应用概率统计 第二十五卷
第四期 2009年8月

Chinese Journal of Applied Probability
and Statistics Vol.25 No.4 Aug. 2009

Wavelet Estimation in Varying-Coefficient Models *

Lu Yiqiang Li Zhilin

(*Institute of Electronic Technology, the PLA Information Engineering University, Zhengzhou, 450004*)

Abstract

This paper is concerned with the estimation of varying-coefficient model that is frequently used in statistical modeling. The wavelet procedures are developed to estimate the coefficient functions. The advantage of this approach is to avoid the restrictive smoothness requirement for nonparametric function of the traditional smoothing approaches for varying-coefficient model, such as kernel and local polynomial methods. Furthermore, the convergence rate of the wavelet estimators is derived and the asymptotic normality is established. Finite sample properties are studied through Monte Carlo simulations.

Keywords: Varying-coefficient models, wavelet, least-square estimation, asymptotic normality, convergence rate.

AMS Subject Classification: Primary 62H12; secondary 62A10.

§1. Introduction

In recent years, many useful data-analytic modeling techniques have been proposed to relax traditional parametric models and to exploit possible hidden structure. For an introduction to these methods, see books by Hastie and Tibshirani (1990), Green and Silverman (1994) and Fan and Gijbels (1996), among others. In dealing with high-dimensional data, many powerful approaches have been incorporated to avoid so-called "curse of dimensionality". Examples include additive modeling (Hastie and Tibshirani, 1990), low-dimensional interaction modeling (Friedman, 1991; Gu and Wahba, 1993; Stone et al., 1997), multiple-index models (Hardle and Stoker, 1989), partially linear models (Green and Silverman, 1994) and their hybrids (Carroll et al., 1997). An important alternative to the additive and other models is the varying-coefficient model, proposed by Hastie and Tishirani (1993), in which the coefficients of classical linear models are replaced by nonparametric functions and hence the regression coefficients are allow to vary as functions of other factors.

The varying-coefficient models is defined as

$$y_i = x_i^\tau \beta(t_i) + \epsilon_i, \tag{1.1}$$

*The project supported by the National Natural Science Foundation of China (10501053).
Received November 9, 2006. Revised April 9, 2007.

where y_i's are responses; $x_i = (x_{i1}, \cdots, x_{ip})^\tau$ and t_i are design points; $\beta(t) = (\beta_1(t), \cdots, \beta_p(t))^\tau$ is p-dimensional vector of unknown functions; ϵ_i's are errors with mean 0 and variance σ^2 and superscript 'τ' denotes the transpose of a vector or matrix.

The varying-coefficient models are simple and useful extension of classical linear models. The appeal of the model (1.1) is that via allowing coefficient $\beta_1, \cdots, \beta_p$ to depend on t, the modeling bias can significantly be reduced and "curse of dimensionality" can be avoided. Another advantage of this model is its interpretability. It is well-recognized that the model (1.1) has extremely wide applications, For example, see Wu et al. (1998) for details on novel applications of varying-coefficient models to longitudinal data. For nonlinear time series applications, see Chen and Tsay (1993) and Cai, Fan and Yao (2000).

Many methods were developed to estimate the coefficient functions in (1.1). These methods were all based on some nonparametric regression techniques, such as local polynomial fitting (Cai, Fan and Li, 2000), kernel methods (Wu et al., 1998), smoothing splines estimation (Chiang et al., 2001) and B-spline approximation (Lu and Mao, 2004), and so on.

In this paper, the wavelet procedure, which was used to estimate the nonparametric curve by Antoniadis et al. (1994), is applied to the varying-coefficient models. The wavelet estimators of coefficient functions are constructed and their large sample properties are derived. The main reason for adopting the wavelet approaches for the varying-coefficient models is that an important assumption by all the existing approaches for coefficient functions $\beta_j(t)$'s is their high smoothness. But in reality, the assumption may not be satisfied. In some practical areas, such as signal and image processing, objects are frequently inhomogeneous. For wavelet approach it is well known that the hypotheses of degrees of smoothness of the underlying function is less restrictive. Due to these abilities to adapt to local features of curves, many authors have also applied wavelet procedures to estimate nonparametric models. See for example recent works by Antoniadis et al. (1994), Donoho and Johnstone (1994), Hall and Patil (1995), Liang et al. (1999), Qian and Cai (1999), Zhou and Yao (2004), and so on.

The paper is organized as follows. In Section 2, the wavelet estimation procedures are introduced. Section 3 establishes the main results. The proofs of the main results are provided in Section 4. In Section 5, some finite sample properties of wavelet and local linear estimators are studied via some simulation examples.

§2. Wavelet Estimation of Varying-Coefficient Models

Suppose that we have a sample $\{y_i, x_i, t_i\}_{i=1}^n$ from model (1.1). For simplicity, throughout this paper we will assume that the design points $\{t_i\}_{i=1}^n$ are fixed, while

$\{x_i\}_{i=1}^n$ are i.i.d. random sample from x. But, by replacing the expectation $\mathsf{E}(\cdot)$ with the conditional expectation $\mathsf{E}(\cdot|t)$, our results can be easily extended to the situation in that $\{t_i\}_{i=1}^n$ are random and x can depend on t. Without loss of generality, let $t \in [0,1]$. The wavelet technique is applied to estimate the coefficient functions in model (1.1). The detailed procedure is summarized below.

