Title

Semiparametric Estimation of Nonseparable Models: A Minimum Distance from Independence Approach

Authors

Komunjer, Ivana
Santos, Andres

Publication Date

2009-03-01

SEMIPARAMETRIC ESTIMATION OF NONSEPARABLE MODELS: A MINIMUM DISTANCE FROM INDEPENDENCE APPROACH

IVANA KOMUNJER AND ANDRES SANTOS

Abstract. This paper focuses on nonseparable structural models of the form $Y = m(X, U, \alpha_0)$ with $U \perp X$ and in which the structural parameter α_0 contains both finite dimensional (θ_0) and infinite dimensional (h_0) unknown components. Our proposal is to estimate α_0 by a minimum distance from independence (MDI) criterion. We show that: (i) our estimator of h_0 is consistent and obtain rates of convergence; (ii) the estimator of θ_0 is $\sqrt{n}$ consistent and asymptotically normally distributed.

1. Introduction

Nonparametric identification of nonlinear structural models is often achieved by assuming that the model's latent variables are independent of the exogenous variables. Examples of such arguments include Brown (1983), Roehrig (1988), Matzkin (1994), Chesher (2003), Matzkin (2003), and Benkard and Berry (2007), among others. Yet the criteria used for estimation in such models rarely involve the independence property. Instead, nonparametric and semiparametric estimation methods typically use the mean independence between the latent and exogenous variables that comes in a form of conditional moment restrictions (see, e.g., Ai and Chen, 2003). Weaker than independence, the mean independence property by itself does

Date: March, 2009.

Affiliations and contact information. Komunjer: Department of Economics, University of California at San Diego, komunjer@ucsd.edu. Santos: Department of Economics, University of California at San Diego, a2santos@ucsd.edu.

Acknowledgments. This paper was presented at the 2008 EC^2 Meeting "Recent Advances in Structural Microeconometrics" in Rome, Italy. Many thanks to the participants of the conference for their suggestions and comments.

not guarantee the identification to hold; hence, nonparametric and semiparametric estimation literature most often simply assumes the models to be identified.

In the present paper we unify the estimation and identification of nonlinear models by employing the same criterion to obtain both: the independence between the models' latent and exogenous variables. We focus on the models of the form:

$$Y = m(X, U, \alpha_0) \qquad U \perp X \qquad U \sim U[0,1]$$

with variables $Y \in \mathbb{R}$ and $X \in \mathbb{R}^{d_x}$ that are observable, a latent disturbance $U \in \mathbb{R}$, and in which the function $m(x, u, \alpha_0)$ strictly increasing in $u \in \mathbb{R}$ for all $x \in \mathbb{R}^{d_x}$. The parameter of the model is $\alpha_0 \equiv (\theta_0, h_0) \in \mathcal{A} \equiv \Theta \times \mathcal{H}$, where $\Theta \subset \mathbb{R}^{d_\theta}$ is finite dimensional and $\mathcal{H}$ is an infinite dimensional set of functions. Requiring $U \sim U[0,1]$ is not a restrictive assumption, as for any random variable ξ with strictly increasing cdf F_ξ we can let $\xi = F_\xi^{-1}(U)$ and consider the function F_ξ^{-1} as part of the nonparametric component h_0.

The key insight of our estimation procedure lies in the following equality implied by the model:

$$P(Y \leq m(X, t_u, \alpha_0);\ X \leq t_x) = t_u \cdot P(X \leq t_x)$$

for all $(t_u, t_x) \in [0,1] \times \mathbb{R}^{d_x}$. We then exploit this relationship between the marginal and joint cdfs, and derive a von-Mises type criterion function:

$$Q(\alpha) \equiv \int \left[P\left(Y \leq m(X, t_u, \alpha);\ X \leq t_x\right) - t_u \cdot P\left(X \leq t_x\right)\right]^2 d\mu(t)$$

where μ a measure on $[0,1] \times \mathbb{R}^{d_x}$. In a sense, the criterion function $Q(\alpha)$ measures the distance from independence of U and X in the model. Hence, we call our estimator $\hat{\alpha}$—which we obtain by minimizing an appropriate sample analogue $Q_n(\alpha)$ of $Q(\alpha)$ above—a minimum distance from independence (MDI) estimator. If α_0 is identified by the assumptions of the model, then α_0 will also be the unique zero of $Q(\alpha)$. Exploiting the standard M-estimation arguments we are then able to: (i) show that the MDI estimator $\hat{\alpha} = (\hat{\theta}, \hat{h})$ is consistent for $\alpha_0 = (\theta_0, h_0)$; (ii) obtain the rate of convergence of the estimator $\hat{h}$ for h_0; and (iii) establish the asymptotic normality of the estimator $\hat{\theta}$ for θ_0.

The approach of minimizing the distance from independence for estimation was originally explored in the seminal work of Manski (1983). In the context of nonlinear parametric simultaneous equations systems, the asymptotic properties of the MDI estimators were derived in Brown and Wegkamp (2002). These results, however, assume that the structural mappings are finitely parameterized and do not allow for the presence of nonparametric parameters, which our approach does.

The remainder of the paper is organized as follows. In Section 2 we illustrate how semiparametric nonseparable models arise naturally in economic analysis by studying a simple version of Berry, Levinsohn, and Pakes's (1995) model of price-setting with differentiated products. Our estimator is presented in Section 3 and its consistency is established. Section 4 obtains the rate of convergence for the estimator of h_0. The asymptotic normality result for the estimator for θ_0 is derived in Section 5. Section 6 concludes the paper. The proofs of all the results stated in the text are relegated to Appendix.

2. Example

We proceed to illustrate how nonseparable structures arise naturally in simple economic models. We consider a basic version of Berry, Levinsohn, and Pakes (1995) (BLP henceforth) model with 2 products and 2 firms. On the demand side, we use a random utility specification à la Hausman and Wise (1978):

$$u_{ij} = -ap_j + b'z_j + \xi_j + \zeta_i + \varepsilon_{ij}, \tag{1}$$

in which u_{ij} is the utility of product j ($j = 1, 2$) to individual i ($i = 1, \ldots, I$) with unobserved characteristics ζ_i ($\zeta_i \in \mathbb{R}$), p_j and z_j are respectively the price and a k-vector of observed characteristics of product j ($p_j \in \mathbb{R}_+$, $z_j \in \mathbb{R}^k$, $k < \infty$); b is a k-vector of coefficients determining the impact of z_j on the utility for j ($b \in \mathbb{R}^k$), and ξ_j is an index of unobserved characteristics of the latter ($\xi_j \in \mathbb{R}$); $-a$ is a taste parameter on the price assumed constant across individuals ($a > 0$); finally, ε_{ij} is an error term that represents the deviations from an average behavior of agents and whose distribution is induced by the characteristics of the individual i and those of product j ($\varepsilon_{ij} \in \mathbb{R}$).

A baseline specification of the random utility in (1) is that ε_{ij} are iid across products j and individuals i. For example, assuming that ε_{ij}'s are Gumbel random variables, the resulting individual choice model is logit. In what follows, we let the difference $\varepsilon_{i2} - \varepsilon_{i1}$ be distributed with known cdf F; while F is assumed to be some known cdf, it need not be logit. When $\varepsilon_{i2} - \varepsilon_{i1}$ has cdf F, the demand for good j, denoted $D_j(p_j, p_{-j})$, is given by:

$$D_j(p_j, p_{-j}) = M \cdot F\big(a(p_{-j} - p_j) + b'(z_{-j} - z_j) + \xi_{-j} - \xi_j\big) \tag{2}$$

where M is the total market size.

Hereafter, we let the $Y \equiv D_1(p_1, p_2)/M$ be the market share for firm 1's good ($Y \in [0, 1]$), $P \equiv p_2 - p_1$, $Z \equiv z_2 - z_1$ and $\xi \equiv \xi_2 - \xi_1$. Then, the structural BLP model of (2) takes the form:

$$Y = F\big(aP + b'Z + \xi\big) \quad \text{with} \quad \xi \perp Z \tag{3}$$

In the model above, prices are endogenous, so even if ξ is independent of Z, we can expect P to depend on ξ. Hence, without further restrictions on ξ and P it is not possible to identify the parameters a and b in (3). We now show how the supply side information may be used to identify these parameters.

We assume that firms compete in prices (à la Bertrand), so each firm chooses prices which maximize its profits:

$$\Pi_j(p_j, p_{-j}) = (p_j - c_j)D_j(p_j, p_{-j})$$

Let the marginal costs $C \equiv (c_1, c_2)$ be observable and let $X \equiv (Z', C')$. The price (p_1, p_2) is then implicitly defined by the solution to the Bertrand game with exogenous variables X. Lemma 1 exploits this relationship to obtain an alternative representation for the model in (3).

Lemma 1. *Assume F is twice continuously differentiable on $\mathbb{R}$ with density f satisfying $f'(x)F(x) < f^2(x)$. If the unobservable ξ in the model* (3) *is continuously distributed, then it follows that:*

$$Y = F(h(X,U) + X'\theta) \tag{4}$$

with $0 < \partial h(X,u)/\partial u < 1$, $U \sim U[0,1]$, and $\theta' = (b', 0, 0)$.

The assumption $f'(x)F(x) < f^2(x)$ guarantees the existence of a unique Nash equilibrium and the Lemma can then be obtained by analyzing the equilibrium strategies.

3. Minimum Distance from Independence Estimation

We now proceed to study our MDI estimator, which applies to a class of models of which (4) is a special case. Consider the following general setup:

$$Y = m(X, U, \alpha_0) \qquad U \perp X \qquad U \sim U[0,1] \tag{5}$$

with observables $Y \in \mathbb{R}$ and $X \in \mathbb{R}^{d_x}$, unobservable $U \in \mathbb{R}$, and where $m(x, u, \alpha)$ is some known real function that is strictly increasing in u for all (x, α). Recall that the parameter of interest α consists of a finite dimensional component $\theta \in \mathbb{R}^{d_\theta}$ and an infinite dimensional one $h \in \mathcal{H}$. We therefore let $(\theta, h) \equiv \alpha \in \mathcal{A} \equiv \Theta \times \mathcal{H}$.

In many models, as in the BLP example discussed in Section 2, the assumption that $U \sim U[0,1]$ is not restrictive as nonuniform latent variables may be transformed to fit this model. When permitted, this re-parameterization is helpful as it allows for a simple characterization of the independence of U and X, as shown in the following Lemma.

Lemma 2. *Let the model* (5) *hold, X be continuously distributed, and $\partial m(x, u, \alpha_0)/\partial u > 0$ for all x. Then, it follows that $U \perp X$ if and only if for all $(t_x, t_u) \in \mathbb{R}^{d_x} \times (0,1)$:*

$$P(Y \leqslant m(X, t_u, \alpha_0);\ X \leqslant t_x) = t_u \cdot P(X \leqslant t_x) \tag{6}$$

Lemma 2 suggests a straightforward way to construct a criterion function through which to estimate α_0. Let $t = (t_u, t_x,)$ and define

$$W(t,\alpha) \equiv P(Y \leqslant m(X,t_u,\alpha);\ X \leqslant t_x) - t_u \cdot P(X \leqslant t_x) \tag{7}$$

Note that under the assumptions of Lemma 2, $U \perp X$ if and only if $W(t,\alpha_0) = 0$ for all t. Hence, a natural candidate for a population criterion function is the Cramer von-Mises type objective:

$$Q(\alpha) \equiv \int W^2(t,\alpha) d\mu(t) \tag{8}$$

where μ is a measure on $\mathbb{R}^{d_x} \times (0,1)$ that is absolutely continuous with respect to Lebesgue measure. The choice of μ is free, though we note that it will influence the asymptotic variance of our estimator for θ.

When the model in (5) is identified by the restriction $U \perp X$, Lemma 2 implies that α_0 is the unique zero of $Q(\alpha)$ and hence we have

$$\alpha_0 = \arg\min_{\mathcal{A}} Q(\alpha) \ .$$

Estimation will then proceed by minimizing an empirical analogue of $Q(\alpha)$. First define the sample analogue to $W(t,\alpha)$:

$$W_n(t,\alpha) \equiv \frac{1}{n}\sum_{i=1}^{n} 1\{y_i \leqslant m(x_i,t_u,\alpha);\ x_i \leqslant t_x\} - t_u \cdot \frac{1}{n}\sum_{i=1}^{n} 1\{x_i \leqslant t_x\} \ , \tag{9}$$

which yields a finite sample criterion function,

$$Q_n(\alpha) = \int W_n^2(t,\alpha) d\mu(t) \tag{10}$$

Since $\mathcal{A}$ contains a nonparametric component, minimizing $Q_n(\alpha)$ to obtain an estimator may not only be computationally difficult, but also undesirable as it may yield slow rates of convergence; see Chen (2006). For this reason we instead sieve the parameter space $\mathcal{A}$. Let $\mathcal{H}_n \subset \mathcal{H}$ be a sequence of approximating spaces, and define the sieve $\mathcal{A}_n = \Theta \times \mathcal{H}_n$. The MDI estimator is then given by,

$$\hat{\alpha} \in \arg\min_{\mathcal{A}_n} Q_n(\alpha) \tag{11}$$

Under the following assumptions, it is possible to establish the consistency of $\hat{\alpha}$.

