A Unified Theory for Functional Coefficients Models

Chunrong Ai and Yong Zhou

April 2, 2010

Abstract

This paper considers a sieve estimation of a general functional coefficients models which allow functional coefficients to depend on different set of covariates and the criterion function to be continuous but not ncessarily continuously differentiable. Under some sufficient conditions, we show that the sieve estimator is asymptotically normally distributed. Our results add to the existing literature by providing the asymptotic distribution for the functional coefficient estimator for the case of nonlinear and nonsmooth criterion function.

1 Introduction

Parametric regression is often the model of choice for analyzing data. A common assumption in the parametric regression literature is that the parameter of interest is finite dimensional and fixed, with the true value minimizing a population criterion function. An example is the linear regression

$$y = x'\alpha_0 + u, E\{u|x\} = 0, \tag{1.1}$$

where the parameter of interest α is finite dimensional and fixed with the true value α_0 maximizing the population criterion function $E\{l(z,\alpha)\} = E\{-(y - x'\alpha)^2\}$, where $z = (y, x)$ denotes the data. A second example is the quantile regression

$$y = x'\alpha_0 + \varepsilon, \Pr(\varepsilon < 0|x\} = 0.5, \tag{1.2}$$

where the parameter of interest α is finite dimensional and fixed, with the true value α_0 maximizing the population criterion function $E\{l(z,\alpha)\} = E\{-|y - x'\alpha|\}$. A third example is the popular binary regression model

$$y = 1\{x'\alpha_0 + u\}, u \sim N(0, 1), \tag{1.3}$$

where the parameter of interest α is finite dimensional and fixed, with the true value α_0 maximizing the population criterion function

$$E\{l(z,\alpha)\} = E\{y \ln(\Phi(x'\alpha)) + (1 - y) \ln(1 - \Phi(x'\alpha))\}.$$

In all of these and many other regressions, the parameter of interest α is interpreted as the marginal effect of the regressor x on the index $x'\alpha$, holding all other factors constant. Thus, the assumption of fixed parameter α implies that the marginal effect is independent of all other factors. This is quite a strong assumption. In many applications, the marginal effect may depend on some or all of those "other factors". Recognizing this limitation of the parametric regression, Hastie and Tibshirani (1991), Cai, Fan and Yao (2000), and Cai, Fan and Li (2000) and many others generalize the parametric regression by allowing the whole or part of the parameter to be unknown functions of some underlying covariates. They call this generalization as the functional coefficients model. It is easy to see that this generalizaion includes the popular partially linear regression model as a special case.

The generalization, though useful, presents a challenge of estimating unknown functions from the finite data points. Perhaps because of this challenge, the generalization has become increasingly popular among theorists. However, the existing literature focuses almost exclusively on the linear regressions (e.g., ...), with few exceptions focusing on the nonlinear regressions (e.g., ...) and the linear quantile regressions (e.g.,...). Most of the existing studies employ the popular local linear regression technique pioneered by Professor Fan to estimate the functional coefficients, and establish the asymptotic properties of the estimated functional coefficients by applying the well-known results of the local linear regression.

The local linear regression approach is convenient if all functional coefficients are functions of the same set of covariates. In applications, it is quite possible that some coefficients are fixed constants while other coefficiens are functional coefficients. It is also possible that the functional coefficients may have different arguments. If we allow for different functional coefficients to depend on different set of arguments, the local linear regression approach is not convenient for imposing such functional form restrictions. On the other hand, the sieve method directly imposes the functional form restrictions on the functional coefficients and allow us to estimate both the fixed coefficients and the functional coefficients jointly. The main objective of this paper is to apply the sieve method to estimate the functional coefficients models and to establish the asymptotic properties of the sieve estimator. Rather than focusing on specific functional coefficients models, the paper studies a general criterion function $l(z,\alpha)$ and presents a unified asymptotic theory.

The paper is organized as follows: Section 2 presents the sieve estimation for a general functional coefficients model, Section 3 derives some large sample properties of the proposed estimator including consistency and asympptotic distributions, Section 4 presents consistent covariance matrices for the proposed estimator, Section 5 presents some simulation results, followed by concluding remarks in Section 6. All technical proofs are relegated to the appendix.

