

# 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  $Y = m_\alpha(X, U)$  with  $U$  uniform on  $(0, 1)$  in which  $m_\alpha$  is a known real function parameterized by a structural parameter  $\alpha$ . We study the case in which  $\alpha$  contains a finite dimensional component  $\theta$  and an infinite dimensional component  $h$ . We assume that the true value  $\alpha_0$  is identified by the restriction  $U \perp X$ . Our proposal is to estimate  $\alpha_0$  by a minimum distance from independence (MDI) criterion. We show that: (i) our estimator for  $h_0$  is consistent and we obtain rates of convergence; (ii) the estimator for  $\theta_0$  is  $\sqrt{n}$  consistent and asymptotically normally distributed.

## 1. INTRODUCTION

Nonparametric identification of nonlinear nonseparable 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

---

*Date:* January 10, 2010.

*Affiliations and contact information.* Komunjer: Department of Economics, University of California at San Diego, [komunj@ucsd.edu](mailto:komunj@ucsd.edu). Santos: Department of Economics, University of California at San Diego, [a2santos@ucsd.edu](mailto:a2santos@ucsd.edu).

*Acknowledgments.* We would like to thank the Editor, Jean-Marc Robin, and two anonymous referees for their suggestions which improved an earlier version of the paper. This paper was presented at the 2008 EC<sup>2</sup> Meeting "Recent Advances in Structural Microeconometrics" in Rome, Italy. Many thanks to the participants of the conference for their comments. All errors are ours.

variables that comes in a form of conditional moment restrictions (see, e.g., Ai and Chen, 2003; Blundell, Chen, and Kristensen, 2007). Weaker than independence, the mean independence property by itself does not guarantee the identification to hold. As a result, this literature most often simply assumes the models to be identified by the conditional moment restrictions.

In the present paper we unify the estimation and identification of nonseparable models by employing the same criterion to obtain both: full independence between the models' latent and exogenous variables. We focus on the models of the form:  $Y = m_\alpha(X, U)$ , with variables  $Y \in \mathbb{R}$  and  $X \in \mathcal{X} \subseteq \mathbb{R}^{d_x}$  that are observable, and a latent disturbance  $U$  that is uniformly distributed on  $(0, 1)$ .<sup>1</sup> We denote by  $\alpha_0$  the true value of the structural parameter  $\alpha$  which consists of (i) a component  $\theta$  in  $\Theta$  that is finite dimensional ( $\Theta \subset \mathbb{R}^{d_\theta}$ ), and (ii) a function  $h$  of  $x$  and  $u$  belonging to an infinite dimensional set of functions  $\mathcal{H}$ . Thus  $\alpha \equiv (\theta, h) \in \mathcal{A} \equiv \Theta \times \mathcal{H}$ . We focus on nonseparable models in which for every value  $x \in \mathcal{X}$  the mapping  $m_\alpha(x, u)$  is strictly increasing in  $u$  on  $(0, 1)$ , and the true value  $\alpha_0$  of  $\alpha$  is identified by the independence restriction  $U \perp X$ .

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

$$(1) \quad P(Y \leq m_{\alpha_0}(X, t_u); X \leq t_x) = t_u \cdot P(X \leq t_x)$$

for all  $t \equiv (t_x, t_u) \in \mathcal{X} \times (0, 1)$ . We exploit this relationship between the marginal and joint cdfs to construct a Cramér-von Mises type criterion function:

$$Q(\alpha) \equiv \int_{\mathcal{X} \times (0, 1)} \left[ P(Y \leq m_\alpha(X, t_u); X \leq t_x) - t_u \cdot P(X \leq t_x) \right]^2 d\mu(t) ,$$

where  $\mu$  a measure on  $\mathcal{X} \times (0, 1)$ . 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  $Q_n(\alpha)$

---

<sup>1</sup>In a semiparametric specification, requiring  $U \sim U(0, 1)$  can often be seen as a normalization on the nonparametric component that does not affect the parametric one. This assumption is also often used for nonparametric identification (see, Matzkin, 2003, for examples of such arguments).

of  $Q(\alpha)$  above—a minimum distance from independence (MDI) estimator. When  $\alpha_0$  is identified by the assumptions of the model, then  $\alpha_0$  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  $\theta_0$ .

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 components, which our approach does. Our paper is also related to the vast literature on estimation of conditional quantiles. Horowitz and Lee (2007) and Chen and Pouzo (2008a), for example, study semiparametric and nonparametric estimation, respectively, in an instrumental variables setting. However, these results concern a finite number of quantile restrictions, while (1) constitutes a continuum of them. Carrasco and Florens (2000) examine efficient GMM estimation under a continuum of restrictions, but their results apply only to finite dimensional parameters. Additional work in nonseparable models concerns identification and estimation of average treatment effects rather than the entire structural parameter as in Altonji and Matzkin (2005), Chernozhukov and Hansen (2005), Florens, Heckman, Meghir, and Vytlacil (2008) and Imbens and Newey (2009) among others.

The remainder of the paper is organized as follows. In Section 2 we present the estimator and establish its consistency while in Section 3 we obtain a rate of convergence. The asymptotic normality result for  $\sqrt{n}(\hat{\theta} - \theta_0)$  is derived in Section 4. In Section 5 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. The same section contains a Monte Carlo experiment that illustrates the properties of our estimator.

Section 6 concludes the paper. The proofs of all the results stated in the text are relegated to an Appendix.

## 2. MINIMUM DISTANCE FROM INDEPENDENCE ESTIMATION

We consider the following nonseparable model:

$$(2) \quad Y = m_\alpha(X, U) \quad \text{and} \quad U \sim U(0, 1)$$

with observables  $Y \in \mathbb{R}$  and  $X \in \mathcal{X} \subseteq \mathbb{R}^{d_x}$ , unobservable  $U \in (0, 1)$ , and structural parameter  $\alpha \in \mathcal{A}$ . In our setup  $\alpha$  consists of an unknown parameter  $\theta \in \Theta$  that is finite dimensional ( $\Theta \subseteq \mathbb{R}^{d_\theta}$ ), as well as an unknown real function  $h : \mathcal{X} \times (0, 1) \rightarrow \mathbb{R}$ . The latter component of  $\alpha$  is infinite dimensional and we assume that  $h \in \mathcal{H}$  where  $\mathcal{H}$  is an infinite dimensional set of real valued functions of  $x$  and  $u$ . We therefore let  $(\theta, h) \equiv \alpha \in \mathcal{A} \equiv \Theta \times \mathcal{H}$ . Hereafter, we assume that the model (2) is correctly specified and we denote by  $\alpha_0$  the true value of the parameter  $\alpha$ .

For every  $\alpha \in \mathcal{A}$ , the structural mapping  $m_\alpha : \mathcal{X} \times (0, 1) \rightarrow \mathbb{R}$  in (2) is a known real function that is continuously differentiable in  $u$  on  $(0, 1)$  for every  $x \in \mathcal{X}$ . Moreover, we assume that for every  $x \in \mathcal{X}$ , we have  $\partial m_{\alpha_0}(x, u)/\partial u > 0$ . In other words, at the true parameter value  $\alpha_0$ , the real function  $m_{\alpha_0}(x, u)$  is assumed to be strictly increasing in  $u$  on  $(0, 1)$  for all values of  $x \in \mathcal{X}$ . In particular, this property guarantees that conditional on  $X$ , the mapping from the unobservables  $U$  to the observables  $Y$  is one-to-one.

Our estimator will be constructed from a sample  $\{y_i, x_i\}_{i=1}^n$  of observations of  $(Y, X)$  drawn according to the model (2) with  $\alpha = \alpha_0$ . We shall assume the following:

**Assumption A.** (i)  $\{y_i, x_i\}_{i=1}^n$  are i.i.d.; (ii)  $X$  is continuously distributed on  $\mathcal{X}$  with density  $f_X(x)$ ; (iii) the densities  $f_{Y|X}(y|x)$  and  $f_X(x)$  are uniformly bounded in  $(y, x)$  on  $\mathcal{S}$  (defined below) and in  $x$  on  $\mathcal{X}$ , respectively.

Assumption A(i) is more likely to hold in cross sectional applications; though extensions to time series context are feasible, we do not pursue them here. Assumptions A(ii,iii) put restrictions on the density of the observables. Combining  $U \sim U(0, 1)$

with  $m_{\alpha_0}(x, \cdot)$  being strictly increasing ensures that conditional on  $X = x$ ,  $Y$  is continuously distributed with support in  $m_{\alpha_0}(x, (0, 1))$ ; we denote by  $f_{Y|X}(\cdot|\cdot)$  its conditional density. Assumption A(ii) then ensures that  $(Y, X)$  are jointly continuously distributed on  $\mathcal{S} \equiv \bigcup_{x \in \mathcal{X}} (m_{\alpha_0}(x, (0, 1)), x)$ . Moreover,  $Y$  is then continuous on  $\mathcal{Y} \equiv \bigcup_{x \in \mathcal{X}} m_{\alpha_0}(x, (0, 1))$ . Note that we allow the support of the dependent variable  $Y$  to depend on the true value  $\alpha_0$  of  $\alpha$ , as in some well known examples of (2) such as the Box-Cox transformation model (see, e.g., Komunjer, 2009).

The key property of the model (2) upon which we base our estimation procedure is that  $\alpha_0$  is nonparametrically identified by an independence restriction:

**Assumption B.** *The true value  $\alpha_0 \in \mathcal{A}$  of the structural parameter  $\alpha$  in model (2) is identified by the restriction:  $U \perp X$ .*

Assumption B requires that the model (2) be identified by an independence restriction. For fully nonparametric specifications, the arguments that lead to this result are well understood (see, e.g., Matzkin, 2003). Identification in semiparametric setups, however, can be more challenging and of course depends on the model specification. In Section 5, we provide more primitive conditions under which the identification Assumption B holds within a simplified BLP model. The following lemma derives a simple characterization of the property in Assumption B.

**Lemma 1.** *Let Assumptions A(ii) and B hold. Then, it follows that:*

$$P(Y \leq m_{\alpha}(X, u); X \leq x) = u \cdot P(X \leq x) \quad \text{for all } (x, u) \in \mathcal{X} \times (0, 1)$$

*if and only if  $\alpha = \alpha_0$ .*

Lemma 1 suggests a straightforward way to construct a criterion function through which to estimate  $\alpha_0$ . Let  $t = (x, u) \in \mathcal{X} \times (0, 1)$  and define

$$(3) \quad W_{\alpha}(t) \equiv P(Y \leq m_{\alpha}(X, u); X \leq x) - u \cdot P(X \leq x) .$$

Under the assumptions of Lemma 1, we have  $W_{\alpha}(t) = 0$  for all  $t \in \mathcal{X} \times (0, 1)$  if and only if  $\alpha = \alpha_0$ . Hence, a natural candidate for a population criterion function is the

Cramer von-Mises type objective:

$$(4) \quad Q(\alpha) \equiv \int_{\mathcal{X} \times (0,1)} W_{\alpha}^2(t) d\mu(t)$$

where  $\mu$  is a measure on  $\mathcal{X} \times (0, 1)$  that is absolutely continuous with respect to Lebesgue measure. The choice of  $\mu$  is free, though we note that it will influence the asymptotic variance of our estimator for  $\theta$ .

When the model in (2) is identified by the restriction  $U \perp X$ , Lemma 1 implies that  $\alpha_0$  is the unique zero of  $Q(\alpha)$  and hence we have:

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

The absolute continuity of  $\mu$  is needed to ensure that  $\alpha_0$  is the unique minimum of  $Q(\alpha)$ . Indeed, if  $\mu$  were to place point masses on some finite number of values  $t_i \in \mathcal{X} \times (0, 1)$  of  $t$  (with  $i \in I$  and  $I$  finite), then the objective function  $Q(\alpha)$  would be minimized at values of  $\alpha$  for which  $W_{\alpha}(t_i) = 0$  for all  $i \in I$ . Therefore, multiple minimizers will exist in specifications where the independence assumption cannot be weakened without losing identification.

Estimation will proceed by minimizing an empirical analogue  $Q_n(\alpha)$  of  $Q(\alpha)$  over an appropriate sieve space. First define the sample analogue to  $W_{\alpha}(t)$ :

$$(5) \quad W_{\alpha,n}(t) \equiv \frac{1}{n} \sum_{i=1}^n 1\{y_i \leq m_{\alpha}(x_i, u); x_i \leq x\} - u \cdot \frac{1}{n} \sum_{i=1}^n 1\{x_i \leq x\},$$

which yields a finite sample criterion function:

$$(6) \quad Q_n(\alpha) \equiv \int_{\mathcal{X} \times (0,1)} W_{\alpha,n}^2(t) d\mu(t).$$

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:

$$(7) \quad \hat{\alpha} \in \arg \min_{\alpha \in \mathcal{A}_n} Q_n(\alpha).$$

For the consistency analysis, we endow  $\mathcal{A}$  with the metric  $\|\alpha\|_c = \|\theta\| + \|h\|_\infty^2$  and impose the following additional assumption:

**Assumption C.** (i)  $\mu$  has full support on  $\mathcal{X} \times (0, 1)$ ; (ii)  $\Theta$  and  $\mathcal{H}$  are compact w.r.t  $\|\cdot\|$  and  $\|\cdot\|_\infty$ ; (iii)  $m_\alpha(x, \cdot) : (0, 1) \rightarrow \mathbb{R}$  is strictly increasing for every  $(\alpha, x) \in \mathcal{A} \times \mathcal{X}$ ; (iv) For every  $x \in \mathcal{X}$ ,  $\sup_{u \in (0, 1)} |m_\alpha(x, u) - m_{\tilde{\alpha}}(x, u)| \leq G(x) \{ \|\theta - \tilde{\theta}\| + \|h - \tilde{h}\|_\infty \}$  with  $E[G^2(X)] < \infty$ ; (v) The entropy  $\int_0^\infty \sqrt{N_{[]}(\eta^3, \mathcal{H}, \|\cdot\|_\infty)} d\eta < \infty$ ; (vi)  $\mathcal{H}_n \subset \mathcal{H}$  are closed in  $\|\cdot\|_\infty$  and for any  $h \in \mathcal{H}$  there exists  $\Pi_n h \in \mathcal{H}_n$  such that  $\|h - \Pi_n h\|_\infty = o(1)$ .

