The Econometrics Journal

Econometrics Journal (2010), volume **13**, pp. S28–S55.
doi: 10.1111/j.1368-423X.2010.00317.x

Semi-parametric estimation of non-separable models: a minimum distance from independence approach

IVANA KOMUNJER[†] AND ANDRES SANTOS[†]

[†]*Department of Economics, University of California at San Diego, 9500 Gilman Drive MS0508, La Jolla, CA 92093, USA.*
E-mails: komunjer@ucsd.edu, a2santos@ucsd.edu

First version received: March 2009; final version accepted: February 2010

Summary This paper studies non-separable structural models that are of the form $Y = m_\alpha(X, U)$ with U uniform on (0, 1) in which m_α is a known real function parametrized by a structural parameter α. We study the case in which α contains a finite dimensional component θ and an infinite dimensional component h. We assume that the true value α_0 is identified by the restriction $U \perp X$. Our proposal is to estimate α_0 by a minimum distance from independence (MDI) criterion. We show that: (a) our estimator for h_0 is consistent and we obtain rates of convergence and (b) the estimator for θ_0 is $\sqrt{n}$ consistent and asymptotically normally distributed.

Keywords: *Identification*, *Non-separable models*, *Semi-parametric estimation*.

1. INTRODUCTION

Non-parametric identification of non-linear non-separable 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 (2006) among others. Yet the criteria used for estimation in such models rarely involve the independence property. Instead, non-parametric and semi-parametric estimation methods typically use the mean independence between the latent and exogenous variables that comes in a form of conditional moment restrictions (see e.g. Ai and Chen, 2003, Blundell et al., 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 this paper, we unify the estimation and identification of non-separable models by employing the same criterion to obtain both: full independence between the models' latent and exogenous variables. We focus on 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).[1] We denote by α_0 the true value of the structural parameter α which consists of (a) a

[1] In a semi-parametric specification, requiring $U \sim U(0, 1)$ can often be seen as a normalization on the non-parametric component that does not affect the parametric one. This assumption is also often used for non-parametric identification (see Matzkin, 2003, for examples of such arguments).

component θ in Θ that is finite dimensional ($\Theta \subset \mathbb{R}^{d_\theta}$) and (b) 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 non-separable 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 α_0 of α 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:

$$P(Y \leqslant m_{\alpha_0}(X, t_u); X \leqslant t_x) = t_u \cdot P(X \leqslant t_x) \tag{1.1}$$

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:

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

where μ is 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 $Q_n(\alpha)$ of $Q(\alpha)$ above—a minimum distance from independence (MDI) estimator. When α_0 is identified by the assumptions of the model, then α_0 will also be the unique zero of $Q(\alpha)$. Exploiting the standard M-estimation arguments we are then able to: (i) show that the MDI estimator $\hat{\alpha} = (\hat{\theta}, \hat{h})$ is consistent for $\alpha_0 = (\theta_0, h_0)$; (ii) obtain the rate of convergence of the estimator $\hat{h}$ for h_0; (iii) establish the asymptotic normality of the estimator $\hat{\theta}$ for θ_0.

The approach of minimizing the distance from independence for estimation was originally explored in the seminal work of Manski (1983). In the context of non-linear 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 parametrized and do not allow for the presence of non-parametric 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 non-parametric and semi-parametric estimation, respectively, in an instrumental variables setting. However, these results concern a finite number of quantile restrictions, while (1.1) 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 non-separable 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 et al. (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 semi-parametric non-separable models arise naturally in economic analysis by studying a simple version of Berry et al. (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 the Appendices.

2. MINIMUM DISTANCE FROM INDEPENDENCE ESTIMATION

We consider the following non-separable model:

$$Y = m_\alpha(X, U) \quad \text{and} \quad U \sim U(0, 1) \tag{2.1}$$

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 α 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 $h : \mathcal{X} \times (0, 1) \to \mathbb{R}$. The latter component of α 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 (2.1) is correctly specified and we denote by α_0 the true value of the parameter α.

For every $\alpha \in \mathcal{A}$, the structural mapping $m_\alpha : \mathcal{X} \times (0, 1) \to \mathbb{R}$ in (2.1) 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 α_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 model (2.1) with $\alpha = \alpha_0$. We assume the following:

ASSUMPTION 2.1. *(a) $\{y_i, x_i\}_{i=1}^n$ are i.i.d.; (b) X is continuously distributed on $\mathcal{X}$ with density $f_X(x)$ and (c) 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 2.1(a) is more likely to hold in cross-sectional applications; though extensions to time-series context are feasible, we do not pursue them here. Assumptions 2.1(b) and (c) 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 2.1(b) then ensures that (Y, X) are jointly continuously distributed on the set $\mathcal{S} \equiv \bigcup_{x \in \mathcal{X}} (m_{\alpha_0}(x, (0, 1)), x)$. 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 α_0 of α, as in some well-known examples of (2.1) such as the Box–Cox transformation model (see e.g. Komunjer, 2009).

The key property of model (2.1) upon which we base our estimation procedure is that α_0 is non-parametrically identified by an independence restriction.

ASSUMPTION 2.2. *The true value $\alpha_0 \in \mathcal{A}$ of the structural parameter α in model (2.1) is identified by the restriction: $U \perp X$.*

Assumption 2.2 requires that model (2.1) be identified by an independence restriction. For fully non-parametric specifications, the arguments that lead to this result are well understood (see e.g. Matzkin, 2003). Identification in semi-parametric 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 2.2 holds within a simplified BLP model. The following lemma derives a simple characterization of the property in Assumption 2.2.

LEMMA 2.1. *Let Assumptions 2.1(b) and 2.2 hold. Then, it follows that:*

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

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

Lemma 2.1 suggests a straightforward way to construct a criterion function through which to estimate α_0. Let $t = (x, u) \in \mathcal{X} \times (0, 1)$ and define

$$W_\alpha(t) \equiv P(Y \leqslant m_\alpha(X, u); X \leqslant x) - u \cdot P(X \leqslant x). \tag{2.2}$$

Under the assumptions of Lemma 2.1, 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:

$$Q(\alpha) \equiv \int_{\mathcal{X}\times(0,1)} W_\alpha^2(t) d\mu(t), \tag{2.3}$$

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

When the model in (2.1) is identified by the restriction $U \perp X$, Lemma 2.1 implies that α_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 μ is needed to ensure that α_0 is the unique minimum of $Q(\alpha)$. Indeed, if μ 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 α 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 $W_\alpha(t)$:

$$W_{\alpha,n}(t) \equiv \frac{1}{n}\sum_{i=1}^{n} 1\{y_i \leqslant m_\alpha(x_i, u); x_i \leqslant x\} - u \cdot \frac{1}{n}\sum_{i=1}^{n} 1\{x_i \leqslant x\}, \tag{2.4}$$

which yields a finite sample criterion function:

$$Q_n(\alpha) \equiv \int_{\mathcal{X}\times(0,1)} W_{\alpha,n}^2(t) d\mu(t). \tag{2.5}$$

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

$$\hat{\alpha} \in \arg\min_{\alpha\in\mathcal{A}_n} Q_n(\alpha). \tag{2.6}$$

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

[2] See the Appendix for details regarding the notations and definitions.

ASSUMPTION 2.3. *(a)* μ *has full support on* $\mathcal{X} \times (0,1)$; *(b)* Θ *and* $\mathcal{H}$ *are compact w.r.t.* $\|\cdot\|$ *and* $\|\cdot\|_\infty$; *(c)* $m_\alpha(x,\cdot):(0,1)\to\mathbb{R}$ *is strictly increasing for every* $(\alpha,x)\in\mathcal{A}\times\mathcal{X}$; *(d) For every* $x\in\mathcal{X}$, $\sup_{u\in(0,1)}|m_\alpha(x,u)-m_{\tilde\alpha}(x,u)| \leqslant G(x)\{\|\theta-\tilde\theta\|+\|h-\tilde h\|_\infty\}$ *with* $E[G^2(X)]<\infty$; *(e) The entropy* $\int_0^\infty \sqrt{N_{[]}(\eta^3,\mathcal{H},\|\cdot\|_\infty)}d\eta<\infty$; *(f)* $\mathcal{H}_n\subset\mathcal{H}$ *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 2.3(a) ensures that $Q(\alpha)$ is uniquely minimized at α_0. Assumptions 2.3(b)–(e) ensure 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 2.3(e), $N_{[]}(\eta^3,\mathcal{H},\|\cdot\|_\infty)$ 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 2.3(e). Finally, Assumption 2.3(f) requires the sieve can approximate the parameter space with respect to the norm $\|\cdot\|_\infty$.

Assumptions 2.1–2.3 are sufficient for establishing the consistency of the MDI estimator under the norm $\|\cdot\|_c$.

THEOREM 2.1. *Under Assumptions 2.1–2.3 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 $h(x,u)$:

$$\|h\|_{L^2}^2=\int_{\mathcal{X}\times(0,1)} h^2(x,u)f_X(x)dx\,du. \tag{3.1}$$

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 (2.1) satisfies $\|m_\alpha\|_{L^2}<\infty$ and define the mapping $m:(\mathcal{A},\|\cdot\|_c)\to L^2$ which to any $\alpha\in\mathcal{A}$ associates $m(\alpha)\equiv m_\alpha$.

Given these definitions, we introduce the following assumption.[3]

ASSUMPTION 3.1. *(a) In a neighbourhood* $\mathcal{N}(\alpha_0)\subset\mathcal{A}$, $m:(\mathcal{A},\|\cdot\|_c)\to L^2$ *is continuously Fréchet differentiable; (b) For every* $(y,x)\in\mathcal{S}$, *the conditional densities satisfy* $|f_{Y|X}(y\,|\,x)-f_{Y|X}(y'\,|\,x)|\leqslant J(x)|y-y'|^\nu$ *with* $E[J^2(X)G^{2\vee 2\nu}(X)]<\infty$; *(c) The marginal density of* μ *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

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

[3] See the Appendix for definitions.

In order to obtain the rates of convergence for $\|\hat{h} - h\|_{L^2}$, it is necessary to examine the local behaviour of $Q(\alpha)$ at α_0. Under Assumptions 2.1(b)–(c), 2.3(d) and 3.1, the Fréchet differentiability of m is inherited by the mapping $Q : (\mathcal{A}, \|\cdot\|_c) \to \mathbb{R}$, which to every $\alpha \in \mathcal{A}$ associates $Q(\alpha)$. To state the form of this Fréchet derivative, we define the linear map $D_{\bar{\alpha}} : (\mathcal{A}, \|\cdot\|_c) \to L^2_\mu$ which to every $\pi \in \mathcal{A}$ associates $D_{\bar{\alpha}}[\pi]$, where $D_{\bar{\alpha}}[\pi] : \mathcal{X} \times (0, 1) \to \mathbb{R}$ maps $t = (x, u) \in \mathcal{X} \times (0, 1)$ into $D_{\bar{\alpha}}[\pi](t)$ given by:

$$D_{\bar{\alpha}}[\pi](t) = \int_{\mathcal{X}} f_{Y\,|\,X}(m_{\bar{\alpha}}(s_x, u)|s_x) \frac{dm_{\bar{\alpha}}(s_x, u)}{d\alpha}[\pi] 1\{s_x \leqslant x\} f_X(s_x) ds_x. \tag{3.2}$$

Lemma 3.1 establishes that $Q(\alpha)$ is twice Fréchet differentiable at α_0.