2 Sieve Estimation

Let z denote the observed data, and let $l(z,\alpha)$ denote a known criterion function that is continuous and is differentiable in the model parameter $\alpha \in \mathcal{A}$ with probability one. The true value $\alpha_0 \in \mathcal{A}$ maximizes the population criterion function $E\{l(z,\alpha)\}$, where the expectation is taken with respect to the true distribution of z. We shall assume that α contains a fixed component θ and a functional component $h = (h_1(x_1), ..., h_m(x_m))$, where for all $j = 1, 2, ..., m$, $h_j(x_j) \in \mathcal{H}_j$ is a function defined over $\mathcal{X}_j$ and x_j is a subset of z. The arguments x_j for all j may contain common variables, but they are not required to be the same. The true value of the fixed coefficient θ_0 is assumed to be an interior point of some compact set Θ, and the true value of the functional coefficient $h_0 = (h_{10}(x_1), ..., h_{m0}(x_m))$ is assumed to be an interior point of some functional space $\mathcal{H} = \mathcal{H}_1 \times \cdots \times \mathcal{H}_m$, defined over $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_m$ and endowed with norm $\|\cdot\|_H$. In applications, $\|\cdot\|_H$ is either the sup-norm or the L_2-norm.

The sieve method approximates the functional coefficient with finite dimensional series approximation. Specifically, for any j, let $\{p_{j1}(x_j), p_{j2}(x_j), ...\}$ denote a series of known basis functions that can approximate any functions in $\mathcal{H}_j$ arbitrarily well. Common basis functions include polynomial series, Fourier series, and splines. For some finite integer k_j, let

$$P_{jk_j}(x_j) = (p_{j1}(x_j), p_{j2}(x_j), ..., p_{jk_j}(x_j))'$$

denote the approximating basis functions, and let $\mathcal{H}_{jk_j} = \{P_{jk_j}(x_j)'\pi_j \in \mathcal{H}_j\}$ denote the approximating space for $\mathcal{H}_j$. Then, for the commonly used sieves, we have

$$\sup_{h_j \in \mathcal{H}_j} \inf_{\widetilde{h}_j \in \mathcal{H}_{jk_j}} \sup_{x_j \in \mathcal{X}_1} \left| h_j(x_j) - \widetilde{h}_j(x_j) \right| = O(k_j^{-\mu_j}), \tag{2.1}$$

where the exponent μ_j is a fixed constant that is inversely related to the dimension of x_j and positively related to the degree of differentiability of the functions. Denote the sieve space as $\mathcal{A}_k = \Theta \times \mathcal{H}_k$, with $\mathcal{H}_k = \mathcal{H}_{1k_1} \times \cdots \times \mathcal{H}_{mk_m}$ and $k = k_1 + ... + k_m$. The sieve space is a compact and finite dimensional parameter space, though the dimension $(k + \dim(\theta))$ shall be allowed to grow with the sample size to obtain consistency of the sieve estimator.

Let $\{z_i, i = 1, 2, ..., N\}$ denote a sample of observations on z. Denote sample observations on x_j by $x_{ji}, i = 1, 2, ..., N$. The proposed estimator for the model parameter is defined as the solution:

$$(\widehat{\theta}, \widehat{\pi}_1, ..., \widehat{\pi}_m) = \arg \max_{\theta \in \Theta, P_{jk_j}(x)'\pi_j \in \mathcal{H}_{jk_j},, j=1,2,...,m} \sum_{i=1}^{N} l(z_i, \theta, P_{1k_1}(x_{1i})'\pi_1, ..., P_{mk_m}(x_{mi})'\pi_m). \tag{2.2}$$

The sieve estimator of the functional coefficient is $\widehat{h}_j(x_j) = P_{jk_j}(x_j)'\widehat{\pi}_j, j = 1, 2, ..., m$ and the parameter of interest is $\widehat{\alpha} = (\widehat{\theta}, \widehat{h}_1(x_1), ..., \widehat{h}_m(x_m))$.

3 Asymptotic properties

We derive the large sample properties of the proposed estimator for the independent and identically distributed samples by applying and extending the techniques developed in Shen (1997) and Ai and Chen (2003). Extensions of our results to weakly dependent and identically distributed samples can be obtained by applying the technique developed in Chen and Shen (1998). Throughout the derivation, we will use the following partially additive quantile regression:

$$y = \theta_0 + h_{10}(x_1) + h_{20}(x_2) + \varepsilon, x = (x_1, x_2)^\tau, \Pr(\varepsilon < 0|x\} = 0.5$$

as an illustrative example. For this example, identification requires that x_1 and x_2 are mutually exclusive and that $E\{h_1(x_1)\} = E\{h_2(x_2)\} = 0$. With $h(x) = (h_1(x_1), h_2(x_2))^\tau$, it is easy to show that the sup-norm is equivalent to

$$\|h\|_H = \sup_{x_1, x_2} |h_1(x_1) + h_2(x_2)|.$$

Moreover, assuming that the support of x is compact and that the joint density function of x is bounded and bounded away from zero and $h(x)$ is continuous in x, then $\|h\|_H$ is equivalent to the mean-squared norm $\|h\|_2 = \sqrt{E\{(h_1(x_1) + h_2(x_2))^2\}}$. With $\alpha = (\theta, h(x)^\tau)^\tau$, the pseudo-norm is denoted as:

$$\|\alpha\|_s^2 = \sup_{x_1, x_2} |\theta + h_1(x_1) + h_2(x_2)|.$$

With $l(z, \alpha) = -|y - \theta - h_1(x_1) - h_2(x_2)|$, it is easy to show that $E\{l(z, \alpha_0)\} > E\{l(z, \alpha)\}$ for any α satisfying $\|\alpha - \alpha_0\|_s \neq 0$.