As already pointed out, Assumption C(i) ensures that  $Q(\alpha)$  is uniquely minimized at  $\alpha_0$ . Assumptions C(ii)-(v) ensures the stochastic process is asymptotically equicontinuous in probability. It is interesting to note that while strict monotonicity of  $m_\alpha(x, \cdot)$  is not needed for identification, imposing it on the parameter space is helpful in the statistical analysis. In Assumption C(v),  $N_{[]}(\eta^3, \mathcal{H}, \|\cdot\|_\infty)$  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 C(v). Finally, Assumption C(vi) requires the sieve can approximate the parameter space with respect to the norm  $\|\cdot\|_\infty$ .

Assumptions A, B and C are sufficient for establishing the consistency of the MDI estimator under the norm  $\|\cdot\|_c$ .

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

### 3. RATE OF CONVERGENCE

In this section we establish 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  $h(x, u)$ :

$$(8) \quad \|h\|_{L^2}^2 = \int_{\mathcal{X} \times (0, 1)} h^2(x, u) f_X(x) dx du.$$

---

<sup>2</sup>See Appendix for details regarding the notations and definitions.

Associated to the norm  $\|h\|_{L^2}$  is the vector space  $L^2 = \{h(x, u) : \|h\|_{L^2} < \infty\}$ . We assume the structural function  $m_\alpha(x, u)$  in (2) satisfies  $\|m_\alpha\|_{L^2} < \infty$  and define the mapping  $m : (\mathcal{A}, \|\cdot\|_c) \rightarrow L^2$  which to any  $\alpha \in \mathcal{A}$  associates  $m(\alpha) \equiv m_\alpha$ .

Given these definitions, we introduce the following assumption.

**Assumption D.** (i) In a neighborhood  $\mathcal{N}(\alpha_0) \subset \mathcal{A}$ ,  $m : (\mathcal{A}, \|\cdot\|_c) \rightarrow L^2$  is continuously Fréchet differentiable;<sup>3</sup> (ii) For every  $(y, x) \in \mathcal{S}$ , the conditional densities satisfy  $|f_{Y|X}(y|x) - f_{Y|X}(y'|x)| \leq J(x)|y - y'|^\nu$  with  $E[J^2(X)G^{2\nu}(X)] < \infty$ ; (iii) The marginal density of  $\mu$  with respect to  $u$  is uniformly bounded on  $(0, 1)$ .

In what follows, we denote by  $\frac{dm}{d\alpha}(\tilde{\alpha})$  the Fréchet derivative of  $m$  evaluated at  $\tilde{\alpha} \in \mathcal{A}$ . For example, consider the structural mapping  $m_\alpha(x, u) = h(x, u) + x'\theta$  and assume that  $\|m_\alpha\|_{L^2} < \infty$ . In this case  $m$  is linear and so it is its own Fréchet derivative, i.e. for any  $\pi = (\pi_h, \pi_\theta) \in \mathcal{A}$  we have  $\frac{dm}{d\alpha}(\alpha)[\pi](x, u) = \pi_h(x, u) + x'\pi_\theta$ . To simplify the notation, we hereafter let:

$$\frac{dm_\alpha(x, u)}{d\alpha}[\pi] \equiv \frac{dm}{d\alpha}(\alpha)[\pi](x, u)$$

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  $\alpha_0$ . Under Assumption A(ii)-(iii), C(iv) and D, the Fréchet differentiability of  $m$  is inherited by the mapping  $Q : (\mathcal{A}, \|\cdot\|_c) \rightarrow \mathbb{R}$  which to every  $\alpha \in \mathcal{A}$  associates  $Q(\alpha)$ . To state the form of this Fréchet derivative, we define the linear map  $D_{\tilde{\alpha}} : (\mathcal{A}, \|\cdot\|_c) \rightarrow L^2_\mu$  which to every  $\pi \in \mathcal{A}$  associates  $D_{\tilde{\alpha}}[\pi]$  where  $D_{\tilde{\alpha}}[\pi] : \mathcal{X} \times (0, 1) \rightarrow \mathbb{R}$  maps  $t = (x, u) \in \mathcal{X} \times (0, 1)$  into  $D_{\tilde{\alpha}}[\pi](t)$  given by:

$$(9) \quad D_{\tilde{\alpha}}[\pi](t) = \int_{\mathcal{X}} f_{Y|X}(m_{\tilde{\alpha}}(s_x, u)|s_x) \frac{dm_{\tilde{\alpha}}(s_x, u)}{d\alpha}[\pi] 1\{s_x \leq x\} f_X(s_x) ds_x$$

Lemma 2 establishes that  $Q(\alpha)$  is twice Fréchet differentiable at  $\alpha_0$ .

**Lemma 2.** Under Assumptions A(ii)-(iii), C(iv) and D(i)-(iii), the map  $Q : (\mathcal{A}, \|\cdot\|_c) \rightarrow \mathbb{R}$  is: (i) continuously Fréchet differentiable in  $\mathcal{N}(\alpha_0)$  with

$$\frac{dQ(\tilde{\alpha})}{d\alpha}[\pi] = \int_{\mathcal{X} \times (0, 1)} W_{\tilde{\alpha}}(t) D_{\tilde{\alpha}}[\pi](t) d\mu(t);$$

---

<sup>3</sup>See Appendix for definitions.

(ii) twice Fréchet differentiable at  $\alpha_0$  with

$$\frac{d^2 Q(\alpha_0)}{d\alpha^2}[\psi, \pi] = \int_{\mathcal{X} \times (0,1)} D_{\alpha_0}[\psi](t) D_{\alpha_0}[\pi](t) d\mu(t).$$

In this model, since  $Q(\alpha)$  is minimized at  $\alpha_0$ , its second derivative at  $\alpha_0$  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 2 we therefore define the inner product and associated norm:

$$(10) \quad \langle \alpha, \tilde{\alpha} \rangle_w \equiv \int_{\mathcal{X} \times (0,1)} D_{\alpha_0}[\alpha](t) D_{\alpha_0}[\tilde{\alpha}](t) d\mu(t) \quad \text{and} \quad \|\alpha\|_w^2 = \langle \alpha, \alpha \rangle_w .$$

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 feasible 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 stronger norm  $\|\alpha\|_s \equiv \|\theta\| + \|h\|_{L^2}$ . 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  $\mathcal{A}_n$ . We impose the following assumptions in order to obtain the rate of convergence of  $\hat{\alpha}$  in the norm  $\|\cdot\|_s$ :

**Assumption E.** (i) In  $\mathcal{N}(\alpha_0)$ ,  $\|\alpha - \alpha_0\|_w^2 \lesssim Q(\alpha) \lesssim \|\alpha - \alpha_0\|_s^2$ ; (ii) The ratio  $\tau_n \equiv \sup_{\mathcal{A}_n} \|\alpha_n\|_s^2 / \|\alpha_n\|_w^2$  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 E(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 E(ii) is crucial in enabling us to obtain rates in  $\|\cdot\|_s$  from rates in  $\|\cdot\|_w$ , and vice versa, which is needed to refine initial estimates of the rate of convergence. The ratio  $\tau_n$  is often referred to as the sieve modulus of continuity (see, e.g., Chen and Pouzo, 2008b). In practice, Assumption E(ii) is requiring the sieve not to grow too fast. Finally Assumption E(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 2.** *Under Assumptions A, B, C, D and E,  $\|\hat{\alpha} - \alpha_0\|_s = o_p(n^{-\frac{1}{4}})$ .*

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

#### 4. 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 Chen and Pouzo (2008a). We proceed in two steps. First, we show that for any  $\lambda \in \mathbb{R}^{d_\theta}$  the linear functional  $F_\lambda(\alpha) = \lambda'\theta$ , which returns a linear combination of the parametric component of the semiparametric specification, is continuous in  $\|\cdot\|_w$ . By appealing to the Riesz Representation theorem it then follows that there is  $v^\lambda$  such that  $\langle v^\lambda, \hat{\alpha} - \alpha_0 \rangle_w = \lambda'(\hat{\theta} - \theta_0)$ . Second, we establish the asymptotic normality of  $\sqrt{n}\langle v^\lambda, \hat{\alpha} - \alpha_0 \rangle_w$  and employ the Cramér-Wold device to conclude the asymptotic normality of  $\sqrt{n}(\hat{\theta} - \theta)$ .

We therefore first aim to establish 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  $\langle \cdot, \cdot \rangle_w$  and that  $\bar{\mathcal{A}}$  is of the form  $\bar{\mathcal{A}} = \mathbb{R}^{d_\theta} \times \bar{\mathcal{H}}$ . For any  $(\alpha - \alpha_0)$  in  $\bar{\mathcal{A}}$ , we can then decompose  $D_{\alpha_0}[\alpha - \alpha_0]$  as:<sup>4</sup>

$$(11) \quad D_{\alpha_0}[\alpha - \alpha_0] \equiv \frac{dW(\alpha_0)}{d\alpha}[\alpha - \alpha_0] = \frac{dW(\alpha_0)}{d\theta'}[\theta - \theta_0] + \frac{dW(\alpha_0)}{dh}[h - h_0].$$

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

$$(12) \quad h_j^* \equiv \arg \min_{h \in \bar{\mathcal{H}}} \int_{\mathcal{X} \times (0,1)} \left( \frac{dW_{\alpha_0}(t)}{d\theta_j} - \frac{dW_{\alpha_0}(t)}{dh}[h] \right)^2 d\mu(t),$$

<sup>4</sup>The first equality in (11) is formally justified in the proof of Lemma 2 in the Appendix, in which it is shown  $D_{\bar{\alpha}}$  is the Fréchet derivative of the mapping  $W : (\mathcal{A}, \|\cdot\|_c) \rightarrow L_\mu^2$  given by  $W : \alpha \mapsto W_\alpha$ , when evaluated at  $\bar{\alpha}$ . Similar to before, we use the notation:

$$\frac{dW_\alpha(t)}{d\alpha}[\pi] \equiv \frac{dW}{d\alpha}(\alpha)[\pi](t).$$

where the minimum in (12) is indeed attained and  $h_j^*$  is well defined due to the Projection Theorem in Hilbert Spaces, (see, e.g., Theorem 3.3.2 in Luenberger, 1969). Similarly define  $h^* \equiv (h_1^*, \dots, h_{d_\theta}^*)$  and let

$$(13) \quad \frac{dW_{\alpha_0}(t)}{dh}[h^*] = \left( \frac{dW_{\alpha_0}(t)}{dh}[h_1^*], \dots, \frac{dW_{\alpha_0}(t)}{dh}[h_{d_\theta}^*] \right).$$

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

$$(14) \quad R_{h^*}(t) = \frac{dW_{\alpha_0}(t)}{d\theta} - \frac{dW_{\alpha_0}(t)}{dh}[h^*],$$

and the associated matrix

$$(15) \quad \Sigma^* \equiv \int_{\mathcal{X} \times (0,1)} R_{h^*}(t) R_{h^*}'(t) d\mu(t).$$

Lemma 3 shows that the functional  $F_\lambda(\alpha) = \lambda'\theta$  is continuous if the matrix  $\Sigma^*$  is positive definite, which may be interpreted as a local identification condition on  $\theta_0$ . Lemma 3 also obtain the formula for the Riesz Representer of  $F_\lambda(\alpha)$ .

**Lemma 3.** *Let  $v_\theta^\lambda \equiv (\Sigma^*)^{-1}\lambda$  and  $v_h^\lambda \equiv -h^*v_\theta^\lambda$ . If  $\Sigma^*$  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  $\|\cdot\|_w$  and the closed form solution or the Riesz Representer  $v^\lambda$  we can study the asymptotic normality of  $\lambda'(\hat{\theta} - \theta)$  by examining  $\sqrt{n}\langle v^\lambda, \hat{\alpha} - \alpha_0 \rangle_w$  instead. The latter representation is simpler to analyze as it is determined by the local behavior of  $Q(\alpha)$  near its minimum  $\alpha_0$ . In order to establish asymptotic normality, we require one final assumption:

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

Assumption F(i) ensures that  $F_\lambda(\alpha) = \lambda'\theta$  is continuous in  $\|\cdot\|_w$ , as shown in Lemma 3. While  $v^\lambda \in \bar{\mathcal{A}}$ , Assumption F(ii) additionally requires  $v^\lambda \in \mathcal{A}$ . As a result  $v^\lambda$  may be approximated by an element  $\Pi_n v^\lambda \in \mathcal{A}_n$  due to Assumption D(iii). The qualification “for  $\|\lambda\|$  small” is due to the compactness assumption on  $\Theta \times \mathcal{H}$  imposing that they be bounded in norm. Finally Assumptions F(iii)-(iv) require  $W_\alpha(t)$  to be twice differentiable and for certain regularity conditions to hold on the derivatives.