LEMMA 3.1. *Under Assumptions 2.1(b)–(c), 2.3(d) and 3.1(a)–(c), $Q : (\mathcal{A}, \|\cdot\|_c) \to \mathbb{R}$ is: (a) continuously Fréchet differentiable in $\mathcal{N}(\alpha_0)$ with*

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

(b) twice Fréchet differentiable at α_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 α_0, its second derivative at α_0 induces a norm on $\mathcal{A}$. This result is analogous to a parametric model, in which if the Hessian H is a positive definite matrix, then $\sqrt{a'Ha}$ is a norm equivalent to the standard Euclidean norm. Guided by Lemma 3.1 we therefore define the inner product and associated norm:

$$\langle \alpha, \tilde{\alpha} \rangle_w \equiv \int_{\mathcal{X}\times(0,1)} D_{\alpha_0}[\alpha](t) D_{\alpha_0}[\tilde{\alpha}](t) d\mu(t) \quad \text{and} \quad \|\alpha\|_w^2 = \langle \alpha, \alpha \rangle_w. \tag{3.3}$$

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 behaviour 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 3.2. *(a) In a neighbourhood $\mathcal{N}(\alpha_0)$, $\|\alpha - \alpha_0\|_w^2 \lesssim Q(\alpha) \lesssim \|\alpha - \alpha_0\|_s^2$; (b) The ratio $\tau_n \equiv \sup_{\mathcal{A}_n} \|\alpha_n\|_s^2 / \|\alpha_n\|_w^2$ satisfies $\tau_n = o(n^\gamma)$ with $\gamma < 1/4$; (c) 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 3.2(a) 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 3.2(b) 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 τ_n is often referred to as the sieve modulus of continuity (see e.g. Chen and Pouzo, 2008b). In practice, Assumption 3.2(b) is requiring the sieve not to grow too fast. Finally, Assumption 3.2(c) refines the requirements of rates of approximation for the sieve $\mathcal{A}_n$.

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

THEOREM 3.1. *Under Assumptions 2.1–2.3, 3.1 and 3.2,* $\|\hat{\alpha} - \alpha_0\|_s = o_p(n^{-\frac{1}{4}})$.

Note that since $\|\hat{\alpha} - \alpha_0\|_s = \|\hat{\theta} - \theta_0\| + \|\hat{h} - h_0\|_{L^2}$, it immediately follows from Theorem 3.1 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 semi-parametric specification, is continuous in $\|\cdot\|_w$. By appealing to the Riesz Representation Theorem it then follows that there is v^λ 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:[4]

$$D_{\alpha_0}[\alpha - \alpha_0] \equiv \frac{dW(\alpha_0)}{d\alpha}[\alpha - \alpha_0] = \frac{dW(\alpha_0)}{d\theta'}[\theta - \theta_0] + \frac{dW(\alpha_0)}{dh}[h - h_0]. \tag{4.1}$$

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

$$h_j^* \equiv \arg\min_{h\in\bar{\mathcal{H}}} \int_{\mathcal{X}\times(0,1)} \left(\frac{dW_{\alpha_0}(t)}{d\theta_j} - \frac{dW_{\alpha_0}(t)}{dh}[h]\right)^2 d\mu(t), \tag{4.2}$$

where the minimum in (4.2) 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 $h^* \equiv (h_1^*, \ldots, h_{d_\theta}^*)$ and let

$$\frac{dW_{\alpha_0}(t)}{dh}[h^*] = \left(\frac{dW_{\alpha_0}(t)}{dh}\left[h_1^*\right], \ldots, \frac{dW_{\alpha_0}(t)}{dh}\left[h_{d_\theta}^*\right]\right). \tag{4.3}$$

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

$$R_{h^*}(t) = \frac{dW_{\alpha_0}(t)}{d\theta} - \frac{dW_{\alpha_0}(t)}{dh}[h^*], \tag{4.4}$$

[4] The first equality in (4.1) is formally justified in the proof of Lemma 3.1 in the Appendix, in which it is shown $D_{\bar{\alpha}}$ is the Fréchet derivative of the mapping $W : (\mathcal{A}, \|\cdot\|_c) \to L^2_\mu$ given by $W : \alpha \mapsto W_\alpha$, when evaluated at $\bar{\alpha}$. Similar to before, we use the notation

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

and the associated matrix

$$\Sigma^* \equiv \int_{\mathcal{X}\times(0,1)} R_{h^*}(t) R'_{h^*}(t) d\mu(t). \tag{4.5}$$

Lemma 4.1 shows that the functional $F_\lambda(\alpha) = \lambda'\theta$ is continuous if the matrix Σ^* is positive definite, which may be interpreted as a local identification condition on θ_0. Lemma 4.1 also obtains the formula for the Riesz Representor of $F_\lambda(\alpha)$.

LEMMA 4.1. *Let $v_\theta^\lambda \equiv (\Sigma^*)^{-1}\lambda$ and $v_h^\lambda \equiv -h^* v_\theta^\lambda$. If Σ^* is positive definite, then for any $\lambda \in \mathbb{R}^{d_\theta}$, $F_\lambda(\alpha - \alpha_0) = \lambda'(\theta - \theta_0)$ is continuous on $\bar{\mathcal{A}}$ under $\|\cdot\|_w$ and in addition we have $F_\lambda(\alpha - \alpha_0) = \langle v^\lambda, \alpha - \alpha_0\rangle_w = \lambda'(\theta - \theta_0)$.*

Having established the continuity of $F_\lambda(\alpha)$ in $\|\cdot\|_w$ and the closed-form solution or the Riesz Representor v^λ 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 analyse as it is determined by the local behaviour of $Q(\alpha)$ near its minimum α_0. In order to establish asymptotic normality, we require one final assumption.

ASSUMPTION 4.1. *(a) The matrix Σ^* is positive definite; (b) $v^\lambda \in \mathcal{A}$ for $\|\lambda\|$ small; (c) 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 and in addition satisfies $\int_{\mathcal{X}\times(0,1)} \sup_{s\in[0,1]} |\frac{dD_{\alpha+\tau\bar{\alpha}}[\pi]}{d\tau}(t)|_{\tau=s}| d\mu(t) \lesssim \|\bar{\alpha}\|_s \|\pi\|_s$ as well as $\int_{\mathcal{X}\times(0,1)} \sup_{s\in[0,1]} \left[\frac{dD_{\alpha+\tau\bar{\alpha}}[\pi]}{d\tau}(t)|_{\tau=s}\right]^2 d\mu(t) \lesssim \|\bar{\alpha}\|_s^2$; (d) 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 4.1(a) ensures that $F_\lambda(\alpha) = \lambda'\theta$ is continuous in $\|\cdot\|_w$, as shown in Lemma 4.1. While $v^\lambda \in \bar{\mathcal{A}}$, Assumption 4.1(b) additionally requires $v^\lambda \in \mathcal{A}$. As a result v^λ may be approximated by an element $\Pi_n v^\lambda \in \mathcal{A}_n$ due to Assumption 3.1(c). 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 4.1(c)–(d) 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 4.1. *Let Assumptions 2.1–4.1 hold. Then, $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{\mathcal{L}} N(0, \Sigma)$, 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

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

5. EXAMPLE AND MONTE CARLO EVIDENCE

5.1. The model

We proceed to illustrate how non-separable structures of the form in (2.1) 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 et al. (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):

$$u_{ij} = -ap_j + b'x_j + \xi_j + \zeta_i + \varepsilon_{ij}, \tag{5.1}$$

in which u_{ij} is the utility of product j ($j = 1, 2$) to individual i ($i = 1, \dots, I$) with unobserved characteristics ζ_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 ξ_j is an index of unobserved characteristics of the latter ($\xi_j \in \mathbb{R}$); $-a$ is a taste parameter on the price assumed constant across individuals ($a > 0$); finally, ε_{ij} is an error term that represents the deviations from an average behaviour 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 (5.1) is that ε_{ij} are i.i.d. across products j and individuals i. For example, assuming that ε_{ij}'s are Gumbel random variables, the resulting individual choice model is logit. In what follows, we let the difference $\varepsilon_{i2} - \varepsilon_{i1}$ be distributed with 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

$$D_j(p_j, p_{-j}) = M \cdot F(-a(p_j - p_{-j}) + b'(x_j - x_{-j}) + \xi_j - \xi_{-j}), \tag{5.2}$$

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 (5.2) takes the form

$$Y = -aP + b'X + \xi \quad \text{with} \quad \xi \perp X. \tag{5.3}$$

In the model above, prices are endogenous, so even if ξ is independent of X, we can expect P to depend on ξ. Hence, without further restrictions on ξ and P it is not possible to identify the parameters a and b in (5.3). We now show how the supply-side information may be used to identify these parameters.

We assume that firms compete in prices (*à la* Bertrand), so each firm chooses 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 5.1 exploits this relationship to obtain an alternative representation for the BLP model (5.3).

LEMMA 5.1. *Assume F is twice continuously differentiable on $\mathbb{R}$ with strictly increasing hazard rate τ. If ξ is continuously distributed, then it follows that:*

$$Y = h(X, U) + X'\theta, \quad U \sim U(0, 1), \tag{5.4}$$

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

Lemma 5.1 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 analysing the equilibrium strategies.

The BLP model (5.4) is clearly a special case of the non-separable structural model in (2.1) with $\alpha \equiv (\theta, h)$ and

$$m_\alpha(X, U) = h(X, U) + X'\theta. \tag{5.5}$$

We now illustrate how to verify other assumptions for this model. If the BLP model variables are i.i.d. with continuous distribution functions, then the continuous differentiability of the demand function guarantees that the sampling Assumption 2.1 holds. We note that in this example the supports of the endogenous and exogenous variables are given by $\mathcal{Y} = \mathbb{R}$ and $\mathcal{X} \subseteq \mathbb{R}^{d_x}$.

But far more difficult to check is the identification Assumption 2.2 for which we now derive more primitive conditions. Our identification result for the BLP model (5.4) is contained in the following theorem.

THEOREM 5.1. *Assume F is strictly increasing, twice continuously differentiable on $\mathbb{R}$ with strictly increasing hazard rate τ. Assume moreover that ξ is continuously distributed, and that we have: (a) $h(0, 1/2) = 0$ and (b) $\partial h(0, 1/2)/\partial x = 1$. Then, the BLP model (5.4) satisfies Assumption 2.2.*

The conditions of Theorem 5.1 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, (a) holds if the distribution F_ξ of the products' unobservables ξ in the BLP model in equation (5.3) 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 $P = 0$.[5] Hence, $-aP + \xi = 0 = h(0, 1/2)$. Requirement (b) fixes the value of the gradient $\partial h(x, u)/\partial x$ at zero. It ensures that the effects of changing θ 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 θ in $Y = (\phi + \theta)'X + r(U)$.