3.1 Consistency

The consistency of the proposed estimator follows from the following three conditions: (i) the criterion function identifies the true value; (ii) the series approximation error of the functional component converges to zero as sample size goes to infinity; and (iii) the sample analogue of the population criterion function converges to the population criterion function in probability in α uniformly over the sieve space.

Condition (i) is the identification condition which must be verified in applications before applying the proposed estimation.This condition is stated formally in Appendix as Assumption A.5. Condition (ii) can be satisfied if the approximating basis functions can approximate any measurable function arbitrarily well and k_j grows with sample sizes (Assumption A.6). Condition (iii) can be established if the criterion function satisfies an envelope condition (Assumption A.2) and a Lipschitz continuity condition (Assumption A.3) and restrict the growth of the sieve space (Assumption A.4). Under all these conditions, it follows from Theorem A.1 that

$$\frac{1}{N}\sum_{i=1}^{N} l(z_i,\alpha) \to E\{l(z,\alpha)\} \text{ in probability uniformly over } \alpha \in \mathcal{A}_k.$$

Hence, we obtain: $\|\widehat{\alpha}-\alpha_0\|_s \to 0$ in probability as $N \to \infty$.

The consistency of the proposed estimator under the sup-norm is assuring but not enough for deriving the asymptotic distribution of the estimator. To derive the asymptotic distribution, we need the convergence rate which in many applications is hard to compute for the sup-norm. Fortunately, Shen (1997) demonstrates that we only need the convergence rate of the proposed estimator under a Fisher-like norm. To define this norm, let x denote the union of $x_j, j = 1, 2, ..., m$, and assume that $E\{l(z,\alpha)|x\}$ has up to the second pathwise derivatives with respect to α. The identification condition ensures that $E\{\alpha' * \frac{\partial^{2E\{l(z,\alpha_0)|x\}}}{\partial\alpha\partial\alpha'} * \alpha\} < 0$ unless $\alpha = 0$. Denote the Fisher-like norm as

$$\|\alpha-\alpha_0\|^2 = -E\{(\alpha-\alpha_0)' * \frac{\partial^{2E\{l(z,\alpha_0)|x\}}}{\partial\alpha\partial\alpha'} * (\alpha-\alpha_0)\}.$$

In general, we can write

$$\frac{\partial^{2E\{l(z,\alpha)|x\}}}{\partial\alpha\partial\alpha'} = A(x)' * B(x,\alpha) * A(x)$$

for some known matrix function $A(x)$, which does not depend on α, and for some matrix $B(x,\alpha)$. For instance, for the partially additive quantile regression, we have

$$\begin{aligned} E\{l(z,\alpha)|x\} &= \int^{\theta+h_1(x_1)+h_2(x_2)} (y-\theta-h_1(x_1)-h_2(x_2))f_0(y|x)dy - \\ &\quad \int_{\theta+h_1(x_1)+h_2(x_2)} (y-\theta-h_1(x_1)-h_2(x_2))f_0(y|x)dy, \end{aligned}$$

where $f_0(y|x)$ denotes the true conditional density of y given x. The parameter enters $E\{l(z,\alpha)|x\}$ through the index function $\theta + h_1(x_1) + h_2(x_2)$. This index structure gives to

$$\frac{\partial^{2E\{l(z,\alpha)|x\}}}{\partial\alpha\partial\alpha'} = \frac{\partial^{2E\{l(z,\alpha)|x\}}}{\partial\theta^2}\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}\frac{\partial^{2E\{l(z,\alpha)|x\}}}{\partial\theta^2}\begin{pmatrix} 1 & 1 & 1 \end{pmatrix}.$$

