The Korean Communications
in Statistics Vol. 6, No. 3, 1999
pp. 849-859

Wavelet Estimation of Regression Functions with Errors in Variables[1)]

Woo-Chul Kim[2)] and Ja-Yong Koo[3)]

Abstract

This paper addresses the issue of estimating regression function with errors in variables using wavelets. We adopt a nonparametric approach in assuming that the regression function has no specific parametric form. To account for errors in covariates, deconvolution is involved in the construction of a new class of linear wavelet estimators. Using the wavelet characterization of Besov spaces, the question of regression estimation with Besov constraint can be reduced to a problem in a space of sequences. Rates of convergence are studied over Besov function classes B_{spq} using L_2 error measure. It is shown that the rates of convergence depend on the smoothness s of the regression function and the decay rate of characteristic function of the contaminating error.

1. Introduction

Let (X, U) denote a pair of random variables and consider the problem of estimating the regression function $\mu(x) = E(U|X = x)$. Due to measuring mechanism or the nature of the environment, the variable X is measured with error and is not directly observable; see Fuller (1987) or Carroll, Ruppert and Stefanski (1995). Instead, X is observed through $Y = X + Y$, where W is a random noise. It is assumed that W has a known distribution, and is independent of (X, U). Given a random sample $(Y_1, U_1), \ldots, (Y_n, U_n)$ from the distribution of (X, U) we want to estimate μ. Here we adopt a nonparametric approach in assuming that μ has no specific parametric form.

A traditional approach to regression estimation is by orthogonal series. Recently, wavelet curve estimation has become a well-known and sound technique for adaptively estimating functions. Optimal rates of convergence have been thoroughly examined for different observation schemes by many authors. Most current wavelet methods focus on ordinary

---

1) The study was supported by the academic research fund (BSRI-97-1418) of Ministry of Education, Republic of Korea

2) Professor, Department of Computer Science and Statistics, Seoul National University, Seoul 151-742, Korea

3) Associate Professor, Department of Statistics, Hallym University, Chunchon, Kangwon-Do 200-702, Korea

regression, see e.g. Donoho and Johnstone (1994). This paper explores the possibility of applying an ordinary nonparametric wavelet regression estimator to the problem of estimating the regression function with errors in covariate.

Regression analysis with errors in variables is evolving rapidly. See, for example, Stefanski (1985), Stefanski and Carroll (1985, 1991), Bickel and Ritov (1987). Recently, for the problem of estimating regression function with errors in covariates, Fan and Truong (1993) address optimal rates of convergence of kernel estimators and Koo and Lee (1998) study rates of convergence of B-spline estimators. One may want to refer to Fan and Truong (1993) for more related references.

The basic idea is closely related to the deconvolution techniques. Consider the deconvolution problem of estimating the density f of X based on $Y_m = X_m + W_m$, $1 \leq m \leq n$. Denote the density of Y by g and let F_W denote the distribution function of W. Then

$$g(y) = \int_{-\infty}^{\infty} f(y-w)dF_W(w)$$

and

$$\Phi_Y(t) = \Phi_W(t)\Phi_X(t),$$

where Φ_U denotes the characteristic function of a random variable U. This suggests that the marginal density f can be estimated by the deconvolution method.

The form of estimator studied in this article is closely related to Fan and Truong (1993) and Koo and Lee (1998). To account for errors in covariates, deconvolution is involved in the construction of a new class of linear wavelet estimators; see, for example, Fan (1991) and Fan and Koo (1999) for studies on deconvolution. A wavelet estimator of regression function μ can be given by

$$\mu^\star = \gamma^\star / f^\star$$

where $\gamma^\star$ is an estimator of the function $\gamma(x) = \mu(x)f(x)$ and and $f^\star$ is a deconvolution density estimator of f. The final estimator has the following form:

$$\mu^\star(x) = \sum_j W_{n,j}(Y_1, \ldots, Y_n)U_j$$

with weights $W_{n,j}(Y_1, \ldots, Y_n)$ constructed to account for measurement errors. Fourier transform and wavelet bases are used for this purpose.

