Efficient Semiparametric Estimation of Nonseparable Models.

Sami STOULI

University College London - IFS/Cemmap

December 10, 2009

Abstract

This paper considers semiparametric estimation of quantile regression models in the presence of endogeneity. Imposing index restrictions in the triangular model framework, an efficient semiparametric estimator for nonseparable models is introduced. Endogeneity is addressed through a control variable approach. The estimator has four main characteristics: 1) It is based on conditional quantile restrictions; 2) it allows for dimension reduction; 3) it addresses endogeneity of one or more covariates; 4) it is efficient. Large-sample properties of the proposed estimator are derived and an empirical application on the effect of property rights on land profits is discussed.

1 Introduction

This paper proposes a flexible estimator for nonseparable models. These models are central in econometrics as economic theory seldom provides as much guidance for empirical model specification as would be required to design an appropriate econometric framework. Indeed, behavioural and equilibrium assumptions from economic theory typically suggests shape and exclusion restrictions but no parametric restrictions and no separability in the disturbances (Matzkin (2007)). However, it can be adventurous to simply supplement theoretical models by imposing parametric structures that functions and distributions may not possess. This applies for instance to models with endogeneity which is a class of models characterized by nonseparable disturbances. As noted by Imbens and Newey (2009), nonseparability in the disturbances is an intrinsic feature of models with endogeneity. For these reasons, identification in nonseparable models has recently attracted a lot of attention (Chesher (2003), Matzkin (2008), Newey and Imbens (2009), Heckman, Matzkin and Nesheim (2009)). Typical such models are of common use in Labor or Development economics and estimation of their structural features usually relies on imposing arbitrary parametric restrictions or assuming separable disturbances - standard instrumental variable methods are of this sort. These empirical approaches have several drawbacks, not the least of them being that they do not allow for the heterogeneity that naturally arises from nonseparability in the disturbances in the economic model. Therefore, a nonparametric analog estimator incorporating nonseparability in the disturbances would strongly weaken the restrictions imposed a priori by practitioneers and significantly reduce the risk of misspecification in commonly used econometric models.

This paper is concerned with a situation where a (continuous) regressor is correlated with the errors because of some unobservable common factor affecting both the outcome and the endogenous variable. Continuity of endogeneous regressors ensures the possibility of using a control variable. Besides, the disturbance in the outcome equation is restricted to be at most two-dimentional. One positive aspect of this restriction is that it allows for identification of individual effects, but at the cost of imposing strong restrictions on the dimensionality of disturbances. Another advantage is that thanks to the local identification approach adopted here, the validity of the model is preserved for some parts of the distribution even when global identification does not hold. Indeed, identification of structural features of the model rely upon local independence conditions whereas Imbens and Newey (2009) require global identification restrictions. Relying on conditional quantile restrictions leads to constructive identification in nonseparable models, as shown by Chesher (2001, 2003). Therefore, this paper proposes a semiparametric quantile estimator for nonseparable models that fully exploits nonparametric identification results, while maintaining parametrization restrictions close to minimal. This is achieved by the use of nonparametric quantile regression estimation methods which offer a comprehensive characterization of the stochastic relationship among variables and are robust to non-Gaussianity of disturbances as well as to misspecification.

Misspecified econometric models lead to biased estimates and this is all the more an issue when the effects of interest depend on preliminary estimation of some of the regressors, as it is naturally the case with control variable approaches. First, consistency of the second-stage estimate will be affected by consistency of the first-stage, which demonstrates the critical nature of specification in the first-stage. Second, the efficiency of a "plug-in" estimator will be adversely affected if the parametrization of the nuisance functions is incorrect. Misspecification can lead to big efficiency losses if parametrization of the second equation is incorrect. Hence, addressing potential misspecification is an important task in order to develop reliable inference procedures in nonseparable models. The semiparametric approach to misspecification in this type of models is to allow for the functional forms of some components of each equation in the triangular model to be unrestricted. Naturally, flexible specification gives rise to the curse of dimensionality and this has to be addressed for estimators to be of any use in practice. Indeed, when it comes to nonseparable models, the curse of dimensionality is all the more inevitable that the structural function has dimension at least two, while in general its dimension will be much larger. Imposing index restrictions fits well the nonseparable framework, encompasses a large class of models and is not nested with additive models. Therefore, the model incorporates index restrictions in order to deal with the curse of dimensionality. Besides, a large class of models naturally falls into this framework such as censored regression quantile models with endogeneity (Blundell and Powell (2007)) and several transformation models.

Efficiency. Chen and Pouzo (2009) - Ai and Chen (2009). Efficiency bounds are of fundamental importance for semiparametric models. Such bounds quantify the efficiency loss that can result from a semiparametric, rather than a parametric, approach. This paper derives the asymptotic efficiency bounds of a control variable estimator for nonseparable models.

Review of the literature. Estimation of parametric triangular models based on quantile regression has been considered by various authors. A first trend in the literature has seen the development of estimators for the class of location-shift models, that is, models in which the effect of covariates is invariant to the particular quantile of the control variable used in the estimation. Amemiya (1982) first introduced a class of two-stage median regression estimators. More recently, Lee (2007) considered a semiparametric quantile regression version of a separable triangular model, where the control variable enters additively and is approximated via series. Removing the separability intrinsic to location models, Ma and Koenker (2006) developed two general classes of estimators for a location-scale form of a parametric model. Building on identification results derived by Chesher (2003), they construct an estimator which allows for a description of the entire stochastic relationship between the endogeneous variable and the outcome. Following their approach Jun (2009) suggested a semiparametric estimator based on a random coefficients model. His model preserves the additivity assumption between covariates of the structural equation while allowing for nonseparability in the disturbances.

Another approach to identification and estimation of nonseparable models has been to consider a single-equation framework (Chernozhukov and Hansen (2005), Chernozhukov et al. (2007), Chen and Pouzo (2008)). The triangular framework is less general and present both advantages and drawbacks compared to the single-equation approach. One such drawback has recently been underlined by Chesher (2009) who clearly defines the natural 'identificational' limits of the triangular framework. Some advantages are that the problem considered here is well-posed and hence leads to simpler asymptotics, and thanks to the model completeness it incorporates the whole of the information available in structural economic model (by specifying the data generating process of the endogenous variable), and hence allows for heterogeneous marginal effects. Indeed the triangular framework allows one to consider the full stochastic behaviour of the model. Last, but not least, a triangular framework naturally arises in various empirical problems where one is interested in the complete stochastic description of the model: consumer demand, labour market, property rights and profits.

Plan of the paper. Section 2 introduces and motivates the model. Section 3 describes the sieve estimator and conditions for consistency and the derivation of convergence rates are discussed. In section 4 I introduce a profiled sieve estimation strategy for quantile regression index models, and consider in some detail the implementation of the proposed estimator. Results of some preliminary simulations are shown. Section 5 describes a potential empirical application. Last, section 6 concludes.