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

**Theorem 3.** *Let Assumptions A-F hold. Then,  $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{L} N(0, \Sigma)$  where*

$$\Sigma \equiv [\Sigma^*]^{-1} \left[ \int_{(\mathcal{X} \times (0,1))^2} R_{h^*}(t) R_{h^*}'(s) \Sigma(t, s) d\mu(t) d\mu(s) \right] [\Sigma^*]^{-1},$$

and for every  $t = (x, u)$  and  $t' = (x', u')$  in  $\mathcal{X} \times (0, 1)$  the kernel  $\Sigma(t, t')$  is given by:

$$\Sigma(t, t') \equiv E \left[ (1\{U \leq u; X \leq x\} - u \cdot 1\{X \leq x\}) (1\{U \leq u'; X \leq x'\} - u' \cdot 1\{X \leq x'\}) \right].$$

## 5. EXAMPLE AND MONTE CARLO EVIDENCE

**5.1. The Model.** We proceed to illustrate how nonseparable structures of the form in (2) arise naturally in simple economic models. We shall also use this example in a small Monte Carlo study of the performance of our estimator. Our example is a basic version of Berry, Levinsohn, and Pakes (1995) (BLP henceforth) model with two products and two firms. On the demand side, we use a random utility specification à la Hausman and Wise (1978):

$$(16) \quad u_{ij} = -ap_j + b'x_j + \xi_j + \zeta_i + \varepsilon_{ij},$$

in which  $u_{ij}$  is the utility of product  $j$  ( $j = 1, 2$ ) to individual  $i$  ( $i = 1, \dots, I$ ) with unobserved characteristics  $\zeta_i$  ( $\zeta_i \in \mathbb{R}$ ),  $p_j$  and  $x_j$  are respectively the price and a  $d_x$ -vector of observed characteristics of product  $j$  ( $p_j \in \mathbb{R}_+$ ,  $x_j \in \mathbb{R}^{d_x}$ ,  $d_x < \infty$ );  $b$  is a  $d_x$ -vector of coefficients determining the impact of  $x_j$  on the utility for  $j$  ( $b \in \mathbb{R}^{d_x}$ ), and  $\xi_j$  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,  $\varepsilon_{ij}$  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 (16) is that  $\varepsilon_{ij}$  are iid across products  $j$  and individuals  $i$ . For example, assuming that  $\varepsilon_{ij}$ '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 some known cdf  $F$  that need not be logit. Note that  $F$  necessarily satisfies  $F(-\varepsilon) = 1 - F(\varepsilon)$ . When  $\varepsilon_{i2} - \varepsilon_{i1}$  has cdf  $F$ , the demand for good  $j$ , denoted  $D_j(p_j, p_{-j})$ , is given by:

$$(17) \quad D_j(p_j, p_{-j}) = M \cdot F(-a(p_j - p_{-j}) + b'(x_j - x_{-j}) + \xi_j - \xi_{-j})$$

where  $M$  is the total market size.

Hereafter, we let the  $Y \equiv F^{-1}(D_1(p_1, p_2)/M)$  be the quantile of the market share for firm 1's good ( $Y \in \mathbb{R}$ ),  $P \equiv p_1 - p_2$ ,  $X \equiv x_1 - x_2$  and  $\xi \equiv \xi_1 - \xi_2$ . Then, the structural BLP model of (17) takes the form:

$$(18) \quad Y = -aP + b'X + \xi \quad \text{with} \quad \xi \perp X$$

In the model above, prices are endogenous, so even if  $\xi$  is independent of  $X$ , we can expect  $P$  to depend on  $\xi$ . Hence, without further restrictions on  $\xi$  and  $P$  it is not possible to identify the parameters  $a$  and  $b$  in (18). 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 the price which maximizes its profit  $\Pi_j(p_j, p_{-j}) = (p_j - c)D_j(p_j, p_{-j})$ . We assume the marginal cost parameter  $c$  to be the same for both firms. The equilibrium prices  $(p_1, p_2)$  are implicitly defined by the solution to the Bertrand game with exogenous variables  $X$ . Lemma 4 exploits this relationship to obtain an alternative representation for the BLP model (18).

**Lemma 4.** *Assume  $F$  is twice continuously differentiable on  $\mathbb{R}$  with strictly increasing hazard rate  $\tau$ . If  $\xi$  is continuously distributed, then it follows that:*

$$(19) \quad Y = h(X, U) + X'\theta, \quad U \sim U(0, 1)$$

*with  $h$  continuously differentiable,  $\partial h(x, u)/\partial u > 0$ , and  $\theta = b$ .*

Lemma 4 assumes the hazard rate  $\tau(\varepsilon) \equiv f(\varepsilon)/[1 - F(\varepsilon)]$  to be strictly increasing on  $\mathbb{R}$ , which is equivalent to requiring that  $f'(\varepsilon)[1 - F(\varepsilon)] + f^2(\varepsilon) > 0$  for all  $\varepsilon \in \mathbb{R}$  (also equivalent to  $f'(\varepsilon)F(\varepsilon) - f^2(\varepsilon) < 0$ ). This assumption guarantees the existence of a unique Nash equilibrium and the Lemma can then be obtained by analyzing the equilibrium strategies.

The BLP model (19) is clearly a special case of the nonseparable structural model in (2) with  $\alpha \equiv (\theta, h)$  and

$$(20) \quad m_\alpha(X, U) = h(X, U) + X'\theta.$$

We now illustrate how to verify other assumptions for this model. If the BLP model variables are iid with continuous distribution functions, then the continuous differentiability of the demand function guarantees that the sampling Assumption A holds. We note that in this example the supports of the endogenous and exogenous variables are given by  $\mathcal{Y} = \mathbb{R}$  and  $\mathcal{X} \subseteq \mathbb{R}^{d_x}$ .

By far more difficult to check is the identification Assumption B for which we now derive more primitive conditions. Our identification result for the BLP model (19) is contained in the following Theorem:

**Theorem 4.** *Assume  $F$  is strictly increasing, twice continuously differentiable on  $\mathbb{R}$  with strictly increasing hazard rate  $\tau$ . Assume moreover that  $\xi$  is continuously distributed, and that we have: (i)  $h(0, 1/2) = 0$  and (ii)  $\partial h(0, 1/2)/\partial x = 1$ . Then, the BLP model (19) satisfies Assumption B.*

The conditions of Theorem 4 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 zero. In particular, (i) holds

if the distribution  $F_\xi$  of the products' unobservables  $\xi$  in the BLP model in Equation (18) is known to satisfy  $F_\xi(0) = 1/2$ , since when  $X = 0$  and  $\xi = 0$  the equilibrium is symmetric ( $x_1 = x_2$ ) which implies  $P = 0$ .<sup>5</sup> Hence,  $-aP + \xi = 0 = h(0, 1/2)$ . Requirement (ii) fixes the value of the gradient  $\partial h(x, u)/\partial x$  at zero. It ensures that the effects of changing  $\theta$  can be separated from those of changing  $h$ . Indeed, if  $h$  is additive 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  $\theta$  in  $Y = (\phi + \theta)'X + r(U)$ .

In the context of the BLP model (19), Assumptions C(ii) and C(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  $\lambda$  be a  $d_x + 1$  dimensional vector of positive integers. Define  $|\lambda| \equiv \sum_{i=1}^{d_x+1} \lambda_i$  and  $D^\lambda = \partial^\lambda / \partial x_1^{\lambda_1} \dots \partial x_{d_x}^{\lambda_{d_x}} \partial u^{\lambda_{d_x+1}}$ . An appropriate set  $\mathcal{H}$  is then:

$$\mathcal{H} = \left\{ h : \max_{|\lambda| \leq \frac{3(d_x+1)}{2} + 1} \left[ \sup_{(x,u) \in \mathcal{X} \times (0,1)} |D^\lambda h(x, u)| \right] \leq M, \inf_{(x,u) \in \mathcal{X} \times (0,1)} \frac{\partial h(x, u)}{\partial x} \geq \epsilon \right\}$$

for some positive  $M$  and  $\epsilon$ . By Theorem 2.7.1 in van der Vaart and Wellner (1996), Assumptions C(i) and C(v) are then satisfied. The definition of  $\mathcal{H}$  also ensures Assumption C(iii) holds, while C(iv) is immediate from (20).

As already noted,  $m_\alpha$  in (20) is linear, and since it is a continuous map from  $(\mathcal{A}, \|\cdot\|_c)$  to  $L^2$ , it is continuously Fréchet differentiable with:

$$\frac{dm_\alpha(x, u)}{d\alpha}[\pi] = \pi_h(x, u) + x'\pi_\theta,$$

which verifies Assumption D(i). In addition, for any  $t = (x, u)$  we then have:

$$(21) \quad D_\alpha[\pi](t) = \int_{\mathcal{X}} f_{Y|X}(h(s_x, u) + s'_x \theta | s_x) (\pi_h(s_x, u) + s'_x \pi_\theta) 1\{s_x \leq x\} f_X(s_x) ds_x.$$

Hence, if  $f_{Y|X}(y|x)$  is uniformly bounded, then  $\mathcal{X}$  and  $\Theta$  compact together with our choice for  $\mathcal{H}$  imply Assumption F(iv) holds. Similarly, by direct calculation we

---

<sup>5</sup>Note that whenever  $\xi_1$  and  $\xi_2$  are identically distributed, the distribution of their difference  $\xi$  satisfies  $F_\xi(0) = 1/2$ .

obtain that in the discussed BLP example, for any  $\alpha = (\theta, h)$  and  $\bar{\alpha} = (\bar{\theta}, \bar{h})$  we have:

$$\begin{aligned} \frac{dD_{\alpha+\tau\bar{\alpha}}[\pi]}{d\tau}(t)\Big|_{\tau=s} &= \int_{\mathcal{X}} f'_{Y|X}(h(s_x, u) + s\bar{h}(s_x, u) + s'_x(\theta + s\bar{\theta})|s_x)(\pi_h(s_x, u) + s'_x\pi_\theta) \\ &\quad \times (\bar{h}(s_x, u) + s'_x\bar{\theta})1\{s_x \leq x\}f_X(s_x)ds_x, \end{aligned}$$

hence Assumption F(iii) is easily verified if  $|f'_{Y|X}(y|x)|$  is bounded in  $(y, x)$  on  $\mathcal{S}$ .

**5.2. Monte Carlo Setup.** We consider the case in which the idiosyncratic errors  $\varepsilon_{i1}$  and  $\varepsilon_{i2}$  in (16) are iid Gumbel random variables, so that the distribution  $F$  of their difference is logit,  $F(\varepsilon) = \exp(\varepsilon)/[1 + \exp(\varepsilon)]$ . The equilibrium prices are solution to the FOC equations:

$$\begin{aligned} \exp(\Delta) + 1 - a(p_1 - c) &= 0 \\ \exp(-\Delta) + 1 - a(p_2 - c) &= 0 \end{aligned}$$

where  $\Delta \equiv -a(p_1 - p_2) + bX + \xi$  as before; in this model  $X$  is scalar ( $d_x = 1$ ). The equilibrium prices obtained by solving the above equations are continuously differentiable functions of  $X$ ; see Lemma 8. A simple application of the Implicit Function Theorem shows that at equilibrium the prices satisfy:

$$\frac{\partial(p_1 - p_2)}{\partial x} = \frac{2b}{3a}$$

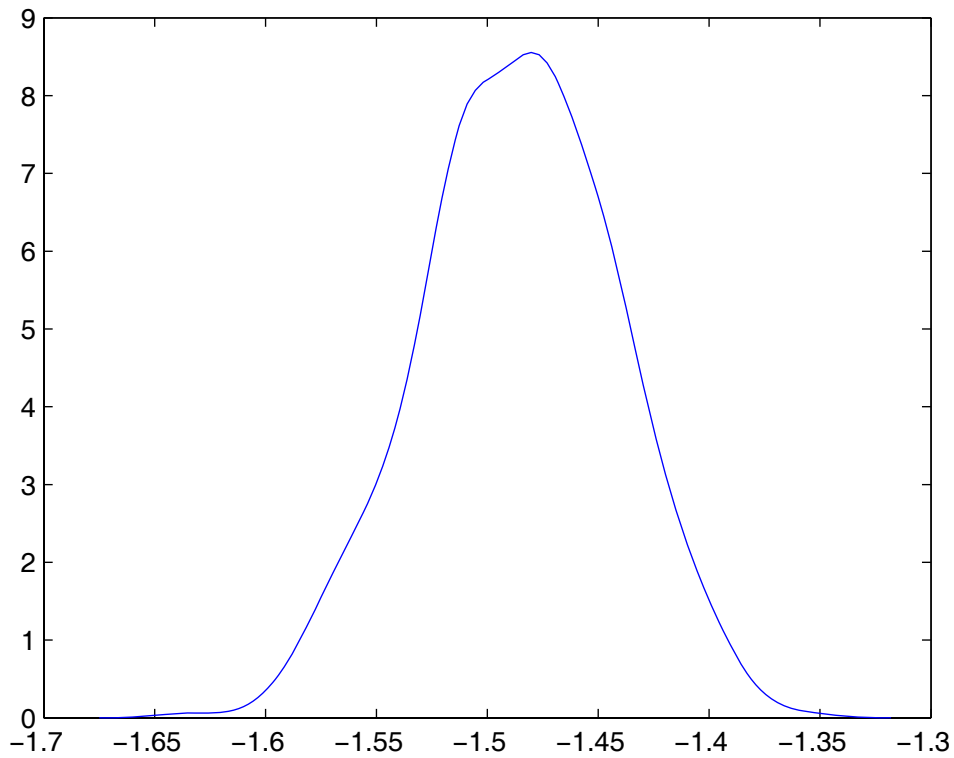
so  $\partial h(0, 1/2)/\partial x = -2b/3$ . We set the true values of the parameters to be  $a = 2.4$ ,  $b = -1.5$  and  $c = 1$ . The variables are drawn as  $X \sim U[-1, 1]$ , and  $\xi \sim N(0, 1)$ , where  $X$  and  $\xi$  are independent.<sup>6</sup> For the sieve we used fully interacted polynomial of order 2 in  $X$  and  $U$ , while the measure  $\mu$  was chosen to be uniform on  $[-1, 1] \times [0, 1]$ .

Table 1 reports the mean, standard deviation, mean squared error and the 10<sup>th</sup>, 50<sup>th</sup> and 90<sup>th</sup> percentile of the proposed estimator  $\hat{\theta}$  for sample sizes  $n = 100, 200, 500$ . The statistics were computed based on  $R = 500$  replications. The estimator performs well, exhibiting only a small downward bias (recall true value is  $b = -1.5$ ) and small mean squared errors for sample sizes of  $n = 200$  and  $n = 500$  observations. For

<sup>6</sup>Note that by setting  $b = -1.5$  we ensure that the identification condition (ii) of Theorem 4 is satisfied. Condition (i) holds since the distribution of  $\xi$  is symmetric around zero.

	Mean	STD	MSE	10%	50%	90%
$n = 100$	-1.493	0.110	0.012	-1.629	-1.498	-1.354
$n = 200$	-1.493	0.072	0.005	-1.583	-1.495	-1.395
$n = 500$	-1.486	0.043	0.002	-1.541	-1.484	-1.432

Table 1: MONTE CARLO RESULTS

Figure 1:  $\hat{\theta}$  KERNEL DENSITY ESTIMATE (n=500)

the latter two sample sizes, the estimator is also within 0.1 of the true value in over eighty percent of the replications. Figure 1 exhibits a Gaussian kernel estimate for the density of  $\hat{\theta}$  obtained with sample size  $n = 500$ . The density is fairly symmetric and centered at the true value  $-1.5$ .