In the context of the BLP model (5.4), Assumptions 2.3(b) and 2.3(e) can be verified by letting $\mathcal{H}$ be a smooth set of functions. For example, suppose x has compact support $\mathcal{X}$ and let λ be a $d_x + 1$ dimensional vector of positive integers. Define $|\lambda| \equiv \sum_{i=1}^{d_x+1} \lambda_i$ and $D^\lambda = \partial^\lambda / \partial x_1^{\lambda_1} \ldots \partial x_{d_x}^{\lambda_{d_x}} \partial u^{\lambda_{d_x+1}}$. An appropriate set $\mathcal{H}$ is then

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

for some positive M and ε. By Theorem 2.7.1 in van der Vaart and Wellner (1996), Assumptions 2.3(a) and 2.3(e) are then satisfied. The definition of $\mathcal{H}$ also ensures Assumption 2.3(c) holds, while 2.3(d) is immediate from (5.5).

As already noted, m_α in (5.5) 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 3.1(a). In addition, for any $t = (x, u)$ we then have

$$D_\alpha[\pi](t) = \int_{\mathcal{X}} f_{Y|X}\big(h(s_x, u) + s_x'\theta | s_x\big)\big(\pi_h(s_x, u) + s_x'\pi_\theta\big) 1\{s_x \leqslant x\} f_X(s_x) ds_x . \tag{5.6}$$

[5] Note that whenever ξ_1 and ξ_2 are identically distributed, the distribution of their difference ξ satisfies $F_\xi(0) = 1/2$.

Hence, if $f_{Y\,|\,X}(y\,|\,x)$ is uniformly bounded, then $\mathcal{X}$ and Θ compact together with our choice for $\mathcal{H}$ imply Assumption 4.1(d) holds. Similarly, by direct calculation we obtain that in the discussed BLP example, for any $\alpha = (\theta, h)$ and $\bar{\alpha} = (\bar{\theta}, \bar{h})$ we have

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

hence Assumption 4.1(c) 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 ε_{i1} and ε_{i2} in (5.1) are i.i.d. Gumbel random variables, so that the distribution F of their difference is logistic. The equilibrium prices are solution to the FOC equations:

$$\exp(\Delta) + 1 - a(p_1 - c) = 0,$$
$$\exp(-\Delta) + 1 - a(p_2 - c) = 0,$$

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 E.1. 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 ξ are independent.[6] For the sieve we used a fully interacted polynomial of order 2 in X and U, while the measure μ was chosen to be uniform on $[-1, 1] \times [0, 1]$.

Table 1 reports the mean, standard deviation, mean squared error and the 10th, 50th and 90th percentile of the proposed estimator $\hat{\theta}$ for sample sizes $n = 100, 200, 500$. The statistics were computed based on 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 200 and 500 observations. For the latter two sample sizes, the estimator is also within 0.1 of the true value in over 80% of the replications. Figure 1 exhibits a Gaussian kernel estimate for the

Table 1. Monte Carlo results.

	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

[6] Note that by setting $b = -1.5$ we ensure that the identification condition (b) of Theorem 5.1 is satisfied. Condition (a) holds since the distribution of ξ is symmetric around zero.

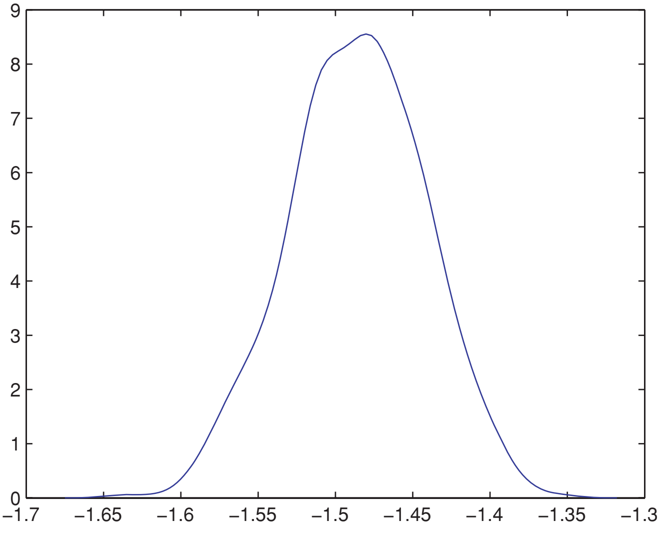

Figure 1. $\hat{\theta}$ Kernel density estimate ($n = 500$).

density of $\hat{\theta}$ obtained with sample size 500. The density is fairly symmetric and centred 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 semi-parametric non-separable models. The resulting estimator converges to the non-parametric 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 et al. (1995) model of price-setting with differentiated products. A small Monte Carlo study illustrates the performance of the proposed estimator within the BLP example.

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^2 Meeting 'Recent Advances in Structural Microeconometrics' in Rome, Italy. Many thanks to the participants of the conference for their comments. All errors are ours.

REFERENCES

Ai, C. and X. Chen (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica 71*, 1795–844.

Akritas, M. and I. van Keilegom (2001). Non-parametric estimation of the residual distribution. *Scandinavian Journal of Statistics 28*, 549–67.

Altonji, J. G. and R. L. Matzkin (2005). Cross section and panel data estimators for nonseparable models with endogenous regressors. *Econometrica 73*, 1053–102.

Benkard, C. L. and S. Berry (2006). On the nonparametric identification of non-linear simultaneous equations models: comment on Brown (1983) and Rhoerig (1988). *Econometrica 74*, 1429–40.

Berry, S. T., J. Levinsohn and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica 63*, 841–90.

Blundell, R., X. Chen and D. Kristensen (2007). Semi-nonparametric IV estimation of shape-invariant Engel curves. *Econometrica 75*, 1613–69.

Brown, B. W. (1983). The identification problem in systems nonlinear in the variables. *Econometrica 51*, 175–96.

Brown, D. J. and M. H. Wegkamp (2002). Weighted minimum mean-square distance from independence estimation. *Econometrica 70*, 2035–51.

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.

Chen, X. and D. Pouzo (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*, 245–61.

Chesher, A. (2003). Identification in nonseparable models. *Econometrica 71*, 1405–41.

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 heterogenous effects. *Econometrica 76*, 1191–206.

Hausman, J. A. and D. A. Wise (1978). A conditional probit model for qualitative choice: discrete decisions recognizing interdependence and heterogeneous preferences. *Econometrica 46*, 403–26.

Horowitz, J. L. and S. Lee (2007). Nonparametric instrumental variable estimation of a quantile regression model. *Econometrica 75*, 1191–208.

Imbens, G. W. and W. K. Newey (2009). Identification and estimation of triangular simultaneous equation models without additivity. *Econometrica 77*, 1481–512.

Komunjer, I. (2009). Global identification of the semiparametric Box–Cox model. *Economics Letters 104*, 53–56.

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

Manski, C. F. (1983). Closest empirical distribution estimation. *Econometrica 51*, 305–19.

Matzkin, R. L. (1994). Restrictions of economic theory in nonparametric methods. In R. F. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics, Volume 4*, 2524–59. Amsterdam: North-Holland.

Matzkin, R. L. (2003). Nonparametric estimation of nonadditive random functions. *Econometrica 71*, 1339–75.

Milgrom, P. and J. Roberts (1990). Rationalizability, learning and equilibrium in games with strategic complementarities. *Econometrica 58*, 1255–77.

Newey, W. K. and J. Powell (2003). Instrumental variables estimation of nonparametric models. *Econometrica 71*, 1565–78.

Roehrig, C. S. (1988). Conditions for identification in nonparametric and parametric models. *Econometrica 56*, 433–47.

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

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

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

APPENDIX A: NOTATION AND DEFINITIONS

The following is a table of the notation and definitions to be used.

$a \lesssim b$	$a \leq Mb$ 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 $\|h\|_\infty \equiv \sup_{(x,u)\in\mathcal{X}\times(0,1)} \lvert h(x,u)\rvert$.
$\|\cdot\|_{L^2}$	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 $t \equiv (x, u)$.
$N_{[\,]}(\epsilon, \mathcal{F}, \|\cdot\|)$	The bracketing numbers of size ε for $\mathcal{F}$ under the norm $\|\cdot\|$.

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

APPENDIX B: PROOFS FOR SECTION 2

Proof of Lemma 2.1: First, consider all values $\bar{\alpha}$ of α 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 $u \mapsto P(X \leqslant \bar{x}; Y \leqslant m_{\bar{\alpha}}(X, u))$ is not strictly increasing on $(0, 1)$; hence, there must exist $\bar{u} \in (0, 1)$ such that $P(X \leqslant \bar{x}; Y \leqslant m_{\bar{\alpha}}(X, \bar{u})) \neq \bar{u} \cdot P(X \leqslant \bar{x})$. Now, consider all values $\tilde{\alpha}$ of α 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 α_0 is an element of that set. Now, for any such $\tilde{\alpha}$, note that for any $(x, u) \in \mathcal{X} \times (0, 1)$ the following holds:

$$\begin{aligned} P(X \leqslant x; Y \leqslant m_{\tilde{\alpha}}(X, u)) &= \int_{s_x \leqslant x} \int_{s_y \leqslant m_{\tilde{\alpha}}(s_x, u)} f_{XY}(s_x, s_y) ds_x ds_y \\ &= \int_{s_x \leqslant x} \int_{s_u \leqslant u} f_{XY}(s_x, m_{\tilde{\alpha}}(s_x, s_u)) \frac{\partial m_{\tilde{\alpha}}(s_x, s_u)}{\partial u} ds_x ds_u \\ &= \int_{s_x \leqslant x} \int_{s_u \leqslant u} f_{X\tilde{U}}(s_x, s_u) ds_x\, ds_u \\ &= P(X \leqslant x; \tilde{U} \leqslant u), \end{aligned} \tag{B.1}$$

where for the second and third equalities follow we made 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 2.2, $\tilde{\alpha} = \alpha_0$ if and only if $\tilde{U} \perp X$, which since

$\tilde{U}$ is uniform on (0, 1) is equivalent to

$$P(\tilde{U} \leqslant u; X \leqslant x) = u \cdot P(X \leqslant x) \tag{B.2}$$

for all $(x, u) \in \mathcal{X} \times (0, 1)$. Combining (B.2) and (B.1) then establishes the lemma. □

LEMMA B.1. *Under Assumptions 2.1(a)–(c) and 2.3(b)–(e), the following class is Donsker:*

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

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

$$\mathcal{F}_u \equiv \{f(y_i, x_i) = 1\{y_i \leqslant m_\alpha(x_i, u)\} : (\alpha, u) \in \mathcal{A} \times (0, 1)\} \tag{B.3}$$

$$\mathcal{F}_x^{(k)} \equiv \left\{ f(x_i) = 1\left\{ x_i^{(k)} \leqslant t \right\} : t \in \mathbb{R} \right\}, \tag{B.4}$$

where $x^{(k)}$ is the kth coordinate of x. Further note that by direct calculation we have

$$\mathcal{F} = \mathcal{F}_u \times \prod_{k=1}^{d_x} \mathcal{F}_x^{(k)}. \tag{B.5}$$