2 The model

This paper considers a triangular structural model which has an outcome equation

$$Y_1 = h_1(X'\theta, U, V) \tag{1}$$

where Y_1 is a scalar continuous random variable, $X \equiv \left[Y_2 \vdots W\right]$ is a k-vector of observed random variables, including Y_2, a continuous endogenous variable and $k-1$ exogenous variables W. Endogeneity of Y_2 arises from its covariation with the unobserved random variable V. θ is a k-component vector of finite-dimensional parameters. Also, h_1 is an unknown function restricted to be strictly monotonic in U. U is a scalar random variable that satisfies the independence condition

$$U \sim U(0,1)|X, V.$$

Besides the outcome equation, completeness of the model is assumed by specifying the structural equation for Y_2 as

$$Y_2 = h_2(Z, V), \tag{2}$$

where Z is a vector of random variables, and V is a scalar continuous random variable, and $h_2(Z,V)$ is strictly monotonic in V. Equation (2) can also be thought of as a reduced form for Y_2.

Remark 2.1. (i) For simplicity $X'\theta$ is assumed to be scalar, however results of the paper easily extends to vector valued $X'\theta$ (this case is considered in the application). (ii) Results in this paper also extends to a triangular model with m equations.

This framework can be motivated by various economic examples. In the recent literature the study of returns to education is the leading motivation for the triangular framework. In this paper, I detail how this framework arises in the study of efficient choice of land use in a fallow farming system. Consider land productivity in Ghana, as in Goldstein and Udry (2008). Individual farmers have to make decisions about the optimal path of fertility and of agricultural output for a given plot. Their decision will depend on two main criteria: (i) the opportunity cost of capital to a particular farmer, and (ii) farmer's confidence in the ability to reestablish cultivation on the plot after fallowing. Assume that the aim of an individual i with control over a set P_i of plots of land (indexed by p) is to manage fertility to maximize the present value of the stream of profits she can claim from this land. The profit per hectare generated by the cultivation of a plot is denoted by π_p, and is strictly concave and increasing in τ (**assumption**). Let the production function of a plot p be $Y_1 = h_1(\tau, \mathbf{X}_p, u, \nu)$, where τ denotes fallow duration, $\mathbf{X}_p$ is a vector of fixed characteristics of plot p, and ν is the unobserved heterogeneity in cost of capital and u is the unobserved heterogeneity in land productivity. Suppose that h_1 is increasing in u and is twice continuously differentiable (**assumption**). The decision facing the individual i is the length of time she should leave each plot fallow before cultivation. Then the individual maximizes the difference between expected revenues given observed and unobserved plot's characteristics and costs of fallowing by solving the following problem:

$$\tau_p^* = \arg\max_{\tau} \mathbb{E}[\pi(\tau, \mathbf{X}_p, u, \nu) \mid Z, \nu] = \arg\max_{\tau} \left[\mathbb{E}[h_1(\tau, \mathbf{X}_p, u, \nu) \mid Z, \nu] - C(\tau, \mathbf{X}_p, \omega_p, \nu)\right], \tag{3}$$

where C is the cost of a particular number of years of fallowing, which depends on ω_p, the likelihood of losing plot p during a year in which it is fallow (equivalently, security of tenure over the plot), plot's characteristics and the unobserved opportunity cost of capital. ω_p will vary according to i's status in local political hierarchies and according to the manner in which she acquired plot p. The optimal fallow duration will then satisfy the first order condition

$$\frac{\partial \mathbb{E}[h_1(\tau, \mathbf{X}_p, u, \nu) \mid Z, \nu]}{\partial \tau} = \frac{\partial C(\tau, \mathbf{X}_p, \omega_p, \nu)}{\partial \tau}, \tag{4}$$

which shows that τ is a function h_2 of $\mathbf{X}_p$, ω_p and ν: $\tau = h_2(Z, \nu)$, with $Z = [\mathbf{X}_p' \vdots \omega_p']$. Assume that $\frac{\partial^2 \mathbb{E}[h_1(\cdot)|Z,\nu]}{\partial\tau\partial\tau'} - \frac{\partial^2 C(\tau,\mathbf{X}_p,\omega_p,\nu)}{\partial\tau\partial\tau'} < 0$ so that the second-order condition is also satisfied (**assumption**). Then by the implicit function theorem $h_2(Z, \nu)$ is monotone in ν. Therefore, this example leads to a triangular model of the form considered in this paper

$$Y_1 = h_1(\tau, \mathbf{X}_p, u, \nu), \tag{5}$$

$$\tau_p = h_2(Z, \nu). \tag{6}$$

The approach adopted here for identification and estimation is based on control variables and has been explored by Chesher (2003). Define $Q_X(\tau \mid Z)$ to be the τ-conditional quantile of X given Z. Besides the restrictions mentioned above (continuity of endogeneous variables, model completeness, U and V are scalars, h_1 and h_2 are restricted to be strictly monotically varying in U and V respectively), the model includes the following restrictions:

Assumption 2.1. Identification in nonseparable models (ITM) (i) (Joint quantile independence) $Q_{U|Y_2,Z}(\tau_1|y_2, z) = Q_{U|V,Z}(\tau_1|v, z) = Q_{U|V}(\tau_1|v)$; (ii) (conditional τ-quantile restriction) $Q_{V|Z}(\tau_2|z) = 0$.

Assumption 2.2. Identification under index restrictions (IR) (i) X does not contain a constant (intercept: θ_0 is identifiable only up to a scale). To get identification, we need some location and scale normalizations: $\|\theta_0\| = 1$. (ii) $h_1(.)$ is differentiable and is not a constant function on the support of $X'\theta_0$. (iii) For the discrete components of x, varying the values of the discrete variables will not divide the support of $x'\theta_0$ into disjoint subsets.

In this version of the paper I add a further restriction on the second equation.