Overall, we find the performance of the estimator on this limited Monte Carlo study to be encouraging.

## 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. A small Monte Carlo study illustrates the performance of the proposed estimator within the BLP example.

## APPENDIX A. NOTATION AND DEFINITIONS

The following is a table of the notation and definitions that will be used throughout the appendix.

$a \lesssim b$	$a \leq Mb$ for some constant $M$ which is universal in the context of the proof.
$\ \cdot\ _c$	The norm $\ \alpha\ _c \equiv \ \theta\  + \ h\ _\infty$ where $\alpha = (\theta, h)$ .
$\ \cdot\ _s$	The norm $\ \alpha\ _s \equiv \ \theta\  + \ h\ _{L^2}$ where $\alpha = (\theta, h)$ .
$\ \cdot\ _\infty$	The norm $\ h\ _\infty \equiv \sup_{(x,u) \in \mathcal{X} \times (0,1)}  h(x, u) $ .
$\ \cdot\ _{L^2}$	The norm $\ h\ _{L^2} \equiv \int_{\mathcal{X} \times (0,1)} h(x, u) f_X(x) dx du$ .
$\ \cdot\ _{L^2_\mu}$	The norm $\ h\ _{L^2_\mu} \equiv \int_{\mathcal{X} \times (0,1)} h(x, u) d\mu(t)$ where $t \equiv (x, u)$ .
$N_{[\cdot]}(\epsilon, \mathcal{F}, \ \cdot\ )$	The bracketing numbers of size $\epsilon$ for $\mathcal{F}$ under the norm $\ \cdot\ $ .

A mapping,  $m : (\mathcal{A}, \|\cdot\|) \rightarrow L^2$  is said to be Fréchet differentiable, if there exists a bounded linear map  $\frac{dm}{d\alpha} : (\mathcal{A}, \|\cdot\|_c) \rightarrow L^2$  such that,

$$\lim_{\|\pi\|_c \searrow 0} \|\pi\|_c^{-1} \left\| m_{\alpha+\pi} - m_\alpha - \frac{dm}{d\alpha}[\pi] \right\|_{L^2} = 0$$

The Fréchet derivative is a natural extension of the standard derivative to general metric spaces.

## APPENDIX B. PROOFS FOR SECTION 2

PROOF OF LEMMA 1: First, consider all values  $\bar{\alpha}$  of  $\alpha$  in  $\mathcal{A}$  such that  $m_{\bar{\alpha}}(x, u)$  is not strictly increasing in  $u$  on  $(0, 1)$  for all values of  $x \in \mathcal{X}$ . Let  $\bar{x} \in \mathcal{X}$  be one such value. Then, the function  $u \mapsto P(X \leq \bar{x}; Y \leq m_{\bar{\alpha}}(X, u))$  is not strictly increasing on  $(0, 1)$ ; hence, there must exist  $\bar{u} \in (0, 1)$  such that  $P(X \leq \bar{x}; Y \leq m_{\bar{\alpha}}(X, \bar{u})) \neq \bar{u} \cdot P(X \leq \bar{x})$ . Now, consider all values  $\tilde{\alpha}$  of  $\alpha$  in  $\mathcal{A}$  such that  $m_{\tilde{\alpha}}(x, u)$  is strictly increasing in  $u$  on  $(0, 1)$  for all values of  $x \in \mathcal{X}$ . Note that  $\alpha_0$  is an element of that set. Now, for any such  $\tilde{\alpha}$ , notice that for any  $(x, u) \in \mathcal{X} \times (0, 1)$  the following holds:

$$\begin{aligned} P(X \leq x; Y \leq m_{\tilde{\alpha}}(X, u)) &= \int_{s_x \leq x} \int_{s_y \leq m_{\tilde{\alpha}}(s_x, u)} f_{XY}(s_x, s_y) ds_x ds_y \\ &= \int_{s_x \leq x} \int_{s_u \leq u} f_{XY}(s_x, m_{\tilde{\alpha}}(s_x, s_u)) \frac{\partial m_{\tilde{\alpha}}(s_x, s_u)}{\partial u} ds_x ds_u \\ &= \int_{s_x \leq x} \int_{s_u \leq u} f_{X\tilde{U}}(s_x, s_u) ds_x ds_u \\ (22) \quad &= P(X \leq x; \tilde{U} \leq u) \end{aligned}$$

where the second and third equalities follow by a change of variable  $(s_x, s_y) = (s_x, m_{\tilde{\alpha}}(s_x, s_u))$  and a change in measure  $Y = m_{\tilde{\alpha}}(X, \tilde{U})$ . Under Assumption B,  $\tilde{\alpha} = \alpha_0$  if and only if  $\tilde{U} \perp X$ , which since  $\tilde{U}$  is uniform on  $(0, 1)$  is equivalent to

$$(23) \quad P(\tilde{U} \leq u; X \leq x) = u \cdot P(X \leq x)$$

for all  $(x, u) \in \mathcal{X} \times (0, 1)$ . Combining (23) and (22) then establishes the Lemma. ■

**Lemma 5.** *Under Assumptions A(i)-(iii) and C(ii)-(v), the following class is Donsker:*

$$\mathcal{F} \equiv \{f(y_i, x_i) = 1\{y_i \leq m_\alpha(x_i, u); x_i \leq x\}, (\alpha, x, u) \in \mathcal{A} \times \mathbb{R}^{d_x} \times (0, 1)\}$$

PROOF: First define the following classes of functions for  $1 \leq k \leq d_x$ :

$$(24) \quad \mathcal{F}_u \equiv \{f(y_i, x_i) = 1\{y_i \leq m_\alpha(x_i, u)\} : (\alpha, u) \in \mathcal{A} \times (0, 1)\}$$

$$(25) \quad \mathcal{F}_x^{(k)} \equiv \{f(x_i) = 1\{x_i^{(k)} \leq t\} : t \in \mathbb{R}\},$$

where  $x^{(k)}$  is the  $k^{\text{th}}$  coordinate of  $x$ . Further notice that by direct calculation we have,

$$(26) \quad \mathcal{F} = \mathcal{F}_u \times \prod_{k=1}^{d_x} \mathcal{F}_x^{(k)}$$

We establish the Lemma by exploiting (26). For any continuously distributed random variable  $V \in \mathbb{R}$  and  $\eta > 0$  we can find  $\{-\infty = t_1, \dots, t_{\lfloor \eta^{-2} \rfloor + 2} = +\infty\}$  such that  $P(t_s \leq V \leq t_{s+1}) \leq \eta^2$ . The brackets  $[1\{v \leq t_s\}, 1\{v \leq t_{s+1}\}]$  then cover  $\{1\{v \leq t\} : t \in \mathbb{R}\}$  and in addition we have,

$$E[(1\{V \leq t_s\} - 1\{V \leq t_{s+1}\})^2] \leq \eta^2$$

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

$$(27) \quad N_{[]}(\eta, \mathcal{F}_x^{(k)}, \|\cdot\|_{L^2}) = O(\eta^{-2})$$

By Assumption C(ii),  $\mathcal{H}$  is compact under  $\|\cdot\|_\infty$  and  $\Theta$  under  $\|\cdot\|$ . Thus, for any  $K_h, K_\theta > 0$  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  $\Theta$  respectively. Defining  $\{\alpha_s\} = \{h_j\} \times \{\theta_l\}$  we then have

$$(28) \quad \#\{\alpha_s\} = N_{[]} (K_h\eta^3, \mathcal{H}, \|\cdot\|_\infty) \times (K_\theta\eta^3)^{-d_\theta}$$

By Assumption C(iii) it then follows that for any  $\alpha \in \mathcal{A}$  there is a  $(\theta_{s^*}, h_{s^*}) \equiv \alpha_{s^*} \in \{\alpha_s\}$  with

$$(29) \quad \begin{aligned} \sup_{u \in (0, 1)} |m_\alpha(x_i, u) - m_{\alpha_{s^*}}(x_i, u)| &\leq G(x_i) \{ \|\theta - \theta_{s^*}\| + \|h - h_{s^*}\|_\infty \} \\ &\leq G(x_i) \{K_\theta + K_h\} \eta^3 \end{aligned}$$

We conclude from (29) that for  $\alpha_s \in \{\alpha_s\}$  brackets of the form

$$(30) \quad [m_{\alpha_s}(x_i, u) - \{K_\theta + K_h\} \eta^3 G(x_i); m_{\alpha_s}(x_i, u) + \{K_\theta + K_h\} \eta^3 G(x)]$$

cover the class  $\{m_\alpha(x_i, u) : \alpha \in \mathcal{A}\}$  for each fixed  $u \in (0, 1)$ . Next note that since  $m_\alpha(x_i, u)$  is strictly increasing in  $u$  for all  $(x_i, \alpha)$  by Assumption C(iii), we may define their inverses:

$$(31) \quad v_\alpha(x_i, t) = u \iff m_\alpha(x_i, u) = t$$

Following Akritas and van Keilegom (2001), for each  $\alpha_s \in \{\alpha_s\}$  we let  $F_s^U(u)$  be as in the first equality in (32) and obtain second equality in (32) from (31).

$$(32) \quad \begin{aligned} F_s^U(u) &\equiv P(Y_i \leq m_{\alpha_s}(X_i, u) + \{K_\theta + K_h\}\eta^3 G(X_i)) \\ &= P(v_{\alpha_s}(X_i, Y_i - \{K_\theta + K_h\}\eta^3 G(X_i)) \leq u) \end{aligned}$$

Arguing as in (27), there is a collection  $\{u_{sk}^U\}$  with  $\#\{u_{sk}^U\} = O(\eta^{-2})$  such that it partitions  $\mathbb{R}$  into segments each with  $F_s^U$  probability at most  $\eta^2/6$ . Similarly, also let

$$(33) \quad F_s^L(u) \equiv P(Y_i \leq m_{\alpha_s}(X_i, u) - \{K_\theta + K_h\}\eta^3 G(X_i))$$

and choose  $\{u_{sk}^L\}$  with  $\#\{u_{sk}^L\} = O(\eta^{-2})$  so that it partitions  $\mathbb{R}$  into segments with  $F_s^L$  probability at most  $\eta^2/6$ . Next, combine  $\{u_{sk}^L\}$  and  $\{u_{sk}^U\}$  into by letting each  $u \in \mathbb{R}$  form the bracket

$$u_{sk_1}^L \leq u \leq u_{sk_2}^U$$

where  $u_{sk_1}^L$  is the largest element of  $\{u_{sk}^L\}$  such that  $u_{sk}^L \leq u$ , and similarly  $u_{sk_2}^U$  is the smallest element in  $\{u_{sk}^U\}$  such that  $u_{sk}^U \geq u$ . We denote this new brackets by  $\{[u_{sk_1}, u_{sk_2}]\}$  and note that

$$(34) \quad \#\{[u_{sk_1}, u_{sk_2}]\} = O(\eta^{-2})$$

It follows from (30) and the strict monotonicity of  $m_\alpha(x, u)$  in  $u$  that for every  $(\alpha, u) \in \mathcal{A} \times (0, 1)$  there exists a  $\alpha_s \in \{\alpha_s\}$  and  $[u_{sk_1}, u_{sk_2}] \in \{[u_{sk_1}, u_{sk_2}]\}$  such that,

$$(35) \quad \begin{aligned} &1\{y_i \leq m_{\alpha_s}(x, u_{sk_1}) - \{K_\theta + K_h\}\eta^3 G(x_i)\} \\ &\leq 1\{y_i \leq m_\alpha(x_i, u)\} \leq 1\{y_i \leq m_{\alpha_s}(x, u_{sk_2}) + \{K_\theta + K_h\}\eta^3 G(x_i)\}, \end{aligned}$$

and hence  $\{[1\{y_i \leq m_{\alpha_s}(x, u_{sk_1}) - \{K_\theta + K_h\}\eta^3 G(x_i)\}, 1\{y_i \leq m_{\alpha_s}(x, u_{sk_2}) + \{K_\theta + K_h\}\eta^3 G(x_i)\}]\}$  form brackets for the class of functions  $\mathcal{F}_u$ .