Using the wavelet characterization of Besov spaces, the question of regression estimation with Besov constraint can be reduced to a problem in a space of sequences.

Rates of convergence are studied over Besov function classes B_{spq} using L_2 error measure. It is shown that the rates of convergence depend on the smoothness s and the decay rate of characteristic function of the contaminating error.

The organization of the paper is as follows. In Section 2 Besov space and wavelets are described for the introduction of notation. Section 3 proposes wavelet estimators for the estimation of regression function with errors in variables. The asymptotic properties of estimators given in Section 3 are proved in Section 4

2. Besov spaces and wavelets

In this section, we recall definitions and set notation for later use. For a function h, let $h_{j,k}(x) = 2^{j/2}h(2^jx-k)$ and let

$$\hat{h}(t) = \int e^{-itx}h(x)dx$$

be the Fourier transform of h. By change of variable,

$$\hat{\varphi}_{j,k}(t) = 2^{-j/2}\exp(-itk/2^j)\hat{\varphi}(t/2^j). \tag{2.1}$$

For a random variable U, let Φ_U denote the characteristic function of U defined by $\Phi_U(t) = Ee^{itU}$. Let $M, M_1, M_2, \cdots$ and C denote positive constants which are independent of n, where C is not necessarily the same at each appearance.

2.1. Multiresolution Analysis and Wavelets

Let us recall that one can construct a function φ such that:

(S1). The sequence $\{\varphi(x-k): k\in Z\}$ is an orthonormal family of $L_2(R)$. Let V_0 be the subspace spanned by $\{\varphi(x-k): k\in Z\}$;

(S2). For all $j\in Z$, $V_j\subset V_{j+1}$ if V_j denotes the space spanned by $\{\varphi_{j,k}; k\in Z\}$;

(S3). φ is of class C^r, the space of r-times continuously differentiable functions.

Under (S3), we have

$$\int (1+|t|)^r|\hat{\varphi}(t)|dt \le C, \tag{2.2}$$

where C is a positive constant. We have $\bigcap_{j\in Z}V_j = \{0\}$ and, furthermore, if $\varphi\in L_2(R)$ and $\int\varphi = 1$, $L_2(R) = \bigcup_{j\in Z}V_j$ and φ is called the scaling function of the multiresolution analysis $(V_j)_{j\in Z}$. From (7.1.23) of Daubechies (1992), we can choose r as large as we want. In addition to (S3), we will assume that

(S4). φ is compactly supported in an interval $[-A\ +A]$.

Under these conditions, define the space W_j by $V_j \oplus W_j$. There exists a function ψ (the 'mother wavelet') such that:

(W1). $\{\psi(x-k):\ k\in Z\}$ is an orthonormal basis of W_0;

(W2). $\{\psi_{j,k}, j, k\in Z\}$ is an orthonormal basis of $L_2(R)$;

(W3). ψ has the same regularity properties as φ.

In addition, we have the decomposition

$$L_2(R) = V_i \oplus W_l \oplus W_{l+1} \oplus \cdots.$$

That is, for all $f \in L_2(R)$,

$$f = \sum_{k\in Z} \alpha_{0,k}\varphi_{i,k} + \sum_{j\ge 0}\sum_{k\in Z}\beta_{j,k}\psi_{j,k},$$

where

$$\alpha_{0,k} = \int f(x)\overline{\varphi_{0,k}(x)}dx, \quad \beta_{j,k} = \int f(x)\overline{\psi_{j,k}(x)}dx.$$

2.2. Besov spaces

We give here the definition of Besov spaces in terms of wavelet coefficients. For the properties of Besov spaces,

Let $r > s$, let E_j be the associated projection operator onto V_j and $D_j = E_{j+1} - E_j$. Besov spaces depend on three parameters $s \ge 1, 1\le p\le\infty$ and $1\le q\le\infty$ and are denoted B_{spq}. Say that $f \in B_{spq}$ if and only if the norm

$$J_{spq}(f) = \|E_0 f\|_{L_p(R)} + \left\{\sum_{j\ge 0}(2^{js}\|D_j f\|_{L_p(R)})^q\right\}^{1/q} < \infty$$