We establish the lemma by exploiting (B.5). For any continuously distributed random variable $V \in \mathbb{R}$ and $\eta > 0$ we can find $\{-\infty = t_1, \ldots, t_{\lfloor \eta^{-2} \rfloor + 2} = +\infty\}$ such that they satisfy $P(t_s \leqslant V \leqslant t_{s+1}) \leqslant \eta^2$. The brackets $[1\{v \leqslant t_s\}, 1\{v \leqslant t_{s+1}\}]$ then cover $\{1\{v \leqslant t\} : t \in \mathbb{R}\}$ and in addition we have

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

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

$$N_{[\,]}\left(\eta, \mathcal{F}_x^{(k)}, \|\cdot\|_{L^2}\right) = O(\eta^{-2}). \tag{B.6}$$

By Assumption 2.3(b), $\mathcal{H}$ is compact under $\|\cdot\|_\infty$ and Θ 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 Θ, respectively. Defining $\{\alpha_s\} = \{h_j\} \times \{\theta_l\}$ we then have:

$$\#\{\alpha_s\} = N_{[\,]}(K_h\eta^3, \mathcal{H}, \|\cdot\|_\infty) \times (K_\theta\eta^3)^{-d_\theta}. \tag{B.7}$$

Hence, by Assumption 2.3(c), for any $\alpha \in \mathcal{A}$ there is a $(\theta_{s^*}, h_{s^*}) \equiv \alpha_{s^*} \in \{\alpha_s\}$ with

$$\begin{aligned}\sup_{u \in (0,1)} |m_\alpha(x_i, u) - m_{\alpha_{s^*}}(x_i, u)| &\leqslant G(x_i)\{\|\theta - \theta_{s^*}\| + \|h - h_{s^*}\|_\infty\} \\ &\leqslant G(x_i)\{K_\theta + K_h\}\eta^3. \end{aligned} \tag{B.8}$$

We conclude from (B.8) that for $\alpha_s \in \{\alpha_s\}$ brackets of the form

$$[m_{\alpha_s}(x_i, u) - \{K_\theta + K_h\}\eta^3 G(x_i); m_{\alpha_s}(x_i, u) + \{K_\theta + K_h\}\eta^3 G(x)] \tag{B.9}$$

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, α) by Assumption 2.3(c), we may define their inverses:

$$v_\alpha(x_i, t) = u \iff m_\alpha(x_i, u) = t. \tag{B.10}$$

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 (B.11) and obtain second equality in (B.11) from (B.10).

$$\begin{aligned} F_s^U(u) &\equiv P(Y_i \leqslant m_{\alpha_s}(X_i, u) + \{K_\theta + K_h\}\eta^3 G(X_i)) \\ &= P(v_{\alpha_s}(X_i, Y_i - \{K_\theta + K_h\}\eta^3 G(X_i)) \leqslant u). \end{aligned} \tag{B.11}$$

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

$$F^L_s(u) \equiv P(Y_i \leqslant m_{\alpha_s}(X_i, u) - \{K_\theta + K_h\}\eta^3 G(X_i)) \tag{B.12}$$

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

$$u^L_{sk_1} \leqslant u \leqslant u^U_{sk_2},$$

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

$$\#\{[u_{sk_1}, u_{sk_2}]\} = O(\eta^{-2}). \tag{B.13}$$

It follows from (B.9) and the strict monotonicity of $m_\alpha(x, u)$ in u that for every $(\alpha, u) \in \mathcal{A} \times (0, 1)$ there exists an $\alpha_s \in \{\alpha_s\}$ and $[u_{sk_1}, u_{sk_2}] \in \{[u_{sk_1}, u_{sk_2}]\}$ such that

$$\begin{aligned} &1\{y_i \leqslant m_{\alpha_s}(x, u_{sk_1}) - \{K_\theta + K_h\}\eta^3 G(x_i)\} \\ &\quad \leqslant 1\{y_i \leqslant m_\alpha(x_i, u)\} \leqslant 1\{y_i \leqslant m_{\alpha_s}(x, u_{ik_2}) + \{K_\theta + K_h\}\eta^3 G(x_i)\}, \end{aligned} \tag{B.14}$$

and hence $\{[1\{y_i \leqslant m_{\alpha_s}(x, u_{sk_1}) - \{K_\theta + K_h\}\eta^3 G(x_i)\}, 1\{y_i \leqslant m_{\alpha_s}(x, u_{ik_2}) + \{K_\theta + K_h\}\eta^3 G(x_i)\}]\}$ 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 $F^U_s(u_{sk_2}) - F^L_s(u_{sk_1})$. The construction of $\{[u_{sk_1}, u_{sk_2}]\}$ in turn implies the first inequality in (B.15) 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 (B.15).

$$\begin{aligned} &F^U_s(u_{sk_2}) - F^L_s(u_{sk_1}) \\ &\quad \leqslant F^U_s(u) - F^L_s(u) + \frac{\eta^2}{3} \\ &\quad \leqslant F^U_s(u; G(X_i) \leqslant M_\eta) - F^L_s(u; G(X_i) \leqslant M_\eta) + 2P(G(X_i) \geq M_\eta) + \frac{\eta^2}{3} \\ &\quad \leqslant F^U_s(u; G(X_i) \leqslant M_\eta) - F^L_s(t; G(X_i) \leqslant M_\eta) + \frac{2}{3}\eta^2. \end{aligned} \tag{B.15}$$

To conclude, note that $M_\eta = \sqrt{6E[G^2(X_i)]}/\eta$ and the Mean Value Theorem imply that

$$\begin{aligned} &F^U_s(u; G(X_i) \leqslant M_\eta) - F^L_s(u; G(X_i) \leqslant M_\eta) \\ &\quad \leqslant P\left(Y_i \leqslant m_{\alpha_s}(X_i, u) + \{K_\theta + K_h\}M_\eta\eta^3\right) - P\left(Y_i \leqslant m_{\alpha_s}(X_i, u) - \{K_\theta + K_h\}M_\eta\eta^3\right) \\ &\quad \leqslant 2\left\{\sup_{y_i, x_i} f_{Y\,|\,X}(y_i \,|\, x_i)\right\}\{K_\theta + K_h\}\sqrt{6E[G^2(X_i)]}\eta^2, \end{aligned}$$

where the resulting expression is finite due to Assumptions 2.1(c) and 2.3(d). Combining the preceding result with that obtained in (B.15) it follows that by choosing

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

the proposed brackets will have L^2 size η. Thus, we have from (B.7) and (B.13),

$$N_{[\,]}(\eta, \mathcal{F}_u, \|\cdot\|_{L^2}) = O(N_{[\,]}(K_h\eta^3, \mathcal{H}, \|\cdot\|_\infty) \times (K_\theta\eta^3)^{-(2+d_\theta)}). \tag{B.16}$$

To conclude note that (B.6), (B.16), Assumption 2.3(d) and Theorem 2.5.6 in van der Vaart and Wellner (1996) imply the classes $\mathcal{F}^{(k)}_x$ 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 equation (B.5) establish the claim of the lemma. □

Proof of Theorem 2.1: By Assumption 2.3(b) and the Tychonoff Theorem, $\mathcal{A}$ is compact with respect to $\|\cdot\|_c$. Furthermore, Lemma B.1 and simple manipulations show,

$$\sup_{t,\alpha} |W_{\alpha,n}(t) - W_\alpha(t)| = o_p(1). \tag{B.17}$$

Exploiting (B.17) and $W_{\alpha,n}(t)$ and $W_\alpha(t)$ being bounded by 1, we obtain that

$$\sup_{\alpha} |\, Q_n(\alpha) - Q(\alpha)| \leqslant \left[\sup_{t,\alpha} |W_{\alpha,n}(t) - W_\alpha(t)|\right] \times \left[\sup_{t,\alpha} |W_{\alpha,n}(t)| + \sup_{t,\alpha} |W_\alpha(t)|\right] = o_p(1). \tag{B.18}$$

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 3.1: Similar to previously, let $W : (\mathcal{A}, \|\cdot\|_c) \to L^2_\mu$ 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 neighbourhood of α_0. Recall that for any $t = (x, u)$,

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

and note that $D_{\bar{\alpha}}[\pi]$ is well defined for every $\bar{\alpha} \in \mathcal{N}(\alpha_0)$ due to Assumption 3.1(a). Next, use $f_{Y\,|\,X}(y\,|\,x)$ uniformly bounded and Jensen's inequality to obtain the first result in (C.1). The second inequality then holds for $\|\cdot\|_o$ the linear operator norm by Assumption 3.1(c).

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

Since Fréchet derivatives are *a fortiori* continuous, (C.1) 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 (C.2) pointwise in $t = (x, u)$.

$$\begin{aligned} &|D_{\bar{\alpha}}[\pi](t) - D_{\tilde{\alpha}}[\pi](t)| \\ &\quad\leqslant \int_{\mathcal{X}} f_{Y\,|\,X}(m_{\bar{\alpha}}(s_x, u)\,|\,s_x) \left|\frac{dm_{\bar{\alpha}}(s_x, u)}{d\alpha}[\pi] - \frac{dm_{\tilde{\alpha}}(s_x, u)}{d\alpha}[\pi]\right| f_X(s_x) ds_x \\ &\quad+ \int_{\mathcal{X}} \left|f_{Y\,|\,X}(m_{\bar{\alpha}}(s_x, u)\,|\,s_x) - f_{Y\,|\,X}(m_{\tilde{\alpha}}(s_x, u)\,|\,s_x)\right| \left|\frac{dm_{\tilde{\alpha}}(s_x, u)}{d\alpha}[\pi]\right| f_X(s_x) ds_x. \end{aligned} \tag{C.2}$$

In turn, the Lipschitz Assumptions 2.3(d) and 3.1(b), $f_{Y\,|\,X}(y\,|\,x)$ uniformly bounded by Assumption 2.1(c) and equation (C.2) yield that pointwise in $t = (x, u)$,

$$\begin{aligned} |D_{\bar{\alpha}}[\pi](t) - D_{\tilde{\alpha}}[\pi](t)| \lesssim \|\bar{\alpha} - \tilde{\alpha}\|^\nu_c \int_{\mathcal{X}} J(s_x) G^\nu(s_x) \left|\frac{dm_{\tilde{\alpha}}(s_x, u)}{d\alpha}[\pi]\right| f_X(s_x) ds_x \\ + \int_{\mathcal{X}} \left|\frac{dm_{\bar{\alpha}}(s_x, u)}{d\alpha}[\pi] - \frac{dm_{\tilde{\alpha}}(s_x, u)}{d\alpha}[\pi]\right| f_X(s_x) ds_x. \end{aligned} \tag{C.3}$$