In order to calculate the size of the proposed brackets, note their  $L^2$  squared norm is equal to  $F_s^U(u_{sk_2}) - F_s^L(u_{sk_1})$ . The construction of  $\{[u_{sk_1}, u_{sk_2}]\}$  in turn implies the first inequality in (36) holds for any  $u \in [u_{sk_1}, u_{sk_2}]$ , while direct calculation yields the second inequality for any constant

$M_\eta > 0$ . Setting  $M_\eta = \sqrt{6E[G^2(X_i)]}/\eta$  and Chebychev's inequality yields the final result in (36).

$$\begin{aligned}
F_s^U(u_{sk_2}) - F_s^L(u_{sk_1}) &\leq F_s^U(u) - F_s^L(u) + \frac{\eta^2}{3} \\
&\leq F_s^U(u; G(X_i) \leq M_\eta) - F_s^L(u; G(X_i) \leq M_\eta) + 2P(G(X_i) \geq M_\eta) + \frac{\eta^2}{3} \\
(36) \quad &\leq F_s^U(u; G(X_i) \leq M_\eta) - F_s^L(t; G(X_i) \leq M_\eta) + \frac{2}{3}\eta^2
\end{aligned}$$

To conclude, notice that  $M_\eta = \sqrt{6E[G^2(X_i)]}/\eta$  and the mean value theorem imply that

$$\begin{aligned}
&F_s^U(u; G(X_i) \leq M_\eta) - F_s^L(u; G(X_i) \leq M_\eta) \\
&\leq P(Y_i \leq m_{\alpha_s}(X_i, u) + \{K_\theta + K_h\}M_\eta\eta^3) - P(Y_i \leq m_{\alpha_s}(X_i, u) - \{K_\theta + K_h\}M_\eta\eta^3) \\
&= E [P(Y_i \leq m_{\alpha_s}(X_i, u) + \{K_\theta + K_h\}M_\eta\eta^3 | X_i) - P(Y_i \leq m_{\alpha_s}(X_i, u) - \{K_\theta + K_h\}M_\eta\eta^3 | X_i)] \\
&\leq 2\{\sup_{y_i, x_i} f_{Y|X}(y_i | x_i)\} \{K_\theta + K_h\} \sqrt{6E[G^2(X_i)]} \eta^2
\end{aligned}$$

where the resulting expression is finite due to Assumptions A(iii) and C(iv). Combining the preceding result with that obtained in (36) it follows that by choosing

$$\{K_\theta + K_h\} \leq (2\{\sup_{y_i, x_i} f_{Y|X}(y_i | x_i)\} \sqrt{6E[G^2(X_i)]})^{-1}$$

the proposed brackets will have  $L^2$  size  $\eta$ . Thus, we have from (28) and (34),

$$(37) \quad N_{[\cdot]}(\eta, \mathcal{F}_u, \|\cdot\|_{L^2}) = O(N_{[\cdot]}(K_h\eta^3, \mathcal{H}, \|\cdot\|_\infty) \times (K_\theta\eta^3)^{-(2+d_\theta)})$$

To conclude note that (27), (37), Assumption C(iv) and Theorem 2.5.6 in van der Vaart and Wellner (1996) imply 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 (26) establishes the claim of the Lemma. ■

PROOF OF THEOREM 1: By Assumption C(ii) and the Tychonoff Theorem,  $\mathcal{A}$  is compact with respect to  $\|\cdot\|_c$ . Furthermore, Lemma 5 and simple manipulations show,

$$(38) \quad \sup_{t, \alpha} |W_{\alpha, n}(t) - W_\alpha(t)| = o_p(1)$$

Exploiting (38) and  $W_{\alpha, n}(t)$  and  $W_\alpha(t)$  being bounded by 1, we obtain that:

$$(39) \quad \sup_\alpha |Q_n(\alpha) - Q(\alpha)| \leq \left[ \sup_{t, \alpha} |W_{\alpha, n}(t) - W_\alpha(t)| \right] \times \left[ \sup_{t, \alpha} |W_{\alpha, n}(t)| + \sup_{t, \alpha} |W_\alpha(t)| \right] = o_p(1)$$

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 3

PROOF OF LEMMA 2: Similar to previously, let  $W : (\mathcal{A}, \|\cdot\|_c) \rightarrow L_\mu^2$  be a mapping which to each  $\alpha \in \mathcal{A}$  associates  $W(\alpha) \equiv W_\alpha$ . We first study the differentiability of  $W$  in a neighborhood of  $\alpha_0$ . Recall that for any  $t = (x, u)$ ,

$$D_{\bar{\alpha}}[\pi](t) = \int_{\mathcal{X}} f_{Y|X}(m_{\bar{\alpha}}(s_x, u)|s_x) \frac{dm_{\bar{\alpha}}(s_x, u)}{d\alpha} [\pi] 1_{\{s_x \leq x\}} f_X(s_x) ds_x$$

and notice that  $D_{\bar{\alpha}}[\pi]$  is well defined for every  $\bar{\alpha} \in \mathcal{N}(\alpha_0)$  due to Assumption D(i). Next, use  $f_{Y|X}(y|x)$  uniformly bounded and Jensen's inequality to obtain the first result in (40). The second inequality then holds for  $\|\cdot\|_o$  the linear operator norm by Assumption D(iii).

$$\begin{aligned} \|D_{\bar{\alpha}}[\pi]\|_{L_\mu^2}^2 &= \int_{\mathcal{X} \times (0,1)} \left[ \int_{\mathcal{X}} f_{Y|X}(m_{\bar{\alpha}}(s_x, u)|s_x) \frac{dm_{\bar{\alpha}}(s_x, u)}{d\alpha} [\pi] 1_{\{s_x \leq x\}} f_X(s_x) ds_x \right]^2 d\mu(t) \\ &\lesssim \int_{\mathcal{X} \times (0,1)} \int_{\mathcal{X}} \left[ \frac{dm_{\bar{\alpha}}(s_x, u)}{d\alpha} [\pi] \right]^2 f_X(s_x) ds_x d\mu(t) \\ (40) \quad &\lesssim \left\| \frac{dm}{d\alpha}(\bar{\alpha}) \right\|_o^2 \|\pi\|_c^2 \end{aligned}$$

Since Fréchet derivative are *a fortiori* continuous, (40) 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 consider  $(\bar{\alpha}, \tilde{\alpha}) \in \mathcal{A}^2$  and use Jensen's inequality to obtain (41) pointwise in  $t = (x, u)$ .

$$\begin{aligned} (41) \quad |D_{\bar{\alpha}}[\pi](t) - D_{\tilde{\alpha}}[\pi](t)| &\leq \int_{\mathcal{X}} f_{Y|X}(m_{\bar{\alpha}}(s_x, u)|s_x) \left| \frac{dm_{\bar{\alpha}}(s_x, u)}{d\alpha} [\pi] - \frac{dm_{\tilde{\alpha}}(s_x, u)}{d\alpha} [\pi] \right| f_X(s_x) ds_x \\ &\quad + \int_{\mathcal{X}} \left| f_{Y|X}(m_{\bar{\alpha}}(s_x, u)|s_x) - f_{Y|X}(m_{\tilde{\alpha}}(s_x, u)|s_x) \right| \left| \frac{dm_{\tilde{\alpha}}(s_x, u)}{d\alpha} [\pi] \right| f_X(s_x) ds_x \end{aligned}$$

In turn, the Lipschitz Assumptions C(iv) and D(ii),  $f_{Y|X}(y|x)$  uniformly bounded by Assumption A(iii) and equation (41) yield the following inequality pointwise in  $t = (x, u)$ ,

$$\begin{aligned} (42) \quad |D_{\bar{\alpha}}[\pi](t) - D_{\tilde{\alpha}}[\pi](t)| &\lesssim \|\bar{\alpha} - \tilde{\alpha}\|_c^\nu \int_{\mathcal{X}} J(s_x) G^\nu(s_x) \left| \frac{dm_{\bar{\alpha}}(s_x, u)}{d\alpha} [\pi] \right| f_X(s_x) ds_x \\ &\quad + \int_{\mathcal{X}} \left| \frac{dm_{\bar{\alpha}}(s_x, u)}{d\alpha} [\pi] - \frac{dm_{\tilde{\alpha}}(s_x, u)}{d\alpha} [\pi] \right| f_X(s_x) ds_x \end{aligned}$$

Using (42), Cauchy-Schwarz and Jensen's inequality and  $E[J^2(X)G^{2\nu}(X)] < \infty$  yields:

$$\begin{aligned} (43) \quad \|D_{\bar{\alpha}}[\pi] - D_{\tilde{\alpha}}[\pi]\|_{L_\mu^2}^2 &\lesssim \|\bar{\alpha} - \tilde{\alpha}\|_c^{2\nu} \int_{\mathcal{X} \times (0,1)} \int_{\mathcal{X}} \left[ \frac{dm_{\bar{\alpha}}(s_x, u)}{d\alpha} [\pi] \right]^2 f_X(s_x) ds_x d\mu(t) \\ &\quad + \int_{\mathcal{X} \times (0,1)} \int_{\mathcal{X}} \left[ \frac{dm_{\bar{\alpha}}(s_x, u)}{d\alpha} [\pi] - \frac{dm_{\tilde{\alpha}}(s_x, u)}{d\alpha} [\pi] \right]^2 f_X(s_x) ds_x d\mu(t) \end{aligned}$$

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 (44), while the first inequality follows from (43). Further, since the

functional  $\left\| \frac{dm}{d\alpha}(\bar{\alpha}) \right\|_o : (\mathcal{N}(\alpha_0), \|\cdot\|_c) \rightarrow \mathbb{R}$  is continuous and  $\mathcal{A}$  is compact under  $\|\cdot\|_c$  it follows that  $\sup_{\mathcal{N}(\alpha_0)} \left\| \frac{dm(\bar{\alpha})}{d\alpha} \right\|_o < \infty$ . The second inequality in (44) then follows.

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

Therefore,  $D_{\bar{\alpha}}$  is continuous in  $\alpha$  by  $m$  being continuously Fréchet differentiable.

We now show  $D_{\bar{\alpha}}$  is indeed the Fréchet derivative of  $W$  at  $\bar{\alpha}$ . Straightforward manipulations imply that for any  $t = (x, u) \in \mathcal{X} \times (0, 1)$  we have

$$(45) \quad W_\alpha(t) = \int_{\mathcal{X}} P(Y \leq m_\alpha(s_x, u) | s_x) 1\{s_x \leq x\} f_X(s_x) ds_x - u \cdot P(X \leq x)$$

Next, using the definition of  $D_{\bar{\alpha}}$  and (45) together with Jensen's inequality we obtain (46) pointwise in  $t$  for any  $\bar{\alpha} \in \mathcal{N}(\alpha_0)$  and  $\pi \in \mathcal{A}$ .

$$\begin{aligned}
(46) \quad &|W_{\bar{\alpha}+\pi}(t) - W_{\bar{\alpha}}(t) - D_{\bar{\alpha}}[\pi](t)| \\
&\lesssim \int_{\mathcal{X}} \left| P(Y \leq m_{\bar{\alpha}+\pi}(s_x, u) | s_x) - P(Y \leq m_{\bar{\alpha}}(s_x, u) | s_x) - f_{Y|X}(m_{\bar{\alpha}}(s_x, u) | s_x) \frac{dm_{\bar{\alpha}}(s_x, u)}{d\alpha}[\pi] \right| f_X(s_x) ds_x
\end{aligned}$$

Applying the mean value theorem inside the integral in (46) then implies

$$\begin{aligned}
(47) \quad &|W_{\bar{\alpha}+\pi}(t) - W_{\bar{\alpha}}(t) - D_{\bar{\alpha}}[\pi](t)| \\
&\lesssim \int_{\mathcal{X}} \left| f_{Y|X}(\bar{m}(s_x, u) | s_x) [m_{\bar{\alpha}+\pi}(s_x, u) - m_{\bar{\alpha}}(s_x, u)] - f_{Y|X}(m_{\bar{\alpha}}(s_x, u) | s_x) \frac{dm_{\bar{\alpha}}(s_x, u)}{d\alpha}[\pi] \right| f_X(s_x) ds_x
\end{aligned}$$

where  $\bar{m}(s_x, u)$  is a convex combination of  $m_{\bar{\alpha}+\pi}(s_x, u)$  and  $m_{\bar{\alpha}}(s_x, u)$ . Therefore, it follows that  $|\bar{m}(s_x, u) - m_{\bar{\alpha}}(s_x, u)| \leq |m_{\bar{\alpha}+\pi}(s_x, u) - m_{\bar{\alpha}}(s_x, u)|$ . The Lipschitz conditions of Assumptions C(iv) and D(ii) then imply the inequality:

$$\begin{aligned}
(48) \quad &\int_{\mathcal{X}} \left| [f_{Y|X}(\bar{m}(s_x, u) | s_x) - f_{Y|X}(m_{\bar{\alpha}}(s_x, u) | s_x)] [m_{\bar{\alpha}+\pi}(s_x, u) - m_{\bar{\alpha}}(s_x, u)] \right| f_X(s_x) ds_x \\
&\leq \|\pi\|_c^{1+\nu} \int J(s_x) G^{1+\nu}(s_x) f_X(s_x) ds_x
\end{aligned}$$

Using (47), (48),  $f_{Y|X}(y|x)$  being bounded and Jensen's inequality in turn establishes the first inequality in (49). The final result in (49) then follows by  $\frac{dm}{d\alpha}$  being the Fréchet derivative of  $m$ .

$$\begin{aligned}
(49) \quad &\|W_{\bar{\alpha}+\pi}(t) - W_{\bar{\alpha}}(t) - D_{\bar{\alpha}}[\pi](t)\|_{L_\mu^2}^2 \\
&\lesssim \|\pi\|_c^{2+2\nu} + \int_{\mathcal{X} \times (0,1)} \int_{\mathcal{X}} \left[ m_{\bar{\alpha}+\pi}(s_x, u) - m_{\bar{\alpha}}(s_x, u) - \frac{dm_{\bar{\alpha}}(s_x, u)}{d\alpha}[\pi] \right]^2 f_X(s_x) ds_x d\mu(t) = o(\|\pi\|_c^2)
\end{aligned}$$

We conclude from (49) and (44) that  $D_{\bar{\alpha}}$  is the Fréchet derivative at  $\bar{\alpha}$  of  $W : (\mathcal{A}, \|\cdot\|_c) \rightarrow L^2_\mu$  and that it is continuous in  $\bar{\alpha}$ . To conclude the proof of the first claim of the Lemma, notice that  $Q(\alpha) = \|W_\alpha(t)\|_{L^2_\mu}^2$ . Since the functional  $\|\cdot\|_{L^2_\mu}^2 : L^2_\mu \rightarrow \mathbb{R}$  is trivially Fréchet differentiable, applying the Chain rule for Fréchet derivatives (see, e.g., Theorem 5.2.5 in Siddiqi, 2004) yields,

$$(50) \quad \frac{dQ(\bar{\alpha})}{d\alpha}[\pi] = \int_{\mathcal{X} \times (0,1)} W_{\bar{\alpha}}(t) D_{\bar{\alpha}}[\pi](t) d\mu(t)$$

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

$$(51) \quad T[\psi, \pi] = \int_{\mathcal{X} \times (0,1)} D_{\alpha_0}[\psi](t) D_{\alpha_0}[\pi](t) d\mu(t)$$

We will show  $T$  is the second Fréchet derivative of  $Q(\alpha)$  at  $\alpha_0$ . Notice that  $T[\psi, \cdot] : \mathcal{A} \rightarrow \mathbb{R}$  is a linear operator. The first requirement of Fréchet differentiability is to show  $T[\psi, \cdot]$  is continuous in  $\psi$ . For this purpose, notice that the first equality in (52) follows by definition while the first and second inequalities are implied by the Cauchy-Schwarz inequality and (40) respectively.