Assumption A. *(i)* $\{y_i, x_i\}$ *are i.i.d.; (ii)* (Y, X, U) *are continuously distributed according to* (5)*; (iii)* $\alpha_0 = \arg\min_{\mathcal{A}} Q(\alpha)$*; (iv) The conditional densities* $f_{Y|X}(y|x)$ *and* $f_X(x)$ *are uniformly bounded in* (y, x)*; (v)* $\mu(t)$ *has full support on* $\mathbb{R}^{d_x} \times (0, 1)$.

Assumption B. *(i)* $m(x, u, \alpha)$ *is strictly increasing in* u $\forall(x, \alpha)$*; (ii)* $\Theta \subset \mathbb{R}^{d_\theta}$ *and* $\mathcal{H}$ *are compact w.r.t* $\|\cdot\|$ *and* $\|\cdot\|_\infty$*; (iii)* $|m(x, t, \alpha) - m(x, t, \tilde{\alpha})| \leqslant G(x)\{\|\theta - \tilde{\theta}\| + \|h - \tilde{h}\|_\infty\}$ *with* $E[G^2(X)] < \infty$*; (iv) The entropy* $\int_0^\infty \sqrt{N_{[]}(\eta^3, \mathcal{H}, \|\cdot\|_\infty)}d\eta < \infty$*; (v)* $\mathcal{H}_n \subset \mathcal{H}$ *are closed in* $||\cdot||_c$ *and for any* $h \in \mathcal{H}$ *there exists a* $\Pi_n h \in \mathcal{H}_n$ *such that* $\|h - \Pi_n h\|_\infty = o(1)$.

Assumption A(iii) requires identification of the model. For fully nonparametric specifications, identification of these models is well understood, see for example Matzkin (2003). Identification in semiparametric setups, however, can be more challenging and of course depends on the model specification. Below, we provide conditions for identification of the BLP example. Assumptions B(i)-(iv) ensures the stochastic process is asymptotically equicontinuous in probability. In Assumption B(iv), $N_{[]}(\eta^3, \mathcal{H}, \|\cdot\|_\infty)$ denotes the bracketing number of $\mathcal{H}$ with respect to $||\cdot||_\infty$, see van der Vaart and Wellner (1996) for details and examples of function classes satisfying Assumption B(iv). Finally, Assumption B(v) requires the sieve can approximate the parameter space with respect to the norm $||\cdot||_\infty$.

For the consistency result, endow $\mathcal{A}$ with the metric $||\alpha||_c = ||\theta|| + ||h||_\infty$.

Theorem 1. *Under Assumptions A and B,* $\|\hat{\alpha} - \alpha\|_c = o_p(1)$.

In the context of the BLP example, Assumption B(v) can be verified by letting $\mathcal{H}$ be a smooth set of functions. For example, suppose x has compact support $\mathcal{X}$ and let λ be a $d_x + 1$ dimensional vector of positive integers. Define $|\lambda| = \sum_{i=1}^{d_x+1} \lambda_i$ and $D^\lambda = \partial^\lambda / \partial x_1^{\lambda_1} \ldots \partial x_{d_x}^{\lambda_{d_x}} \partial u^{\lambda_{d_x+1}}$. An appropriate set $\mathcal{H}$ is then given by:

$$\mathcal{H} = \{h(x, u) : \max_{|\lambda| \leq \frac{3(d_x+1)}{2}+1} \sup_{x,u} |D^\lambda h(x, u)| \leq M,\ \inf_{x,u} \frac{\partial h(x, u)}{\partial u} \geq \epsilon,\ \epsilon > 0\} .$$

By Theorem 2.7.1 in van der Vaart and Wellner (1996), Assumption B(iv) is then satisfied. For approximating sieves $\mathcal{H}_n$ that satisfy B(v) appropriate options are splines or polynomials, see Chen (2006) for further examples and discussion. Assumption B(iii) follows by the mean value theorem if we assume F has bounded derivative. To verify Assumption B(ii), note the set of functions $\{\frac{\partial h(x,u)}{\partial u} : h(x,u) \in \mathcal{H}\}$ is compact in $||\cdot||_\infty$ due to them having uniformly bounded derivatives and the Arzela Ascoli Theorem. Compactness for all of $\mathcal{H}$ is then implied by the same arguments.

The hardest condition to verify for the discussed BLP model is identification, for which we provide primitive conditions in the following Theorem:

Theorem 2. *Let $h(x,u)$ be continuously differentiable, (i) $\partial h(x,u)/\partial u > 0$ for all x; (ii) $h(0,1/2) = 0$ and (iii) $\partial h(0,1/2)/\partial x = 1$. If F is a known cdf that is strictly increasing, then the parameters (θ, h) of the model (4) are identified by the restrictions $U \perp X$ and $U \sim U[0,1]$.*

Combining the results of Lemma 1 and Theorem 2 then shows that the parameters (θ, h) in the BLP model in (4) are identified.

The conditions of the Theorem fix the values of the unknown function h and of its gradient with respect to x, denoted by $\partial h(x,u)/\partial x$, at one point. In particular, (ii) holds if the distribution F_ξ of the products' unobservable ξ in the BLP model in Equation (3) is known to satisfy $F_\xi(0) = 1/2$, since when $X = 0$ and $\xi = 0$ the equilibrium is symmetric which implies $P = 0$. Hence, $aP + \xi = 0 = h(0,1/2)$. Requirement (iii) fixes the value of the gradient $\partial h(x,u)/\partial x$ at the same point. It ensures that the effects of changing θ can be separated from those of changing h. Indeed, if h is additively separable in x as in: $h(x,u) = \phi' x + r(u)$, then (ii) holds if $\phi = 1$. This restriction is as we would expect since it would be otherwise impossible to identify θ in $Y = F((\phi+\theta)'X + r(U))$.

4. Rate of Convergence

This section examines the rate of convergence of $\hat{h}$. This result is not only interesting in its own right, but is also instrumental in deriving the asymptotic normality of $\sqrt{n}(\hat{\theta} - \theta)$. We focus on the following norm for $h(x, u)$:

$$\|h\|_{L^2}^2 = \int h^2(x, t_u) f_X(x) dx d\mu(t) \tag{12}$$

Associated to the norm $||h||_{L^2}$ is the vector space $L^2 = \{h(x, u) : ||h||_{L^2} < \infty\}$.

In order to obtain the rates of convergence for $||\hat{h} - h||_{L^2}$, it is necessary to examine the local behavior of $Q(\alpha)$ at α_0. Consider the function $m(x, t_u, \alpha)$, and notice that it can be thought of as a mapping $m : (\mathcal{A}, || \cdot ||_c) \to L^2$. As a mapping, $m(x, t_u, \alpha)$ is said to be Frechet differentiable, if there exists a bounded linear map $\frac{dm(x,t_u,\alpha)}{d\alpha} : (\mathcal{A}, || \cdot ||_c) \to L^2$ such that,

$$\lim_{||\pi||_c \searrow 0} \frac{||m(x, t_u, \alpha + \pi) - m(x, t_u, \alpha) - \frac{dm(x,t_u,\alpha)}{d\alpha}[\pi]||_{L^2}}{||\pi||_c} = 0$$

The Frechet derivative is a natural extension of the standard derivative to arbitrary metric spaces. To illustrate these concepts, notice that in the BLP example the mapping $m(x, t_u, \alpha) = h(x, t_u) + x'\theta$. Further, since in this case $m(x, t_u, \alpha)$ is linear, the mapping is its own Frechet derivative, i.e. for $\pi = (h_\pi, \theta_\pi)$ we have:

$$\frac{dm(x, t_u, \alpha)}{d\alpha}[\pi] = h_\pi(x, t_u) + x'\theta_\pi$$

Given these definitions, we introduce the following assumption.

Assumption C. *(i) In a neighborhood $\mathcal{N}(\alpha_0) \subset \mathcal{A}$, $m : (\mathcal{A}, \| \cdot \|_c) \to L^2$ is continuously Frechet differentiable; (ii) The conditional densities $|f_{Y|X}(y|x) - f_{Y|X}(y'|x)| \leqslant J(x)|y - y'|^\nu$ with $E[J^2(X)G^{2\vee 2\nu}(X)] < \infty$.*

Under Assumption A(iv), B(iii) and C, the Frechet differentiability of $m(x, t_u, \alpha)$ is inherited by $Q(\alpha)$ as a mapping $Q : (\mathcal{A}, || \cdot ||_c) \to \mathbb{R}$. To state the form of this Frechet derivative, we define the linear map:

$$D_{\bar{\alpha}}[\pi] = \int f_{Y|X}(m(x, t_u, \bar{\alpha})|x) \frac{dm(x, t_u, \bar{\alpha})}{d\alpha}[\pi] 1\{x \leqslant t_x\} f_X(x) dx \tag{13}$$

Lemma 3 establishes that $Q(\alpha)$ is twice Frechet differentiable at α_0.

Lemma 3. *Under Assumption A(iv), B(iii) and C(i)-(ii), $Q(\alpha) : (\mathcal{A}, \|\cdot\|_c) \to \mathbb{R}$ is: (i) continuously Frechet differentiable in $\mathcal{N}(\alpha_0)$ with $\frac{dQ(\bar{\alpha})}{d\alpha}[\pi] = \int W(t,\bar{\alpha}) D_{\bar{\alpha}}[\pi] d\mu(t)$ (ii) and twice Frechet differentiable at α_0 with $\frac{d^2 Q(\alpha_0)}{d^2\alpha}[\psi,\pi] = \int D_{\alpha_0}[\psi] D_{\alpha_0}[\pi] d\mu(t)$.*

In this model, since $Q(\alpha)$ is minimized at α_0, its second derivative at α_0 induces a norm on $\mathcal{A}$. This result is analogous to a parametric model, in which if the Hessian H is a positive definite matrix, then $\sqrt{a'Ha}$ is a norm equivalent to the standard Euclidean norm. Guided by Lemma 3 we hence define the inner product and associated norm,

$$\langle \alpha, \tilde{\alpha}\rangle_w = \int D_{\alpha_0}[\alpha] D_{\alpha_0}[\tilde{\alpha}] d\mu(t) \qquad \|\alpha\|_w^2 = \langle \alpha, \alpha\rangle_w \tag{14}$$

The advantage of the norm $||\cdot||_w$ is that through a Taylor expansion it is often possible to show $||\alpha - \alpha_0||_w^2 \lesssim Q(\alpha)$, which makes it easy to obtain rates of convergence in $||\cdot||_w$. However, the norm $||\cdot||_w$ may not be of interest in itself. We instead aim to obtain a rate of convergence in the norm:

$$||\alpha||_s = ||\theta|| + ||h||_{L^2} \ .$$

It is possible to obtain a rate of convergence for $||\hat{\alpha} - \alpha_0||_s$ by understanding the behavior of the ratio $||\cdot||_s / ||\cdot||_w$ on the sieve $\mathcal{A}_n$. The assumptions we impose to obtain the rate of convergence for $||\hat{\alpha} - \alpha_0||_s$ are:

Assumption D. *(i) In $\mathcal{N}(\alpha_0)$, $||\alpha - \alpha_0||_w^2 \lesssim Q(\alpha) \lesssim ||\alpha - \alpha_0||_s^2$; (ii) The ratio $\tau_n = \sup_{\mathcal{A}_n} \|\alpha_n\|_s^2 / \|\alpha_n\|_w^2$ satisfies $\tau_n = o(n^\gamma)$ with $\gamma < 1/4$; (iii) For any $h \in \mathcal{H}$ there exists $\Pi_n h \in \mathcal{H}_n$ with $\|h - \Pi_n h\|_s = o(n^{-\frac{1}{2}})$ and $\|h - \Pi_n h\|_c = o(n^{-\frac{1}{4}})$.*

Assumption D(i) requires $||\alpha - \alpha_0||_w \lesssim Q(\alpha)$. As discussed, this is often verified through a Taylor expansion and allows us to obtain a rate of convergence in $||\cdot||_w$. In our model, $||\cdot||_w$ is too weak and $Q(\alpha)$ is often not continuous in this norm. We impose instead $Q(\alpha) \lesssim ||\alpha - \alpha_0||_s^2$. Assumption D(ii) is crucial in enabling us to obtain rates in $||\cdot||_s$ from rates in $||\cdot||_w$, and vise versa, which is needed to refine initial estimates of the rate of convergence. The ratio τ_n is often referred to as the sieve modulus of continuity, see for example Chen and Pouzo (2008). In practice, Assumption D(ii) is requiring the sieve not to grow too fast. Finally Assumption D(iii) refines the requirements of rates of approximation for the sieve $\mathcal{A}_n$.

Given these assumptions we obtain the following rate of convergence result:

Theorem 3. *Under Assumptions A, B, C and D,* $\|\hat{\alpha} - \alpha_0\|_s = o_p(n^{-\frac{1}{4}})$.