(with usual modification for $q = \infty$). Using now the decomposition of f:

$$E_j f = \sum_{k\in Z}\alpha_{j,k}\varphi_{j,k},$$

$$D_j f = \sum_{k\in Z}\beta_{j,k}\psi_{j,k},$$

we may say that $f \in B_{spq}$ if and only if the equivalent norm

$$J_{spq}(f) = \|\alpha_{0\cdot}\|_{l_p} + \left\{\sum_{j\ge 0}(2^{j(s+1/2-1/p)}\|\beta_{j\cdot}\|_{l_p})^q\right\}^{1/q} < \infty$$

[we have set $\|\alpha_{0\cdot}\|_{l_p} = (\sum_{k\in Z}|\alpha_{0,k}|^p)^{1/p}$ and $\|\beta_{j\cdot}\|_{l_p} = (\sum_{k\in Z}|\beta_{j,k}|^p)^{1/p}$]. Abusing the notation slightly, we will also write $\|\beta\|_{spq}$ for the above sequence norm applied to coefficients. Set also $B_{spq}(M) = \{\alpha : \|\beta\|_{spq} \le M\}$.

One may want to refer to Devore and Popov (1988) and Donoho, Johnstone, Kerkyacharian and Picard (1996). Well-known cases of the Besov spaces, which are used in statistical literature, include the Hilbert-Sobolev spaces $H^s = B_{s22}$, the set of bounded s-Lipschitz functions $B_{s\infty\infty}$ with non-integer s. The function spaces considered in Fan and Truong (1993) and Koo and Lee (1998) are particular cases of the Besov spaces.

3. Linear wavelet estimators

In this section, we define the linear wavelet regression estimators and state asymptotic results of the proposed estimators.

3.1 Notations

Let $\gamma(x) = \mu(x)f(x) = \int f_{XU}(x, u)\,du$ where f_{XU} and f are the joint density function of (X, U) and the marginal density function of X, respectively.
From now on, it is assumed that X takes values in the interval $I_a = [-a, 1+a]$ for a positive constant a. We consider functions on the interval $I = [0, 1]$. For approximation on I, we need only the wavelets $\varphi_{j,k}$ which do not vanish identically on I. We let K_j denote the sets of k for which $\varphi_{j,k}$ which do not vanish identically on I. Let us assume that the wavelet basis is derived from $\{\varphi_{j,k} : k \in K_j\}$.

3.2. Estimator of the marginal density

For the estimation of a density function, we will consider the following class of linear wavelet density estimators. Let E denote the kernel associated with a projection on the space V_0 of a multiresolution analysis associated with a wavelet:

$$E(x, y) = \sum_{k\in} \varphi(x-k)\overline{\varphi(y-k)}.$$

Let us put $E_j(x, y) = 2^j E(2^j x, 2^j y)$ and $Ef(x) = \int E(x, y)f(y)dy$. Let $E_j^\star(x)$ be the estimator of f at the level j based on X_m's:

$$E_j^\star(x) = \int E_j(x, y)dF_n(y) = \frac{1}{n}\sum_{m=1}^{n} 2^j E(2^j x, 2^j X_m)$$

where F_n is the empirical cumulative distribution function based on X_m's. We can express the estimator $E_j^{\star}(x)$, which is a linear density estimator, as follows. A linear wavelet density estimate has the form

$$f^{*}(x) = \sum_k \tilde{\alpha}_{j,k} \varphi_{j,k}(x).$$

Since the variables $X_1, \ldots, X_n$ are $\tilde{\alpha}_{j,k}$ can be replaced by an estimator using the Fourier transform and the relation

$$\Phi_Y(t) = \Phi_X(t)\Phi_W(t),$$

which follows from the independence of X and W.

To develop upper bounds, we assume that

(A1). $\Phi_W(t) \neq 0$ for any t.