$$(52) \quad \begin{aligned} \|T[\psi, \cdot]\|_o^2 &= \sup_{\pi \in \bar{\mathcal{A}}_c} \|\pi\|_c^{-2} T^2[\psi, \pi] \\ &\leq \int_{\mathcal{X} \times (0,1)} D_{\alpha_0}^2[\psi](t) d\mu(t) \times \sup_{\pi \in \bar{\mathcal{A}}_c} \|\pi\|_c^{-2} \int_{\mathcal{X} \times (0,1)} D_{\alpha_0}^2[\pi](t) d\mu(t) \\ &\lesssim \left\| \frac{dm(\alpha_0)}{d\alpha} \right\|_o^4 \|\psi\|_c^2 \end{aligned}$$

It follows from (52) that  $T[\psi, \cdot]$  is continuous in  $\psi \in \mathcal{A}$ . Next, we verify  $T$  is the second Fréchet derivative of  $Q(\alpha)$  at  $\alpha_0$ . In (53) use (50) and  $W_{\alpha_0}(t) = 0$  for all  $t$  to notice  $\frac{dQ(\alpha_0)}{d\alpha} = 0$  and obtain:

$$(53) \quad \begin{aligned} &\left\| \frac{dQ(\alpha_0 + \psi)}{d\alpha} - \frac{dQ(\alpha_0)}{d\alpha} - T[\psi, \cdot] \right\|_o^2 \\ &= \sup_{\pi \in \bar{\mathcal{A}}_c} \|\pi\|_c^{-2} \left( \int_{\mathcal{X} \times (0,1)} W_{\alpha_0 + \psi}(t) D_{\alpha_0 + \psi}[\pi](t) d\mu(t) - \int_{\mathcal{X} \times (0,1)} D_{\alpha_0}[\psi](t) D_{\alpha_0}[\pi](t) d\mu(t) \right)^2 \end{aligned}$$

Next, use the Cauchy-Schwarz inequality to obtain the first inequality in (54) and  $D_{\alpha_0}$  being the Fréchet derivative of  $W : (\mathcal{A}, \|\cdot\|_c) \rightarrow L^2_\mu$  at  $\alpha_0$  for the second.

$$(54) \quad \begin{aligned} &\sup_{\pi \in \bar{\mathcal{A}}_c} \|\pi\|_c^{-2} \left( \int_{\mathcal{X} \times (0,1)} (W_{\alpha_0 + \psi}(t) - W_{\alpha_0}(t) - D_{\alpha_0}[\psi](t)) D_{\alpha_0 + \psi}[\pi](t) d\mu(t) \right)^2 \\ &\leq \|W_{\alpha_0 + \psi}(t) - W_{\alpha_0}(t) - D_{\alpha_0}[\psi](t)\|_{L^2_\mu}^2 \times \sup_{\pi \in \bar{\mathcal{A}}_c} \|\pi\|_c^{-2} \|D_{\alpha_0 + \psi}[\pi]\|_{L^2_\mu}^2 \leq o(\|\psi\|_c^2) \times \|D_{\alpha_0 + \psi}\|_o^2 \end{aligned}$$

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

$$(55) \quad \begin{aligned} &\sup_{\pi \in \bar{\mathcal{A}}_c} \|\pi\|_c^{-2} \left( \int_{\mathcal{X} \times (0,1)} D_{\alpha_0}[\psi](t) (D_{\alpha_0 + \psi}[\pi](t) - D_{\alpha_0}[\pi](t)) d\mu(t) \right)^2 \\ &\leq \|D_{\alpha_0}\|_o^2 \|\psi\|_c^2 \times \sup_{\pi \in \bar{\mathcal{A}}_c} \|\pi\|_c^{-2} \|D_{\alpha_0 + \psi}[\pi] - D_{\alpha_0}[\pi]\|_{L^2_\mu}^2 \leq \|D_{\alpha_0}\|_o^2 \|\psi\|_c^2 \times \|D_{\alpha_0 + \psi} - D_{\alpha_0}\|_o^2 \end{aligned}$$

To conclude, combine (53), (54), (55) and  $W_{\alpha_0}(t) = 0$  for all  $t$  to derive the first inequality in (56). As argued in (44), however,  $\|D_{\bar{\alpha}}\|_o$  is bounded in a neighborhood of  $\alpha_0$ . Thus, the continuity of  $D_{\bar{\alpha}}$  in  $\bar{\alpha}$  for  $\bar{\alpha} \in \mathcal{N}(\alpha_0)$  implies the final result in (56).

$$(56) \quad \left\| \frac{dQ(\alpha_0 + \psi)}{d\alpha} - \frac{dQ(\alpha_0)}{d\alpha} - T[\psi, \cdot] \right\|_o^2 \\ \leq o(\|\psi\|_c^2) \times \|D_{\alpha_0 + \psi}\|_o^2 + \|\psi\|_c^2 \|D_{\alpha_0}\|_o^2 \|D_{\alpha_0 + \psi} - D_{\alpha_0}\|_o^2 = o(\|\psi\|_c^2)$$

It follows from (56) that  $T$  is the second Fréchet derivative of  $Q(\alpha)$  at  $\alpha_0$ . ■

PROOF OF THEOREM 2: Let  $\Pi_n \alpha_0 = \arg \min_{\mathcal{A}_n} \|\alpha_0 - \alpha\|_s$ . By Theorem 1,  $\hat{\alpha} \in \mathcal{N}(\alpha_0)$  with probability tending to one and hence Assumptions E(i) and E(iii), imply that with probability tending to one we have that:

$$(57) \quad \|\hat{\alpha} - \alpha_0\|_w^2 \lesssim Q(\hat{\alpha}) - Q(\Pi_n \alpha_0) + Q(\Pi_n \alpha_0) \\ = Q(\hat{\alpha}) - Q(\Pi_n \alpha_0) + o(n^{-1})$$

By Theorem 1 and  $\|\cdot\|_s \lesssim \|\cdot\|_c$ , there is a  $\delta_n \rightarrow 0$  such that  $P(\|\hat{\alpha} - \alpha_0\|_s > \delta_n) \rightarrow 0$ . Letting  $\mathcal{A}_0^{\delta_n} = \{\alpha \in \mathcal{A} : \|\alpha - \alpha_0\|_s \leq \delta_n\}$  then yields the first inequality in (58). Noticing that  $Q_n(\hat{\alpha}) \leq Q_n(\Pi_n \alpha_0)$  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 5 which implies  $\sqrt{n}(W_{\alpha,n}(t) - W_\alpha(t))$  is tight in  $L^\infty(\mathbb{R}^{d_t} \times \mathcal{A})$  together with the definition of  $Q(\alpha)$ .

$$(58) \quad Q(\hat{\alpha}) - Q(\Pi_n \alpha_0) \leq Q_n(\hat{\alpha}) - Q_n(\Pi_n \alpha_0) + 2 \sup_{\mathcal{A}_0^{\delta_n}} |Q_n(\alpha) - Q(\alpha)| \\ \leq 2 \sup_{(t,\alpha) \in \mathbb{R}^{d_t} \times \mathcal{A}} |W_{\alpha,n}(t) - W_\alpha(t)| \times \left[ \sup_{\mathcal{A}_0^{\delta_n}} \int_{\mathcal{X} \times (0,1)} (W_{\alpha,n}(t) + W_\alpha(t))^2 d\mu(t) \right]^{\frac{1}{2}} \\ \leq O_p(n^{-\frac{1}{2}}) \times \left[ \sup_{(t,\alpha) \in \mathbb{R}^{d_t} \times \mathcal{A}} (W_{\alpha,n}(t) - W_\alpha(t))^2 + \sup_{\mathcal{A}_0^{\delta_n}} 4 \int_{\mathcal{X} \times (0,1)} W_\alpha^2(t) d\mu(t) \right]^{\frac{1}{2}} \\ \leq O_p(n^{-\frac{1}{2}}) \times [O_p(n^{-1}) + \sup_{\mathcal{A}_0^{\delta_n}} 4Q(\alpha)]^{\frac{1}{2}}$$

By Assumption E(i),  $\sup_{\mathcal{A}_0^{\delta_n}} Q(\alpha) \lesssim \delta_n^2 = o(1)$ . Therefore, combining (57) and (58):

$$(59) \quad \|\hat{\alpha} - \alpha_0\|_w^2 \lesssim O_p(n^{-\frac{1}{2}}) \times o_p(1) + o(n^{-1}) = o_p(n^{-\frac{1}{2}})$$

To obtain a rate with respect to  $\|\cdot\|_s$ , we use Assumption E(iii) for the first and second inequalities in (60). Further, it follows from (59) and Assumption E(iii) that  $\|\hat{\alpha} - \Pi_n \alpha_0\|_w^2 = o_p(n^{-\frac{1}{2}})$  which together with Assumption E(ii) implies the equality in (60).

$$(60) \quad \|\hat{\alpha} - \alpha_0\|_s^2 \leq \|\hat{\alpha} - \Pi_n \alpha_0\|_s^2 + o(n^{-1}) \leq \sup_{\alpha \in \mathcal{A}_n} \frac{\|\alpha\|_s^2}{\|\alpha\|_w^2} \times \|\hat{\alpha} - \Pi_n \alpha_0\|_w^2 + o(n^{-1}) = o_p(n^{-\frac{1}{2} + \gamma})$$

We can now exploit the local behavior of the objective function to improve on the obtained rate of convergence. Notice that due to (60) it is possible to choose  $\delta_n = o(n^{-\frac{1}{4} + \frac{\gamma}{2}})$  such that  $P(\hat{\alpha} \in \mathcal{A}_0^{\delta_n}) \rightarrow 1$ . Repeating the steps in (58) we obtain (61) with probability approaching one.

$$(61) \quad \begin{aligned} Q(\hat{\alpha}) - Q(\Pi_n \alpha_0) &\leq O_p(n^{-\frac{1}{2}}) \times [O_p(n^{-1}) + \sup_{\mathcal{A}_0^{\delta_n}} 4Q(\alpha)]^{\frac{1}{2}} \\ &= O_p(n^{-\frac{1}{2}}) \times o_p(n^{-\frac{1}{4} + \frac{\gamma}{2}}) \end{aligned}$$

From (57) and (61) and Assumption E(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 (60) 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  $\|\hat{\alpha} - \alpha_0\|_s^2 = o_p(n^{(\gamma - \frac{1}{2})(1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \dots)})$ . 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 4

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

$$(62) \quad Q_n^s(\alpha) = \int_{\mathcal{X} \times (0,1)} (W_{\alpha_0,n}(t) + W_\alpha(t))^2 d\mu(t)$$

Throughout the proofs we will exploit the following Lemma,

**Lemma 6.** *If Assumptions A, C, D and E hold, then:  $Q_n^s(\hat{\alpha}) \leq \inf_{\mathcal{A}_n} Q_n^s(\alpha) + o_p(n^{-1})$ .*

PROOF: Since  $\|\hat{\alpha} - \alpha_0\|_c = o_p(1)$  and  $W_{\alpha_0}(t) = 0$  for all  $t \in \mathcal{X} \times (0, 1)$ , Lemma 5 implies that

$$\sup_{t \in \mathcal{X} \times (0,1)} |W_{\hat{\alpha},n}(t) - W_{\hat{\alpha}}(t) - W_{\alpha_0,n}(t)| = o_p(n^{-\frac{1}{2}}).$$

By simple manipulations we therefore obtain:

$$(63) \quad \begin{aligned} Q_n^s(\hat{\alpha}) &\leq \int_{\mathcal{X} \times (0,1)} (|W_{\alpha_0,n}(t) + W_{\hat{\alpha}}(t) - W_{\hat{\alpha},n}(t)| + |W_{\hat{\alpha},n}(t)|)^2 d\mu(t) \\ &= \int_{\mathcal{X} \times (0,1)} W_{\hat{\alpha},n}^2(t) d\mu(t) + o_p(n^{-\frac{1}{2}}) \times \int_{\mathcal{X} \times (0,1)} |W_{\hat{\alpha},n}(t)| d\mu(t) + o_p(n^{-1}) \end{aligned}$$

Next, apply Jensen's inequality and  $Q_n(\hat{\alpha}) \leq Q_n(\Pi_n \alpha_0)$  to obtain the first and second inequalities in (64). By Lemma 5,  $\sup_{t, \alpha} |W_{\alpha,n}(t) - W_\alpha(t)| = O_p(n^{-\frac{1}{2}})$ . Together with Assumption E(i), the

final two inequalities in (64) then immediately follow.

$$\begin{aligned}
\int_{\mathcal{X} \times (0,1)} |W_{\hat{\alpha},n}(t)| d\mu(t) &\leq \left[ \int_{\mathcal{X} \times (0,1)} W_{\hat{\alpha},n}^2(t) d\mu(t) \right]^{\frac{1}{2}} \\
&\leq \left[ \int_{\mathcal{X} \times (0,1)} W_{\Pi_n \alpha_0, n}^2(t) d\mu(t) \right]^{\frac{1}{2}} \\
&\leq \left[ 2 \int_{\mathcal{X} \times (0,1)} (W_{\Pi_n \alpha_0, n}(t) - W_{\Pi_n \alpha_0}(t))^2 d\mu(t) + 2 \int_{\mathcal{X} \times (0,1)} W_{\Pi_n \alpha_0}^2(t) d\mu(t) \right]^{\frac{1}{2}} \\
(64) \quad &\lesssim [O_p(n^{-1}) + \|\Pi_n \alpha_0 - \alpha_0\|_s^2]^{\frac{1}{2}}
\end{aligned}$$