Notice that since $||\hat{\alpha} - \alpha_0||_s = ||\hat{\theta} - \theta_0|| + ||\hat{h} - h_0||_{L^2}$, it immediately follows from Theorem 3 that $||\hat{h} - h_0||_{L^2} = o_p(n^{-\frac{1}{4}})$, as desired.

5. Asymptotic Normality

In this section we establish the asymptotic normality of $\sqrt{n}(\hat{\theta} - \theta)$. The approach of the proof is similar to that of Ai and Chen (2003) and proceeds in two steps:

(i) For any $\lambda \in \mathbb{R}^{d_\theta}$ show the functional $F_\lambda(\alpha) = \lambda'\theta$, which returns the parametric component of the semiparametric specification is continuous in $\|\cdot\|_w$.

(ii) By the Riesz Representation theorem there is v^λ such that $\langle v^\lambda, \hat{\alpha} - \alpha_0\rangle_w = \lambda'(\hat{\theta} - \theta_0)$. We then establish the asymptotic normality of $\sqrt{n}\langle v^\lambda, \hat{\alpha} - \alpha_0\rangle_w$.

We therefore begin by establishing the continuity of $F_\lambda(\alpha) = \lambda'\theta$ in $||\cdot||_w$. Let $\bar{\mathcal{A}}$ denote the closure of the linear span of $\mathcal{A} - \alpha_0$ under $\|\cdot\|_w$, and observe that $(\bar{\mathcal{A}}, \|\cdot\|_w)$ is a Hilbert Space with inner product $\langle\cdot,\cdot\rangle_w$. Notice that $\bar{\mathcal{A}} = \mathbb{R}^{d_\theta} \times \bar{\mathcal{H}}$, with $\bar{\mathcal{H}}$ the closure of the linear span of $\mathcal{H} - h_0$ under $\|\cdot\|_w$. For any $(\alpha - \alpha_0)$ in $\bar{\mathcal{A}}$, we can then decompose $D_{\alpha_0}[\alpha - \alpha_0]$ according to:[1]

$$\begin{aligned} D_{\alpha_0}[\alpha - \alpha_0] &= \frac{dW(t,\alpha_0)}{d\alpha}[\alpha - \alpha_0] \\ &= \frac{dW(t,\alpha_0)}{d\theta'}[\theta - \theta_0] + \frac{dW(t,\alpha_0)}{dh}[h - h_0] \end{aligned} \tag{15}$$

For each component θ_i of θ, $1 \leqslant i \leqslant d_\theta$, let $h_j^* \in \bar{\mathcal{H}}$ be defined by:

$$h_j^* = \arg\min_{h\in\bar{\mathcal{H}}} \int \left(\frac{dW(t,\alpha_0)}{d\theta_j} - \frac{dW(t,\alpha_0)}{dh}[h]\right)^2 d\mu(t) \tag{16}$$

[1]The first equality in (15) is formally justified in the proof of Lemma 3 in the Appendix, in which it is shown D_{α_0} is the Frechet derivative of $W(t, \alpha)$.

where the minimum in (16) is indeed attained and h_j^* is well defined due to the Projection Theorem in Hilbert Spaces, see for example Theorem 3.3.2 in (Luenberger, 1969). Similarly define $h^* = (h_1^*, \ldots, h_{d_\theta}^*)$ and let,

$$\frac{dW(t,\alpha_0)}{dh}[h^*] = \left(\frac{dW(t,\alpha_0)}{dh}[h_1^*], \ldots, \frac{dW(t,\alpha_0)}{dh}[h_{d_\theta}^*]\right) \tag{17}$$

As a final piece of notation, we also need to denote the vector of residuals,

$$R_{h^*}(t) = \frac{dW(t,\alpha_0)}{d\theta} - \frac{dW(t,\alpha_0)}{dh}[h^*] \tag{18}$$

and the associated matrix:

$$\Sigma^* = \int R_{h^*}(t) R'_{h^*}(t) d\mu(t) \tag{19}$$

As Lemma 4 shows, the continuity of the functional $F_\lambda(\alpha) = \lambda'\theta$ depends on the matrix Σ^* being positive definite. Lemma 4 also obtain the formula for the Riesz Representor of $F_\lambda(\alpha)$.

Lemma 4. *Let $v_\theta^\lambda = (\Sigma^*)^{-1}\lambda$, $v_h^\lambda = -h^* v_\theta^\lambda$ and $(v_\theta^\lambda, v_h^\lambda) \in \bar{\mathcal{A}}$. If Σ^* is positive-definite, then for any $\lambda \in \mathbb{R}^{d_\theta}$, $F_\lambda(\alpha - \alpha_0) = \lambda'(\theta - \theta_0)$ is continuous on $\bar{\mathcal{A}}$ under $\|\cdot\|_w$ and in addition we have $F_\lambda(\alpha - \alpha_0) = \langle v^\lambda, \alpha - \alpha_0\rangle_w = \lambda'(\theta - \theta_0)$.*

Having established the continuity of $F_\lambda(\alpha)$ in $||\cdot||_w$, we can now show the asymptotic normality of $\sqrt{n}\langle v^\lambda, \hat{\alpha} - \alpha_0\rangle_w$. For this purpose we require one final assumption.

Assumption E. *(i) Σ^* is positive definite; (ii) $v^\lambda \in \mathcal{A}$ for $||\lambda||$ small; (iii) $\forall\ \alpha \in \mathcal{N}(\alpha_0)$, and $\pi, \bar{\alpha} \in \mathcal{A}$, the pathwise derivative $\frac{dD_{\alpha+\tau\bar{\alpha}}[\pi]}{d\tau}$ exists with $\int \sup_{s\in[0,1]} \left| \frac{dD_{\alpha+\tau\bar{\alpha}}[\pi]}{d\tau}\Big|_{\tau=s} \right| d\mu(t) \lesssim \|\bar{\alpha}\|_s \|\pi\|_s$ and $\int \sup_{s\in[0,1]} \left(\frac{dD_{\alpha+\tau\bar{\alpha}}[\pi]}{d\tau}\Big|_{\tau=s} \right)^2 d\mu(t) \lesssim \|\bar{\alpha}\|_s^2$; (iv) $\forall\ \alpha \in \mathcal{N}(\alpha_0)$ and $\pi \in \mathcal{A}$, $|D_\alpha[\pi]|$ is bounded.*

Assumption E(i) ensures that $F_\lambda(\alpha) = \lambda'\theta$ is continuous in $||\cdot||_w$, as shown in Lemma 4. While $v^\lambda \in \bar{\mathcal{A}}$, Assumption E(ii) additionally requires $v^\lambda \in \mathcal{A}$. As a result v^λ may be approximated by an element $\Pi_n v^\lambda \in \mathcal{A}_n$. The qualification "for $||\lambda||$ small" is due to the compactness assumption on $\Theta \times \mathcal{H}$ requiring them to be bounded in norm. Finally Assumptions E(iii)-(iv) require $W(t, \alpha)$ to be twice differentiable and for some regularity conditions on the derivatives. For example, in the discussed BLP example for $\alpha = (\theta_\alpha, h_\alpha)$ we have:

$$D_\alpha[\pi] = \int f_{Y|X}(x'\theta_\alpha + h_\alpha(x,u)|x)(x'\theta_\pi + h_\pi(x,t_u))1\{x \le t_x\} f_X(x)dx \tag{20}$$

Hence, Assumption E(iv) is easily verified by requiring $\sup_{t_u} E[|x'\theta + h(x,t_u)|] < \infty$ uniformly in $(\theta, h) \in \mathcal{A}$. Similarly, by direct calculation we obtain that in the discussed BLP example,

$$\frac{dD_{\alpha+\tau\bar{\alpha}}[\pi]}{d\tau}\Big|_{\tau=s} = \int f'_{Y|X}(x'(\theta_\alpha + s\bar{\alpha}) + h_\alpha(x,t_u) + sh_{\bar{\alpha}}(x,t_u))(x'\theta_\pi + h_\pi(t_u,x)) \times (x'\theta_{\bar{\alpha}} + h_{\bar{\alpha}}(t_u,x))1\{x \le t_x\} f_X(x)dx$$

and hence Assumption E(iii) is easily verified if $|f'_{Y|X}(y|x)|$ is bounded in (y, x).

We are now ready to establish the asymptotic normality of $\sqrt{n}(\hat{\theta} - \theta_0)$.

Theorem 4. *Let Assumptions A, B, C, D and E hold. Then,* $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{\mathcal{L}} N(0, \Sigma)$ *where* $\Sigma = [\Sigma^*]^{-1} \left[\int R_{h^*}(t) R'_{h^*}(s)\Sigma(t,s)d\mu(t)d\mu(s)\right] [\Sigma^*]^{-1}$

6. Conclusion

We have proposed a general estimation framework for a large class of semiparametric nonseparable models. The resulting estimator converges to the nonparametric component at a $o_p(n^{-\frac{1}{4}})$ rate, and yields an asymptotically normal estimator for the parametric component. Some of the Assumptions must be verified in a model specific basis, which we have done in an example motivated by Berry, Levinsohn, and Pakes's (1995) model of price-setting with differentiated products.

Appendix A. Details of the BLP example

In this Appendix, we give the proofs of Lemma 1 and Theorem 2. We start with an auxiliary Lemma whose result will be useful later on.

Lemma 5. *Assume F is twice continuously differentiable with $f'(x)F(x)/f^2(x) < 1$. Then the BLP equilibrium prices exist, are unique, and the map $(\xi_2-\xi_1, z_2-z_1, c_1, c_2) \mapsto (p_2-p_1)$ is twice continuously differentiable with:*

$$-\frac{1}{a} < \frac{\partial(p_2-p_1)}{\partial(\xi_2-\xi_1)} < 0$$

Proof of Lemma 5: Under the assumption $f'(x)F(x)/f^2(x) < 1$ the goods are substitutes and the elasticity of demand is a decreasing function of the other firm's prices, i.e.

$$\frac{\partial^2 \ln D_j(p_j, p_{-j})}{\partial p_j \partial p_{-j}} > 0 \ .$$

It follows that the (log-transformed) Bertrand duopoly played by the firms is supermodular; hence, there exist a pure Nash equilibrium to the game (see, e.g., Milgrom and Roberts, 1990). We now show that this equilibrium is unique. For this purpose note that

$$-\frac{\partial^2 \ln \Pi_j(p_j, p_{-j})}{\partial p_j^2} - \frac{\partial^2 \ln \Pi_j(p_j, p_{-j})}{\partial p_j \partial p_{-j}} = \frac{1}{(p_j-c_j)^2} > 0$$

so that the "dominant diagonal" condition of Milgrom and Roberts (1990) holds; this guarantees that the equilibrium is unique.

Since $f'(x)F(x)/f^2(x) < 1$ we also have $\partial^2 \ln D_j(p_j, p_{-j})/\partial p_j^2 < 0$, which implies that $\partial^2 \ln \Pi_j(p_j, p_{-j})/\partial p_j^2 < 0$, and the Nash equilibrium (p_1^*, p_2^*) is the unique solution to the first order conditions $\Phi(p_1, p_2, \xi) = 0$, where we have let $\xi = \xi_2 - \xi_1$ and

$$\Phi(p_1, p_2, \xi) = \begin{bmatrix} \frac{1}{p_1-c_1} + \frac{\partial \ln D_1(p_1,p_2)}{\partial p_1} \\ \frac{1}{p_2-c_2} + \frac{\partial \ln D_2(p_1,p_2)}{\partial p_2} \end{bmatrix}$$

Note that the map Φ is continuously differentiable and we have:

$$\mathrm{D}_{(p_1,p_2)}\Phi = \begin{bmatrix} -\frac{1}{(p_1-c_1)^2} + \frac{\partial^2 \ln D_1(p_1,p_2)}{\partial p_1^2} & \frac{\partial^2 \ln D_1(p_1,p_2)}{\partial p_1 \partial p_2} \\ \frac{\partial^2 \ln D_2(p_1,p_2)}{\partial p_1 \partial p_2} & -\frac{1}{(p_2-c_2)^2} + \frac{\partial^2 \ln D_2(p_1,p_2)}{\partial p_2^2} \end{bmatrix}$$

In addition, note that the demand function in (2) satisfies:

$$-\frac{\partial^2 \ln D_j(p_j, p_{-j})}{\partial p_j^2} = \frac{\partial^2 \ln D_j(p_j, p_{-j})}{\partial p_j \partial p_{-j}} = \alpha \frac{\partial^2 \ln D_j(p_j, p_{-j})}{\partial p_j \partial(\xi_{-j}-\xi_j)} > 0 \tag{21}$$

where the last inequality follows from Assumption $f'(x)F(x)/f^2(x) < 1$. Therefore,

$$\det \mathrm{D}_{(p_1,p_2)}\Phi = \frac{1}{(p_1-c_1)^2(p_2-c_2)^2} - \frac{1}{(p_1-c_1)^2}\frac{\partial^2 \ln D_2(p_1,p_2)}{\partial p_2^2} - \frac{1}{(p_2-c_2)^2}\frac{\partial^2 \ln D_1(p_1,p_2)}{\partial p_1^2} > 0$$