Hence $A(x) = \begin{pmatrix} 1 & 1 & 1 \end{pmatrix}$ and $B(x,\alpha) = \frac{\partial^{2E\{l(z,\alpha)|x\}}}{\partial\theta^2}$. This factorization is useful because $\frac{\partial^{2E\{l(z,\alpha)|x\}}}{\partial\theta^2}$ may be singular but $B(x,\alpha)$ is not. We shall assume that the largest and lowest eignvalues of $-B(x,\alpha)$

are bounded and bounded away from zero for all x and all α satisfying $\|\alpha-\alpha_0\|_s = o(1)$, then $\|\alpha-\alpha_0\|$ is equivalent to the mean-squared norm

$$\|\alpha-\alpha_0\|_2^2 = E\{(\alpha-\alpha_0)^\tau A(x)^\tau A(x)(\alpha-\alpha_0)\},$$

which is equivalent to the pseudo-norm

$$\|\alpha-\alpha_0\|_s = \sup_x |A(x)(\alpha-\alpha_0)|.$$

Assume that the third directional derivative of $E\{l(z,\alpha)|x\}$ in the direction $\alpha-\alpha_0$ exists and is bounded by $O(\|\alpha-\alpha_0\|_s^3)$ for all x and for all α satisfying $\|\alpha-\alpha_0\|_\infty = o(1)$. Then we can easily show that for all α satisfying $\|\alpha-\alpha_0\|_s = o(1)$ we have

$$E\{l(z,\alpha_0)\} - E\{l(z,\alpha)\} = \|\alpha-\alpha_0\|^2 + o\left(\|\alpha-\alpha_0\|^2\right),$$

and Assumption A.7 is satisfied. To illustrate with the partially additive quantile regression again, the third directional derivative of $E\{l(z,\alpha)|x\}$ in the direction $\alpha-\alpha_0$ is

$$\frac{\partial^3 E\{l(z,\alpha)|x\}}{\partial\theta^3}(\theta + h_1(x_1) + h_2(x_2) - \theta_0 - h_{10}(x_1) - h_{20}(x_2))^3,$$

which is bounded by $c\|\alpha-\alpha_0\|_s^3$ if $\left|\frac{\partial^3 E\{l(z,\alpha)|x\}}{\partial\theta^3}\right| \le c$.

Assumption A.8 is satisfied with $\delta_N = N^{-1/4}$ if $A(x)$ is bounded and $k_j^{-\mu_j}N^{1/4} - > 0$ as $N \to +\infty$. Assumption A.4 and A.9 are satisfied with $\delta_N = N^{-1/4}$ if $\frac{\sqrt{N}}{k_j} \to +\infty$. Applying Theorem A.3, we obtain

$$\|\widehat{\alpha}-\alpha_0\| = o_p(N^{-1/4}),\ \ \|\widehat{\alpha}-\alpha_0\|_2 = o_p(N^{-1/4}), \text{ and } \|\widehat{\alpha}-\alpha_0\|_s = o_p(N^{-1/4}).$$

3.2 Asymptotic Distribution

The convergence rate established above allows us to prove the following key result

$$\begin{aligned} &\frac{1}{N}\sum_{i=1}^N l(z_i,\alpha) - \frac{1}{N}\sum_{i=1}^N l(z_i,\alpha_0) \\ =\ & \frac{1}{N}\sum_{i=1}^N \frac{\partial l(z_i,\alpha_0)}{\partial a'}(a-a_0) + \frac{1}{2}E\left\{(a-a_0)'\frac{\partial^2 E\{l(z,\alpha_0)|x\}}{\partial a\partial a'}(a-a_0)\right\} + O_p(\|\alpha-\alpha_0\|_s^3)\end{aligned}$$

for all $\alpha \in \mathcal{A}$ satisfying $\|\alpha-\alpha_0\|_s = o(N^{-1/4})$. Denote

$$P_k(x) = \begin{pmatrix} p_{1k_1}(x_1)' & 0 & 0 & 0 & 0 \\ 0 & p_{2k_2}(x_2)' & 0 & 0 & 0 \\ . & . & . & . & . \\ 0 & 0 & . & . & p_{mk_m}(x_m)' \end{pmatrix},$$

and

$$\pi = \begin{pmatrix} \pi_1 \\ \pi_2 \\ . \\ \pi_m \end{pmatrix} \text{ and } \pi_{0k} = \begin{pmatrix} \pi_{1k_1} \\ \pi_{2k_2} \\ . \\ \pi_{mk_m} \end{pmatrix},$$

where $p_{jk_j}(x_j)'\pi_{jk_j}$ denotes the projection of $h_{j0}(x_j)$ on $p_{jk_j}(x_j)$. Under some sufficient conditions, we can show