Using (C.3), Cauchy–Schwarz and Jensen's inequality and $E[J^2(X)G^{2\nu}(X)] < \infty$ yields

$$\|D_{\bar{\alpha}}[\pi] - D_{\tilde{\alpha}}[\pi]\|_{L^2_\mu}^2 \lesssim \|\bar{\alpha} - \tilde{\alpha}\|_c^{2\nu} \int_{\mathcal{X}\times(0,1)} \int_{\mathcal{X}} \left[\frac{dm_{\bar{\alpha}}(s_x, u)}{d\alpha}[\pi]\right]^2 f_X(s_x) ds_x d\mu(t)$$
$$+ \int_{\mathcal{X}\times(0,1)} \int_{\mathcal{X}} \left[\frac{dm_{\bar{\alpha}}(s_x, u)}{d\alpha}[\pi] - \frac{dm_{\tilde{\alpha}}(s_x, u)}{d\alpha}[\pi]\right]^2 f_X(s_x) ds_x d\mu(t). \quad \text{(C.4)}$$

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 (C.5), while the first inequality follows from (C.4). Further, since the functional $\left\|\frac{dm}{d\alpha}(\bar{\alpha})\right\|_o$: $(\mathcal{N}(\alpha_0), \|\cdot\|_c) \to \mathbb{R}$ 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 (C.5) then follows.

$$\|D_{\bar{\alpha}} - D_{\tilde{\alpha}}\|_o^2 = \sup_{\pi \in \bar{\mathcal{A}}_c} \|\pi\|_c^{-2} \|D_{\bar{\alpha}}[\pi] - D_{\tilde{\alpha}}[\pi]\|_{L^2_\mu}^2$$
$$\lesssim \|\bar{\alpha} - \tilde{\alpha}\|_c^{2\nu} \left\|\frac{dm}{d\alpha}(\bar{\alpha})\right\|_o^2 + \left\|\frac{dm}{d\alpha}(\bar{\alpha}) - \frac{dm}{d\alpha}(\tilde{\alpha})\right\|_o^2$$
$$\lesssim \|\bar{\alpha} - \tilde{\alpha}\|_c^{2\nu} + \left\|\frac{dm}{d\alpha}(\bar{\alpha}) - \frac{dm}{d\alpha}(\tilde{\alpha})\right\|_o^2. \quad \text{(C.5)}$$

Therefore, $D_{\bar{\alpha}}$ is continuous in α 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

$$W_\alpha(t) = \int_{\mathcal{X}} P(Y \leqslant m_\alpha(s_x, u) \mid s_x) 1\{s_x \leqslant x\} f_X(s_x) ds_x - u \cdot P(X \leqslant x). \quad \text{(C.6)}$$

Next, using the definition of $D_{\bar{\alpha}}$ and (C.6) together with Jensen's inequality we obtain (C.7) pointwise in t for any $\bar{\alpha} \in \mathcal{N}(\alpha_0)$ and $\pi \in \mathcal{A}$.

$$|W_{\bar{\alpha}+\pi}(t) - W_{\bar{\alpha}}(t) - D_{\bar{\alpha}}[\pi](t)| \lesssim \int_{\mathcal{X}} |P(Y \leqslant m_{\bar{\alpha}+\pi}(s_x, u)|s_x)$$
$$- P(Y \leqslant m_{\bar{\alpha}}(s_x, u)|s_x) - f_{Y|X}(m_{\bar{\alpha}}(s_x, u)|s_x) \frac{dm_{\bar{\alpha}}(s_x, u)}{d\alpha}[\pi]| f_X(s_x) ds_x. \quad \text{(C.7)}$$

Applying the Mean Value Theorem inside the integral in (C.7) then implies

$$|W_{\bar{\alpha}+\pi}(t) - W_{\bar{\alpha}}(t) - D_{\bar{\alpha}}[\pi](t)| \lesssim \int_{\mathcal{X}} |f_{Y|X}(\bar{m}(s_x, u) \mid s_x)[m_{\bar{\alpha}+\pi}(s_x, u) - m_{\bar{\alpha}}(s_x, u)]$$
$$- f_{Y|X}(m_{\bar{\alpha}}(s_x, u) \mid s_x) \frac{dm_{\bar{\alpha}}(s_x, u)}{d\alpha}[\pi]| f_X(s_x) ds_x, \quad \text{(C.8)}$$

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 $|\bar{m}(s_x, u) - m_{\bar{\alpha}}(s_x, u)| \leqslant |m_{\bar{\alpha}+\pi}(s_x, u) - m_{\bar{\alpha}}(s_x, u)|$. The Lipschitz conditions of Assumptions 2.3(d) and 3.1(b) then imply the inequality:

$$\int_{\mathcal{X}} |[f_{Y|X}(\bar{m}(s_x, u) \mid s_x) - f_{Y|X}(m_{\bar{\alpha}}(s_x, u) \mid s_x)][m_{\bar{\alpha}+\pi}(s_x, u) - m_{\bar{\alpha}}(s_x, u)]| f_X(s_x) ds_x$$
$$\leqslant \|\pi\|_c^{1+\nu} \int J(s_x) G^{1+\nu}(s_x) f_X(s_x) ds_x. \quad \text{(C.9)}$$

Using (C.8), (C.9), $f_{Y\mid X}(y\mid x)$ being bounded and Jensen's inequality in turn establishes the first inequality in (C.10). The final result in (C.10) then follows by $\frac{dm}{d\alpha}$ being the Fréchet derivative of m.

$$\|W_{\bar\alpha+\pi}(t)-W_{\bar\alpha}(t)-D_{\bar\alpha}[\pi](t)\|^2_{L^2_\mu}\lesssim\|\pi\|_c^{2+2\nu}$$
$$+\int_{\mathcal{X}\times(0,1)}\int_{\mathcal{X}}\left[m_{\bar\alpha+\pi}(s_x,u)-m_{\bar\alpha}(s_x,u)-\frac{dm_{\bar\alpha}(s_x,u)}{d\alpha}[\pi]\right]^2 f_X(s_x)ds_x d\mu(t)=o\left(\|\pi\|_c^2\right). \tag{C.10}$$

We conclude from (C.10) and (C.5) that $D_{\bar\alpha}$ is the Fréchet derivative at $\bar\alpha$ of the map $W:(\mathcal{A},\|\cdot\|_c)\to L^2_\mu$ and that it is continuous in $\bar\alpha$. To conclude the proof of the first claim of the lemma, note that $Q(\alpha)=\|W_\alpha(t)\|^2_{L^2_\mu}$. Since the functional $\|\cdot\|^2_{L^2_\mu}:L^2_\mu\to\mathbb{R}$ is trivially Fréchet differentiable, applying the Chain rule for Fréchet derivatives (see e.g. Theorem 5.2.5 in Siddiqi, 2004) yields

$$\frac{dQ(\bar\alpha)}{d\alpha}[\pi]=\int_{\mathcal{X}\times(0,1)}W_{\bar\alpha}(t)D_{\bar\alpha}[\pi](t)d\mu(t). \tag{C.11}$$

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

$$T[\psi,\pi]=\int_{\mathcal{X}\times(0,1)}D_{\alpha_0}[\psi](t)D_{\alpha_0}[\pi](t)d\mu(t). \tag{C.12}$$

We will show T is the second Fréchet derivative of $Q(\alpha)$ at α_0. Note that $T[\psi,\cdot]:\mathcal{A}\to\mathbb{R}$ is a linear operator. The first requirement of Fréchet differentiability is to show $T[\psi,\cdot]$ is continuous in ψ. For this purpose, note that the first equality in (C.13) follows by definition while the first and second inequalities are implied by the Cauchy–Schwarz inequality and (C.1), respectively.

$$\begin{aligned}\|T[\psi,\cdot]\|_o^2&=\sup_{\pi\in\bar{\mathcal{A}}_c}\|\pi\|_c^{-2}T^2[\psi,\pi]\\&\leqslant\int_{\mathcal{X}\times(0,1)}D^2_{\alpha_0}[\psi](t)d\mu(t)\times\sup_{\pi\in\bar{\mathcal{A}}_c}\|\pi\|_c^{-2}\int_{\mathcal{X}\times(0,1)}D^2_{\alpha_0}[\pi](t)d\mu(t)\\&\lesssim\left\|\frac{dm(\alpha_0)}{d\alpha}\right\|_o^4\|\psi\|_c^2.\end{aligned} \tag{C.13}$$

It follows from (C.13) that $T[\psi,\cdot]$ is continuous in $\psi\in\mathcal{A}$. Next, we verify T is the second Fréchet derivative of $Q(\alpha)$ at α_0. In (C.14) use (C.11) and $W_{\alpha_0}(t)=0$ for all t to note $\frac{dQ(\alpha_0)}{d\alpha}=0$ and obtain

$$\begin{aligned}&\left\|\frac{dQ(\alpha_0+\psi)}{d\alpha}-\frac{dQ(\alpha_0)}{d\alpha}-T[\psi,\cdot]\right\|_o^2\\&=\sup_{\pi\in\bar{\mathcal{A}}_c}\|\pi\|_c^{-2}\left(\int_{\mathcal{X}\times(0,1)}(W_{\alpha_0+\psi}(t)D_{\alpha_0+\psi}[\pi](t)-D_{\alpha_0}[\psi](t)D_{\alpha_0}[\pi](t))d\mu(t)\right)^2.\end{aligned} \tag{C.14}$$

Next, use the Cauchy–Schwarz inequality to obtain the first inequality in (C.15) and D_{α_0} being the Fréchet derivative of $W:(\mathcal{A},\|\cdot\|_c)\to L^2_\mu$ at α_0 for the second.

$$\begin{aligned}&\sup_{\pi\in\bar{\mathcal{A}}_c}\|\pi\|_c^{-2}\left(\int_{\mathcal{X}\times(0,1)}(W_{\alpha_0+\psi}(t)-W_{\alpha_0}(t)-D_{\alpha_0}[\psi](t))D_{\alpha_0+\psi}[\pi](t)d\mu(t)\right)^2\\&\leqslant\|W_{\alpha_0+\psi}(t)-W_{\alpha_0}(t)-D_{\alpha_0}[\psi]\|^2_{L^2_\mu}\times\sup_{\pi\in\bar{\mathcal{A}}_c}\|\pi\|_c^{-2}\|D_{\alpha_0+\psi}[\pi]\|^2_{L^2_\mu}\\&\leqslant o\Big(\|\psi\|_c^2\Big)\times\|D_{\alpha_0+\psi}\|_o^2.\end{aligned} \tag{C.15}$$

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

$$\begin{aligned}\sup_{\pi\in\bar{\mathcal{A}}_c} \|\pi\|_c^{-2} &\left(\int_{\mathcal{X}\times(0,1)} D_{\alpha_0}[\psi](t)(D_{\alpha_0+\psi}[\pi](t) - D_{\alpha_0}[\pi](t))d\mu(t)\right)^2 \\ &\leqslant \|D_{\alpha_0}\|_o^2\|\psi\|_c^2 \times \sup_{\pi\in\bar{\mathcal{A}}_c} \|\pi\|_c^{-2}\|D_{\alpha_0+\psi}[\pi] - D_{\alpha_0}[\pi]\|_{L^2_\mu}^2 \\ &\leqslant \|D_{\alpha_0}\|_o^2\|\psi\|_c^2 \times \|D_{\alpha_0+\psi} - D_{\alpha_0}\|_o^2. \end{aligned} \tag{C.16}$$

