution is not symmetric around x^*, and x^* is not the mode either. When $h(x^*) = 1$, the error structure is the same as in (S26). In the simulation, we use $\sigma(x^*) = 0.5\exp(-x^*)$ and $h(x^*) = \exp(-0.1x^*)$. The simulation results are in Table SI.

Dept. of Economics, Johns Hopkins University, 440 Mergenthaler Hall, 3400 N. Charles Street, Baltimore, MD 21218, U.S.A.; yhu@jhu.edu
and
Dept. of Economics, University of Chicago, 1126 East 59th Street, Chicago, IL 60637, U.S.A.; smschenn@uchicago.edu.

REFERENCES

AI, C., AND X. CHEN (2003): "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71, 1795–1843.

BEGUN, J., W. HALL, W. HUANG, AND J. WELLNER (1983): "Information and Asymptotic Efficiency in Parametric–Nonparametric Models," *The Annals of Statistics*, 11, 432–452.

BLUNDELL, R., X. CHEN, AND D. KRISTENSEN (2007): "Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves," *Econometrica*, 75, 1613–1669.

CHEN, X., AND X. SHEN (1998): "Sieve Extremum Estimates for Weakly Dependent Data," *Econometrica*, 66, 289–314.

CHEN, X., L. HANSEN, AND J. SCHEINKMAN (1997): "Shape-Preserving Estimation of Diffusions," Working Paper, University of Chicago.

CHEN, X., H. HONG, AND E. TAMER (2005): "Measurement Error Models with Auxiliary Data," *Review of Economic Studies*, 72, 343–366.

FAN, J. (1991): "On the Optimal Rates of Convergence for Nonparametric Deconvolution Problems," *The Annals of Statistics*, 19, 1257–1272.

GALLANT, A., AND D. NYCHKA (1987): "Semi-Nonparametric Maximum Likelihood Estimators," *Econometrica*, 55, 363–390.

NEWEY, W. K. (1995): "Convergence Rates for Series Estimators," in *Statistical Methods of Economics and Quantitative Economics: Essays in Honor of C. R. Rao*, ed. by G. Maddalla, P. Phillips, and T. Srinavasan. Cambridge, MA: Blackwell, 254–275.

——— (1997): "Convergence Rates and Asymptotic Normality for Series Estimators," *Journal of Econometrics*, 79, 147–168.

NEWEY, W. K., AND J. L. POWELL (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71, 1565–1578.

SCHENNACH, S. M. (2004): "Estimation of Nonlinear Models with Measurement Error," *Econometrica*, 72, 33–75.

SHEN, X. (1997): "On Methods of Sieves and Penalization," *The Annals of Statistics*, 25, 2555–2591.

STEIN, C. (1956): "Efficient Nonparametric Testing and Estimation," in *Proceedings of the Third Berkeley Symposium on Mathematics, Statistics and Probability*, 1. Berkeley, CA: University of California, 187–195.

TAUPIN, M.-L. (1998): "Estimation in the Nonlinear Errors-in-Variables Model," *Comptes Rendus de l'Académie des Science, Série I, Mathématiques*, 326, 885–890.