Hence, by the Implicit Function Theorem (see, e.g., Theorem 9.28 in Rudin, 1976), the equation $\Phi(p_1, p_2, \xi) = 0$ defines in a neighborhood of the point (p_1^*, p_2^*, ξ) a mapping $\xi \mapsto p_j$ that is continuously differentiable, and whose derivative at this point equals:

$$\frac{\partial p_1}{\partial \xi} = -\frac{1}{a}\frac{1}{\det \mathrm{D}_{(p_1,p_2)}\Phi}\frac{1}{(p_2-c_2)^2}\frac{\partial^2 \ln D_1(p_1,p_2)}{\partial p_1^2} \tag{22}$$

$$\frac{\partial p_2}{\partial \xi} = \frac{1}{a}\frac{1}{\det \mathrm{D}_{(p_1,p_2)}\Phi}\frac{1}{(p_1-c_1)^2}\frac{\partial^2 \ln D_2(p_1,p_2)}{\partial p_2^2} \tag{23}$$

where the first equality uses (21) and the fact that

$$\frac{\partial^2 \ln D_2(p_1,p_2)}{\partial p_2^2}\frac{\partial^2 \ln D_1(p_1,p_2)}{\partial p_1 \partial \xi} - \frac{\partial^2 \ln D_1(p_1,p_2)}{\partial p_1 \partial p_2}\frac{\partial^2 \ln D_2(p_1,p_2)}{\partial p_2 \partial \xi} = 0$$

while the second exploits (21) and the fact that

$$\frac{\partial^2 \ln D_1(p_1,p_2)}{\partial p_1^2}\frac{\partial^2 \ln D_2(p_1,p_2)}{\partial p_2 \partial \xi} - \frac{\partial^2 \ln D_2(p_1,p_2)}{\partial p_1 \partial p_2}\frac{\partial^2 \ln D_1(p_1,p_2)}{\partial p_1 \partial \xi} = 0$$

¿From (22) we then have the desired result:

$$-\frac{1}{a} < \frac{\partial(p_2 - p_1)}{\partial \xi} < 0$$

which concludes the proof of the Lemma. ■

Proof of Lemma 1: It follows from Lemma 5 that the demand for good 1 in (2) is an increasing function of $\xi_2 - \xi_1$. To see this, note that:

$$\frac{\partial D_1(p_1,p_2)}{\partial(\xi_2-\xi_1)} = M \cdot \left[a\frac{\partial(p_2-p_1)}{\partial(\xi_2-\xi_1)} + 1\right] \cdot f\big(a(p_2-p_1) + b'(z_2-z_1) + \xi_2 - \xi_1\big)$$

which together with Proposition 1 yields:

$$\frac{\partial D_1(p_1,p_2)}{\partial(\xi_2-\xi_1)} > 0$$

Since ξ is continuously distributed, it has a strictly increasing cdf, which we denote F_ξ. Noting that $F_\xi(\xi) \sim U[0,1]$, we may define:

$$h(X,U) \equiv a(p_2-p_1) + F_\xi^{-1}(U)$$

and the claim of the Lemma follows immediately from Equation (3). ■

Proof of Theorem 2: Let $F_{Y|X}(\cdot|\cdot;\mathcal{S})$ denote the conditional distribution of Y given X that is induced by the structure $\mathcal{S} \equiv (\theta, h)$. Fix $x \in \mathbb{R}^{d_x}$ and let $v : \mathbb{R}^{d_x+1} \to \mathbb{R}$ be such that for any $u \in \mathbb{R}$, we have: $h(x,u) = t$ if and only if $u = v(t,x)$. Note that $v(\cdot, x)$ is well defined since by (i) we have $\partial h(x,u)/\partial u > 0$. Then, for any $y \in \mathbb{R}$,

$$
\begin{aligned}
F_{Y|X}(y|x;\mathcal{S}) &= P\big(Y \leqslant y \big| X = x\big) \\
&= P\big(h(X,U) \leqslant F^{-1}(y) - \theta' x \big| X = x\big) \\
&= P\big(U \leqslant v(F^{-1}(y) - \theta' x, x) \big| X = x\big) \\
&= P\big(U \leqslant v(F^{-1}(y) - \theta' x, x)\big) \\
&= v(F^{-1}(y) - \theta' x, x) \qquad (24)
\end{aligned}
$$

where the second equality uses the fact that $h(x,u)$ is strictly increasing in u, the third exploits the independence of U and X, and the last follows from U being uniform.

Since $h(x,u)$ is continuously differentiable on $\mathbb{R}^{d_x} \times (0,1)$ and such that $\partial h(x,u)/\partial u > 0$ on $\mathbb{R}^{d_x} \times (0,1)$, $v(t,x)$ is continuously differentiable on $\mathbb{R}^{d_x+1}$ with:

$$
\begin{aligned}
\frac{\partial v}{\partial x}(t,x) &= -\frac{\partial h}{\partial x}(x, v(t,x)) \left[\frac{\partial h}{\partial u}(x, v(t,x))\right]^{-1} \\
\frac{\partial v}{\partial t}(t,x) &= \left[\frac{\partial h}{\partial u}(x, v(t,x))\right]^{-1} \qquad (25)
\end{aligned}
$$

Further, for any $(y,x) \in \mathbb{R}^{d_x+1}$ let $\Phi(y,x) \equiv P\big(Y \leqslant y \big| X = x\big)$. Under our assumptions on F, $\Phi(y,x)$ is continuously differentiable on $\mathbb{R}^{d_x+1}$ and we have:

$$
\begin{aligned}
\frac{\partial \Phi}{\partial y}(y,x) &= \frac{\partial v}{\partial t}(F^{-1}(y) - \theta' x, x) \frac{1}{f(F^{-1}(y))} \\
\frac{\partial \Phi}{\partial x}(y,x) &= -\theta \frac{\partial v}{\partial t}(F^{-1}(y) - \theta' x, x) + \frac{\partial v}{\partial x}(F^{-1}(y) - \theta' x, x) \qquad (26)
\end{aligned}
$$

In particular, $\partial \Phi(y,x)/\partial y > 0$ on $\mathbb{R}^{d_x+1}$. Combining (25) and (26) we then obtain:

$$
-\frac{1}{f(F^{-1}(y))} \left[\frac{\partial \Phi}{\partial x}(y,x)\right] \left[\frac{\partial \Phi}{\partial y}(y,x)\right]^{-1} = \theta + \frac{\partial h}{\partial x}\big(x, v(F^{-1}(y) - \theta' x, x)\big) \qquad (27)
$$

Evaluate the left hand side of (27) at $x = 0 \in \mathbb{R}^{d_x}$ and $y = F(0) \in (0,1)$. For these values of x and y, we have: $F^{-1}(y) - \theta' x = 0$ so by using Assumption (ii), $v(0,0) = 1/2$.

Combining the latter with Assumption (iii) then gives:

$$\theta = -\frac{1}{f(0)}\left[\frac{\partial \Phi}{\partial x}\big(F(0), 0\big)\right]\left[\frac{\partial \Phi}{\partial y}\big(F(0), 0\big)\right]^{-1} - 1$$

from which it follows that θ is identified. The identification of $v(t, x)$, and hence $h(x, u)$ then immediately follows from (24). ■

Appendix B. Proofs for Section 3

Proof of Lemma 2: Since U is uniform on $(0, 1)$ it immediately follows that $U \perp X$ if and only if for all $(t_x, t_u) \in \mathbb{R}^{d_x} \times (0, 1)$:

$$P(U \leqslant t_u;\ X \leqslant t_x) = t_u \cdot P(X \leqslant t_x) \tag{28}$$

Further, notice that for any $(t_x, t_u) \in \mathbb{R}^{d_x} \times (0, 1)$ the following holds:

$$\begin{aligned} P(X \leqslant t_x;\ Y \leqslant m(X, t_u, \alpha_0)) &= \int_{s_x \leqslant t_x} \int_{s_y \leqslant m(s_x, t_u, \alpha_0)} f_{XY}(s_x, s_y) ds_x ds_y \\ &= \int_{s_x \leqslant t_x} \int_{s_u \leqslant t_u} f_{XY}(s_x, m(s_x, s_u, \alpha_0)) \frac{\partial m(s_x, s_u, \alpha_0)}{\partial u} ds_x ds_u \\ &= \int_{s_x \leqslant t_x} \int_{s_u \leqslant t_u} f_{XU}(s_x, s_u) \mathrm{d}s_x ds_u \\ &= \int_{s_x \leqslant t_x} f_X(s_x) ds_x \int_{s_u \leqslant t_u} f_U(s_u) ds_u \qquad (29) \end{aligned}$$

where the second and third equalities follow by a change of variable $(s_x, s_y) = (s_x, m(s_x, s_u, \alpha_0))$ and a change in measure. The final equality in (29) then follows by $U \perp X$. Combining (28), (29) and the fact that U is unform on $(0, 1)$ then establishes the claim of the Lemma. ■

Lemma 6. *Under Assumptions A and B, the following class is Donsker:*

$$\mathcal{F} = \big\{1\{y \leqslant m(x, t_u, \alpha);\ x \leqslant t_x\}, (\alpha, t_x, t) \in \mathcal{A} \times \mathbb{R}^{d_x} \times (0, 1)\big\}$$

Proof: First define the following classes of functions for $1 \leqslant k \leqslant d_x$:

$$\mathcal{F}_u = \big\{1\{y \leqslant m(x, t, \alpha)\}\big\} : (\alpha, t) \in \mathcal{A} \times (0, 1)\} \tag{30}$$

$$\mathcal{F}_x^{(k)} = \{1\{x^{(k)} \leqslant t\} : t \in \mathbb{R}\} \tag{31}$$

Further notice that by direct calculation we have,

$$\mathcal{F} = \mathcal{F}_u \times \prod_{k=1}^{d_x} \mathcal{F}_x^{(k)} \tag{32}$$

We will establish the Lemma by exploiting (32). Notice that for any continuously distributed random variable $V \in \mathbb{R}$ and $\eta > 0$ we can find $\{-\infty = t_1, t_2, \ldots, t_{\lfloor \eta^{-2} \rfloor + 2} = +\infty\}$ such that $P(t_i \leqslant V \leqslant t_{i+1}) \leqslant \eta^2$. The brackets $[1\{v \leqslant t_i\},\ 1\{v \leqslant t_{i+1}\}]$ then cover $\{1\{v \leqslant t\} : t \in \mathbb{R}\}$ and in addition by construction we have,

$$E[(1\{V \leqslant t_i\} - 1\{V \leqslant t_{i+1}\})^2] \leqslant \eta^2$$

Therefore, we immediately establish that for all $1 \leqslant k \leqslant d_x$:

$$N_{[\,]}(\eta, \mathcal{F}_x^{(k)}, \|\cdot\|_{L^2}) = O(\eta^{-2}) \tag{33}$$

By assumption, $\mathcal{H}$ is compact under $\|\cdot\|_\infty$ and Θ under $\|\cdot\|$. Thus, there exists a collection $\{h_j\}$ and $\{\theta_l\}$ such that the open balls of size $K_h\eta^3$ around $\{h_j\}$ and of size $K_\theta\eta^3$ around $\{\theta_l\}$ cover $\mathcal{H}$ and Θ respectively. Define the collection $\{\alpha_i\} = \{h_j\} \times \{\theta_l\}$ and note that,

$$\#\{\alpha_i\} = N_{[\,]}(K_h\eta^3, \mathcal{H}, \|\cdot\|_\infty) \times (K_\theta\eta^3)^{-d_\theta} \tag{34}$$

Furthermore, it then follows that for any $\alpha \in \mathcal{A}$ there exists a $\alpha_i \in \{\alpha_i\}$ such that

$$\begin{aligned} |m(x,t,\alpha) - m(x,t,\alpha_i)| &\leqslant G(x)\{\|\theta - \theta_i\| + \|h - h_i\|_\infty\} \\ &\leqslant G(x)\{K_\theta + K_h\}\eta^3 \end{aligned} \tag{35}$$

We conclude from (35) that for $\alpha_i \in \{\alpha_i\}$, brackets of the form

$$[m(x,t,\alpha_i) - \{K_\theta + K_h\}\eta^3 G(x);\ m(x,t,\alpha_i) + \{K_\theta + K_h\}\eta^3 G(x)]$$

cover the class $\{m(x,t,\alpha) : \alpha \in \mathcal{A}\}$ for each fixed t. Next note that since $m(x,u,\alpha)$ is strictly increasing in u for all (x,α), we may define their inverses:

$$v(u,x,\alpha) = t \iff m(x,t,\alpha) = u \tag{36}$$

Following Akritas and van Keilegom (2001), for each $\alpha_i \in \{\alpha_i\}$ we let $F_i^U(t)$ be as in the first equality in (37) and obtain second equality in (37) from (36).

$$\begin{aligned} F_i^U(t) &\equiv P(Y \leqslant m(X,t,\alpha_i) + \{K_\theta + K_h\}\eta^3 G(X)) \\ &= P(v(Y - \{K_\theta + K_h\}\eta^3 G(X), X, \alpha_i) \leqslant t) \end{aligned} \tag{37}$$