To conclude, combine (C.14), (C.15) and (C.16) and $W_{\alpha_0}(t) = 0$ for all t to derive the first inequality in (C.17). As argued in (C.5), however, $\|D_{\bar{\alpha}}\|_o$ is bounded in a neighbourhood of α_0. Thus, the continuity of $D_{\bar{\alpha}}$ in $\bar{\alpha}$ for $\bar{\alpha} \in \mathcal{N}(\alpha_0)$ implies the final result in (C.17).

$$\begin{aligned}&\left\|\frac{dQ(\alpha_0+\psi)}{d\alpha} - \frac{dQ(\alpha_0)}{d\alpha} - T[\psi,\cdot]\right\|_o^2 \\ &\quad\leqslant o\left(\|\psi\|_c^2\right) \times \|D_{\alpha_0+\psi}\|_o^2 + \|\psi\|_c^2\|D_{\alpha_0}\|_o^2\|D_{\alpha_0+\psi} - D_{\alpha_0}\|_o^2 = o\left(\|\psi\|_c^2\right).\end{aligned} \tag{C.17}$$

It follows from (C.17) that T is the second Fréchet derivative of $Q(\alpha)$ at α_0. □

Proof of Theorem 3.1: Let $\Pi_n\alpha_0 = \arg\min_{\mathcal{A}_n} \|\alpha_0 - \alpha\|_s$. By Theorem 2.1, $\hat{\alpha} \in \mathcal{N}(\alpha_0)$ with probability tending to one and hence Assumptions 3.2(a) and 3.2(c), imply that with probability tending to one we have that

$$\begin{aligned}\|\hat{\alpha} - \alpha_0\|_w^2 &\lesssim Q(\hat{\alpha}) - Q(\Pi_n\alpha_0) + Q(\Pi_n\alpha_0) \\ &= Q(\hat{\alpha}) - Q(\Pi_n\alpha_0) + o(n^{-1}).\end{aligned} \tag{C.18}$$

By Theorem 2.1 and $\|\cdot\|_s \lesssim \|\cdot\|_c$, there is a $\delta_n \to 0$ such that $P(\|\hat{\alpha} - \alpha_0\|_s > \delta_n) \to 0$. Letting $\mathcal{A}_0^{\delta_n} = \{\alpha \in \mathcal{A} : \|\alpha - \alpha_0\|_s \leqslant \delta_n\}$ then yields the first inequality in (C.19). Noticing that $Q_n(\hat{\alpha}) \leqslant Q_n(\Pi_n\alpha_0)$ by virtue of $\hat{\alpha}$ minimizing $Q_n(\alpha)$ over $\mathcal{A}_n$ and using the Cauchy–Schwarz inequality gives us the second inequality. For the third and fourth inequalities we use Lemma B.1 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 $Q(\alpha)$.

$$\begin{aligned}Q(\hat{\alpha}) &- Q(\Pi_n\alpha_0) \\ &\leqslant Q_n(\hat{\alpha}) - Q_n(\Pi_n\alpha_0) + 2\sup_{\mathcal{A}_0^{\delta_n}} |Q_n(\alpha) - Q(\alpha)| \\ &\leqslant 2 \sup_{(t,\alpha)\in\mathbb{R}^{d_t}\times\mathcal{A}} |W_{\alpha,n}(t) - W_\alpha(t)| \times \left[\sup_{\mathcal{A}_0^{\delta_n}} \int_{\mathcal{X}\times(0,1)} (W_{\alpha,n}(t) + W_\alpha(t))^2 d\mu(t)\right]^{\frac{1}{2}} \\ &\leqslant O_p(n^{-\frac{1}{2}}) \times \left[\sup_{(t,\alpha)\in\mathbf{R}^{d_t}\times\mathcal{A}} (W_{\alpha,n}(t) - W_\alpha(t))^2 + \sup_{\mathcal{A}_0^{\delta_n}} 4\int_{\mathcal{X}\times(0,1)} W_\alpha^2(t)d\mu(t)\right]^{\frac{1}{2}} \\ &\leqslant O_p(n^{-\frac{1}{2}}) \times \left[O_p(n^{-1}) + \sup_{\mathcal{A}_0^{\delta_n}} 4Q(\alpha)\right]^{\frac{1}{2}}.\end{aligned} \tag{C.19}$$

By Assumption 3.2(a), $\sup_{\mathcal{A}_0^{\delta_n}} Q(\alpha) \lesssim \delta_n^2 = o(1)$. Therefore, combining (C.18) and (C.19):

$$\|\hat{\alpha} - \alpha_0\|_w^2 \lesssim O_p(n^{-\frac{1}{2}}) \times o_p(1) + o(n^{-1}) = o_p(n^{-\frac{1}{2}}). \tag{C.20}$$

To obtain a rate with respect to $\|\cdot\|_s$, we use Assumption 3.2(c) for the first and second inequalities in (C.21). It follows from (C.20) and Assumption 3.2(c) that $\|\hat{\alpha} - \Pi_n\alpha_0\|_w^2 = o_p(n^{-\frac{1}{2}})$ which together with Assumption 3.2(b) implies the equality in (C.21).

$$\|\hat{\alpha} - \alpha_0\|_s^2 \leqslant \|\hat{\alpha} - \Pi_n\alpha_0\|_s^2 + o(n^{-1}) \leqslant \sup_{\alpha\in\mathcal{A}_n} \frac{\|\alpha\|_s^2}{\|\alpha\|_w^2} \times \|\hat{\alpha} - \Pi_n\alpha_0\|_w^2 + o(n^{-1}) = o_p(n^{-\frac{1}{2}+\gamma}). \tag{C.21}$$

We can now exploit the local behaviour of the objective function to improve on the obtained rate of convergence. Note that due to (C.21) it is possible to choose $\delta_n = o(n^{-\frac{1}{4}+\frac{\gamma}{2}})$ such that $P(\hat{\alpha} \in \mathcal{A}_0^{\delta_n}) \to 1$. Repeating the steps in (C.19) we obtain (C.22) with probability approaching one.

$$\begin{aligned} Q(\hat{\alpha}) - Q(\Pi_n\alpha_0) &\leqslant O_p(n^{-\frac{1}{2}}) \times \left[O_p(n^{-1}) + \sup_{\mathcal{A}_0^{\delta_n}} 4Q(\alpha) \right]^{\frac{1}{2}} \\ &= O_p(n^{-\frac{1}{2}}) \times o_p(n^{-\frac{1}{4}+\frac{\gamma}{2}}). \end{aligned} \tag{C.22}$$

From (C.18), (C.22) and Assumption 3.2(b), 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 (C.21) 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 $\|\hat{\alpha} - \alpha_0\|_s^2 = o_p(n^{(\gamma-\frac{1}{2})(1+\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\cdots)})$. 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 α, it is convenient to define

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

Throughout the proofs we will exploit the following lemma:

LEMMA D.1. *If Assumptions 2.1, 2.3, 3.1, 3.2 hold, then:* $Q_n^s(\hat{\alpha}) \leqslant \inf_{\mathcal{A}_n} Q_n^s(\alpha) + o_p(n^{-1})$.

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

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

By simple manipulations we therefore obtain

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

Next, apply Jensen's inequality and $Q_n(\hat{\alpha}) \leqslant Q_n(\Pi_n\alpha_0)$ to obtain the first and second inequalities in (D.3). By Lemma B.1, $\sup_{t,\alpha} |W_{\alpha,n}(t) - W_\alpha(t)| = O_p(n^{-\frac{1}{2}})$. Together with Assumption 3.2(a), the final two inequalities in (D.3) then immediately follow.

$$
\begin{aligned}
\int_{\mathcal{X}\times(0,1)} & |W_{\hat{\alpha},n}(t)| d\mu(t) \\
&\leqslant \left[\int_{\mathcal{X}\times(0,1)} W^2_{\hat{\alpha},n}(t) d\mu(t)\right]^{\frac{1}{2}} \\
&\leqslant \left[\int_{\mathcal{X}\times(0,1)} W^2_{\Pi_n\alpha_0,n}(t) d\mu(t)\right]^{\frac{1}{2}} \\
&\leqslant \left[2\int_{\mathcal{X}\times(0,1)} (W_{\Pi_n\alpha_0,n}(t) - W_{\Pi_n\alpha_0}(t))^2 d\mu(t) + 2\int_{\mathcal{X}\times(0,1)} W^2_{\Pi_n\alpha_0}(t) d\mu(t)\right]^{\frac{1}{2}} \\
&\lesssim \left[O_p(n^{-1}) + \|\Pi_n\alpha_0 - \alpha_0\|_s^2\right]^{\frac{1}{2}} .
\end{aligned} \tag{D.3}
$$

By Assumption 3.2(c), $\|\Pi_n\alpha_0 - \alpha_0\|_s = o(n^{-\frac{1}{2}})$ and hence combining (D.2) and (D.3),

$$
Q_n^s(\hat{\alpha}) \leqslant Q_n(\hat{\alpha}) + o_p(n^{-1}). \tag{D.4}
$$

Let $\tilde{\alpha} \in \arg\min_{\mathcal{A}_n} Q_n^s(\alpha)$, and note that Lemma B.1 and the same arguments as in Theorem 2.1 imply $\|\alpha_0 - \tilde{\alpha}\|_c = o_p(1)$. The same arguments as in (D.2) then imply that $Q_n(\tilde{\alpha})$ is bounded above by

$$
\begin{aligned}
&\int_{\mathcal{X}\times(0,1)} (|W_{\tilde{\alpha},n}(t) - W_{\alpha_0,n}(t) - W_{\tilde{\alpha}}(t)| + |W_{\alpha_0,n}(t) + W_{\tilde{\alpha}}(t)|)^2 d\mu(t) \\
&= \int_{\mathcal{X}\times(0,1)} (W_{\alpha_0,n}(t) + W_{\tilde{\alpha}}(t))^2 d\mu(t) + o_p(n^{-\frac{1}{2}}) \times \int_{\mathcal{X}\times(0,1)} |W_{\alpha_0,n}(t) + W_{\tilde{\alpha}}(t)| d\mu(t) + o_p(n^{-1}).
\end{aligned} \tag{D.5}
$$

Proceeding as in (D.3), Jensen's inequality and $Q_n^s(\tilde{\alpha}) \leqslant Q_n^s(\Pi_n\alpha_0)$ imply the first and second inequalities in (D.6). The last two results in (D.6) then follow by Assumption 3.2(a) and by noting that Lemma B.1 implies $\sup_t |W_{\alpha_0,n}(t)| = O_p(n^{-\frac{1}{2}})$,

$$
\begin{aligned}
\int_{\mathcal{X}\times(0,1)} & |W_{\alpha_0,n}(t) + W_{\tilde{\alpha}}(t)| d\mu(t) \\
&\leqslant \left[\int_{\mathcal{X}\times(0,1)} (W_{\alpha_0,n}(t) + W_{\tilde{\alpha}}(t))^2 d\mu(t)\right]^{\frac{1}{2}} \\
&\leqslant \left[\int_{\mathcal{X}\times(0,1)} (W_{\alpha_0,n}(t) + W_{\Pi_n\alpha_0}(t))^2 d\mu(t)\right]^{\frac{1}{2}} \\
&\leqslant \left[2\int_{\mathcal{X}\times(0,1)} W^2_{\alpha_0,n}(t) d\mu(t) + 2\int_{\mathcal{X}\times(0,1)} W^2_{\Pi_n\alpha_0}(t) d\mu(t)\right]^{\frac{1}{2}} \\
&\lesssim \left[O_p(n^{-1}) + \|\alpha_0 - \Pi_n\alpha_0\|_s^2\|\right]^{\frac{1}{2}} .
\end{aligned} \tag{D.6}
$$