By Assumption E(iii),  $\|\Pi_n \alpha_0 - \alpha_0\|_s = o(n^{-\frac{1}{2}})$  and therefore combining (63) and (64),

$$(65) \quad Q_n^s(\hat{\alpha}) \leq Q_n(\hat{\alpha}) + o_p(n^{-1})$$

Let  $\tilde{\alpha} \in \arg \min_{\mathcal{A}_n} Q_n^s(\alpha)$ , and notice that Lemma 5 and the same arguments as in Theorem 1 imply that  $\|\alpha_0 - \tilde{\alpha}\|_c = o_p(1)$ . The same arguments as in (63) then imply that:

$$\begin{aligned}
Q_n(\tilde{\alpha}) &\leq \int_{\mathcal{X} \times (0,1)} (|W_{\tilde{\alpha},n}(t) - W_{\alpha_0,n}(t) - W_{\tilde{\alpha}}(t)| + |W_{\alpha_0,n}(t) + W_{\tilde{\alpha}}(t)|)^2 d\mu(t) \\
(66) \quad &= \int_{\mathcal{X} \times (0,1)} (W_{\alpha_0,n}(t) + W_{\tilde{\alpha}}(t))^2 d\mu(t) + o_p(n^{-\frac{1}{2}}) \times \int_{\mathcal{X} \times (0,1)} |W_{\alpha_0,n}(t) + W_{\tilde{\alpha}}(t)| d\mu(t) + o_p(n^{-1})
\end{aligned}$$

Proceeding as in (64), Jensen's inequality and  $Q_n^s(\tilde{\alpha}) \leq Q_n^s(\Pi_n \alpha_0)$  imply the first and second inequalities in (67). The last two results in (67) then follow by Assumption E(i) and by noting that Lemma 5 implies  $\sup_t |W_{\alpha_0,n}(t)| = O_p(n^{-\frac{1}{2}})$ .

$$\begin{aligned}
\int_{\mathcal{X} \times (0,1)} |W_{\alpha_0,n}(t) + W_{\tilde{\alpha}}(t)| d\mu(t) &\leq \left[ \int_{\mathcal{X} \times (0,1)} (W_{\alpha_0,n}(t) + W_{\tilde{\alpha}}(t))^2 d\mu(t) \right]^{\frac{1}{2}} \\
&\leq \left[ \int_{\mathcal{X} \times (0,1)} (W_{\alpha_0,n}(t) + W_{\Pi_n \alpha_0}(t))^2 d\mu(t) \right]^{\frac{1}{2}} \\
&\leq \left[ 2 \int_{\mathcal{X} \times (0,1)} W_{\alpha_0,n}^2(t) d\mu(t) + 2 \int_{\mathcal{X} \times (0,1)} W_{\Pi_n \alpha_0}^2(t) d\mu(t) \right]^{\frac{1}{2}} \\
(67) \quad &\lesssim [O_p(n^{-1}) + \|\alpha_0 - \Pi_n \alpha_0\|_s^2]^{\frac{1}{2}}
\end{aligned}$$

Since  $\|\Pi_n \alpha_0 - \alpha_0\|_s = o(n^{-\frac{1}{2}})$  by Assumption E(iii), (66) and (67) imply,

$$(68) \quad Q_n(\tilde{\alpha}) \leq Q_n^s(\tilde{\alpha}) + o_p(n^{-1})$$

Hence, since  $Q_n(\hat{\alpha}) \leq Q_n(\tilde{\alpha})$ , the definition of  $\tilde{\alpha}$  together with (65) and (68) establish

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

which establishes the claim of the Lemma. ■

PROOF OF LEMMA 3: The arguments closely follow those of Ai and Chen (2003). We first establish continuity. Since  $F_\lambda$  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 (69) by using (12), while the second equality is definitional.

$$(69) \quad \min_{h \in \mathcal{H}} \int_{\mathcal{X} \times (0,1)} \left( \frac{dW(\alpha_0)}{d\theta'}[\theta](t) - \frac{dW(\alpha_0)}{dh}[h](t) \right)^2 d\mu(t) \\ = \int_{\mathcal{X} \times (0,1)} \left( \left[ \frac{dW(\alpha_0)}{d\theta}(t) - \frac{dW(\alpha_0)}{dh}[h^*](t) \right]' \theta \right)^2 d\mu(t) = \theta' \Sigma^* \theta$$

In order to show  $F_\lambda$  is bounded we need to establish the left hand side of (70) is finite. Using (69) immediately implies the first equality in (70). For the second equality notice the optimization problem is solved at  $\theta^* = (\Sigma^*)^{-1}\lambda$  and plug in  $\theta^*$ .

$$(70) \quad \sup_{0 \neq \alpha \in \bar{\mathcal{A}}} \frac{F_\lambda^2(\alpha)}{\|\alpha\|_w^2} = \sup_{0 \neq \theta \in \mathbb{R}^{d_\theta}} \frac{(\lambda' \theta)^2}{\theta' \Sigma^* \theta} = \lambda' (\Sigma^*)^{-1} \lambda$$

Since by assumption  $\Sigma^*$  is positive-definite, (70) is finite and hence  $F_\lambda$  is bounded which establishes continuity. For the second claim of the Lemma, notice the following orthogonality condition must hold as a result of (11) and (12):

$$(71) \quad \int_{\mathcal{X} \times (0,1)} \left( \frac{dW(\alpha_0)}{d\theta}(t) - \frac{dW(\alpha_0)}{dh}[h^*](t) \right) \frac{dW(\alpha_0)}{dh}[h](t) d\mu(t) = 0$$

for all  $h \in \bar{\mathcal{H}}$ . Therefore, employing result (71) and the definition of  $\langle \alpha - \alpha_0, v^\lambda \rangle$  we obtain:

$$\langle \alpha - \alpha_0, v^\lambda \rangle_w \\ = \int_{\mathcal{X} \times (0,1)} \left[ \frac{dW(\alpha_0)}{d\theta}[\theta - \theta_0](t) + \frac{dW(\alpha_0)}{dh}[h - h_0](t) \right] \left[ \frac{dW(\alpha_0)}{d\theta}[v^\lambda](t) + \frac{dW(\alpha_0)}{dh}[v_h^\lambda](t) \right] d\mu(t) \\ = (\theta - \theta_0)' \left\{ \int_{\mathcal{X} \times (0,1)} \left[ \frac{dW(\alpha_0)}{d\theta}(t) - \frac{dW(\alpha_0)}{dh}[h^*](t) \right] \left[ \frac{dW(\alpha_0)}{d\theta}(t) - \frac{dW(\alpha_0)}{dh}[h^*](t) \right]' d\mu(t) \right\} v_\theta^\lambda,$$

Hence, since  $v_\theta^\lambda = (\Sigma^*)^{-1}\lambda$ , the second claim of the Lemma follows. ■

**Lemma 7.** *Let Assumption A, C, D, E and F hold, and let  $v_n^\lambda = \Pi_n v^\lambda$ . Then:*

- (i)  $\int_{\mathcal{X} \times (0,1)} W_{\alpha_0,n}(t) D_{\hat{\alpha}}[v_n^\lambda](t) d\mu(t) = \int_{\mathcal{X} \times (0,1)} W_{\alpha_0,n}(t) D_{\alpha_0}[v^\lambda](t) d\mu(t) + o_p(n^{-\frac{1}{2}})$
- (ii)  $\int_{\mathcal{X} \times (0,1)} (W_{\hat{\alpha}}(t) - W_{\alpha_0}(t)) D_{\hat{\alpha}}[v_n^\lambda](t) d\mu(t) = \int_{\mathcal{X} \times (0,1)} D_{\alpha_0}[\hat{\alpha} - \alpha_0](t) D_{\alpha_0}[v^\lambda](t) d\mu(t) + o_p(n^{-\frac{1}{2}})$
- (iii)  $\sqrt{n} W_{\alpha_0,n}(t) \xrightarrow{\mathcal{L}} G(t)$  where  $G(t)$  a Gaussian process with covariance:

$$\Sigma(t, t') = E[(1\{U \leq u; X \leq x\} - u1\{X \leq x\})(1\{U \leq u'; X \leq x'\} - u'1\{X \leq x'\})]$$

PROOF: To establish the first claim apply the Cauchy-Schwarz inequality, the definition of the operator norm, Theorem 1 and Lemma 5 implying  $\sup_t |W_{\alpha_0, n}(t)| = O_p(n^{-\frac{1}{2}})$  to obtain that with probability approaching one we have:

$$(72) \quad \left| \int_{\mathcal{X} \times (0,1)} W_{\alpha_0, n}(t) D_{\hat{\alpha}}[v_n^\lambda - v^\lambda](t) d\mu(t) \right| \leq \left[ \int_{\mathcal{X} \times (0,1)} W_{\alpha_0, n}^2(t) d\mu(t) \right]^{\frac{1}{2}} \times \|D_{\hat{\alpha}}[v_n^\lambda - v^\lambda]\|_{L_\mu^2} \\ \leq O_p(n^{-\frac{1}{2}}) \times \sup_{\alpha \in \mathcal{N}(\alpha_0)} \|D_\alpha\|_o \times \|v_n^\lambda - v^\lambda\|_c$$

As argued in (44),  $\sup_{\alpha \in \mathcal{N}(\alpha_0)} \|D_\alpha\|_o < \infty$ . Further, Assumption F(ii) and C(vi) imply that  $\|v^\lambda - v_n^\lambda\|_c = o(1)$ . Therefore, we obtain from (72) that,

$$(73) \quad \int_{\mathcal{X} \times (0,1)} W_{\alpha_0, n}(t) D_{\hat{\alpha}}[v_n^\lambda](t) d\mu(t) = \int_{\mathcal{X} \times (0,1)} W_{\alpha_0, n}(t) D_{\hat{\alpha}}[v^\lambda](t) d\mu(t) + o_p(n^{-\frac{1}{2}})$$

Similarly, the derivations in (72) imply the inequality in (74). The equality is a result of the continuity of  $D_\alpha$  in  $\alpha$  under  $\|\cdot\|_c$ , as established in the proof of Lemma 2.

$$(74) \quad \left| \int_{\mathcal{X} \times (0,1)} W_{\alpha_0, n}(t) (D_{\hat{\alpha}}[v^\lambda](t) - D_{\alpha_0}[v^\lambda](t)) d\mu(t) \right| \\ \leq O_p(n^{-\frac{1}{2}}) \times \|D_{\hat{\alpha}} - D_{\alpha_0}\|_o \times \|v^\lambda\|_c = o_p(n^{-\frac{1}{2}})$$

Together, (73) and (74) establish the first claim of the Lemma.

For the second claim of the Lemma, notice that Assumption F(iii) allows us to do a second order Taylor expansion to obtain (75) pointwise in  $t \in \mathcal{X} \times (0, 1)$ ,

$$(75) \quad W_{\hat{\alpha}}(t) = W_{\alpha_0}(t) + D_{\alpha_0}[\hat{\alpha} - \alpha_0](t) + \frac{1}{2} \frac{dD_{\alpha_0 + \tau(\hat{\alpha} - \alpha_0)}[\hat{\alpha} - \alpha_0](t)}{d\tau} \Big|_{\tau=s(t)}$$

The first equality in (76) then follows from (75), while the second one is implied by Assumptions F(iii) and F(iv). The final equality in turn follows from Theorem 2.

$$(76) \quad \int_{\mathcal{X} \times (0,1)} (W_{\hat{\alpha}}(t) - W_{\alpha_0}(t) - D_{\alpha_0}[\hat{\alpha} - \alpha_0](t)) D_{\hat{\alpha}}[v_n^\lambda](t) d\mu(t) \\ = \frac{1}{2} \int_{\mathcal{X} \times (0,1)} \left( \frac{dD_{\alpha_0 + \tau(\hat{\alpha} - \alpha_0)}[\hat{\alpha} - \alpha_0](t)}{d\tau} \Big|_{\tau=s(t)} \right) \times D_{\hat{\alpha}}[v_n^\lambda](t) d\mu(t) \lesssim \|\hat{\alpha} - \alpha_0\|_s^2 = o_p(n^{-\frac{1}{2}})$$

Next, apply the Cauchy-Schwarz inequality and a Taylor expansion to obtain the first inequality in (77). The second inequality then follows by Assumption F(iii),  $\|\hat{\alpha} - \alpha_0\|_w \lesssim \|\hat{\alpha} - \alpha_0\|_s$  in a neighborhood of  $\alpha_0$ , as implied by Assumption E(i), and Theorem 2.

$$(77) \quad \left| \int_{\mathcal{X} \times (0,1)} D_{\alpha_0}[\hat{\alpha} - \alpha_0](t) (D_{\hat{\alpha}}[v_n^\lambda](t) - D_{\alpha_0}[v_n^\lambda](t)) d\mu(t) \right| \\ \leq \|\hat{\alpha} - \alpha_0\|_w \times \left[ \int_{\mathcal{X} \times (0,1)} \left( \frac{dD_{\alpha_0 + \tau(\hat{\alpha} - \alpha_0)}[v_n^\lambda](t)}{d\tau} \Big|_{\tau=s(t)} \right)^2 d\mu(t) \right]^{\frac{1}{2}} \lesssim \|\hat{\alpha} - \alpha_0\|_s^2 = o_p(n^{-\frac{1}{2}})$$

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

$$(78) \quad \left| \int_{\mathcal{X} \times (0,1)} D_{\alpha_0}[\hat{\alpha} - \alpha_0](t) (D_{\alpha_0}[v_n^\lambda](t) - D_{\alpha_0}[v^\lambda](t)) d\mu(t) \right| \\ \leq \|\hat{\alpha} - \alpha_0\|_w \times \|D_{\alpha_0}\|_o \times \|v_n^\lambda - v^\lambda\|_c = o_p(n^{-\frac{1}{2}})$$

Combining results (76), (77) and (78) establishes the second claim of the Lemma. The third claim of the Lemma is immediate from  $W_{\alpha_0, n}(t)$  being a Donsker class due to Lemma 5 and regular central limit theorem. ■

PROOF OF THEOREM 3: Let  $u^* = \pm v^\lambda$ ,  $u_n^* = \Pi_n u^*$  and  $0 < \epsilon_n = o(n^{-\frac{1}{2}})$  be such that it satisfies  $Q_n^s(\hat{\alpha}) \leq \inf_{\mathcal{A}_n} Q_n^s(\alpha) + O_p(\epsilon_n^2)$ , which is possible due to Lemma 6. Define  $\alpha(\tau) = \hat{\alpha} + \tau \epsilon_n u_n^*$  and note that by Assumption D(i) and Lemma 1, with probability tending to one  $\alpha(\tau) \in \mathcal{A}_n$  for  $\tau \in [0, 1]$ . Therefore, Lemma 6 establishes the first equality in (79). A second order Taylor expansion around  $\tau = 0$  yields the equality in (79) for some  $s \in [0, 1]$ .