$$\begin{aligned}\widehat{\pi}-\pi_{0k} &= -\left(E\left\{P(x)'*\frac{\partial^2 E\{l(z,\alpha_0)|x\}}{\partial h\partial h'}*P(x)\right\}\right)^{-1}* \\ &\quad \left[\frac{1}{N}\sum_{i=1}^{N}P(x_i)'\frac{\partial l(z_i,\alpha_0)}{\partial h}+E\left\{P(x)'*\frac{\partial^2 E\{l(z,\alpha_0)|x\}}{\partial h\partial \theta'}\right\}(\widehat{\theta}-\theta_0)\right]+o_p(k*N^{-1/2})\end{aligned}$$

and $\widehat{\theta}-\theta_0=$

$$\left(\ E\left\{w^*(x)'\tfrac{\partial^2 E\{l(z,\alpha_0)|x\}}{\partial\alpha\partial\alpha'}w^*(x)\right\}\ \right)^{-1}*\frac{1}{N}\sum_{i=1}^{N}\left[w^*(x_i)'\frac{\partial l(z_i,\alpha_0)}{\partial\alpha}\right]+o_p(N^{-1/2}),$$

where $w^*(x)'=(I,w_1^*(x))$ solves

$$\sup_{w(x)'=(I,w_1(x))}E\left\{w(x)'\frac{\partial^2 E\{l(z,\alpha_0)|x\}}{\partial\alpha\partial\alpha'}w(x)\right\}.$$

Applying a central limit theorem, we obtain $\sqrt{N}(\widehat{\theta}-\theta_0)$ is asymptotically normally distributed with mean zero and covariance given by

$$\begin{aligned}\Omega &= \left(\ E\left\{w^*(x)'\tfrac{\partial^2 E\{l(z,\alpha_0)|x\}}{\partial\alpha\partial\alpha'}w^*(x)\right\}\ \right)^{-1}\left(\ E\left\{w^*(x)'\tfrac{\partial l(z,\alpha_0)}{\partial\alpha}\tfrac{\partial l(z,\alpha_0)}{\partial\alpha'}w^*(x)\right\}\ \right) \\ &\quad *\left(\ E\left\{w^*(x)'\tfrac{\partial^2 E\{l(z,\alpha_0)|x\}}{\partial\alpha\partial\alpha'}w^*(x)\right\}\ \right)^{-1}.\end{aligned}$$

Therefore,

$$\begin{aligned}\widehat{h}(x)-h_0(x) &= P(x)\widehat{\pi}-h_0(x)=-P(x)\left(E\left\{P(x)'*\frac{\partial^2 E\{l(z,\alpha_0)|x\}}{\partial h\partial h'}*P(x)\right\}\right)^{-1}* \\ &\quad \left[\frac{1}{N}\sum_{i=1}^{N}P(x_i)'\frac{\partial l(z_i,\alpha_0)}{\partial h}\right]+o_p(k*N^{-1/2}).\end{aligned}$$

Denote

$$\begin{aligned}\Gamma(x) &= P(x)\left(E\left\{P(x)'*\frac{\partial^2 E\{l(z,\alpha_0)|x\}}{\partial h\partial h'}*P(x)\right\}\right)^{-1}E\left\{P(x)'\frac{\partial l(z,\alpha_0)}{\partial h}\frac{\partial l(z,\alpha_0)}{\partial h'}P(x)\right\} \\ &\quad *\left(E\left\{P(x)'*\frac{\partial^2 E\{l(z,\alpha_0)|x\}}{\partial h\partial h'}*P(x)\right\}\right)^{-1}P(x)'.\end{aligned}$$

Then under conditions similar to those imposed by newey (1997), we have $\Gamma(x)^{1/2}\left(\widehat{h}(x)-h_0(x)\right)$ is asymppttotically normally distributed with mean zero and identity covariance matrix.

4 Covariance Estimator

To be continued

5 Simulation

To be continued

6 Conclusion

To be continued

APPENDIX A: Useful Convergence Results

First, we prove some uniform convergence results that are useful for deriving the asymptotic properties of the sieve estimator. Let $\{z_i, i = 1, 2, ..., N\}$ denote a random sam[ple of size N. Let $\mathcal{A}_N \subset \mathcal{A}$ denote a sieve approximating space of $\mathcal{A}$ and let $||\alpha||_s$ denote some metric defined over $\alpha \in \mathcal{A}$. Let $N(\varepsilon, \mathcal{A}_N, ||\cdot||_s)$ denote the minimal number of ε-radius covering balls of $\mathcal{A}_N$ under the metric $||\cdot||_s$. Let $l(z, \alpha)$ denote a generic measurable function of the data $z \in \mathcal{Z}$ and the parameter $\alpha \in \mathcal{A}$. The sieve estimator is defined as

$$\widehat{\alpha} = \arg\max_{\alpha \in \mathcal{A}_N} \sum_{i=1}^{N} l(z_i, \alpha).$$

Throughout the appendix, we impose the following conditions:

ASSUMPTION A.1. $\{z_i, i = 1, 2, ..., N\}$ *is a sample of observations drawn ndependently from the distribution of* z.