Since $\|\Pi_n\alpha_0 - \alpha_0\|_s = o(n^{-\frac{1}{2}})$ by Assumption 3.2(c), (D.5) and (D.6) imply

$$
Q_n(\tilde{\alpha}) \leqslant Q_n^s(\tilde{\alpha}) + o_p(n^{-1}). \tag{D.7}
$$

Hence, since $Q_n(\hat{\alpha}) \leqslant Q_n(\tilde{\alpha})$, the definition of $\tilde{\alpha}$ together with (D.4) and (D.7) establish

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

which establishes the claim of the lemma. □

Proof of Lemma 4.1: The arguments closely follow those of Ai and Chen (2003). We first establish continuity. Since F_λ is linear, it is only necessary to establish that it is bounded. For any $\theta \in \mathbb{R}^{d_\theta}$, we can obtain the first equality in (D.8) by using (4.2), while the second equality is definitional.

$$\min_{h\in\bar{\mathcal{H}}} \int_{\mathcal{X}\times(0,1)} \left(\frac{dW(\alpha_0)}{d\theta'}[\theta](t) - \frac{dW(\alpha_0)}{dh}[h](t)\right)^2 d\mu(t)$$
$$= \int_{\mathcal{X}\times(0,1)} \left(\left[\frac{dW(\alpha_0)}{d\theta}(t) - \frac{dW(\alpha_0)}{dh}[h^*](t)\right]'\theta\right)^2 d\mu(t) = \theta'\Sigma^*\theta. \tag{D.8}$$

In order to show F_λ is bounded we need to establish the left-hand side of (D.9) is finite. Using (D.8) immediately implies the first equality in (D.9). For the second equality note the optimization problem is solved at $\theta^* = (\Sigma^*)^{-1}\lambda$ and plug in θ^*.

$$\sup_{0\neq\alpha\in\bar{\mathcal{A}}} \frac{F_\lambda^2(\alpha)}{\|\alpha\|_w^2} = \sup_{0\neq\theta\in\mathbb{R}^{d_\theta}} \frac{(\lambda'\theta)^2}{\theta'\Sigma^*\theta} = \lambda'(\Sigma^*)^{-1}\lambda. \tag{D.9}$$

Since by assumption Σ^* is positive definite, (D.9) is finite and hence F_λ is bounded which establishes continuity. For the second claim of the lemma, note the following orthogonality condition must hold as a result of (4.1) and (4.2):

$$\int_{\mathcal{X}\times(0,1)} \left(\frac{dW(\alpha_0)}{d\theta}(t) - \frac{dW(\alpha_0)}{dh}[h^*](t)\right)\frac{dW(\alpha_0)}{dh}[h](t)d\mu(t) = 0 \tag{D.10}$$

for all $h \in \bar{\mathcal{H}}$. Therefore, employing result (D.10) we obtain $\langle \alpha - \alpha_0, v^\lambda\rangle$ equals:

$$(\theta-\theta_0)'\left\{\int_{\mathcal{X}\times(0,1)} \left[\frac{dW(\alpha_0)}{d\theta}(t) - \frac{dW(\alpha_0)}{dh}[h^*](t)\right]\left[\frac{dW(\alpha_0)}{d\theta}(t) - \frac{dW(\alpha_0)}{dh}[h^*](t)\right]' d\mu(t)\right\} v_\theta^\lambda.$$

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

LEMMA D.2. *Let Assumptions 2.1, 2.3, 3.1, 3.2 and 4.1 hold, and let* $v_n^\lambda = \Pi_n v^\lambda$. *Then: (a)* $\int_{\mathcal{X}\times(0,1)} W_{\alpha_0,n}(t)D_{\hat\alpha}[v_n^\lambda](t)d\mu(t) = \int_{\mathcal{X}\times(0,1)} W_{\alpha_0,n}(t)D_{\alpha_0}[v^\lambda](t)d\mu(t) + o_p(n^{-\frac{1}{2}})$; *also (b)* $\int_{\mathcal{X}\times(0,1)}(W_{\hat\alpha}(t) - W_{\alpha_0}(t))D_{\hat\alpha}[v_n^\lambda](t)d\mu(t) = \int_{\mathcal{X}\times(0,1)} D_{\alpha_0}[\hat\alpha - \alpha_0](t)D_{\alpha_0}[v^\lambda](t)d\mu(t) + o_p(n^{-\frac{1}{2}})$; *and (c)* $\sqrt{n}W_{\alpha_0,n}(t) \xrightarrow{\mathcal{L}} G(t)$, *where $G(t)$ is a Gaussian process with covariance:*

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

Proof: To establish the first claim apply the Cauchy–Schwarz inequality, the definition of the operator norm, Theorem 2.1 and Lemma B.1 implying $\sup_t |W_{\alpha_0,n}(t)| = O_p(n^{-\frac{1}{2}})$ to obtain that with probability approaching one we have

$$\left|\int_{\mathcal{X}\times(0,1)} W_{\alpha_0,n}(t)D_{\hat\alpha}[v_n^\lambda - v^\lambda](t)d\mu(t)\right| \leqslant \left[\int_{\mathcal{X}\times(0,1)} W_{\alpha_0,n}^2(t)d\mu(t)\right]^{\frac{1}{2}} \times \|D_{\hat\alpha}\left[v_n^\lambda - v^\lambda\right]\|_{L_\mu^2}$$
$$\leqslant O_p(n^{-\frac{1}{2}}) \times \sup_{\alpha\in\mathcal{N}(\alpha_0)} \|D_\alpha\|_o \times \|v_n^\lambda - v^\lambda\|_c. \tag{D.11}$$

As argued in (C.5), $\sup_{\alpha\in\mathcal{N}(\alpha_0)}\|D_\alpha\|_o < \infty$. Further, Assumptions 4.1(b) and 2.3(f) imply that $\|v^\lambda - v_n^\lambda\|_c = o(1)$. Therefore, we obtain from (D.11) that

$$\int_{\mathcal{X}\times(0,1)} W_{\alpha_0,n}(t)D_{\hat\alpha}\left[v_n^\lambda\right](t)d\mu(t) = \int_{\mathcal{X}\times(0,1)} W_{\alpha_0,n}(t)D_{\hat\alpha}[v^\lambda](t)d\mu(t) + o_p(n^{-\frac{1}{2}}). \tag{D.12}$$

Similarly, the derivations in (D.11) imply the inequality in (D.13). The equality is a result of the continuity of D_α in α under $\|\cdot\|_c$, as established in the proof of Lemma 3.1.

$$\left|\int_{\mathcal{X}\times(0,1)} W_{\alpha_0,n}(t)(D_{\hat{\alpha}}[v^\lambda](t) - D_{\alpha_0}[v^\lambda](t))d\mu(t)\right| \leqslant O_p(n^{-\frac{1}{2}}) \times \|D_{\hat{\alpha}} - D_{\alpha_0}\|_o \times \|v^\lambda\|_c = o_p(n^{-\frac{1}{2}}). \tag{D.13}$$

Together, equations (D.12) and (D.13) establish the first claim of the lemma.

For the second claim of the lemma, note that Assumption 4.1(c) allows us to do a second-order Taylor expansion to obtain (D.14) pointwise in $t \in \mathcal{X} \times (0, 1)$,

$$W_{\hat{\alpha}}(t) = W_{\alpha_0}(t) + D_{\alpha_0}[\hat{\alpha} - \alpha_0](t) + \frac{1}{2}\frac{dD_{\alpha_0+\tau(\hat{\alpha}-\alpha_0)}[\hat{\alpha}-\alpha_0](t)}{d\tau}\bigg|_{\tau=s(t)}. \tag{D.14}$$

The first equality in (D.15) then follows from (D.14), while the second one is implied by Assumptions 4.1(c) and 4.1(d). The final equality in turn follows from Theorem 3.1.

$$\begin{aligned}&\int_{\mathcal{X}\times(0,1)} (W_{\hat{\alpha}}(t) - W_{\alpha_0}(t) - D_{\alpha_0}[\hat{\alpha}-\alpha_0](t))D_{\hat{\alpha}}\left[v_n^\lambda\right](t)d\mu(t)\\ &= \frac{1}{2}\int_{\mathcal{X}\times(0,1)}\left(\frac{dD_{\alpha_0+\tau(\hat{\alpha}-\alpha_0)}[\hat{\alpha}-\alpha_0](t)}{d\tau}\bigg|_{\tau=s(t)}\right)D_{\hat{\alpha}}\left[v_n^\lambda\right](t)d\mu(t) \lesssim \|\hat{\alpha}-\alpha_0\|_s^2 = o_p(n^{-\frac{1}{2}}).\end{aligned} \tag{D.15}$$

Next, apply the Cauchy–Schwarz inequality and a Taylor expansion to obtain the first inequality in (D.16). The second inequality then follows by Assumption 4.1(c), $\|\hat{\alpha}-\alpha_0\|_w \lesssim \|\hat{\alpha}-\alpha_0\|_s$ in a neighbourhood of α_0 by Assumption 3.2(a) and Theorem 3.1.

$$\begin{aligned}&\left|\int_{\mathcal{X}\times(0,1)} D_{\alpha_0}[\hat{\alpha}-\alpha_0](t)(D_{\hat{\alpha}}\left[v_n^\lambda\right](t) - D_{\alpha_0}\left[v_n^\lambda\right](t))d\mu(t)\right|\\ &\leqslant \|\hat{\alpha}-\alpha_0\|_w \times \left[\int_{\mathcal{X}\times(0,1)}\left(\frac{dD_{\alpha_0+\tau(\hat{\alpha}-\alpha_0)}\left[v_n^\lambda\right](t)}{d\tau}\bigg|_{\tau=s(t)}\right)^2 d\mu(t)\right]^{\frac{1}{2}} \lesssim \|\hat{\alpha}-\alpha_0\|_s^2 = o_p(n^{-\frac{1}{2}}).\end{aligned} \tag{D.16}$$

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 3.2(b) we are able to conclude,

$$\left|\int_{\mathcal{X}\times(0,1)} D_{\alpha_0}[\hat{\alpha}-\alpha_0](t)(D_{\alpha_0}\left[v_n^\lambda\right](t) - D_{\alpha_0}[v^\lambda](t))d\mu(t)\right| \leqslant \|\hat{\alpha}-\alpha_0\|_w \times \|D_{\alpha_0}\|_o \times \|v_n^\lambda - v^\lambda\|_c = o_p(n^{-\frac{1}{2}}). \tag{D.17}$$