Assumption 2.3. Additivity and heteroskedasticity (AH). Let $\xi = \sigma(Z).V$, then

$$Y_2 = h_2(Z) + \xi.$$

Note that this assumption preserves nonseparability in V although being more restrictive than the general case. Assumption ITM(i) is similar to Assumption (2) in Lee (2007). The first equality in ITM(i) holds when v is the value of V that satisfies $v = y_2 - h_2(z)$. The second equality in ITM(i) assumes a τ_1-quantile independence of U on Z conditional on V. Under the previous assumptions ITM and IR and AH, taking the conditional quantiles of Y_1 on $V = v$, $W = w$ and $Z = z$, and of Y_2 on $Z = z$ yields

$$Q_{Y_1|V,W,Z}(\tau_1|v,w,z) = h_1(x'\theta, Q_{U|V,W}(\tau_1|v,w), v) \tag{7}$$

$$Q_{Y_2|Z}(\tau_2|z) = h_2(z) + Q_{\xi|Z}(\tau_2|z) \tag{8}$$

Set $v \equiv y_2 - h(z)$ and note that conditioning on $V = v$ and $Z = z$ is identical to conditioning on $Y_2 = y_2$ and $Z = z$, we can rewrite (7) and (8):

$$Q_{Y_1|X,Z}(\tau_1|x,z) = h_1(x'\theta, Q_{U|V,W}(\tau_1|y_2 - h(z), w), y_2 - h_2(z)) \tag{9}$$

$$Q_{Y_2|Z}(\tau_2|z) = h_2(z) \tag{10}$$

This system of equation suggests a control variable analogue estimator in which the control variable enters non-additively. Indeed, looking at units with common values of V allows one to identify the variations in Y_1 induced by variations in Y_2. As equation (10) suggests, an estimate $\hat{h}_2(z)$ can be obtained from nonparametric quantile regression of Y_2 on Z. This estimates will then be used in the control function. The description of the estimator is given in the next section.

3 Estimation using sieve minimum-distance

3.1 The sieve estimator

Building on identification results, this paper develops a sieve-based analog estimator (e.g Grenander (1981), Shen (1997), Chen (2007), Chen and Pouzo (2009)) and derive its asymptotic properties. For simplicity, $X'\theta$ is considered to be a scalar, although the treatment can be extended to multivariate settings. The support of all variables is allowed to be unbounded, i.e., to be the whole real line.

Define $D \equiv (y, x, z)$ for $y = (y_1, y_2) \in \mathcal{Y}$, $x \in \mathcal{X}$, and $z \in \mathcal{Z}$. Let $\alpha_0 = (\theta_0, h_0) \in \mathcal{A} \equiv \Theta \times \mathcal{H}$ denote the true parameters of interest, where Θ is a compact subset of $\mathcal{R}^{d_x}$ and $h_0 = (h_{01}, h_{02}) \in \mathcal{H}$ are real-valued measurable functions of the data. Let $V = Y_2 - h_2(Z)$. From equations (9) and (10), one can define the following two residual functions $\rho_1(D_i, \alpha) = 1[Y_{i1} \leq h_1(X_i'\theta, s(\tau_1, V)] - \tau_1$, where $s(\tau_1, V)$ simply makes explicit that h_1 will depend on V and a particular τ_1, and $\rho_2(D_i, \alpha_2) = 1[Y_{i2} \leq h_2(Z_i)] - \tau_2$. These two residual functions lead to a vector of two semiparametric conditional moment restrictions $m_1(D_1, \alpha) = E[\rho_1(D_1, \alpha)|D_1]$ and $m_2(D_i, \alpha_2) = E[\rho_2(D_2, \alpha_2)|D_2]$. These two moment conditions can be stacked into a vector $m(D, \alpha) = (m_1(D_1, \alpha_1), m_2(D_2, \alpha_2))'$. Based on these moment condition a stacked sieve minimum distance estimator for α_0 can be defined as

$$\hat{\alpha}_n = \arg \min_{\alpha \in \mathcal{A}_{k(n)}} \frac{1}{n} \sum_{i=1}^{n} \hat{\rho}(d_i, \alpha)' \hat{\sum}(d_i)^{-1} \hat{\rho}(d_i, \alpha) \quad (11)$$

$$= \arg \min_{\alpha \in \mathcal{A}_{k(n)}} \frac{1}{n} \sum_{i=1}^{n} \{\hat{\rho}_1(d_i, \alpha_1)' \hat{\sum}_1(d_i)^{-1} \hat{\rho}_1(d_i, \alpha_1) + \hat{\rho}_2(d_i, \alpha_2)' \hat{\sum}_2 \hat{\rho}_2(d_i, \alpha_2)\}, \quad (12)$$

with $\mathcal{A}_{k(n)} \equiv \Theta \times \mathcal{H}_{k(n)}$, and $\sum(D) = (\sum_1(D_1), \sum_2(D_2))'$, where $\sum_1(D_1) \equiv Var(\rho_1(D_1, \theta, h_1)|D_1)$ and $\sum_2(D_2) \equiv Var(\rho_2(D_2, h_2)|D_2)$ are positive-definite weighting matrices. Stack GMM/SMD has been considered in a general context by Newey (1984) or Ai and Chen (2007) for instance.

In practice, one would consider the following three-step procedure:

Step1. Obtain an initial consistent estimator:

$$\arg \min_{\alpha \in \mathcal{A}_{k(n)}} \frac{1}{n} \sum_{i=1}^{n} \{\hat{\rho}_1(d_i, \alpha)' \hat{\rho}_1(d_i, \alpha) + \hat{\rho}_2(d_i, h_2)' \hat{\rho}_2(d_i, h_2)\}.$$

Step2. Obtain a consistent estimator $\hat{\sum}(D)$ of the optimal weighting matrix$\sum_0(D)$ using $\hat{\alpha}_n$ and any nonparametric regression procedures (such as kernel, nearest-neighbor or linear sieves).

Step3. Obtain the optimally weighted estimator $\tilde{\alpha}_n = (\tilde{\theta}_n, \tilde{h}_n(\hat{\nu}))$ by repeating the previous sieve-based estimation of step 1 in order to solve

$$\min_{\alpha \in \mathcal{A}_n} E[\rho(D, \alpha)' [\hat{\sum}_0(D)]^{-1} \rho(D, \theta, \alpha)].$$

3.2 Consistency and convergence rates

3.2.1 Consistency

The choice of sieve bases is motivated by the kind of smoothness one wants to impose and the support of the function to be approximated. The asymptotic analysis in this paper relies on standard smoothness restrictions on the unknown functions h_1 and h_2. Suppose that $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \ldots \times \mathcal{X}_d$ is the Cartesian product of compact intervals $\mathcal{X}_1, \ldots, \mathcal{X}_d$. Let $0 < \gamma \leq 1$. A real-valued function h on $\mathcal{X}$ is said to satisfy a Hölder condition with exponent γ if there is a positive number c such that $|h(x) - h(y)| \leq c\,|x - y|_e^\gamma$ for all $x, y \in \mathcal{X}$; here $|x|_e = (\sum_{l=1}^d x_l^2)^{1/2}$ is the Euclidean norm of $x = (x_1, \ldots, x_d) \in \mathcal{X}$. Given a d-tuple $\alpha = (\alpha_1 + \ldots + \alpha_d)$ of nonnegative integers, set $[\alpha] = \alpha_1 + \ldots + \alpha_d$ and let D^α denote the differential operator defined by

$$D^\alpha = \frac{\partial^{[\alpha]}}{\partial x_1^{\alpha_1} \ldots \partial x_2^{\alpha_d}}. \tag{13}$$