Arguing as in (33), it follows that it is possible to pick t_{ik}^U with $k = 1, \ldots, O(\eta^{-2})$ such that they partition $\mathbb{R}$ into segments each with F_i^U probability at most $\eta^2/6$. Also let

$$F_i^L(t) = P(Y \leqslant m(X,t,\alpha_i) - \{K_\theta + K_h\}\eta^3 G(X)) \tag{38}$$

and choose t_{ik}^L with $k = 1, \ldots, O(\eta^{-2})$ such that they partition $\mathbb{R}$ into segments each with F_i^L probability at most $\eta^2/6$. Next combine $\{t_{ik}^L\}$ and $\{t_{ik}^U\}$ into a single bracket, by letting each $t \in \mathbb{R}$ form the bracket

$$t_{ik_1}^L \leqslant t \leqslant t_{ik_2}^U$$

where $t_{ik_1}^L$ is the largest element of $\{t_{ik}^L\}$ such that $t_{ik}^L \leqslant t$, and similarly $t_{ik_2}^U$ is the smallest element in $\{t_{ik}^U\}$ such that $t_{ik}^U \geq t$. We denote this new brackets by $\{[t_{ik_1}, t_{ik_2}]\}$ and note that direct calculation shows

$$\#\{[t_{ik_1}, t_{ik_2}]\} = O(\eta^{-2}) \tag{39}$$

It follows from the strict monotonicity of $m(x,t,\alpha)$ in t that for every $(\alpha, t) \in \mathcal{A} \times (0,1)$ there exists a $\alpha_i \in \{\alpha_i\}$ and $[t_{ik_1}, t_{ik_2}] \in \{[t_{ik_1}, t_{ik_2}]\}$ such that,

$$\begin{aligned} 1\{y \leqslant m(x, t_{ik_1}, \alpha_i) - \{K_\theta + K_h\}\eta^3 G(x)\} &\leqslant 1\{y \leqslant m(x,t,\alpha)\} \\ &\leqslant 1\{y \leqslant m(x, t_{ik_2}, \alpha_i) + \{K_\theta + K_h\}\eta^3 G(x)\} \end{aligned} \tag{40}$$

In order to calculate the size of the proposed brackets, note their L^2 squared norm is $F_i^U(t_{ik_2}) - F_i^L(t_{ik_1})$. The construction of $\{[t_{ik_1}, t_{ik_2}]\}$ in turn implies the first inequality in (41) holds for any $t \in [t_{ik_1}, t_{ik_2}]$, while direct calculation yields the second inequality for any constat M_η. Setting $M_\eta = \sqrt{6E[G^2(X)]}/\eta$ and Chebychev's inequality yields (41).

$$\begin{aligned} F_i^U(t_{ik_2}) - F_i^L(t_{ik_1}) &\leqslant F_i^U(t) - F_i^L(t) + \frac{\eta^2}{3} \\ &\leqslant F_i^U(t; G(X) \leqslant M_\eta) - F_i^L(t; G(X) \leqslant M_\eta) + 2P(G(X) \geq M_\eta) + \frac{\eta^2}{3} \\ &\leqslant F_i^U(t; G(X) \leqslant M_\eta) - F_i^L(t; G(X) \leqslant M_\eta) + \frac{2}{3}\eta^2 \end{aligned} \tag{41}$$

To conclude, notice that the first inequality and equality in (42) follows by direct calculation. The second inequality is implied by the mean value theorem and $M_\eta = \sqrt{6E[G^2(Y,X)]}/\eta$. The resulting expression is finite due to Assumptions A and B.

$$\begin{aligned}
&F_i^U(t; G(X) \leqslant M_\eta) - F_i^L(t; G(X) \leqslant M_\eta) \\
&\quad \leqslant P\big(Y \leqslant m(X,t,\alpha_i) + \{K_\theta + K_h\}M_\eta \eta^3\big) - P\big(Y \leqslant m(X,t,\alpha_i) - \{K_\theta + K_h\}M_\eta \eta^3\big) \\
&\quad = E\left[P\big(Y \leqslant m(x,t,\alpha_i) + \{K_\theta + K_h\}M_\eta \eta^3 \big| x\big) - P\big(Y \leqslant m(x,t,\alpha_i) - \{K_\theta + K_h\}M_\eta \eta^3 \big| x\big)\right] \\
&\quad \leqslant 2\{\sup_{y,x} f_{Y|X}(y|x)\}\{K_\theta + K_h\}\sqrt{6E[G^2(Y,X)]}\eta^2
\end{aligned} \tag{42}$$

It follows from (41) and (42), that by choosing

$$\{K_\theta + K_h\} \leqslant \big(2\{\sup_{y,x} f_{Y|X}(y|x)\}\sqrt{6E[G^2(X)]}\big)^{-1}$$

the proposed brackets will have L^2 size η. Thus, we have from (34) and (39),

$$N_{[\,]}(\eta, \mathcal{F}_u, \|\cdot\|_{L^2}) = O(N_{[\,]}(K_h\eta^3, \mathcal{H}, \|\cdot\|_\infty) \times (K_\theta \eta^3)^{-(2+d_\theta)}) \tag{43}$$

To conclude note that (33), (43), Assumption B and Theorem 2.5.6 in van der Vaart and Wellner (1996) implies the classes $\mathcal{F}_x^{(k)}$ and $\mathcal{F}_u$ are Donsker. In turn, since all classes are uniformly bounded by 1, Theorem 2.10.6 in van der Vaart and Wellner (1996) and (32) establishes the claim of the Lemma. ■

Proof of Theorem 1: By Assumption B and the Tychonoff Theorem, $\mathcal{A}$ is compact with respect to $\|\cdot\|_c$. Furthermore, Lemma 6 and simple manipulations show,

$$\sup_{t,\alpha} |W_n(t,\alpha) - W(t,\alpha)| \xrightarrow{p} 0 \tag{44}$$

By direct calculation, exploiting (44) and noticing $W_n(t,\alpha)$ and $W(t,\alpha)$ are uniformly bounded by 1, we then obtain

$$\sup_{\alpha} |Q_n(\alpha) - Q(\alpha)| \leqslant \sup_{t,\alpha} |W_n(t,\alpha) - W(t,\alpha)| \times \left\{\sup_{t,\alpha} |W_n(t,\alpha)| + \sup_{t,\alpha} |W(t,\alpha)|\right\} \xrightarrow{p} 0 \tag{45}$$

The result then follows by Lemma A1 in Newey and Powell (2003) and noticing that their requirement that $Q_n(\alpha)$ being continuous can be substituted by $\hat{\alpha}$ being an element of the argmin correspondence. ■

Appendix C. Proofs for Section 4

Throughout the proofs in this Appendix, it is useful to define the norm

$$||\xi||^2_{L^2_\mu} = \int \xi^2(t) d\mu(t)$$

and study the associated vector space $L^2_\mu = \{\xi(t) : ||\xi||_{L^2_\mu} < \infty\}$.

Proof of Lemma 3: We first study the differentiability of $W(t,\alpha) : (\mathcal{A}, \|\cdot\|_c) \to L^2_\mu$ in a neighborhood of α_0. Notice that $D_{\bar{\alpha}}[\pi]$ is well defined for every $\bar{\alpha} \in \mathcal{N}(\alpha_0)$ due to Assumption C(i). Next, use $f_{Y|X}(y|x)$ uniformly bounded and Jensen's inequality to obtain the first result in (46). The second inequality holds for $\|\cdot\|_o$ the linear operator norm; see Chapter 6 in (Luenberger, 1969).

$$\|D_{\bar{\alpha}}[\pi]\|^2_{L^2_\mu} \lesssim \int \left(\frac{dm(x,t_u,\bar{\alpha})}{d\alpha}[\pi] \right)^2 f_X(x) dx d\mu(t)$$

$$\leqslant \left|\left| \frac{dm(x,t_u,\bar{\alpha})}{d\alpha} \right|\right|^2_o \|\pi\|^2_c \tag{46}$$

Since Frechet derivative are *a fortiori* continuous, (46) implies $D_{\bar{\alpha}}[\pi]$ is continuous in $\pi \in \mathcal{A}$ for all $\bar{\alpha} \in \mathcal{N}(\alpha_0)$. To examine continuity of $D_{\bar{\alpha}}$ in $\bar{\alpha} \in \mathcal{N}(\alpha_0)$, we use Jensen's inequality to obtain (47) pointwise in t_u.

$$|D_{\bar{\alpha}}[\pi] - D_{\tilde{\alpha}}[\pi]| \leqslant \int \left| f_{Y|X}(m(x,t_u,\bar{\alpha})|x) - f_{Y|X}(m(x,t_u,\tilde{\alpha})|x) \right| \left| \frac{dm(x,t_u,\bar{\alpha})}{d\alpha}[\pi] \right| f_X(x) dx$$
$$+ \int f_{Y|X}(m(x,t_u,\tilde{\alpha})|x) \left| \frac{dm(x,t_u,\bar{\alpha})}{d\alpha}[\pi] - \frac{dm(x,t_u,\tilde{\alpha})}{d\alpha}[\pi] \right| f_X(x) dx \tag{47}$$

In turn, the Lipschitz Assumptions B(iii) and C(ii), $f_{Y|X}(y|x)$ uniformly bounded by Assumption A(iv) and equation (47) yield the following inequality,

$$|D_{\bar{\alpha}}[\pi] - D_{\tilde{\alpha}}[\pi]| \lesssim \|\bar{\alpha} - \tilde{\alpha}\|^\nu_c \int J(x) G^\nu(x) \left| \frac{dm(x,t_u,\bar{\alpha})}{d\alpha}[\pi] \right| f_X(x) dx$$
$$+ \int \left| \frac{dm(x,t_u,\bar{\alpha})}{d\alpha}[\pi] - \frac{dm(x,t_u,\tilde{\alpha})}{d\alpha}[\pi] \right| f_X(x) dx \tag{48}$$

Using (48), Markov's and Jensen's inequality and $E[J^2(X) G^{2\nu}(X)] < \infty$ then implies:

$$\|D_{\bar{\alpha}}[\pi] - D_{\tilde{\alpha}}[\pi]\|^2_{L^2_\mu} \lesssim \|\bar{\alpha} - \tilde{\alpha}\|^{2\nu}_c \int \left(\frac{dm(x,t_u,\bar{\alpha})}{d\alpha}[\pi] \right)^2 f_X(x) dx d\mu(t)$$
$$+ \int \left(\frac{dm(x,t_u,\bar{\alpha})}{d\alpha}[\pi] - \frac{dm(x,t_u,\tilde{\alpha})}{d\alpha}[\pi] \right)^2 f_X(x) dx d\mu(t) \tag{49}$$

Let $\bar{\mathcal{A}}_c$ denote the completion of the linear span of $\mathcal{A}$ under $\|\cdot\|_c$. The definition of $\|\cdot\|_o$ then implies the first equality in (50), while the first inequality follows from (49). Further, since $\|D_{\bar\alpha}\|_o : (\mathcal{N}(\alpha_0), \|\cdot\|_c) \to \mathbb{R}$ is a continuous functional and $\mathcal{A}$ is compact under $\|\cdot\|_c$ it follows that $\sup_{\mathcal{N}(\alpha_0)} \|D_{\bar\alpha}\|_o < \infty$. The second inequality in (50) then follows.

$$\begin{aligned}
\|D_{\bar\alpha} - D_{\tilde\alpha}\|_o^2 &= \sup_{\pi \in \bar{\mathcal{A}}_c} \|\pi\|_c^{-2} \|D_{\bar\alpha}[\pi] - D_{\tilde\alpha}[\pi]\|_{L^2_\mu}^2 \\
&\lesssim \|\bar\alpha - \tilde\alpha\|_c^{2\nu} \left\|\frac{dm(x,t_u,\bar\alpha)}{d\alpha}\right\|_o^2 + \left\|\frac{dm(x,t_u,\bar\alpha)}{d\alpha} - \frac{dm(x,t_u,\tilde\alpha)}{d\alpha}\right\|_o^2 \\
&\lesssim \|\bar\alpha - \tilde\alpha\|_c^{2\nu} + \left\|\frac{dm(x,t_u,\bar\alpha)}{d\alpha} - \frac{dm(x,t_u,\tilde\alpha)}{d\alpha}\right\|_o^2 \qquad (50)
\end{aligned}$$

Since $m(x, t_u, \alpha)$ is continuously Frechet differentiable on $\mathcal{N}(\alpha_0)$, (50) implies $D_{\bar\alpha}$ is continuous in α.