ASSUMPTION A.2. There exists a positive function $C_1(z)$ *satisfying* $E\{C_1(z)^p\} < +\infty$ *for some* $p > 2$ *and* $|l(z, \alpha)| \leq C_1(z)$.

ASSUMPTION A.3. For any pair $\alpha \in \mathcal{A}_N$ and $\beta \in \mathcal{A}_N$, *there exist* $C_2(z)$ *satisfying* $E\{C_2(z)\} < \infty$ *and a positive constant* κ *such that*

$$|l(z, \alpha) - l(z, \beta)| \leq C_2(z) \|\alpha - \beta\|_s^{\kappa}.$$

ASSUMPTION A.4. For some positive value δ_N *that either equals to one or converges to zero as sample size* N *goes to infinity and*

$$\frac{N \times \delta_N^2}{\ln[N(\delta_N^{1/\kappa}, \mathcal{A}_N, ||\cdot||_s)]} \to +\infty \text{ and } N\delta_N^2 \to +\infty \text{ as } N \to +\infty.$$

The following uniform convergence result is important for establishing the large sample properties of the sieve estimator.

THEOREM A.1: *Under Assumptions A.1-A.4, we obtain* $\frac{1}{N}\sum_{i=1}^{N} l(z_i, \alpha) - E\{l(z, \alpha)\} = o_p(\delta_N)$ *uniformly over* $\alpha \in \mathcal{A}_N$ *as* $N \to +\infty$.

PROOF: Let c denote a generic constant which may have different values in different expressions. For any pair $\alpha \in \mathcal{A}_N$ and $\beta \in \mathcal{A}_N$, Assumption A.3 gives

$$\left|\varepsilon(z, \alpha^1) - \varepsilon(z, \alpha^2)\right| \leq (C_2(z) + E\{C_2(z)\}) \|\alpha - \beta\|_s^{\kappa}.$$

It follows that

$$\left|\frac{1}{N}\sum_{i=1}^{N}[\varepsilon(z_i, \alpha) - \varepsilon(z_i, \beta)]\right| \leq \frac{1}{N}\sum_{i=1}^{N}(C_2(z_i) + E\{C_2(z)\}) \|\alpha - \beta\|_s^{\kappa}.$$

For arbitrarily small $\eta > 0$, there exists a constant c such that

$$\Pr\left(\frac{1}{N}\sum_{i=1}^{N}(C_2(z_i)+E\{C_2(z)\})>c\right)=\eta.$$

For any small $\epsilon > 0$, partition $\mathcal{A}_N$ into b_N mutually exclusive subsets $\mathcal{A}_{Nm}, m = 1, 2, \cdots, b_N$, so that with $\alpha \in \mathcal{A}_{Nm}$ and $\beta \in \mathcal{A}_{Nm}$ imply $\|\alpha - \beta\|_s^{\kappa} \leq \epsilon\delta_N/c$. Then with probability approaching one, we have:

$$\left|\frac{1}{N}\sum_{i=1}^{N}[\varepsilon(z_i,\alpha)-\varepsilon(z_i,\beta)]\right|\leq\epsilon\delta_N.$$

Let α^m denote a fixed point in $\mathcal{A}_{Nm}$. For any $\alpha \in \mathcal{A}_N$, there exists m such that $\|\alpha - \alpha^m\|_s^{\kappa} \leq \epsilon\delta_N/c$. Then, with probability approaching one,

$$\sup_{\alpha\in\mathcal{A}_n}\left|\frac{1}{N}\sum_{i=1}^{N}\varepsilon(z_i,\alpha)\right|\leq\epsilon\delta_N+\max_m\left|\frac{1}{N}\sum_{i=1}^{N}\varepsilon(z_i,\alpha^m)\right|.$$

Hence

$$\Pr\left(\sup_{\alpha\in\mathcal{A}_N}\left|\frac{1}{N}\sum_{i=1}^{N}\varepsilon(z_i,\alpha)\right|>3\epsilon\delta_N\right)<\eta+\Pr\left(\max_m\left|\frac{1}{N}\sum_{i=1}^{N}\varepsilon(z_i,\alpha^m)\right|>2\epsilon\delta_N\right)$$