Let m be a nonnegative integer and set $p = m + \gamma$. A real-valued function h on $\mathcal{X}$ by $\Lambda_c^p(\mathcal{X})$ (called a Hölder class), and the space of all m-times continuously differentiable real-valued functions on $\mathcal{X}$ by $\mathcal{C}^m(\mathcal{X})$. Define a Hölder ball with smoothness $p = m + \gamma$ as

$$\Lambda_{\mathcal{C}}^p(\mathcal{X}) = \left\{ h \in \mathcal{C}^m(\mathcal{X}) : \sup_{[\alpha] \leq m} \sup_{x \in \mathcal{X}} |D^\alpha h(x)| \leq c, \sup_{[\alpha] = m} \sup_{x,y \in \mathcal{X}, x \neq y} \left| \frac{D^\alpha h(x) - D^\alpha h(y)}{|x - y|_e^\gamma} \right| \leq c \right\}. \tag{14}$$

Functions belonging to a Hölder ball are well approximated by linear sieves. Linear sieves form a large class of sieves useful for sieve estimation. Tensor-product construction is a standard way to generate linear sieves of multivariate functions from linear sieves of univariate functions. For instance, for a bivariate tensor product spaces: let $\mathcal{U}_l, l = 1, 2$, be compact sets in Euclidean spaces and $\mathcal{U} = \mathcal{U}_1 \times \mathcal{U}_2$ be their Cartesian product. Let $\mathbb{G}_l$ be a linear space of functions on $\mathcal{U}_l$ for $l = 1, 2$, each of which can be any of the sieve spaces described below, among others. The tensor product, $\mathbb{G}$, of $\mathbb{G}_1, \mathbb{G}_2$ is defined as the space of functions on $\mathcal{U}$ spanned by the functions $g_1(x_1) \times g_2(x_2)$, where $g_l \in \mathbb{G}_l$ for $l = 1, 2$. To facilitate the treatment of functions on noncompact domain, the technique suggested in Chen, Hong and Tamer (2005) is applied introducing a weighting function as follows:

$$\|\alpha\|_s = \|\theta\|_E + \|h\|_{\infty,\omega} \tag{15}$$

where the weighted Holder norm is defined as $\|g\|_{\infty,\omega} \equiv sup_\xi |g(\xi)\omega(\xi)|$, where $\omega(\xi) = (1 + \|\xi\|_E^2)^{-\varsigma/2}$, $\varsigma > p > 0$.

Assumption 3.1. (i) The data $(Y_i, X_i, Z_i) : i = 1, 2, ..., n$ are i.i.d.; (ii) $f_{Y_1|Y_2,X,V}(y_1 \mid y_2, x, v)$ is continuous in (y_1, y_2, x, v), and $\sup_{y_1} f_{Y_1|Y_2,X,V}(y_1) \leq const. < \infty$ for almost all Y_2, X, V. (iv) $f_{Y_2|Z}(y_2 \mid z)$ is continuous in (y_2, z), and $\sup_{y_2} f_{Y_2|Z}(y_2) \leq const. < \infty$ for almost all Z; (v) $E[F_{Y_1|X,V}(h_1(X'\theta, V)) \mid X = x, V = \nu] \in \Lambda_1^{p_1}(\mathcal{D}^2)$, $E[F_{Y_2|X}(h_2(Z)) \mid Z = z] \in \Lambda_1^{p_2}(\mathcal{D})$.

Assumption 3.2. (i) $h \in \mathcal{H}$ a Holder space under the weighted metric $\|h\|_{\infty,\omega}$, $\alpha_0 \in \mathcal{A} \equiv \Theta \times \mathcal{H}$; (ii) $\mathcal{A}_n \equiv \Theta \times \mathcal{H}_n, n \geq 1$, are the sieve spaces satisfying $\mathcal{H}_n \subseteq \mathcal{H}_{n+1} \subseteq \mathcal{H}$, and for any $\alpha \in \mathcal{A}$ there exists $\pi_n\alpha \in \mathcal{A}_n$ such that $\|\pi_n\alpha_0 - \alpha_0\|_s = o(1)$ (as $k_n \to \infty$) and $k_n/n \to 0$.

Assumption 3.3. $\hat{\sum}(D) = \sum(D) + o_p(1)$ uniformly over $D \in \mathcal{D}$.

Assumption 3.4. (i) There is a function $b(.)$ such that $b(\delta) \to 0$ as $\delta \to 0$ and

$E[\sup_{\|\alpha-\alpha'\|<\delta} \|m(D_i, \alpha) - m(D_i, \alpha')\|^2] \leq b(\delta)$ for all small positive value δ;

(ii) $E[\sup_{\alpha\in\mathcal{A}_n} |\rho(D, \alpha)'\rho(D, \alpha)|^2] < \infty$.

Discussion of assumptions. One needs to ensure that the control variable estimate $\hat{\nu}$ is close to its true value ν_0 for n large enough. This will be verified by standard nonparametric estimators. One needs to show that the approximation error arising from the first step estimation of the control variable is $o(1)$, which is a necessary condition to show consistency (see Andrews (1994)). The compactness assumption 3.2(i) can be replaced by a more general assumption allowing for noncompact parameter space and a penalty assumption. Indeed, Assumption 3.2(i) can be replaced by the following high-level condition: Assumption 3.2. (i') $h_0 \in \mathcal{H}$ where $\mathcal{H}$ is a separable Banach space under the metric $\|h\|_c$, $\alpha_0 \in \mathcal{A} \equiv \Theta \times \mathcal{H}$. This would allow for functions belonging to a Sobolev space. Since Sobolev spaces are usually not compact under standard norms, one would have to penalize the estimator (see Shen (1997), Chen and Pouzo (2009)). Assumption 3.2(ii) means that the true parameter α_0 is in the approximating sieve space $\mathcal{A}_n$ and it guarantees that the sieve space approximates the true function space well. This is similar to the denseness condition in Gallant and Nychka (1987) or Ai and Chen (2003). Verification of this assumption depends on the particular sieve used to approximate h. Assumption 3.4 is an enveloppe condition which together with Holder continuity of the objective function serves to verify uniform convergence.

Lemma 3.1. *Let $\hat{\alpha}$ be the sieve extremum estimator (11). Let Assumptions 2.1-2.3 and 3.1-3.4 hold. Then $\|\hat{\alpha}_n - \alpha_0\|_s = o(1)$.*

Proof: see appendix.