Let $\alpha_{j,k} = \int \varphi_{j,k}(x) f(x)\, dx$ from now on. It can be seen that

$$\alpha_{j,k} = \frac{1}{2\pi} \int \hat{\varphi}_{j,k}(t) \frac{\Phi_Y(t)}{\Phi_W(t)}\, dt. \tag{3.1}$$

Here the relation (3.1) is interpreted as a formal relation. Let

$$\hat{\Phi}_Y(t) = n^{-1} \sum_{m=1}^{n} \exp(itY_m).$$

From (3.1),

$$\hat{\alpha}_{j,k} = \frac{1}{2\pi} \int \hat{\varphi}_{j,k}(t) \frac{\hat{\Phi}_Y(t)}{\Phi_W(t)}\, dt$$

is an unbiased estimators of $\alpha_{j,k}$. Let the transform K^j of a function h be defined by

$$(K^j h)(y) = \frac{1}{2\pi} \mathrm{Re}\left[\int e^{ity} \frac{\hat{h}(t)}{\Phi_W(2^j t)}\, dt \right], \quad y \in R \tag{3.2}$$

where $Re[\zeta]$ denotes the real part of a complex number ζ. Then it follows from (2.1) and (3.2) that

$$\hat{\alpha}_{j,k} = \frac{1}{n} \sum_{m=1}^{n} (K^j \varphi)_{j,k}(Y_m).$$

Given $\hat{\alpha}_{j,k}$, we can estimate f by $f^{\star} = \sum_k \hat{\alpha}_{j,k} \varphi_{j,k}$.

3.3. Estimator of regression function

Define

$$\Phi_{UY}(t) = EUe^{itY}$$

and let $\beta_{j,k}=v\int\psi_{j,k}(x)f(x)dx$ from now on. Then we have that

$$\beta_{j,k}\equiv\int\psi_{j,k}(x)f(x)dx=\int\varphi_{j,k}(x)\gamma(x)dx.$$

Let us note that

$$\beta_{j,k}=\frac{1}{2\pi}\int\widehat{\varphi}_{j,k}(t)\frac{\Phi_{ZY}(t)}{\Phi_W(t)}dt. \tag{3.3}$$

Here the relation (3.3) is also interpreted as a formal relation.

Let

$$\widehat{\Phi_{ZY}}(t)=n^{-1}\sum_{m=1}^{n}U_m\exp(itY_m)$$

From (3.3)

$$\widehat{\beta}_{j,k}=\frac{1}{2\pi}\int\widehat{\varphi}_{j,k}(t)\frac{\widehat{\Phi}_{ZY}(t)}{\Phi_W(t)}dt \tag{3.4}$$

is unbiased estimators of $\beta_{j,k}$. By (2.1) and (3.4), we have

$$\widehat{\beta}_{j,k}=\frac{1}{n}\sum_{m=1}^{n}(K^j\varphi)_{j,k}(Y_m)Z_m$$

Now we propose an estimator of μ by

$$\mu^{\star}=\frac{\sum_{k\in K_j}\widehat{\beta}_{j,k}\varphi_{j,k}}{\sum_{k\in K_j}\widehat{\alpha}_{j,k}\varphi_{j,k}}. \tag{3.5}$$

Then (3.5) can be interpreted as $\mu^{\star}=\gamma^{\star}/f^{\star}$, where $\gamma^{\star}=\sum_k\widehat{\beta}_{j,k}\varphi_{j,k}$ and $f^{\star}=\sum_k\widehat{\alpha}_{j,k}\varphi_{j,k}$.

3.4 Asymptotic properties

In this section, we state asymptotic results for the estimator described above. In order to derive rates of convergence of $\widehat{\mu}$, we assume the following conditions in addition to (A1). Let E and V denote the expectation and variance.

(A2). There is a positive constant M_1 such that $M_1^{-1}\le f(x)\le M_1$ for $x\in I_a$;

(A3). $\sigma^2(x)\le M_2$ for $x\in I_a$ where $\sigma^2(x)=V(Z\mid X=x)$ is the conditional variance of Z given $X=x$;

(A4). $|\Phi_W(t)|\ge M_3(1+|t|)^{-d}$ for a nonnegative number d;

(A5). $r > d+1$.

Remark 3.1. There are two examples of the conditional distribution of U given $X=x$ satisfying (A3): (i) Normal distribution with mean $\mu(x)$ and finite variance $\sigma^2(x)$, and (ii) Poisson distribution with finite $\mu(x)$.