For some constant $c > 0$ and $M > 0$, denote $d_i = 1\{C_1(z_i) \leq M\}$. Denote $l_1(z_i, \alpha) = d_i l(z_i, \alpha)$, $l_2(z_i, \alpha) = (1 - d_i)l(z_i, \alpha)$, $\varepsilon_1(z_i, \alpha) = l_1(z_i, \alpha) - E\{l_1(z, \alpha)\}$, and $\varepsilon_2(z_i, \alpha) = l_2(z_i, \alpha) - E\{l_2(z, \alpha)\}$. We obtain

$$\begin{aligned}
&\Pr\left(\max_m\left|\frac{1}{N}\sum_{i=1}^{N}\varepsilon(z_i,\alpha^m)\right|>2\epsilon\delta_N\right)\\
\leq\;&\Pr\left(\max_m\left|\frac{1}{N}\sum_{i=1}^{N}\varepsilon_1(z_i,\alpha^m)\right|>\epsilon\delta_N\right)\\
&+\Pr\left(\max_m\left|\frac{1}{N}\sum_{i=1}^{n}\varepsilon_2(z_i,\alpha^m)\right|>\epsilon\delta_N\right)\equiv P_1+P_2.
\end{aligned}$$

Applying Markov inequality we obtain

$$P_2\leq\frac{E\left\{\max_m\left|\frac{1}{N}\sum_{i=1}^{N}\varepsilon_2(z_i,\alpha^m)\right|\right\}}{\epsilon\delta_N}.$$

Note that

$$\begin{aligned}
|\varepsilon_2(z_i,\alpha^m)|&\leq(1-d_i)C_1(z_i)+E\{(1-d_i)C_1(z_i)\}\\
&\leq(1-d_i)C_1(z_i)+\sqrt{E\{(1-d_i)\}}\sqrt{E\{C_1(z)^2\}}.
\end{aligned}$$

Then

$$E\left\{\max_m\left|\frac{1}{N}\sum_{i=1}^{N}\varepsilon_2(z_i,\alpha^m)\right|\right\}\leq$$
$$\leq 2\sqrt{E\{(1-d_i)\}}\sqrt{E\{C_1(z)^2\}}\leq\frac{c}{M^{p/2}}$$

for some constant c as $E\{(1 - d_i)\} = \Pr(C_1(z) > M) = O(M^{-p})$. Set $M = \left(\frac{c}{\delta_N\epsilon\eta}\right)^{2/p}$. We obtain $P_2 \leq \frac{c}{M^{p/2}\epsilon\delta_N} = \eta$.

Note that

$$\sigma_m^2\equiv N\times E\{[\frac{1}{N}\sum_{i=1}^{N}\varepsilon_1(z_i,\alpha^m)]^2\}=E\{\varepsilon_1(z_i,\alpha^m)^2\}\leq E\{C_1(z)^2\}$$

and $|\frac{1}{N}\sum_{i=1}^{N}\varepsilon_1(z_i,\alpha^m)| \leq 2M$. Applying the Bernstein inequality for independent processes, we obtain:

$$\Pr\left(\left|\frac{1}{N}\sum_{i=1}^{N}\varepsilon_1(z_i,\alpha^m)\right| > \epsilon\delta_N\right) \leq 2\exp\left(-N\varepsilon^2\delta_N^2/(\sigma_m^2 + M\varepsilon\delta_N)\right).$$

Hence

$$P_1 < 2b_n \exp\left(-N\varepsilon^2\delta_{1n}^2/(E\{C_1(z)^2\} + M\varepsilon\delta_N)\right).$$

Since $M\delta_N = O(\delta_N^{1-2/p}) \to 0$ or bounded. For some positive constant c, we have

$$\begin{aligned} P_1 &< 2b_N\exp(-cN\delta_N^2) = 2\exp(\ln(b_N) - cN\delta_N^2) \\ &= 2\exp(-cN\delta_N^2)\exp(1 - \frac{\ln(b_N)}{cN\delta_N^2}) \to 0 \end{aligned}$$

if

$$\frac{N\times\delta_N^2}{\ln(b_N)} = \frac{N\times\delta_N^2}{\ln[N(\delta_N^{1/\kappa},\mathcal{A}_N,||\cdot||_s)]} \to +\infty \text{ and } N\delta_N^2 \to +\infty.$$

This completes the proof. *Q.E.D.*

Assumption A.4 implies that the uniform convergence rate established in Theorem A.1 is arbitrarily close to but always slower than $N^{-1/2}$.Next, we impose the identification condition and restriction on the approximation error.

ASSUMPTION A.5. $\alpha^* = \arg\max_{\alpha\in\mathcal{A}} E\{l(z,\alpha)\}$ *if and only if* $\|\alpha^* - \alpha_0\|_s = 0$.