3.2.2 Convergence rates

Let $\{\delta_n\}$ denote a positive sequence that decreases to zero as $n \to \infty$. $\hat{\alpha}_n$ is said to be consistent for α_0 at a rate (strictly) faster than δ_n if $\|\hat{\alpha}_n - \alpha_0\| / \delta_n \to 0$ in probability, denoted as $\|\hat{\alpha}_n - \alpha_0\| = o_p(\delta_n)$. The goal is to show that $\hat{\alpha}_n$ is consistent at a rate faster than $n^{-1/4}$ as this will be useful to show asymptotic normality. Together with assumptions for consistency, (i) the sieve approximation error rate, $\|\alpha_0 - \pi\alpha_0\|$, have to approach zero suitably fast, and (ii) the sieve space, Θ_n, must not be too complex. The following set of assumptions sufficient in order to obtain rates of convergence for α.

Assumption 3.5. There is $C_1 > 0$ such that for all small $\varepsilon > 0$,

$$\sup_{\{\alpha \in \mathcal{A}_n : \|\alpha_0 - \alpha\| \leq \delta\}} E\left(m(\theta, D_t) - m(\theta_0, D_t)\right)^2 \leq C_1 \varepsilon^2. \tag{16}$$

Assumption 3.6. For any $\delta > 0$, there exists a constant $s \in (0,2)$ such that

$$\sup_{\{\alpha \in \mathcal{A}_n : \|\alpha_0 - \alpha\| \leq \delta\}} |l(\theta, D_t) - l(\theta_0, D_t)| \leq \delta^S U(D_t), \tag{17}$$

with $E([U(D_t)]^\gamma) \leq C_2$ for some $\gamma \geq 2$.

Theorem 3.1. *Let $\hat{\alpha}$ be the sieve extremum estimator (13). Let Assumptions IR, ITM, 3.1-3.6 hold. Then $\|\hat{\alpha}_n - \alpha_0\|_s = o(n^{-1/4})$.*

Proof: This assumptions still need to be verified.

Remark 3.1. (i) Imposing index restrictions in the first-stage might be necessary to insure fast enough convergence rates. Take $k_{2n} = O_p\left(n^{\frac{1}{2p+d_z}}\right)$. With $p = 2$, $k_{2n} = n^{\frac{1}{4+d_z}}$, and hence $\|\hat{\nu} - \nu_0\| = O_p\left(n^{-\frac{2}{4+d_z}}\right)$. Thus, **if** $d_z \leq 2$**,** $\|\hat{\nu} - \nu_0\| = o_p\left(n^{-1/4}\right)$**,** if $d_z > 2$, impose index restrictions to get $\|\hat{\nu} - \nu_0\| = o_p\left(n^{-1/4}\right)$. (ii) An alternative specification would be to consider the outcome equation $Y_1 = h_1(X'\theta, \varepsilon)$ where $X'\theta$ is a vector of small dimension M, and X_m $(m = 1, ..., M)$ is a subvector of X. The approach presented in this paper applies to this more general framework. The constraint on the dimensionality of $X'\theta$ will then come from the convergence rates.

3.3 Asymptotic normality and semiparametric efficiency [TO BE ADDED]

Asymptotic normality and efficiency will be shown by verifying theorem 3.2 in Chen and Pouzo (2009).

4 Computation and Simulation

4.1 Sieve extremum estimation

4.1.1 Computational efficiency.

Quantile regression estimators have the remarkable feature that they can be rewritten as a linear programming problem. From a computational point of view, this is extremely convenient and efficient procedures have been developed by Koenker (2005) and others. Unfortunately the GMM framework for quantile regression does not allow one to use these procedures. In order to take advantage of these computational features, I suggest to first consider a consistent estimator which guarantees computational efficiency, and then use the resulting estimates as starting values for the asymptotically efficient estimator. Estimation of quantile regression models with single-index restrictions has previously been considered in the literature by Chaudhuri et al. (1997) and Khan (2001). Since the model considered in this paper allows for discrete regressors and the link function is not necessarily monotonous in the index, the estimation procedure developed here cannot rely on the average derivative estimator introduced by Chaudhuri et al. (1997), nor can it be based on the rank estimator developed by Khan (2001). Ichimura and Lee (2006) and Wu et al. (2009) have proposed single-index quantile regression estimators based on local linear polynomials only. This paper extends their approach to quantile regression models with multivariate link function and to general sieve estimators of the link function.

First, the unknown functions $h \in \mathcal{H}$ are approximated by $h_n \in \mathcal{H}_n$, where again $\mathcal{H}_n$ is some sieve space, that is, some finite-dimensional approximation spaces (e.g Fourier series, splines, power series,...) which becomes dense in $\mathcal{H}$ as sample size $n \to \infty$. Then the vector of finite-dimensional parameters θ and the unknown sieve coefficients are estimated by a profiled sieve least absolute deviation procedure. In order to estimate the finite-dimensional parameter vector θ by semiparametric profiled estimation, using the sample criterion function $n^{-1}\sum_{i=1}^{n} m(D_i, \theta, \hat{h}_1(\hat{\nu}))$, it is much easier if $\hat{h}_1(\hat{\nu})$ were a sieve estimate obtained by $\max_{h \in \mathcal{H}_n} \hat{Q}_n(\theta, h_1(\hat{\nu})) = n^{-1}\sum_{i=1}^{n} m(\theta, h_1(\hat{\nu}), D_i)$ for any arbitrarily fixed θ, rather than a kernel estimate. Besides, in order to impose shape restrictions sieves, i.e splines, are much more appropriate (see note 3 pp.2 in Chen and Pouzo (2009a)). I follow this recommendation and propose a profiled semiparametric sieve extremum estimation procedure which is consistent with the recent literature on sieve estimation (for similar remarks and a comprehensive review, see Chen (2006)). Hence, consider the following asymptotically inefficient but computationally convenient minimization problem:

$$\arg\min_{h_1, \theta} \sum_{i=1}^{n} \rho_{\tau_1} \left\{ y_i - h_1(x_i'\theta, \hat{\nu}_i) \right\} \tag{18}$$

Approximate $h_1(\cdot)$ by a sieve tensor product: $h(\cdot) = \sum_{j=1}^{k_1}\sum_{l=1}^{k_2} \beta_{ijl}\psi_j(x'\hat{\theta}^{(k)})\psi_l(\hat{\nu})$, where ψ is a univariate basis; note that series estimators offer an alternative to the tensor-product approach suggested here (see for instance Newey (1994)).

Step1. Obtain a consistent estimator of the control variable ν_0.

Step1.1 Estimate h_2 as the solution to

$$\arg\min_{h_2} \sum_{i=1}^{n} \rho_{\tau_2} \left\{ y_{2i} - h_2(z_i) \right\} \tag{19}$$

Step1.2 Get an estimate of the residuals $\hat{\nu}_i = y_{2i} - \hat{y}_{2i}$.

Step2 *Semiparametric estimation.*

Step2.0 Get initial values for $\theta^{(0)}$.