Remark 3.2. The contaminating error W is said to be ordinary smooth of order $d$according to Fan and Truong (1993). Ordinary smooth distributions include gamma and double exponential distributions. The order d determines the rate of convergence. The larger d becomes, the harder is the problem of estimating the regression function

Remark 3.3. The condition (A5) is necessary for the formal definition in (3.2) and (3.5).

Now we state the main results whose proofs are given in Section 4. Let J denote the number of elements in the set K_j. It can be noted that $J \asymp 2^j$, where for two sequences of positive numbers $a_n \asymp b_n$ means that $C^{-1} \leq a_n/b_n \leq C$

The following theorem gives an upper bound on the L_2 rate of convergence for the estimator $\hat{\mu}$. We can note that the rate of convergence depends on the smoothness s of f and γ in addition to the smoothness d of the error distribution.

Theorem 1 Under (A1)-(A5), if f and γ belong to $B_{spq}(M)$ with $s > 1/2$, $p \geq 2$ and $q \geq 1$, then

$$\|\mu_\star - \mu\|_2 = O_p(n^{-s/(2s+2d+1)})$$

for $J \asymp n^{1/(2s+2d+1)}$.

4. Proofs

In this section, we provide the proofs of the asymptotic results presented in Section 3.

Using the characterization of Besov spaces and the Sobolev embeddings, it can be seen that

$$sup_{x\in I_a}|f(x)| \leq C,\ \ sup_{x\in I_a}|\gamma(x)| \leq C \tag{4.1}$$

and

$$\|E_j f - f\|_2 = O(J^{-s}),\ \ \|E_j\gamma - \gamma\|_2 = O(J^{-s}). \tag{4.2}$$

Let g be the density function of Y which is given by $\int f(y-x)dF_W(x)$. Since the convolution operator is an averaging operator, we have that

$$|g(y)| \leq C$$

by (4.1)

Since $\alpha_{j,k} = E(K^j\varphi)_{j,k}(Y)$ we have, by Parseval's identity,

$$\begin{aligned} E(\hat{\alpha}_{j,k} - \alpha_{j,k})^2 &= \frac{1}{n} V((K^j\varphi)_{j,k}(Y)) \\ &\leq \frac{1}{n}\int |(K^j\varphi)_{j,k}(y)|^2 g(y)dy \\ &= \frac{1}{n}\int |(K^j\varphi(y)|^2 g((y+k)/2^j)dy \\ &\leq \frac{C}{n}\int |(K^j\varphi)(y)|^2 dy \\ &= \frac{C}{n}\int \left|\frac{\hat{\varphi}(t)}{\Phi_W(2^j t)}\right| dt \end{aligned}$$

It follows from (A5) that

$$|\Phi_W(2^j t)| - 1 = O(J^d(1+|t|))$$

and

$$\int (1+|t|)^{2d} |\hat{\varphi}(t)|^2 dt = O(1).$$

Therefore, we can obtain

$$\sum_{x \in K} (\hat{\alpha}_{j,k} - \alpha_{j,k})^2 = O_P(J^{2d+1}/n). \tag{4.3}$$

Let $\tau^2(x) = E(U^2|X=x)$. Since σ^2 and μ are bounded on I_a, $\tau^2(x)$ is bounded for $x \in I_a$. By independence of W and (X, U), we have

$$\begin{aligned} E(\hat{\beta}_{j,k} - \beta_{j,k})^2 &\leq \frac{1}{n} E\left|(K^j\varphi)_{j,k}(X+W)\right|^2 \tau^2(X) \\ &\leq \frac{1}{n} \sup \tau^2(x) E\left|(K^j\varphi)_{j,k}(Y)\right|^2 \\ &= O(J^{2d}/n) \end{aligned}$$