Combining results (D.15)–(D.17) 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 B.1 and regular Central Limit Theorem. □

Proof of Theorem 4.1: Let $u^* = \pm v^\lambda$, $u_n^* = \Pi_n u^*$ and $0 < \varepsilon_n = o(n^{-\frac{1}{2}})$ be such that it satisfies $Q_n^s(\hat{\alpha}) \leqslant \inf_{\mathcal{A}_n} Q_n^s(\alpha) + O_p(\varepsilon_n^2)$, which is possible due to Lemma D.1. Define $\alpha(\tau) = \hat{\alpha} + \tau\varepsilon_n u_n^*$ and note that by Assumption 3.1(a) and Lemma 2.1, with probability tending to one $\alpha(\tau) \in \mathcal{A}_n$ for $\tau \in [0, 1]$. Therefore, Lemma D.1 establishes the first equality in (D.18). A second-order Taylor expansion around $\tau = 0$ yields the equality in (D.18) for some $s \in [0, 1]$.

$$\begin{aligned}0 &\leqslant Q_n^s(\alpha(1)) - Q_n^s(\alpha(0)) + O_p\left(\varepsilon_n^2\right)\\ &= 2\varepsilon_n\int_{\mathcal{X}\times(0,1)} (W_{\alpha_0,n}(t) + W_{\hat{\alpha}}(t))D_{\hat{\alpha}}\left[u_n^*\right](t)d\mu(t) + \frac{1}{2}\frac{d^2 Q_n(\alpha(\tau))}{d\tau^2}\bigg|_{\tau=s},\end{aligned} \tag{D.18}$$

where by direct calculation we have that

$$\left.\frac{d^2 Q_n(\alpha(\tau))}{d\tau^2}\right|_{\tau=s} = \varepsilon_n^2 \int_{\mathcal{X}\times(0,1)} \left(D_{\alpha(s)}\left[u_n^*\right](t)\right)^2 d\mu(t) + \int_{\mathcal{X}\times(0,1)} (W_{\alpha_0,n}(t) + W_{\alpha(s)}(t)) \left.\frac{dD_{\hat{\alpha}+\tau\varepsilon_n u_n^*}\left[\varepsilon_n u_n^*\right](t)}{d\tau}\right|_{\tau=s} d\mu(t). \tag{D.19}$$

As shown in (C.5), $\sup_{\alpha\in\mathcal{N}(\alpha_0)} \|D_\alpha\|_o < \infty$, and hence, since $\|v^\lambda\|_c < \infty$, we obtain that

$$\int_{\mathcal{X}\times(0,1)} \left(D_{\alpha(s)}\left[u_n^*\right](t)\right)^2 d\mu(t) \leqslant \sup_{\alpha\in\mathcal{N}(\alpha_0)} \|D_\alpha\|_o^2 \times \left\|u_n^*\right\|_c^2 = O(1). \tag{D.20}$$

Since $W_{\alpha,n}(t)$ and $W_\alpha(t)$ are both bounded by 1, Assumption 4.1(c) establishes

$$\left|\int_{\mathcal{X}\times(0,1)} (W_{\alpha_0,n}(t) + W_{\alpha(s)}(t)) \left.\frac{dD_{\hat{\alpha}+\tau\varepsilon_n u_n^*}\left[\varepsilon_n u_n^*\right](t)}{d\tau}\right|_{\tau=s} d\mu(t)\right| \lesssim \left\|\varepsilon_n u_n^*\right\|_s^2 = O\left(\varepsilon_n^2\right). \tag{D.21}$$

Therefore, by combining (D.18)–(D.21), $u_n^* = \pm v_n^\lambda$ and $\varepsilon_n = o(n^{-\frac{1}{2}})$, it follows that:

$$\int_{\mathcal{X}\times(0,1)} (W_{\alpha_0,n}(t) + W_{\hat{\alpha}}(t)) D_{\hat{\alpha}}\left[u_n^*\right](t) d\mu(t) = o_p(n^{-\frac{1}{2}}). \tag{D.22}$$

To conclude, in (D.23) use Lemma 4.1 for the first equality, Lemma D.2(b) for the second equality, $W_{\alpha_0}(t) = 0$ and (D.22) for the third one and Lemma D.2(a) for the final result.

$$\begin{aligned}\sqrt{n}\lambda'(\hat{\theta} - \theta_0) &= \sqrt{n} \int_{\mathcal{X}\times(0,1)} D_{\alpha_0}[\hat{\alpha} - \alpha_0](t) D_{\alpha_0}[v^\lambda](t) d\mu(t) \\ &= \sqrt{n} \int_{\mathcal{X}\times(0,1)} (W_{\hat{\alpha}}(t) - W_{\alpha_0}(t)) D_{\hat{\alpha}}\left[v_n^\lambda\right](t) d\mu(t) + o_p(1) \\ &= \sqrt{n} \int_{\mathcal{X}\times(0,1)} W_{\alpha_0,n}(t) D_{\hat{\alpha}}\left[v_n^\lambda\right](t) d\mu(t) + o_p(1) \\ &= \sqrt{n} \int_{\mathcal{X}\times(0,1)} W_{\alpha_0,n}(t) D_{\alpha_0}[v^\lambda](t) d\mu(t) + o_p(1).\end{aligned} \tag{D.23}$$

Hence, applying Lemma D.2(c) we are able to conclude from (D.23) that

$$\sqrt{n}\lambda'(\hat{\theta} - \theta) \xrightarrow{\mathcal{L}} N(0, \Omega_\lambda), \tag{D.24}$$

where $\Omega_\lambda = \int D_{\alpha_0}[v^\lambda](t) D_{\alpha_0}[v^\lambda](s) \Sigma(t, s) d\mu(t) d\mu(s)$. Using the closed form for v^λ, obtained in Lemma 4.1, and the definition of $R_{h^*}(t)$ in turn imply

$$\begin{aligned}D_{\alpha_0}[v^\lambda](t) &= \left[\frac{dW(\alpha_0)}{d\theta}(t) - \frac{dW(\alpha_0)}{dh}[h^*](t)\right] [\Sigma^*]^{-1}\lambda \\ &= R_{h^*}(t)[\Sigma^*]^{-1}\lambda.\end{aligned} \tag{D.25}$$

The Cramér–Wold device, (D.24) and (D.25) then establish the claim of the theorem. □

APPENDIX E: DETAILS OF THE BLP EXAMPLE

In this appendix, we give the proofs of Lemma 5.1 and Theorem 5.1. We start with an auxiliary lemma whose result will be useful later on.

LEMMA E.1. *Assume F is twice continuously differentiable on $\mathbb{R}$ with strictly increasing hazard rate τ. 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: 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

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

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 exists 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

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

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

$$\Phi(p_1, p_2, \xi) = \begin{bmatrix} \dfrac{1}{p_1 - c} + \dfrac{\partial \ln D_1(p_1, p_2)}{\partial p_1} \\ \dfrac{1}{p_2 - c} + \dfrac{\partial \ln D_2(p_1, p_2)}{\partial p_2} \end{bmatrix}.$$

Note that the map Φ is continuously differentiable and we have

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

In addition, note that the demand function in (5.2) satisfies

$$-\frac{\partial^2 \ln D_j(p_j, p_{-j})}{\partial p_j^2} = \frac{\partial^2 \ln D_j(p_j, p_{-j})}{\partial p_j \partial p_{-j}} = a\frac{\partial^2 \ln D_j(p_j, p_{-j})}{\partial p_j \partial(\xi_j - \xi_{-j})} > 0, \tag{E.1}$$

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

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

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

$$\begin{pmatrix} \dfrac{\partial p_1}{\partial \xi} \\ \dfrac{\partial p_2}{\partial \xi} \end{pmatrix} = -[\mathrm{D}_{(p_1,p_2)}\Phi(p_1, p_2, \xi)]^{-1}\mathrm{D}_{\xi}\Phi(p_1, p_2, \xi).$$

Thus,

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

$$\frac{\partial p_2}{\partial \xi} = \frac{1}{a}\frac{1}{\det \mathrm{D}_{(p_1,p_2)}\Phi}\frac{1}{(p_1 - c)^2}\frac{\partial^2 \ln D_2(p_1, p_2)}{\partial p_2^2}, \tag{E.3}$$

where the first equality uses (E.1) 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 (E.1) 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 (E.2) to (E.3) 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 5.1: Since $\xi = \xi_1 - \xi_2$ is continuously distributed, it has a strictly increasing cdf, which we denote F_ξ. Noting that $F_\xi(\xi) \sim U(0, 1)$, we may define

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

so that

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

Since that by Lemma E.1 h is continuously differentiable 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 5.1. □

Proof of Theorem 5.1: Consider the BLP model in (5.4) and let $F_{Y|X}(\cdot|\cdot)$ denote the conditional distribution of Y given X that is induced by the structure (θ, h). Fix $x \in \mathcal{X}$ and let $v : \mathbb{R} \times \mathcal{X} \to (0, 1)$ 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 (a) we have $\partial h(x,u)/\partial u > 0$. Then, for any $(y,x) \in \mathcal{S}$,

$$\begin{aligned} F_{Y\,|\,X}(y|x) &= P(Y \leqslant y \mid X = x) \\ &= P(h(X,U) \leqslant y - \theta' x \mid X = x) \\ &= P(U \leqslant v(y - \theta' x, x) \mid X = x) \\ &= P(U \leqslant v(y - \theta' x, x)) \\ &= v(y - \theta' x, x), \end{aligned} \tag{E.4}$$

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

$$\begin{aligned} \frac{\partial v}{\partial x}(t,x) &= -\frac{\partial h}{\partial x}(x, v(t,x)) \left[\frac{\partial h}{\partial u}(x, v(t,x)) \right]^{-1}, \\ \frac{\partial v}{\partial t}(t,x) &= \left[\frac{\partial h}{\partial u}(x, v(t,x)) \right]^{-1}. \end{aligned} \tag{E.5}$$

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

$$\begin{aligned} \frac{\partial \Phi}{\partial y}(y,x) &= \frac{\partial v}{\partial t}(y - \theta' x, x), \\ \frac{\partial \Phi}{\partial x}(y,x) &= -\theta \frac{\partial v}{\partial t}(y - \theta' x, x) + \frac{\partial v}{\partial x}(y - \theta' x, x). \end{aligned} \tag{E.6}$$

In particular, $\partial \Phi(y,x)/\partial y > 0$ on $\mathcal{S}$. Combining (E.5) and (E.6) we then obtain

$$-\left[\frac{\partial \Phi}{\partial x}(y,x) \right] \left[\frac{\partial \Phi}{\partial y}(y,x) \right]^{-1} = \theta + \frac{\partial h}{\partial x}\Big(x, v(y - \theta' x, x)\Big). \tag{E.7}$$

Evaluate the left-hand side of (E.7) 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 5.1, $v(0,0) = 1/2$. Combining the latter with condition (b) 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 θ is identified. The identification of v, and hence h, then immediately follows from (E.4). □