Step2.1 Given $\hat{\theta}^{(k)}$, obtain $\hat{\beta}^{(k)}$ by solving

$$\min_{\beta} \sum_{i=1}^{n} \rho_{\tau_1} \left\{ y_i - \sum_{j=1}^{k_{1n}} \sum_{l=1}^{k_{2n}} \beta_{ijl} \psi_j(x'\hat{\theta}^{(k)}) \psi_l(\hat{\nu}) \right\}, \tag{20}$$

Step2.2 Given $\hat{\beta}^{(k)}$, obtain $\hat{\theta}^{(k+1)}$ by solving the following optimization problem:

$$\min_{\theta} \sum_{i=1}^{n} \rho_{\tau_1} \left\{ y_i - \sum_{j=1}^{k_{1n}} \sum_{l=1}^{k_{2n}} \hat{\beta}_{ijl}^{(k)} \psi_j(x'\theta) \psi_l(\hat{\nu}) \right\}, \tag{21}$$

Step 3: *Repeat Steps 1 and 2 until convergence to* $\hat{\theta}*$.

One can then estimate $\hat{h}$ given $\hat{\theta}^{*}$. Note that the optimization problem reduces to successive linear quantile regressions, which is computationally extremely fast.

Remark 4.1. The objective function might not be well-behaved, need to try various initial values (see Newey and McFadden (1994) and Bates and Watts (1988)).

Remark 4.2. The estimation strategy can be extended to a model with m endogeneous variables by adding a linear index of control variables. The estimator would then become a double-index model.

4.2 Experimental simulations

In this subsection I show the results of two Monte-Carlo simulations in order to explore the finite sample properties of the computationally efficient estimator. Note that these simulations do not include the asymptotically efficient estimator and hence provide a lower bound to the expected performance of the asymptotically efficient estimator. The nonparametric estimator used is quantile smoothing splines as in Koenker et al. (1994).

4.2.1 Single-index quantile regression

The first Monte-Carlo simulation shows that the procedure introduced in this paper produces results comparable to the finite sample properties of the single-index quantile regression estimator based on local linear polynomials suggested by Wu, Yu and Yu (2009). I reproduce one of their experimental setting. The DGP is:

$$y = sin\left(\frac{\pi(u-A)}{C-A}\right) + 0.1Z, \tag{22}$$

where $u = x^T\theta$, $x = (x_1, x_2, x_3)^T$, and theoretical values are given by $\theta_0 = \frac{1}{\sqrt{3}}(1,1,1)^T$, $A = \frac{\sqrt{3}}{2} - \frac{1.645}{\sqrt{12}}$; $\frac{\sqrt{3}}{2} + \frac{1.645}{\sqrt{12}}$; $x_i \hookrightarrow Unif(0,1)$, $i = 1,2,3$, $Z \hookrightarrow N(0,1)$, x_i's and Z are mutually independent. In the table t denotes quantiles. Results are shown in table 1.

	Theoretical Value	Coeffcient	Bias	Std. Error	MSE
t = 0.1					
theta 1	0.5774	0.5767	-0.0006	0.0243	0.0006
theta 2	0.5774	0.5737	-0.0036	0.0239	0.0006
theta 3	0.5774	0.5801	0.0028	0.0227	0.0005
t = 0.3					
theta 1	0.5774	0.5727	-0.0047	0.0181	0.0003
theta 2	0.5774	0.5756	-0.0017	0.0203	0.0004
theta 3	0.5774	0.5828	0.0054	0.0183	0.0004
t = 0.5					
theta 1	0.5774	0.5758	-0.0016	0.0172	0.0003
theta 2	0.5774	0.5740	-0.0034	0.0173	0.0003
theta 3	0.5774	0.5815	0.0041	0.0184	0.0004
t = 0.7					
theta 1	0.5774	0.5749	-0.0024	0.0194	0.0004
theta 2	0.5774	0.5750	-0.0024	0.0189	0.0004
theta 3	0.5774	0.5812	0.0038	0.0174	0.0003
t = 0.9					
theta 1	0.5774	0.5728	-0.0045	0.0211	0.0005
theta 2	0.5774	0.5735	-0.0039	0.0238	0.0006
theta 3	0.5774	0.5843	0.0070	0.0230	0.0006

Table 1: Comparison with Wu, Yu and Yu (2009) - n = 100, r = 100

4.2.2 Additive triangular model with index restrictions

This Monte-Carlo simulation considers the estimation of a triangular model with index restrictions. Y_2 is the endogenous variable. The DGP is:

$$Y_1 = (x^T\theta_0)^2 + U$$

$$Y_2 = 1 + 2.x_1 + 3.z + V$$

$$U = \lambda.cos(V)^2 + \epsilon$$

where u=$x^T\theta$, $x = (x_1, x_2, y_2)^T$,θ_0=(1,4.5,-3)^T; λ=3; $x_1 \hookrightarrow$N(0,3^2), $x_3 \hookrightarrow$N(2,4^2), Z$\hookrightarrow$N(3,5^2), $\epsilon \hookrightarrow$Unif(0,1), V$\hookrightarrow$Unif(0,1), x_is and Z are mutually independent. Results are for $\tau_1 = .5$. Results are shown in table 2.

	Theoretical Value	Coeffcient	Bias	Std. Error	MSE
t = 0.5					
theta 1	1.0000	1.0000	0.0000	0.0000	0.0000
theta 2	4.5000	4.5243	0.0243	0.3307	0.1100
theta 3	-3.0000	-3.0095	-0.0095	0.2360	0.0558

Table 2: Simulation for A Separable Triangular Model - n = 300, r = 100

4.2.3 Simulation to be added.

Results for a nonseparable triangular model will be added. Nonetheless, as the two previous simulations already show, the sieve profiled estimation procedure developed in this paper should be easily adapted for the nonseparable model.

5 Empirical application: Profits and property rights.

The role of institutions is currently a major center of interest in development economics. After a first generation of cross-country regressions of economic growth on various measures of institutional quality (Hall and Jones (1999), Acemoglu et al. (2001), Rodrik et al. (2004) among others), the economic profession seems to have reached the consensus that institutions in general, and property rights in particular, matter for growth and development. Following these macro studies the question of the mechanisms through which property rights affect economic performance has come under the scrutiny of more recent work (Banerjee and Iyer (2005), Pande and Udry (2006)). Several mechanisms can be underlying the connection between particular institutions and investment or productivity.

Goldstein and Udry (2008) have recently investigated one particular mechanism that could potentially link property rights over land to agricultural investment, and in turn, to farmers' profits. Investment incentives depend on expectations of rights over the returns to that investment and hence on the nature of property rights. Goldstein and Udry (2008) examine the connection from a set of complex and explicitly negotiable property rights over land to agricultural investment, and in turn, to agricultural productivity. There exists several mechanisms through which property rights over land might influence investment in agriculture.