We now show $D_{\bar\alpha}$ is indeed the Frechet derivative of $W(t, \alpha) : (\mathcal{A}, \|\cdot\|_c) \to L^2_\mu$. Straightforward manipulations imply,

$$W(t,\alpha) = \int P(Y \leqslant m(x,t_u,\alpha)|x) 1\{x \leqslant t_x\} f_X(x)dx - t_u P(X \leqslant t_x) \qquad (51)$$

Next, using the definition of $D_{\bar\alpha}$ and (51) together with Jensen's inequality we obtain (52) pointwise in t for any $\bar\alpha \in \mathcal{N}(\alpha_0)$ and $\pi \in \mathcal{A}$.

$$\begin{aligned}
|W(t,\bar\alpha+\pi) - W(t,\bar\alpha) - D_{\bar\alpha}[\pi]| \leqslant 2 \int |P(Y \leqslant m(x,t_u,\bar\alpha+\pi)|x) \\
- P(Y \leqslant m(x,t_u,\bar\alpha)|x) - f_{Y|X}(m(x,t_u,\bar\alpha)|x) \frac{dm(x,t_u,\bar\alpha)}{d\alpha}[\pi]| f_X(x)dx
\end{aligned} \qquad (52)$$

Applying the mean value theorem inside the integral in (52) then implies

$$\begin{aligned}
|W(t,\bar\alpha+g) - W(t,\bar\alpha) - D_{\bar\alpha}[\pi]| \leqslant 2 \int |f_{Y|X}(\bar m(x,t_u)|x)(m(x,t_u,\bar\alpha+\pi) - m(x,t_u,\bar\alpha)) \\
- f_{Y|X}(m(x,t_u,\bar\alpha)|x) \frac{dm(x,t_u,\bar\alpha)}{d\alpha}[\pi]| f_X(x)dx
\end{aligned} \qquad (53)$$

where $\bar m(x, t_u)$ is a convex combination of $m(x, t_u, \bar\alpha+\pi)$ and $m(x, t_u, \bar\alpha)$. Hence, $|\bar m(x, t_u) - m(x, t_u, \bar\alpha)| \leqslant |m(x, t_u, \bar\alpha+\pi) - m(x, t_u, \bar\alpha)|$. The Lipschitz conditions of Assumptions B(iii) and C(ii) imply the inequality:

$$\begin{aligned}
\int |(f_{Y|X}(\bar m(x,t_u)|x) - f_{Y|X}(m(x,t_u,\bar\alpha)|x))(m(x,t_u,\bar\alpha+\pi) - m(x,t_u,\bar\alpha))| f_X(x)dx \\
\leqslant \|\pi\|_c^{1+\nu} \int J(x) G^{1+\nu}(x) f_X(x)dx
\end{aligned} \qquad (54)$$

Using (53), (54) and Jensen's inequality establishes the first inequality in (55). The final result in (55) then follows by $\frac{dm(x,t_u,\bar{\alpha})}{d\alpha}[\pi]$ being the Frechet derivative of $m(x, t_u, \bar{\alpha})$.

$$\|W(t,\bar{\alpha}+\pi) - W(t,\bar{\alpha}) - D_{\bar{\alpha}}[\pi]\|^2_{L^2_\mu} \lesssim \|\pi\|_c^{2+2\nu} + \int \left(m(x,t_u,\bar{\alpha}+\pi) - m(x,t_u,\bar{\alpha}) - \frac{dm(x,t_u,\bar{\alpha})}{d\alpha}[\pi]\right)^2 f_X(x)dxd\mu(t) = o(\|\pi\|_c^2) \tag{55}$$

We conclude from (55) that $D_{\bar{\alpha}}$ is the Frechet derivative of $W(t,\alpha) : (\mathcal{A}, \|\cdot\|_c) \to L^2_\mu$ and that it is continuous in $\bar{\alpha}$. To conclude the proof of the first claim of the Lemma, notice that $Q(\alpha) = \|W(t,\alpha)\|^2_{L^2_\mu}$. Since the functional $\|\cdot\|^2_{L^2_\mu} : L^2_\mu \to \mathbb{R}$ is trivially Frechet differentiable, applying the Chain rule for Frechet derivatives (see for example Theorem 5.2.5 in (Siddiqi, 2004)) yields,

$$\frac{dQ(\bar{\alpha})}{d\alpha}[\pi] = \int W(t,\bar{\alpha})D_{\bar{\alpha}}[\pi]d\mu(t) \tag{56}$$

To establish the second claim of the Lemma, define the bilinear form $T : \mathcal{A} \times \mathcal{A} \to \mathbb{R}$,

$$T[\psi,\pi] = \int D_{\alpha_0}[\psi]D_{\alpha_0}[\pi]d\mu(t) \tag{57}$$

We will show T is the second Frechet derivative of $Q(\alpha)$ at α_0. Notice that $T[\psi,\cdot] : \mathcal{A} \to \mathbb{R}$ is a linear operator. The first requirement of Frechet differentiability is to show $T[\psi,\cdot]$ is continuous in ψ. For this purpose, notice that the first equality in (58) follows by definition while the first and second inequalities are implied by the Cauchy-Schwarz inequality and (46) respectively.

$$\begin{aligned}\|T[\psi,\cdot]\|_o^2 &= \sup_{\pi\in\bar{\mathcal{A}}_c} \|\pi\|^{-2}T^2[\psi,\pi] \\ &\leqslant \int D^2_{\alpha_0}[\psi]d\mu(t) \times \sup_{\pi\in\bar{\mathcal{A}}_c} \int D^2_{\alpha_0}[\pi]d\mu(t) \\ &\lesssim \left\|\frac{dm(x,t_u,\alpha_0)}{d\alpha}\right\|_o^4 \|\psi\|_c^2\end{aligned} \tag{58}$$

It follows from (58) that $T[\psi,\cdot]$ is continuous in $\psi \in \mathcal{A}$. Next, we verify T is the second Frechet derivative of $Q(\alpha)$ at α_0. In (59) use (56) and $W(t,\alpha_0) = 0$ for all t to notice

$\frac{dQ(\alpha_0)}{d\alpha} = 0$ and obtain the equality:

$$\left|\left|\frac{dQ(\alpha_0+\psi)}{d\alpha} - \frac{dQ(\alpha_0)}{d\alpha} - T[\psi,\cdot]\right|\right|_o^2 = \sup_{\pi\in\bar{\mathcal{A}}_c} \|\pi\|^{-2}\left(\int W(t,\alpha_0+\psi)D_{\alpha_0+\psi}[\pi]d\mu(t) - \int D_{\alpha_0}[\psi]D_{\alpha_0}[\pi]d\mu(t)\right)^2 \tag{59}$$

In (60) we control one of the terms in the right hand side of (59). Use the Cauchy-Schwarz inequality to obtain the first inequality in (60) and D_{α_0} being the Frechet derivative of $W(t,\alpha_0) : (\mathcal{A}, \|\cdot\|_c) \to L^2_\mu$ for the second.

$$\begin{aligned}
&\sup_{\pi\in\bar{\mathcal{A}}_c} \|\pi\|^{-2}\Big(\int (W(t,\alpha_0+\psi) - W(t,\alpha_0) - D_{\alpha_0}[\psi])D_{\alpha_0+\psi}[\pi]d\mu(t)\Big)^2 \\
&\qquad \leqslant \|W(t,\alpha_0+\psi) - W(t,\alpha_0) - D_{\alpha_0}[\psi]\|^2_{L^2_\mu} \times \sup_{\pi\in\bar{\mathcal{A}}_c} \|\pi\|^{-2}\|D_{\alpha_0+\psi}[\pi]\|^2_{L^2_\mu} \\
&\qquad \leqslant o(\|\psi\|^2_c) \times \|D_{\alpha_0+\psi}\|^2_o
\end{aligned} \tag{60}$$

Similarly, we use the Cauchy-Schwarz and the definition of $||\cdot||_o$ to obtain,

$$\begin{aligned}
&\sup_{\pi\in\bar{\mathcal{A}}_c} \|\pi\|_c^{-2}\Big(\int D_{\alpha_0}[\psi](D_{\alpha_0+\psi}[\pi] - D_{\alpha_0}[\pi])d\mu(t)\Big)^2 \\
&\qquad \leqslant \|D_{\alpha_0}\|^2_o\|\psi\|^2_c \times \sup_{\pi\in\bar{\mathcal{A}}_c} \|\pi\|_c^{-2}\|D_{\alpha_0+\psi}[\pi] - D_{\alpha_0}[\pi]\|^2_{L^2_\mu} \\
&\qquad \leqslant \|D_{\alpha_0}\|^2_o\|\psi\|^2_c \times ||D_{\alpha_0+\psi} - D_{\alpha_0}||^2_o
\end{aligned} \tag{61}$$

To conclude, combine (59), (60), (61) and $W(t,\alpha_0) = 0$ for all t to derive the first inequality in (62). As argued in (50), however, $\|D_{\bar{\alpha}}\|_o$ is bounded in a neighborhood of α_0. Thus, the continuity of $D_{\bar{\alpha}}$ in $\bar{\alpha}$ for $\bar{\alpha} \in \mathcal{N}(\alpha_0)$ implies the final result in (62).

$$\begin{aligned}
&\left|\left|\frac{dQ(\alpha_0+\psi)}{d\alpha} - \frac{dQ(\alpha_0)}{d\alpha} - T[\psi,\cdot]\right|\right|_o^2 \\
&\qquad \leqslant o(\|\psi\|^2_c) \times \|D_{\alpha_0+\psi}\|^2_o + \|\psi\|^2_c\|D_{\alpha_0}\|^2_o\,||D_{\alpha_0+\psi} - D_{\alpha_0}||^2_o = o(\|\psi\|^2_c)
\end{aligned} \tag{62}$$

It follows T is the second Frechet derivative of $Q(\alpha)$ at α_0. ■

Proof of Theorem 3: Let $\Pi_n\alpha_0 = \arg\min_{\mathcal{A}_n}$. By Theorem 1, $\hat{\alpha} \in \mathcal{N}(\alpha_0)$ with probability tending to one. Therefore, Assumptions D(i) and (iii), imply that with probability tending to one we have,

$$\|\hat{\alpha}-\alpha_0\|_w^2 \lesssim Q(\hat{\alpha}) - Q(\Pi_n\alpha_0) + Q(\Pi_n\alpha_0)$$
$$= Q(\hat{\alpha}) - Q(\Pi_n\alpha_0) + o(n^{-1}) \tag{63}$$

Let $\delta_n \to 0$ sufficiently slow such that $P(\|\hat{\alpha}-\alpha_0\|_s > \delta_n) \to 0$, which is possible due to Theorem 1 and $\|\cdot\|_s \lesssim \|\cdot\|_c$. Letting $\mathcal{A}_0^{\delta_n} = \{\alpha \in \mathcal{A} : \|\alpha - \alpha_0\|_s \leqslant \delta_n\}$ then yields the first inequality in (64). Noticing that $Q_n(\hat{\alpha}) \leqslant Q_n(\Pi_n\alpha_0)$ by virtue of $\hat{\alpha}$ minimizing $Q_n(\alpha)$ over $\mathcal{A}_n$ and using the Cauchy-Schwarz inequality gives us the second inequality. For the third and fourth inequalities we use Lemma 6 which implies $\sqrt{n}(W_n(t,\alpha) - W(t,\alpha))$ is tight in $L^\infty(\mathbb{R}^{d_t} \times \mathcal{A})$ together with the definition of $Q(\alpha)$.

$$\begin{aligned}
Q(\hat{\alpha}) - Q(\Pi_n\alpha_0) &\leqslant Q_n(\hat{\alpha}) - Q_n(\Pi_n\alpha_0) + 2\sup_{\mathcal{A}_0^{\delta_n}} |Q_n(\alpha) - Q(\alpha)| \\
&\leqslant 2 \sup_{(t,\alpha)\in\mathbb{R}^{d_t}\times\mathcal{A}} |W_n(t,\alpha) - W(t,\alpha)| \times [\sup_{\mathcal{A}_0^{\delta_n}} \int (W_n(t,\alpha) + W(t,\alpha))^2 d\mu(t)]^{\frac{1}{2}} \\
&\leqslant O_p(n^{-\frac{1}{2}}) \times [\sup_{(t,\alpha)\in\mathbf{R}^{d_t}\times\mathcal{A}} (W_n(t,\alpha) - W(t,\alpha))^2 + \sup_{\mathcal{A}_0^{\delta_n}} 4 \int W^2(t,\alpha) d\mu(t)]^{\frac{1}{2}} \\
&\leqslant O_p(n^{-\frac{1}{2}}) \times [O_p(n^{-1}) + \sup_{\mathcal{A}_0^{\delta_n}} 4Q(\alpha)]^{\frac{1}{2}}
\end{aligned} \tag{64}$$

By Assumption D(i), $\sup_{\mathcal{A}_0^{\delta_n}} Q(\alpha) \lesssim \delta_n^2 = o(1)$. Therefore, combining (63) and (64):

$$\|\hat{\alpha}-\alpha_0\|_w^2 \lesssim O_p(n^{-\frac{1}{2}}) \times o_p(1) + o(n^{-1}) = o_p(n^{-\frac{1}{2}}) \tag{65}$$

To obtain a rate with respect to $\|\cdot\|_s$, we use Assumption D(iii) for the first and second inequalities in (66). Further, it follows from (65) and Assumption D(iii) that $\|\hat{\alpha}-\Pi_n\alpha_0\|_w^2 = o_p(n^{-\frac{1}{2}})$ which together with Assumption D(ii) implies the equality in (66).