$$(79) \quad 0 \leq Q_n^s(\alpha(1)) - Q_n^s(\alpha(0)) + O_p(\epsilon_n^2) \\ = 2\epsilon_n \int_{\mathcal{X} \times (0,1)} (W_{\alpha_0, n}(t) + W_{\hat{\alpha}}(t)) D_{\hat{\alpha}}[u_n^*](t) d\mu(t) + \frac{1}{2} \frac{d^2 Q_n(\alpha(\tau))}{d\tau^2} \Big|_{\tau=s}$$

where by direct calculation we have that:

$$(80) \quad \frac{d^2 Q_n(\alpha(\tau))}{d\tau^2} \Big|_{\tau=s} = \epsilon_n^2 \int_{\mathcal{X} \times (0,1)} (D_{\alpha(s)}[u_n^*](t))^2 d\mu(t) \\ + \int_{\mathcal{X} \times (0,1)} (W_{\alpha_0, n}(t) + W_{\alpha(s)}(t)) \frac{dD_{\hat{\alpha} + \tau \epsilon_n u_n^*}[\epsilon_n u_n^*](t)}{d\tau} \Big|_{\tau=s} d\mu(t)$$

As shown in (44),  $\sup_{\alpha \in \mathcal{N}(\alpha_0)} \|D_\alpha\|_o < \infty$ , and hence, since  $\|v^\lambda\|_c < \infty$  we obtain that:

$$(81) \quad \int_{\mathcal{X} \times (0,1)} (D_{\alpha(s)}[u_n^*](t))^2 d\mu(t) \leq \sup_{\alpha \in \mathcal{N}(\alpha_0)} \|D_\alpha\|_o^2 \times \|u_n^*\|_c^2 = O(1)$$

Since  $W_{\alpha_0, n}(t)$  and  $W_\alpha(t)$  are both bounded by 1, Assumption F(iii) establishes:

$$(82) \quad \left| \int_{\mathcal{X} \times (0,1)} (W_{\alpha_0, n}(t) + W_{\alpha(s)}(t)) \frac{dD_{\hat{\alpha} + \tau \epsilon_n u_n^*}[\epsilon_n u_n^*](t)}{d\tau} \Big|_{\tau=s} d\mu(t) \right| \lesssim \|\epsilon_n u_n^*\|_s^2 = O(\epsilon_n^2)$$

Therefore, by combining (79)-(82),  $u_n^* = \pm v_n^\lambda$  and  $\epsilon_n = o(n^{-\frac{1}{2}})$ , it follows that:

$$(83) \quad \int_{\mathcal{X} \times (0,1)} (W_{\alpha_0, n}(t) + W_{\hat{\alpha}}(t)) D_{\hat{\alpha}}[u_n^*](t) d\mu(t) = o_p(n^{-\frac{1}{2}})$$

To conclude, in (84) use Lemma 3 for the first equality, Lemma 7(ii) for the second equality,  $W_{\alpha_0}(t) = 0$  and (83) for the third one and Lemma 7(i) for the final result.

$$\begin{aligned}
\sqrt{n}\lambda'(\hat{\theta} - \theta_0) &= \sqrt{n} \int_{\mathcal{X} \times (0,1)} D_{\alpha_0}[\hat{\alpha} - \alpha_0](t) D_{\alpha_0}[v^\lambda](t) d\mu(t) \\
&= \sqrt{n} \int_{\mathcal{X} \times (0,1)} (W_{\hat{\alpha}}(t) - W_{\alpha_0}(t)) D_{\hat{\alpha}}[v_n^\lambda](t) d\mu(t) + o_p(1) \\
&= \sqrt{n} \int_{\mathcal{X} \times (0,1)} W_{\alpha_0,n}(t) D_{\hat{\alpha}}[v_n^\lambda](t) d\mu(t) + o_p(1) \\
(84) \qquad \qquad \qquad &= \sqrt{n} \int_{\mathcal{X} \times (0,1)} W_{\alpha_0,n}(t) D_{\alpha_0}[v^\lambda](t) d\mu(t) + o_p(1)
\end{aligned}$$

Hence, applying Lemma 7(iii) we are able to conclude from (84) that:

$$(85) \qquad \qquad \qquad \sqrt{n}\lambda'(\hat{\theta} - \theta) \xrightarrow{L} N(0, \Omega_\lambda)$$

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  $v^\lambda$ , obtained in Lemma 3, and the definition of  $R_{h^*}(t)$  in turn imply:

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

The Cramer-Wold device, (85) and (86) in turn establish the claim of the Theorem. ■

#### APPENDIX E. DETAILS OF THE BLP EXAMPLE

In this Appendix, we give the proofs of Lemma 4 and Theorem 4. We start with an auxiliary Lemma whose result will be useful later on.

**Lemma 8.** *Assume  $F$  is twice continuously differentiable on  $\mathbb{R}$  with strictly increasing hazard rate  $\tau$ . Then the BLP equilibrium prices exist, are unique, and the map  $(\xi_1 - \xi_2, x_1 - x_2, c) \mapsto (p_1 - p_2)$  is twice continuously differentiable with:*

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

PROOF OF LEMMA 8: Under the strictly increasing hazard rate assumption the goods are substitutes, and since  $f'(\varepsilon)[1 - F(\varepsilon)] + f^2(\varepsilon) > 0$  and  $f'(\varepsilon)F(\varepsilon) - f^2(\varepsilon) < 0$  we have that

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

i.e. the elasticity of demand is a decreasing function of the other firm's prices. 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

$$\frac{\partial^2 \ln \Pi_j(p_j, p_{-j})}{\partial p_j^2} < 0, \quad \frac{\partial^2 \ln \Pi_j(p_j, p_{-j})}{\partial p_j \partial p_{-j}} > 0$$

and

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

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

Since under the strictly increasing hazard rate assumption we have  $f'(\varepsilon)[1 - F(\varepsilon)] + f^2(\varepsilon) > 0$  and  $f'(\varepsilon)F(\varepsilon) - f^2(\varepsilon) < 0$  it also holds that  $\partial^2 \ln D_j(p_j, p_{-j})/\partial p_j^2 < 0$ , 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  $\xi = \xi_1 - \xi_2$  and

$$\Phi(p_1, p_2, \xi) = \left[ \begin{array}{c} \frac{1}{p_1 - c} + \frac{\partial \ln D_1(p_1, p_2)}{\partial p_1} \\ \frac{1}{p_2 - c} + \frac{\partial \ln D_2(p_1, p_2)}{\partial p_2} \end{array} \right]$$

Note that the map  $\Phi$  is continuously differentiable and we have:

$$D_{(p_1, p_2)} \Phi = \left[ \begin{array}{cc} -\frac{1}{(p_1 - c)^2} + \frac{\partial^2 \ln D_1(p_1, p_2)}{\partial p_1^2} & \frac{\partial^2 \ln D_1(p_1, p_2)}{\partial p_1 \partial p_2} \\ \frac{\partial^2 \ln D_2(p_1, p_2)}{\partial p_1 \partial p_2} & -\frac{1}{(p_2 - c)^2} + \frac{\partial^2 \ln D_2(p_1, p_2)}{\partial p_2^2} \end{array} \right]$$

In addition, note that the demand function in (17) satisfies:

$$(87) \quad -\frac{\partial^2 \ln D_j(p_j, p_{-j})}{\partial p_j^2} = \frac{\partial^2 \ln D_j(p_j, p_{-j})}{\partial p_j \partial p_{-j}} = a \frac{\partial^2 \ln D_j(p_j, p_{-j})}{\partial p_j \partial (\xi_j - \xi_{-j})} > 0$$

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

$$\det D_{(p_1, p_2)} \Phi = \frac{1}{(p_1 - c)^2 (p_2 - c)^2} - \frac{1}{(p_1 - c)^2} \frac{\partial^2 \ln D_2(p_1, p_2)}{\partial p_2^2} - \frac{1}{(p_2 - c)^2} \frac{\partial^2 \ln D_1(p_1, p_2)}{\partial p_1^2} > 0$$

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  $(p_1^*, p_2^*, \xi)$  a mapping  $\xi \mapsto (p_1, p_2)$  that is continuously differentiable, and whose derivative at this point equals:

$$\left[ \begin{array}{c} \frac{\partial p_1}{\partial \xi} \\ \frac{\partial p_2}{\partial \xi} \end{array} \right] = -[D_{(p_1, p_2)} \Phi(p_1, p_2, \xi)]^{-1} D_\xi \Phi(p_1, p_2, \xi).$$

Thus,

$$(88) \quad \frac{\partial p_1}{\partial \xi} = -\frac{1}{a} \frac{1}{\det D_{(p_1, p_2)} \Phi} \frac{1}{(p_2 - c)^2} \frac{\partial^2 \ln D_1(p_1, p_2)}{\partial p_1^2}$$

$$(89) \quad \frac{\partial p_2}{\partial \xi} = \frac{1}{a} \frac{1}{\det D_{(p_1, p_2)} \Phi} \frac{1}{(p_1 - c)^2} \frac{\partial^2 \ln D_2(p_1, p_2)}{\partial p_2^2}$$

where the first equality uses (87) 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 (87) 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 (88)-(89) we then have the desired result:

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

which concludes the proof of the Lemma. ■

PROOF OF LEMMA 4: Since  $\xi = \xi_1 - \xi_2$  is continuously distributed, it has a strictly increasing cdf, which we denote  $F_\xi$ . Noting that  $F_\xi(\xi) \sim U(0, 1)$ , we may define:

$$h(X, U) \equiv -a(p_1 - p_2) + F_\xi^{-1}(U), \quad \text{with } X \equiv x_1 - x_2$$

so that

$$Y \equiv F^{-1} \left( \frac{D_1(p_1, p_2)}{M} \right) = h(X, U) + \theta' X, \quad \text{where } \theta \equiv b$$

Now, note that by Lemma 8  $h$  is continuously differentiable and we have for all  $(x, u) \in \mathcal{X} \times (0, 1)$ :

$$\frac{\partial h(x, u)}{\partial u} = \left[ -a \frac{\partial(p_1 - p_2)}{\partial \xi} + 1 \right] \frac{1}{f_\xi(F_\xi^{-1}(u))} > 0$$

which completes the proof of Lemma 4. ■

PROOF OF THEOREM 4: Consider the BLP model in (19) and let  $F_{Y|X}(\cdot|\cdot)$  denote the conditional distribution of  $Y$  given  $X$  that is induced by the structure  $(\theta, h)$ . Fix  $x \in \mathcal{X}$  and let  $v : \mathbb{R} \times \mathcal{X} \rightarrow (0, 1)$  be such that for any  $u \in (0, 1)$ , 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  $(y, x) \in \mathcal{S}$ ,

$$\begin{aligned}
F_{Y|X}(y|x) &= P(Y \leq y | X = x) \\
&= P(h(X, U) \leq y - \theta'x | X = x) \\
&= P(U \leq v(y - \theta'x, x) | X = x) \\
&= P(U \leq v(y - \theta'x, x)) \\
(90) \qquad &= v(y - \theta'x, x)
\end{aligned}$$

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

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

Further, for any  $(y, x) \in \mathcal{S}$  let  $\Phi(y, x) \equiv F_{Y|X}(y|x)$ . Under our assumptions on  $F$ ,  $\Phi(y, x)$  is continuously differentiable on  $\mathcal{S}$  and we have:

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

In particular,  $\partial \Phi(y, x)/\partial y > 0$  on  $\mathcal{S}$ . Combining (91) and (92) we then obtain:

$$(93) \qquad -\left[ \frac{\partial \Phi}{\partial x}(y, x) \right] \left[ \frac{\partial \Phi}{\partial y}(y, x) \right]^{-1} = \theta + \frac{\partial h}{\partial x}(x, v(y - \theta'x, x))$$

Evaluate the left hand side of (93) at  $x = 0 \in \mathcal{X}$  and  $y = 0$ . For these values of  $x$  and  $y$ , we have:  $y - \theta'x = 0$  so by using condition (i) of Theorem 4,  $v(0, 0) = 1/2$ . Combining the latter with condition (ii) then gives:

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

from which it follows that  $\theta$  is identified. The identification of  $v$ , and hence  $h$  then immediately follows from (90). ■

## 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.
- ALTONJI, J. G., AND R. L. MATZKIN (2005): “Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors,” *Econometrica*, 73(4), 1053–1102.
- 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.
- BLUNDELL, R., X. CHEN, AND D. KRISTENSEN (2007): “Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves,” *Econometrica*, 75, 1613–1669.
- 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.
- CARRASCO, M., AND J.-P. FLORENS (2000): “Generalization of GMM to a Continuum of Moment Conditions,” *Econometric Theory*, 16, 797–834.
- CHEN, X. (2006): “Large Sample Sieve Estimation of Semi-Nonparametric Models,” Working paper, Yale University.
- CHEN, X., AND D. POUZO (2008a): “Efficient Estimation of Semiparametric Conditional Moment Models with Possibly Nonsmooth Moments,” Working paper, Yale University.
- (2008b): “Estimation of Nonparametric Conditional Moment Models with Possibly Nonsmooth Moments,” Working paper, Yale University.

- CHERNOZHUKOV, V., AND C. HANSEN (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73(1), 245–261.
- CHESHER, A. (2003): "Identification in Nonseparable Models," *Econometrica*, 71, 1405–1441.
- FLORENS, J. P., J. J. HECKMAN, C. MEGHIR, AND E. VYTLACIL (2008): "Identification of Treatment Effects Using Control Functions in Models with Continuous, Endogenous Treatment and Heterogeneous Effects," *Econometrica*, 76(5), 1191–1206.
- 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.
- HOROWITZ, J. L., AND S. LEE (2007): "Nonparametric Instrumental Variable Estimation of a Quantile Regression Model," *Econometrica*, 75(4).
- IMBENS, G. W., AND W. K. NEWEY (2009): "Identification and Estimation of Triangular Simultaneous Equation Models Without Additivity," *Econometrica*, (forthcoming).
- KOMUNJER, I. (2009): "Global Identification of the Semiparametric Box-Cox Model," *Economics Letters*, 104.
- 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.

NEWBY, 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.