Consider an environment in which fertilizer is expensive, land is relatively abundant, and crop returns are low. Hence, fallowing is the primary mechanism by which farmers increase their yields: it is the most important investment in land productivity. Farmers who lack political power are not confident of maintaining their land rights over a long fallow which leads to shorter than optimal fallow durations. This framework can be applied to the model considered in this paper as shown in section 2.

The survey consists of a 2-year rural survey in Ghana. There were four village clusters. Within each village cluster, 60 married couples selected. Each head and spouse was intervied 15 times. Multiple plots and each plot can be identified with a particular individual who controls it. Several questions about plots (plot-level inputs, harvests, sales, credit + plot rights and history questionnaires). Additional information about soil fertility. Data on education and individual wealth.

The econometric strategy followed in the present paper is to examine the effect of an individual's position in local political and social hierarchies on his or her fallowing choices on a plot, conditional on plot characteristics. In turn, productivity effects of (endogenous) fallowing choices are estimated using the individual's political and social position as instruments for the fallowing choice. Equations (5) and (6) suggests the following semiparametric econometric model specification

$$Y_1 = h_1(\tau_p, \mathbf{X}_p'\theta_1, u, v) \tag{23}$$

$$\tau_p = h_2(Z'\theta_2, \nu) = h_2(Z'\theta_2) + \lambda(Z, v). \tag{24}$$

Note that in this particular case one would want to impose index restrictions in the second equation as well. Endogeneity arises because τ_p is not independent of the error term since it may respond to the same unobserved attributes of the plot that influence profits. An appropriate instrument is ω_p, which is unobserved. Instead, a set of variables Z that represent the cultivator's position in local social and political hierarchies is used as instruments. Estimation of structural feature will yield local quantile effects of fallow duration on profits. This more general empirical approach attempts at providing a more thorough understanding of the effect of property rights on land profits. This endeavour should render possible to investigate and bring to light two aspects of the relationship between property rights and productivity: heterogeneity and nonlinearities.

6 Conclusion

This paper makes an attempt at extending currently existing parametric or semiparametric estimators for quantile regression models in the presence of endogeneity. The first contribution is to develop a sieve based estimator that generalizes considerably previous estimators. A second contribution is to discuss implementation issues for efficient estimation of semi-nonparametric quantile regression estimators. Last, it is hoped that the application will lead to new results on the understanding of the interaction between institutions and economic development.

7 References

Acemoglu D, Johnson S, Robinson J., 2001, The colonial origins of comparative development: An empirical investigation. *American Economic Review* 91:1369–1401.

Ai, C. and X. Chen, 2003, Efficient estimation of models with conditional moment restrictions containing unknown functions, *Econometrica*, 71, 1795-1843.

Ai, C. and X. Chen, 2007, Estimation of possibly misspecified semiparametric conditional moment restriction models with different conditioning variables, *Journal of Econometrics*, 141, 5-43.

Ai, C. and X. Chen, 2009, Semiparametric efficiency bound for models of sequential moment restrictions containing unknown functions, cemmap working paper CWP28/09.

Banerjee, A. and L. Iyer, 2005, History, institutions and economic performance: The legacy of colonial land tenure systems in india. *American Economic Review*.

Bates, D. M., and D. G. Watts, *Nonlinear Regression Analysis and Its Applications*, John Wiley, New York, 1988.

Blundell, R. and J. Powell, 2003, Endogeneity in Nonparametric and Semiparametric Regression Models, In Dewatripont, Mathias; Hansen, Lars Peter and Turnovsky, Stephen J. (eds.). *Advances in Economics and Econometrics*, Vol. II, Econometric Society Monograph Series, Eighth World Congress, Cambridge, UK: Cambridge University Press.

Blundell, R. and J. Powell, 2007, Censored Regression Quantiles with Endogenous Regressors, *Journal of Econometrics*.

Chaudhuri, P., Doksum, K., Samarov, A., 1997, On average derivative quantile regression, *Annals of Statistics* 25, 715744.

Chen, X., 2007, Large-sample sieve estimation of semi-nonparametric models, *Handbook of Econometrics* 6B.

Chen, X. and D.Pouzo, 2008, Estimation of nonparametric conditional moment models with possibly nonsmooth moments, Working Paper.

Chen, X. and D.Pouzo, 2009, Efficient Estimation of semiparametric conditional moment models with possibly nonsmooth moments, *Journal of Econometrics*.

Chen, X., H. Hong and E. Tamer, 2005, Measurement error models with auxiliary data, *Review* *of Economic Studies*, 72, 343-366.

Chernozhukov, V., Hansen, C., 2005, An IV model of quantile treatment effects. *Econometrica* 73 (1), 245–262.

Chernozhukov, V., Hansen, C., 2006, Instrumental Quantile Regression Inference for Structural and Treatment Effect Models, *Journal of Econometrics*, 132, 491–525.

Chernozhukov, V., Imbens, G., Newey, W., 2007, Instrumental variable estimation of nonseparable models. *Journal of Econometrics* 139 (1), 4–14.

Chesher, A., 2001, Exogenous Impact and Conditional Quantile Functions. *Centre for Microdata Methods and Practice.* Working Paper 01/01.

Chesher, A., 2003, Identification in nonseparable models. *Econometrica* 71 (5).

Chesher, A., 2009, Single equation endogenous binary response models, Cemmap Working Paper.

Gallant, A. R., and Nychka, 1987, Semi-nonparametric maximum likelihood estimation, *Econometrica*, 55, 363-390.

Goldstein, M. and C. Udry, 2005, The profits of power: Land rights and agricultural investment in ghana. *Journal of Political Economy.*

Grenander, U., 1981, *Abstract Inference*, New York: Wiley Series.

Hall RE., Jones C., 1999, Why do some countries produce so much more output per worker than others?, *Quarterly Journal of Economics,* CXIV:83–116.

Heckman, J., Matzkin, R., and Nesheim, L., 2004, Simulation and Estimation of Hedonic Models. *Frontiers of Applied General Equilibrium Modeling: Essays in Honour of Herbert Scarf.* North Holland.

Ichimura, H. and S. Lee ,2006, Characterization of the Asymptotic Distribution of Semiparametric M-Estimators, Cemmap Working Paper.

Imbens, G. and W. Newey, 2009, I dentification and estimation of triangular simultaneous equations models without additivity, *Econometrica* 77 (5), 1481-1512.

Jun, S.-J., 2009, Local structural quantile effects in a model with a nonseparable control variable, *Journal of Econometrics,* 151(1), 82-97.

Khan, S., 2001, Two-stage rank estimation of quantile index models. *Journal of Econometrics*

100, 319355.

Koenker, R., 2005, *Quantile Regression*, Econometric Society Monographs No. 38, Cambridge University Press.