$$\|\hat{\alpha}-\alpha_0\|_s^2 \leqslant \|\hat{\alpha}-\Pi_n\alpha_0\|_s^2 + o(n^{-1}) \leqslant \sup_{\alpha\in\mathcal{A}_n} \frac{\|\alpha\|_s^2}{\|\alpha\|_w^2} \times \|\hat{\alpha}-\Pi_n\alpha_0\|_w^2 + o(n^{-1}) = o_p(n^{-\frac{1}{2}+\gamma}) \tag{66}$$

We can now exploit the local behavior of the objective function to improve on the obtained rate of convergence. Notice that due to (66) it is possible to choose $\delta_n = o(n^{-\frac{1}{4}+\frac{\gamma}{2}})$ such that $P(\hat{\alpha} \in \mathcal{A}_0^{\delta_n}) \to 1$. Repeating the steps in (64) we obtain (67) with probability approaching one.

$$Q(\hat{\alpha}) - Q(\Pi_n\alpha_0) \leqslant O_p(n^{-\frac{1}{2}}) \times [O_p(n^{-1}) + \sup_{\mathcal{A}_0^{\delta_n}} 4Q(\alpha)]^{\frac{1}{2}}$$

$$= O_p(n^{-\frac{1}{2}}) \times o_p(n^{-\frac{1}{4}+\frac{\gamma}{2}}) \tag{67}$$

¿From (63) and (67) and Assumption D(ii), we then obtain $\|\hat{\alpha}-\alpha_0\|_w^2 = o_p(n^{-\frac{1}{2}-\frac{1}{4}+\frac{\gamma}{2}})$ and similarly that $\|\hat{\alpha} - \Pi_n\alpha_0\|_w^2 = o_p(n^{-\frac{1}{2}-\frac{1}{4}+\frac{\gamma}{2}})$. In turn, by repeating the argument in (66) we obtain the improved rate $\|\hat{\alpha} - \alpha_0\|_s^2 = o_p(n^{(\gamma-\frac{1}{2})(1+\frac{1}{2})})$. Proceeding in this fashion we get $\|\hat{\alpha}-\alpha_0\|_s^2 = o_p(n^{(\gamma-\frac{1}{2})(1+\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\ldots)})$. Since $\gamma - 1/2 < -1/4$, repeating this argument a possibly large, but finite, number of times yields the desired conclusion $\|\hat{\alpha}-\alpha_0\|_s^2 = o_p(n^{-\frac{1}{2}})$ thus establishing the claim of the Theorem. ■

Appendix D. Proofs for Section 5

Because the criterion function $Q_n(\alpha)$ is not smooth in α, it is convenient to define the alternate criterion:

$$Q_n^s(\alpha) = \int (W_n(t,\alpha_0) + W(t,\alpha))^2 d\mu(t) \tag{68}$$

Throughout the proofs we will exploit the following Lemma,

Lemma 7. *If Assumptions A, B, C and D hold, then* $Q_n^s(\hat{\alpha}) \leqslant \inf_{\mathcal{A}_n} Q_n^s(\alpha) + o_p(n^{-1})$.

Proof: Since $\|\hat{\alpha} - \alpha_0\|_c = o_p(1)$, Lemma 6 implies

$$\sup_t |W_n(t,\hat{\alpha}) - W(t,\hat{\alpha}) - W_n(t,\alpha_0)| = o_p(n^{-\frac{1}{2}}) \ .$$

By simple manipulations we therefore obtain:

$$Q_n^s(\hat{\alpha}) \leqslant \int \left(|W_n(t,\alpha_0) + W(t,\hat{\alpha}) - W_n(t,\hat{\alpha})| + |W_n(t,\hat{\alpha})|\right)^2 d\mu(t)$$

$$= \int W_n^2(t,\hat{\alpha})d\mu(t) + o_p(n^{-\frac{1}{2}}) \times \int |W_n(t,\hat{\alpha})|d\mu(t) + o_p(n^{-1}) \tag{69}$$

Next, apply Jensen's inequality and $Q_n(\hat{\alpha}) \leqslant Q_n(\Pi_n\alpha_0)$ to obtain the first and second inequalities in (70). By Lemma 6, $\sup_{t,\alpha} |W_n(t,\alpha) - W(t,\alpha)| = O_p(n^{-\frac{1}{2}})$. Together with

Assumption D(i), the final two inequalities in (70) then immediately follow.

$$\begin{aligned}\int |W_n(t,\hat{\alpha})|d\mu(t) &\leqslant \left[\int W_n^2(t,\hat{\alpha})d\mu(t)\right]^{\frac{1}{2}} \\ &\leqslant \left[\int W_n^2(t,\Pi_n\alpha_0)d\mu(t)\right]^{\frac{1}{2}} \\ &\leqslant \left[2\int (W_n(t,\Pi_n\alpha_0)-W(t,\Pi_n\alpha_0))^2 d\mu(t) + 2\int W^2(t,\Pi_n\alpha_0)d\mu(t)\right]^{\frac{1}{2}} \\ &\lesssim \left[O_p(n^{-1}) + \|\Pi_n\alpha_0-\alpha_0\|_s^2\right]^{\frac{1}{2}} \qquad (70)\end{aligned}$$

By Assumption D(iii), $\|\Pi_n\alpha_0-\alpha_0\|_s = o(n^{-\frac{1}{2}})$ and therefore combining (69) and (70),

$$Q_n^s(\hat{\alpha}) \leqslant Q_n(\hat{\alpha}) + o_p(n^{-1}) \qquad (71)$$

Let $\tilde{\alpha} \in \arg\min_{\mathcal{A}_n} Q_n^s(\alpha)$, and notice that Lemma 6 and the same arguments as in Theorem 1 imply that $\|\alpha_0-\tilde{\alpha}\|_c = o_p(1)$. The same arguments as in (69) then imply (72).

$$\begin{aligned}Q_n(\tilde{\alpha}) &\leqslant \int \left(|W_n(t,\tilde{\alpha})-W_n(t,\alpha_0)-W(t,\tilde{\alpha})| + |W_n(t,\alpha_0)+W(t,\tilde{\alpha})|\right)^2 d\mu(t) \\ &= \int \left(W_n(t,\alpha_0)+W(t,\tilde{\alpha})\right)^2 d\mu(t) + o_p(n^{-\frac{1}{2}})\int |W_n(t,\alpha_0)+W(t,\tilde{\alpha})|d\mu(t) + o_p(n^{-1}) \qquad (72)\end{aligned}$$

Proceeding as in (70), Jensen's inequality and $Q_n^s(\tilde{\alpha}) \leqslant Q_n^s(\Pi_n\alpha_0)$ imply the first and second inequalities in (73). The last two results in (73) then follow by Assumption D(i) and by noting that Lemma 6 implies $\sup_t |W_n(t,\alpha_0)| = O_p(n^{-\frac{1}{2}})$.

$$\begin{aligned}\int |W_n(t,\alpha_0)+W(t,\tilde{\alpha})|d\mu(t) &\leqslant \left[\int \left(W_n(t,\alpha_0)+W(t,\tilde{\alpha})\right)^2 d\mu(t)\right]^{\frac{1}{2}} \\ &\leqslant \left[\int \left(W_n(t,\alpha_0)+W(t,\Pi_n\alpha_0)\right)^2 d\mu(t)\right]^{\frac{1}{2}} \\ &\leqslant \left[2\int W_n^2(t,\alpha_0)d\mu(t) + 2\int W^2(t,\Pi_n\alpha_0)d\mu(t)\right]^{\frac{1}{2}} \\ &\lesssim \left[O_p(n^{-1}) + \|\alpha_0-\Pi_n\alpha_0\|_s^2\|\right]^{\frac{1}{2}} \qquad (73)\end{aligned}$$

Since $\|\Pi_n\alpha_0-\alpha_0\|_s = o(n^{-\frac{1}{2}})$ by Assumption D(iii), (72) and (73) imply,

$$Q_n(\tilde{\alpha}) \leqslant Q_n^s(\tilde{\alpha}) + o_p(n^{-1}) \qquad (74)$$

Hence, since $Q_n(\hat{\alpha}) \leqslant Q_n(\tilde{\alpha})$, the definition of $\tilde{\alpha}$ together with (71) and (74) establish

$$Q_n^s(\hat{\alpha}) \leqslant Q_n(\tilde{\alpha}) + o_p(n^{-1}) \leqslant \inf_{\mathcal{A}_n} Q_n^s(\alpha) + o_p(n^{-1})$$

which establishes the claim of the Lemma. ■

Proof of Lemma 4: The arguments in this Lemma follow those of Ai and Chen (2003). We first establish continuity of F_λ. Since F_λ is linear, it is only necessary to establish that it is bounded. For any $\theta \in \mathbb{R}^{d_\theta}$, we can obtain the first equality in (75) by using (16), while the second equality is definitional.

$$\min_{w\in\bar{\mathcal{H}}} \int \left(\frac{dW(t,\alpha_0)}{d\theta'}[\theta] - \frac{dW(t,\alpha_0)}{dh}[h]\right)^2 d\mu(t) = \int \left(\left[\frac{dW(t,\alpha_0)}{d\theta} - \frac{dW(t,\alpha_0)}{dh}[h^*]\right]' \theta\right)^2 d\mu(t) = \theta'\Sigma^*\theta \tag{75}$$

In order to show F_λ is bounded we need to establish the left hand side of (76) is finite. Using (75) immediately implies the first equality in (76). For the second equality notice the optimization problem is solved at $\theta^* = (\Sigma^*)^{-1}\lambda$ and plug in θ^*.

$$\sup_{0\neq\alpha\in\bar{\mathcal{A}}} \frac{F_\lambda^2(\alpha)}{\|\alpha\|_w^2} = \sup_{0\neq\theta\in\mathbb{R}^{d_\theta}} \frac{(\lambda'\theta)^2}{\theta'\Sigma^*\theta} = \lambda'(\Sigma^*)^{-1}\lambda \tag{76}$$

Since by assumption Σ^* is positive-definite, (76) is finite and hence F_λ is bounded which establishes continuity. For the second claim of the Lemma, notice the following orthogonality condition must hold as a result of (16),

$$\int \left(\frac{dW(t,\alpha_0)}{d\theta} - \frac{dW(t,\alpha_0)}{dh}[h^*]\right) \frac{dW(t,\alpha_0)}{dh}[h] d\mu(t) = 0 \tag{77}$$

for all $h \in \bar{\mathcal{H}}$. Hence, the first equality in (78) is definitional, while the second one is implied by (77). Plugging in the definition of v_θ^λ establishes the third inequality in (78).

$$\begin{aligned}
\langle \alpha - \alpha_0, v^\lambda\rangle_w &= \int \left[\frac{dW(t,\alpha_0)}{d\theta}[\theta-\theta_0] + \frac{dW(t,\alpha_0)}{dh}[h-h_0]\right]\left[\frac{dW(t,\alpha_0)}{d\theta}[v^\lambda] + \frac{dW(t,\alpha_0)}{dh}[v_h^\lambda]\right] d\mu(t) \\
&= (\theta-\theta_0)'\left\{\int \left[\frac{dW(t,\alpha_0)}{d\theta} - \frac{dW(t,\alpha_0)}{dh}[h^*]\right]\left[\frac{dW(t,\alpha_0)}{d\theta} - \frac{dW(t,\alpha_0)}{dh}[h^*]\right]' d\mu(t)\right\} v_\theta^\lambda \\
&= (\theta-\theta_0)'\lambda
\end{aligned} \tag{78}$$

which verifies the second claim of the Lemma. ■

Lemma 8. *Let Assumption A, B, C, D and E hold, and let* $v_n^\lambda = \Pi_n v^\lambda$. *Then:*

(i) $\int W_n(t,\alpha_0) D_{\hat{\alpha}}[v_n^\lambda] d\mu(t) = \int W_n(t,\alpha_0) D_{\alpha_0}[v^\lambda] d\mu(t) + o_p(n^{-\frac{1}{2}})$

(ii) $\int (W(t,\hat{\alpha}) - W(t,\alpha_0)) D_{\hat{\alpha}}[v_n^\lambda] d\mu(t) = \int D_{\alpha_0}[\hat{\alpha}-\alpha_0] D_{\alpha_0}[v^\lambda] d\mu(t) + o_p(n^{-\frac{1}{2}})$

(iii) $\sqrt{n}W_n(t,\alpha_0) \xrightarrow{\mathcal{L}} G(t)$ *where* $G(t)$ *a Gaussian process with covariance:*

$$\Sigma(t,t') = E[(1\{U \leqslant t_u; X \leqslant t_x\} - t_u 1\{X \leqslant t_x\})(1\{U \leqslant t'_u; X \leqslant t'_x\} - t'_u 1\{X \leqslant t'_x\})]$$

PROOF: To establish the first claim apply the Cauchy-Schwarz inequality, the definition of the operator norm and Lemma 6 implying $\sup_t |W_n(t,\alpha_0)| = O_p(n^{-\frac{1}{2}})$ to obtain the inequalities in (79).

$$\left|\int W_n(t,\alpha_0) D_{\hat{\alpha}}[v_n^\lambda - v^\lambda] d\mu(t)\right| \leqslant \left[\int W_n^2(t,\alpha_0) d\mu(t)\right]^{\frac{1}{2}} \times \|D_{\hat{\alpha}}[v_n^\lambda - v^\lambda]\|_{L^2_\mu}$$
$$\leqslant O_p(n^{-\frac{1}{2}}) \times \sup_{\alpha \in \mathcal{N}(\alpha_0)} \|D_\alpha\|_o \times \|v_n^\lambda - v^\lambda\|_c \tag{79}$$

As argued in (50), $\sup_{\alpha\in\mathcal{N}(\alpha_0)} \|D_\alpha\|_o < \infty$. Further, Assumption E(ii) and B(v) imply $\|v^\lambda - v_n^\lambda\|_c = o(1)$. Therefore, we obtain from (79) that,

$$\int W_n(t,\alpha_0) D_{\hat{\alpha}}[v_n^\lambda] d\mu(t) = \int W_n(t,\alpha_0) D_{\hat{\alpha}}[v^\lambda] d\mu(t) + o_p(n^{-\frac{1}{2}}) \tag{80}$$

Similarly, the derivations in (79) imply the inequality in (81). The equality is a result of the continuity of D_α in α under $\|\cdot\|_c$, as established in the proof of Lemma 3.

$$\left|\int W_n(t,\alpha_0)(D_{\hat{\alpha}}[v^\lambda] - D_{\alpha_0}[v^\lambda]) d\mu(t)\right| \leqslant O_p(n^{-\frac{1}{2}}) \times \|D_{\hat{\alpha}} - D_{\alpha_0}\|_o \times \|v^\lambda\|_c = o_p(n^{-\frac{1}{2}}) \tag{81}$$

Together, (80) and (81) establish the first claim of the Lemma.