ASSUMPTION A.6. For any $\alpha \in \mathcal{A}$, *there exists* $\Pi_N\alpha \in \mathcal{A}_N$ *satisfying* $\|\Pi_N\alpha - \alpha\|_s \to 0$ *as* $N \to +\infty$.

These and Theorem A.1 imply the following consistency result:

THEOREM A.2: *Under Assumptions A.1-A.6, we obtain* $\|\widehat{\alpha} - \alpha_0\|_s = o_p(1)$ *as* $N \to +\infty$.

The consistency result of Theorem A.2 is useful but not enough for establishing the asymptotic distribution of the sieve estimator. To establish the asymptotic distribution, we need convergence rates. However, under the pseudo norm $\|\cdot\|_s$, obtaining a convergence rate for the sieve estimator may not be possible. In the following, we will compute the convergence rate for a norm $\|\cdot\|$ that is locally equivalent to the criterion function:

ASSUMPTION A.7. There exist some positive constants $c_1 < c_2$ *such that*

$$c_1\|\alpha - \alpha_0\|^2 \leq E\{l(z,\alpha_0)\} - E\{l(z,\alpha)\} \leq c_2\|\alpha - \alpha_0\|^2$$

holds for any α *satisfying* $\|\alpha - \alpha_0\|_s = o(1)$.

Assumption A.3 implies that the norm $\|\cdot\|$ is generally weaker than the pseudo norm $\|\cdot\|_s$, except for some special cases where the two norms are equivalent. Within the neighborhood of α_0 defined by $\|\alpha - \alpha_0\|_s = o(1)$, we shall impose some additional conditions on the size of the sieve space and the sieve approximation errors.

ASSUMPTION A.8. $\|\Pi_N\alpha - \alpha\| = o(\delta_N)$ *holds uniformly over any* α *satisfying* $\|\alpha - \alpha_0\|_s = o(1)$.

ASSUMPTION A.9. With $k = \dim(\mathcal{A}_N)$, $\ln[N(\varepsilon^{1/\kappa},\mathcal{A}_N,||\cdot||_s)] = O(k\ln(\frac{1}{\varepsilon}))$

Notice that all conditions of Theorem 1 of Shen and Wong (1994, pp.583) are trivially satisfied. Applying that theorem, we obtain:

Theorem A.3: *Under Assumptions A.1-A.9, we obtain* $\|\widehat{\alpha} - \alpha_0\| = o_p(\delta_N)$ *as* $N \to +\infty$.

APPENDIX B: Asymptotic Distributions

To be continued

Reference

Bhattacharya, P. K. and Gangopadhyay, A. K. (1990). Kernel and nearest-neighbor estimation of a conditional quantile. *Ann. Statist.*, **18**, 1400–1415.

Cai, Z., Fan, J. and Li, R. (2000). Efficient estimation and inferences for varying-coefficient models. *J. Amer. Statist. Assoc.*, **95**, 888–902.

Cai, Z., Fan, J. and Yao, Q. (2000). Functional-coefficient regression models for nonlinear time series. *J. Amer. Statist. Assoc.*, **95**, 941–956.

Chaudhuri, P. (1991a). Global nonparametric estimation of conditional quantile functions and their derivatives. *J. Multivariate Anal.*, **39**, 246–269.

Chaudhuri, P. (1991b). Nonparametric estimates of regression quantiles and their local Bahadur representation. *Ann. Statist.*, **19**, 760–777.

He, X. and Zhu, L. X. (2003). A lack-of-fit test for quantile regression. *J. Amer. Statist. Assoc.*, **98**, 1013–1022.

He, X. and Liang, H. (2000). Quantile regression estimates for a class of linear and partially linear errors-in-variables models. *Statist. Sinica*, **10**, 129–140.

He, X. and Ng, P. (1999). Quantile splines with several covariates. *J. Statist. Plann. Inference*, **75**, 343–352.

He, X. and Shao, Q.-M. (1996). A general Bahadur representation of M-estimators and its application to linear regression with nonstochastic designs. *Ann. Statist.*, **24**, 2608–2630.

He, X., Ng, P. and Portnoy, S. (1998). Bivariate quantile smoothing splines. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **60**, 537–550.

Koenker, R., and Bassett, G. Jr. (1978). Regression quantiles. *Econometrica*, **46**, 33–50.

Koenker, R., Ng, P. and Portnoy, S. (1994). Quantile smoothing splines. *Biometrika*, **81**, 673–680.

Ying, Z., Jung, S. H. and Wei, L. J. (1995). Survival analysis with median regression models. *J. Amer. Statist. Assoc.*, **90**, 178–184.

Yu, K. and Jones, M. C. (1998). Local linear quantile regression. *J. Amer. Statist. Assoc.* **93**, 228–237.