Now choose J such that $J \asymp n^{1/(2s+2d+1)}$. It follows from (4.2) and (4.3) that

$$\begin{aligned} \|\hat{f} - f\|_2^2 &\leq 2\|\hat{f} - E_J f\|_2^2 + 2\|E_J f - f\|_2^2 \\ &= O\Big(\sum_{k \in K_j} (\hat{\alpha}_{j,k} - \alpha_{j,k})^2\Big) + O(J^{-2s}) \\ &= O_P(J^{2d+1}/n) + O(J^{-2s}) \\ &= O_P(n^{-2s/(2s+2d+1)}) \end{aligned} \tag{4.4}$$

Similarly, we have

$$\|\gamma^{\star}-\gamma\|_2 = O_P(n^{-s/(2s+2d+1)}) \tag{4.5}$$

By Taylor expansion, we have

$$\begin{aligned}\mu^{\star}-\mu &= f^{-1}\gamma^{\star}(1+f^{-1}(f^{\star}-f))^{-1}-\mu \\ &= f^{-1}(\gamma^{\star}-\gamma)-f^{-1}\mu(f^{\star}-f) \\ &\quad -f^{-1}(\gamma^{\star}f^{-1}-\mu)(f^{\star}-f) \\ &\quad +\frac{(1+\xi)^3}{4}(f^{-1}(f^{\star}-f))^2,\end{aligned} \tag{4.6}$$

where $|\xi|\leq|f^{-1}(f^{\star}-f)|$. It follows from (4.4) and (4.5) that

$$\|f^{-1}(f^{\star}-f)\|_{\infty}=O(\|f^{\star}-f\|_{\infty})=O(J^{1/2}\|f^{\star}-f\|_2)=o_P(1) \tag{4.7}$$

and

$$\|\gamma^{\star}f^{-1}-\mu\|_{\infty}=O(\|\gamma^{\star}-\gamma\|_{\infty})=O(J^{1/2}\|\gamma^{\star}-\gamma\|_2)=o_P(1) \tag{4.8}$$

Combining (4.6)-(4.8), we get that

$$\|\mu^{\star}-\mu\|_2=(1+o_P(1)\;O(\|\gamma^{\star}-\gamma\|_2+\|f^{\star}-f\|_2$$

By (4.4) and (4.5) we have the desired result.

References

[1] Bickel, P. J. and Ritov, Y. (1987). Efficient estimation in the errors invariables model. *The Annals of Statistics,* 15, 513-540.

[2] Carroll, R. J., Ruppert, D. and Stefanski, L. A. (1995). *Measurement Error in Nonlinear Models.* Chapman and Hall, London.

[3] Daubechies, I. (1992). *Ten Lectures on Wavelets.* SIAM, Philadelphia.

[4] DeVore, R. and Popov, V. (1988). Interpolation of Besov spaces. *American Mathematical Society.* 305, 397-414.

[5] Donoho, D. L. and Johnstone, I. M. (1994). Ideal spatial adaptation via wavelet shrinkage. *Boimetrika,* 81, 425-455.

[6] Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D. (1996). Density estimation by wavelet thresholding. *The Annals of Statistics,* 24, 508-539.

[7] Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *The Annals of Statistics,* 19, 1257-1272.

[8] Fan, J. and Koo, J. -Y. (1999). Wavelet deconvolution. Manuscript.

[9] Fan, J. and Truong, Y. K. (1993). Nonparametric regression with errors in variables. *The Annals of Statistics,* 21, 1900-1925.

[10] Fuller, W. A. (1987). *Measurement Error Models.* (Wiley, New York).

[11] Koo, J. -Y. and Kim, W. -C. (1996). Wavelet density estimation by approximation of log-densities. *Statistics and Probability Letters,* 26, 271-278.

[12] Koo, J. -Y. and Lee, K. -W. (1998). B-spline estimation of regression functions with errors in variable. *Statistics and Probability Letters,* 40, 57-66.

[13] Stefanski, L. A. (1985). The effects of measurement error on parameter estimation. *Biometrika,* 74, 385-391.

[14] Stefanski, L. A. and Carroll, R. J. (1985). Covariate measurement error in logistic regression. *The Annals of Statistics,* 13,1335-1351.

[15] Stefanski, L. A. and Carroll, R. J. (1991). Deconvolution-based score tests in measurement error models. *The Annals of Statistics,* 19, 249-259.