For the second claim of the Lemma, notice that Assumption E(iii) allows us to do a second order Taylor expansion to obtain (82),

$$W(t,\hat{\alpha}) = W(t,\alpha_0) + D_{\alpha_0}[\hat{\alpha} - \alpha_0] + \frac{1}{2}\frac{dD_{\alpha_0+\tau(\hat{\alpha}-\alpha_0)}[\hat{\alpha}-\alpha_0]}{d\tau}\Big|_{\tau=s(t)} \tag{82}$$

The first equality in (83) then follows from (82), while the second one is implied by Assumptions E(iii) and E(iv). The final equality is in turn implied by Theorem 3.

$$\int (W(t,\hat{\alpha}) - W(t,\alpha_0) - D_{\alpha_0}[\hat{\alpha}-\alpha_0]) D_{\hat{\alpha}}[v_n^\lambda]$$
$$= \frac{1}{2}\int \frac{dD_{\alpha_0+\tau(\hat{\alpha}-\alpha_0)}[\hat{\alpha}-\alpha_0]}{d\tau}\Big|_{\tau=s(t)} D_{\hat{\alpha}}[v_n^\lambda] d\mu(t) \lesssim \|\hat{\alpha}-\alpha_0\|_s^2 = o_p(n^{-\frac{1}{2}}) \tag{83}$$

Next, apply the Cauchy-Schwarz inequality and a Taylor expansion to obtain the first inequality in (84). The second inequality then follows by Assumption E(iii), $\|\hat{\alpha} - \alpha_0\|_w \lesssim$

$\|\hat{\alpha}-\alpha_0\|_s$ in a neighborhood of α_0, as implied by Assumption D(i), and Theorem 3.

$$(84)\quad \left|\int D_{\alpha_0}[\hat{\alpha}-\alpha_0](D_{\hat{\alpha}}[v_n^\lambda]-D_{\alpha_0}[v_n^\lambda])d\mu(t)\right|$$
$$\leqslant \|\hat{\alpha}-\alpha_0\|_w\left[\int\left(\frac{dD_{\alpha_0+\tau(\hat{\alpha}-\alpha_0)}[v_n^\lambda]}{d\tau}\Big|_{\tau=s(t)}\right)^2 d\mu(t)\right]^{\frac{1}{2}} \lesssim \|\hat{\alpha}-\alpha_0\|_s^2 = o_p(n^{-\frac{1}{2}})$$

Similarly, applying the Cauchy-Schwarz inequality, $\|\hat{\alpha}-\alpha_0\|_w = o_p(n^{-\frac{1}{4}})$ and similarly $\|v_n^\lambda - v^\lambda\|_c = o(n^{-\frac{1}{4}})$ by Assumption D(ii) we derive,

$$(85)\quad \left|\int D_{\alpha_0}[\hat{\alpha}-\alpha_0](D_{\alpha_0}[v_n^\lambda]-D_{\alpha_0}[v^\lambda])d\mu(t)\right| \leqslant \|\hat{\alpha}-\alpha_0\|_w\times\|D_{\alpha_0}\|_o\|v_n^\lambda-v^\lambda\|_c = o_p(n^{-\frac{1}{2}})$$

Combining results (83), (84) and (85) establishes the second claim of the Lemma. The third claim of the Lemma is immediate from $W_n(t,\alpha_0)$ being a Donsker class due to Lemma 6 and regular central limit theorem. Hence all claims of the Lemma are shown. ■

Proof of Theorem 4: Let $u^* = \pm v^\lambda$, $u_n^* = \Pi_n u^*$ and $0 < \epsilon_n = o(n^{-\frac{1}{2}})$ be such that $Q_n^s(\hat{\alpha}) \leqslant \inf_{\mathcal{A}_n} Q_n^s(\alpha) + O_p(\epsilon_n^2)$, which is possible due to Lemma 7. Define $\alpha(\tau) = \hat{\alpha}+\tau\epsilon_n u_n^*$ and note that by Assumption C(i) and Lemma 1, with probability tending to one $\alpha(l)\in\mathcal{A}_n$ for $l\in[0,1]$. Therefore, Lemma 7 establishes the first equality in (86). A second order Taylor expansion around $\tau = 0$ yields the equality in (86) for some $s\in[0,1]$.

$$\begin{aligned} & 0 \leqslant Q_n^s(\alpha(1)) - Q_n^s(\alpha(0)) + O_p(\epsilon_n^2) \\ (86)\qquad & = 2\epsilon_n\int (W_n(t,\alpha_0)+W(t,\hat{\alpha}))D_{\hat{\alpha}}[u_n^*]d\mu(t) + \frac{1}{2}\frac{d^2Q_n(\alpha(l))}{d\alpha^2}\Big|_{l=s}\end{aligned}$$

where by direct calculation we have that:

$$\begin{aligned} (87)\quad & \frac{d^2Q_n(\alpha(s))}{d\alpha^2}\Big|_{l=s} \\ & = \epsilon_n^2\int\left(D_{\alpha(s)}[u_n^*]\right)^2 d\mu(t) + \int (W_n(t,\alpha_0)+W(t,\alpha(s)))\frac{dD_{\hat{\alpha}+\tau\epsilon_n u_n^*}[\epsilon_n u_n^*]}{d\tau}\Big|_{\tau=s} d\mu(t)\end{aligned}$$

As shown in (50), $\sup_\alpha \|D_\alpha\|_o < \infty$, and hence, since $\|v^\lambda\|_c<\infty$ we obtain,

$$(88)\qquad \int\left(D_{\alpha(s)}[u_n^*]\right)^2 d\mu(t) \leqslant \sup_{\alpha\in\mathcal{N}(\alpha_0)}\|D_\alpha\|_o\times\|u_n^*\|_c^2 = O(1)$$

Since $W_n(t,\alpha)$ and $W(t,\alpha)$ are both bounded by 1, Assumption E(iii) establishes:

$$(89)\qquad \left|\int (W_n(t,\alpha_0)+W(t,\alpha(s)))\frac{dD_{\hat{\alpha}+\tau\epsilon_n u_n^*}[\epsilon_n u_n^*]}{d\tau}\Big|_{\tau=s} d\mu(t)\right| \lesssim \|\epsilon_n u_n^*\|_s^2 = O(\epsilon_n^2)$$

Therefore, by combining (86)-(89), $u_n^* = \pm v_n^\lambda$ and $\epsilon_n = o(n^{-\frac{1}{2}})$, it follows that:

$$\int (W_n(t,\alpha_0) + W(t,\hat{\alpha})) D_{\hat{\alpha}}[u_n^*] d\mu(t) = o_p(n^{-\frac{1}{2}}) \tag{90}$$

To conclude, in (91) use Lemma 4 for the first equality, Lemma 8(ii) for the second equality, $W(t,\alpha_0) = 0$ and (90) for the third one and Lemma 8(i) for the final result.

$$\begin{aligned}
\sqrt{n}\lambda'(\hat{\theta} - \theta_0) &= \sqrt{n} \int D_{\alpha_0}[\hat{\alpha} - \alpha_0] D_{\alpha_0}[v^\lambda] \\
&= \sqrt{n} \int (W(t,\hat{\alpha}) - W(t,\alpha_0)) D_{\hat{\alpha}}[v_n^\lambda] d\mu(t) + o_p(1) \\
&= \sqrt{n} \int W_n(t,\alpha_0) D_{\hat{\alpha}}[v_n^\lambda] d\mu(t) + o_p(1) \\
&= \sqrt{n} \int W_n(t,\alpha_0) D_{\alpha_0}[v^\lambda] d\mu(t) + o_p(1) \qquad (91)
\end{aligned}$$

Hence, applying Lemma 8(iii) we obtain from (91) that,

$$\sqrt{n}\lambda'(\hat{\theta} - \theta) \xrightarrow{\mathcal{L}} N(0, \Omega_\lambda) \tag{92}$$

where $\Omega_\lambda = \int D_{\alpha_0}[v^\lambda](t) D_{\alpha_0}[v^\lambda](s) \Sigma(t,s) d\mu(t) d\mu(s)$. Using the closed form for v^λ, obtained in Lemma 4, and the definition of $R_{h^*}(t)$ yields the results in (93).

$$\begin{aligned}
D_{\alpha_0}[v^\lambda] &= \left[\frac{dW(t,\alpha_0)}{d\theta} - \frac{dW(t,\alpha_0)}{dh}[h^*] \right] [\Sigma^*]^{-1} \lambda \\
&= R_{h^*}(t) [\Sigma^*]^{-1} \lambda \qquad (93)
\end{aligned}$$

The Cramer-Wald device, (92) and (93) in turn establish the claim of the Theorem. ■

References

Ai, C., and X. Chen (2003): "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71, 1795–1844.

Akritas, M., and I. van Keilegom (2001): "Non-Parametric Estimation of the Residual Distribution," *Scandinavian Journal of Statistics*, 28, 549–567.

Benkard, C. L., and S. Berry (2007): "On the Nonparametric Identification of Non-Linear Simultaneous Equations Models: Comment on Brown (1983) and Rhoerig (1988)," *Econometrica*, forthcoming.

Berry, S. T., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63(4), 841–890.

Brown, B. W. (1983): "The Identification Problem in Systems Nonlinear in the Variables," *Econometrica*, 51, 175–196.

Brown, D. J., and M. H. Wegkamp (2002): "Weighted Minimum Mean-Square Distance from Independence Estimation," *Econometrica*, 70, 2035–2051.

Chen, X. (2006): "Large Sample Sieve Estimation of Semi-Nonparametric Models," Working paper, Yale University.

Chen, X., and D. Pouzo (2008): "Estimation of Nonparametric Conditional Moment Models with Possibly Nonsmooth Moments," Working paper, Yale University.

Chesher, A. (2003): "Identification in Nonseparable Models," *Econometrica*, 71, 1405–1441.

Hausman, J. A., and D. A. Wise (1978): "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences," *Econometrica*, 46, 403–426.

Luenberger, D. G. (1969): *Optimization by Vector Space Methods*. Wiley, New York.

Manski, C. F. (1983): "Closest Empirical Distribution Estimation," *Econometrica*, 51, 305–319.

Matzkin, R. L. (1994): "Restrictions of Economic Theory in Nonparametric Methods," in *Handbook of Econometrics Volume*, ed. by R. F. Engle, and D. L. McFadden, pp. 2524–2559. North-Holland.

——— (2003): "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, 71, 1339–1375.

Milgrom, P., and J. Roberts (1990): "Rationalizability, Learning and Equilibrium in Games with Strategic Complementarities," *Econometrica*, 58(6), 1255–1277.

Newey, W. K., and J. Powell (2003): "Instrumental Variables Estimation of Nonparametric Models," *Econometrica*, 71, 1565–1578.

Roehrig, C. S. (1988): "Conditions for Identification in Nonparametric and Parametric Models," *Econometrica*, 56, 433–447.

Rudin, W. (1976): *Principles of Mathematical Analysis*. McGraw-Hill.

Siddiqi, A. H. (2004): *Applied Functional Analysis*. Marcel Dekker, Inc., New York.

van der Vaart, A. W., and J. A. Wellner (1996): *Weak Convergence and Empirical Processes: with Applications to Statistics*. Springer, New York.