Koenker, R., P. Ng and S. Portnoy, 1994, Quantile Smoothing Splines, *Biometrika*, 81, 673-680.

Lee, S., 2007, Endogeneity in Quantile Regression Models: A Control Function Approach, *Journal of Econometrics* 141, 1131–1158.

Ma, L. and Roger Koenker, 2006, Quantile regression methods for recursive structural equation models. *Journal of Econometrics*, 134, 471–506.

Matzkin, R., 2007, Nonparametric Identification. In *Handbook of Econometrics*, Vol 6B.

Newey, W. and D. McFadden, 1994, Large Sample Estimation and Hypothesis Testing, in: *Handbook of Econometrics*, Vol.4, Ch.36, pp. 2113–2148.

Newey, W., and J. Powell, 2003, Instrumental variable estimation of nonparametric models, *Econometrica*, 71(5), 1565-1578.

Pande, R. and C. Udry, 2005, Institutions and development: A view from below. mimeo.

Rodrik, D, Subramanian, A. and Francesco Trebbi, 2004, Institutions Rule: The Primacy of Institutions over Geography and Integration in Economic Development. *Journal of Economic Growth* 9(2):131–65.

Shen, X., 1997, On the method of sieves and penalization, *Annals of Statistics*, 25, 2555-2591.

Wu, Yu and Yu, 2009, Single-Index Quantile Regression, Working Paper.

A On smoothness

One issue arising in the estimation and the derivation of asymptotic properties of semi-nonparametric models is the decision over of the smoothness of the functions h_0. This is a critical step because it has implications on all aspects of semiparametric estimation. There are three main aspects that should inform the choice of smoothness for h_0:

1. ***(Economic) Theory***. The smoothness of h_0 should ideally follow from economic theory or, more generally, from smoothness properties intrinsic to the object it represents (whether a density or a utility function).

2. ***Implementation***. The type and the smoothness of the estimator will be made according to the smoothness assumptions on h_0. For instance, it is known that if $\Theta = \Lambda^p([0,1]^d)$ is the Holder space of p-smooth functions with $p > d/2$, then functions in this space are well aproximated by linear sieves and one can for example use series estimator. It is not the case however if $\Theta = W_1^1([0,1]^d)$: one should rather take the sieve space, Θ_n, to be the nonlinear Gaussian radial basis ANN. Therefore, although ease of computation should not be the only concern when one decides how smooth h_0 should be, in practice it will be taken into account (see Chen (2007) for more on this).

3. ***Large-sample properties of the estimator***.

   (a) The smoothness assumption has major implications in terms of *convergence rates.* This plays an important role in semiparametric models in particular, as $\sqrt{n}$-consistent estimation of the parametric part of the model often relies on the assumption that $\left\|\hat{h} - h\right\| = o(n^{-\frac{1}{4}})$.

   (b) Compactness of the parameter space $\mathcal{H}$, or at least of the sieve space $\mathcal{H}_n$, is a key requirement for *consistency.* A natural consequence of this fact is that one should make sure that she/he will be able to verify the compactness assumption for the parameter space considered.

B Consistency and convergence rates

B.1 Consistency

[PRELIMINARY]

Consistency is obtained by application of theorem 3.1 in Chen (2007).

Condition 3.1 (Identification) The identification condition is verified by assumptions ITM and IR which ensures that for $j = 1, 2$, $E[\rho_j(D, \alpha)] = 0$ implies $\theta = \theta_0$ and $h = h_0$. $(\theta_0, h_0) \in int(\Theta) \times \mathcal{H}$ satisfies the model. In particular, condition 3.1 (ii) is satisfied since the semiparametric model is well-posed.

Condition 3.2 (Sieve spaces) Given the choice of sieve spaces $\mathcal{H}_n$ and the definition of the norm $\|\cdot\|_{\infty,\omega}$, $\sup_{h\in\mathcal{H}} \|\alpha - \Pi_n\alpha\| = o(1)$ (see Chen, Hansen and Scheinkman (1997) or Ai and Chen (2003)).

Condition 3.3 (Continuity) Condition 3.3(ii) is implied by conditions 3.2 and 3.3(ii)'. The proof of condition 3.3(ii)' is devided into two parts: (i) $E[\rho_1(D, \alpha_1)' \sum_1(D)\rho_1(D, \alpha_1)] = o(1)$, and (ii) $E[\rho_2(D, \alpha_2)'\rho_2(D, \alpha_2)] = o(1)$.

- (i) To verify condition 3.3(ii)', note that for any $h(.) \in \mathcal{H}$,

$$
\begin{aligned}
|m(D,\alpha) - m(D,\alpha_0)| &= |E[\rho(D,\alpha) - \rho(D,\alpha_0)]| \\
&= \left|E[F_{Y_1|X,V}(h(X'\theta, V)) - F_{Y_1|X,V}(h_0(X'\theta_0, V))]\right| \\
&= \left|E[f_{Y_1|X,V}(\bar{h}(X'\bar{\theta}, V))[h(X'\theta, V) - h_0(X'\theta_0, V)]]\right| \\
&\leq E[f_{Y_1|X,V}(\bar{h}(X'\bar{\theta}, V))] \times \sup_{x,v} \left|h(X'\theta, V) - h_0(X'\theta_0, V)\right|
\end{aligned}
$$

where $\bar{h}(X'\bar{\theta}, V)$ is in between $h(X'\theta, V)$ and $h_0(X'\theta_0, V)$. Thus assumption on smoothness of f and $\|\pi_n h_0 - h_0\|_\infty = o(1)$ imply

$$
E[|m(D,\alpha)|] \leq E[f_{Y_1|X,V}(\bar{h}(X'\bar{\theta}, V)) \mid Z] \times \|\pi_n\alpha_0 - \alpha_0\|_\infty = o(1). \tag{25}
$$

- (ii) $E[\rho_2(D,\alpha_2)'\rho_2(D,\alpha_2)] = o(1)$ follows from similar steps.

Hence (i) and (ii) verify condition 3.3(ii)', and hence imply condition 3.3(ii)

Condition3.4 (Compact sieve spaces) Note that $\mathcal{A} = \Theta \times \mathcal{H}$, with $\mathcal{H} = \Lambda^p(\mathcal{R}^d)$, is compact under the norm $\|\alpha\|_s = \|\theta\|_E + \|h\|_{\infty,\omega}$, and $\|h\|_{\infty,\omega} \equiv sup_\xi |h(\xi)\omega(\xi)|$, where $\omega(\xi) = (1+\|\xi\|_E^2)^{-\varsigma/2}$, $\varsigma > p > 0$. Hence condition 3.4 is verified.

Condition3.5 (Uniform convergence) Uniform convergence over sieves follows from continuity, from simple convergence of Q_n and from Condition 3.4 of Chen (2006).

B.2 Convergence rates

[TO BE ADDED]