For convenience, we first introduce some symbols and definitions along the line Antoniadis et al. (1994). (For more information of wavelet analysis, see Vidakovic, 1999). Suppose that $\phi(\cdot)$ is given scaling function in Schwarz space with order l. A multiresolution analysis of $L^2(R)$ consists of an increasing sequence of closed subspace $\{V_m\}$, $m = \cdots, -2, -1, 0, 1, 2, \cdots$, where $L^2(R)$ is the set of square integral functions over real line. The associated integral kernel of V_m is given by

$$E_m(t,s) = 2^m \sum_{k\in Z} \phi(2^m t - k)\phi(2^m s - k),$$

where Z denotes the set of integers. The projection of $h(t)$ onto V_m is $\displaystyle\int_R E_m(\cdot,t)h(t)\mathrm{d}t$. Let $A_i = [s_{i-1}, s_i]$ be a partition of the interval $[0,1]$ with $t_i \in A_i$. Then the wavelet estimator of $\beta(t) = (\beta_1(t), \cdots, \beta_p(t))^\tau$ is the solution of minimizing the following weighted local least-squares equation

$$\sum_{i=1}^n (y_i - \sum_{j=1}^p x_{ij} b_j)^2 \int_{A_i} E_m(t,s)\mathrm{d}s$$

with respect to $b = (b_1, \cdots, b_p)^\tau$.

Let $x_i = (x_{i1}, \cdots, x_{ip})^\tau$, $X = (x_1, \cdots, x_n)^\tau$, $Y = (y_1, \cdots, y_n)^\tau$ and

$$W(t) = \mathrm{diag}\Big(\int_{A_1} E_m(t,s)\mathrm{d}s, \cdots, \int_{A_n} E_m(t,s)\mathrm{d}s\Big).$$

The estimator of of $\beta(t) = (\beta_1(t), \cdots, \beta_p(t))^\tau$ is given by

$$\widehat{\beta}(t) = (\widehat{\beta}_1(t), \cdots, \widehat{\beta}_p(t))^\tau = (X^\tau W(t) X)^{-1} X^\tau W(t) Y. \tag{2.1}$$

In the above wavelet estimator m act as a tuning parameter, such as the bandwidth does for standard kernel smoothers. A key aspect of wavelet estimators is that the tuning parameter ranges over a much more limited set of values than is common with other nonparametric regression techniques. In practice only a small number of values of m (say three or four) need to be considered. The optimal m can be selected by cross-validation procedure.

The weight $\displaystyle\int_{A_i} E_m(t,s)\mathrm{d}s$ can be calculated by the cascade algorithm given by Antoniadis et al. (1994). Thus, the wavelet estimator can be easily calculated.

§3. Asymptotic Properties of the Proposed Estimators

We begin with the following assumptions required to derive the large sample properties of the proposed estimators in Section 2.

A1. $x_1, \cdots, x_n$ are i.i.d. samples from x and $\mathsf{E}(xx^\tau)$ is no-singular. ϵ_i's are i.i.d random errors with mean 0 and variance σ^2. Furthermore, suppose that each element of xx^τ and $x\epsilon$ has finite $(3+\delta)$th moments $(\delta > 0)$.

A2. $\beta_j(\cdot)'s$ belong to Sobolev space with order $\upsilon > 1/2$.

A3. $\beta_j(\cdot)'s$ satisfy the Lipschitz of order condition of order $\gamma > 0$.

A4. $\phi(\cdot)$ is in the Schwarz space with order $l \geq \upsilon$, satisfies the Lipschtz condition with order l and has a compact support. Furthermore, $\widehat{\phi}(\xi) - 1 = O(\xi)$ as $\xi \to 0$, where $\widehat{\phi}(\cdot)$ is the Fourier transform of $\phi(\cdot)$.

A5. $\max(|s_i - s_{i-1}|) = O(1/n)$.

A6. Furthermore, we also assume that for some Lipschitz function $k(\cdot)$,

$$\rho(n) = \max\left|s_i - s_{i-1} - \frac{k(s_i)}{n}\right| = o(n^{-1}).$$

The above conditions are mild and easily satisfied. The similar conditions are assumed in the nonparametric regression of Antoniadis et al. (1994). In addition, since the functions belonging to Sobolev space with order $\upsilon > 3/2$ are continuously differential, A3 is redundant when $\upsilon > 3/2$ and A2 is weaker than smoothness.

Now let's establish the large sample properties of the estimators described in the Section 2.

Theorem 3.1 Suppose that Assumption A1-A5 hold. Then for any $t \in (0,1)$

$$\max_{1\leq j\leq p} |\widehat{\beta}_j(t) - \beta_j(t)| = O(n^{-1/3}\log(n)) \qquad \text{a.s.} \tag{3.1}$$

provided $\upsilon > 3/2$, $\gamma > 1/3$ and $2^m = O(n^{1/3})$.

Remark 1 When $t = t_i$ is the design point, (3.1) admits the result of Theorem 3.3 of Zhou and You (2004). Theorem 3.1 gives the convergence rate of the estimators $\widehat{\beta}_j(t)$'s for any $t \in (0,1)$. Theorem 3.1 also states that the convergence rate of wavelet estimators of coefficient functions is comparable with the optimal convergence rate of the nonparametric estimation in nonparametric models.

To obtain asymptotic normality result, we need to consider an approximation to $\widehat{\beta}_j(t)$ based on its values at dyadic points of order m. That is, define

$$\widehat{\beta}_j^d(t)=\widehat{\beta}_j(t^{(m)}),$$

where $t^{(m)}=[2^m t]/2^m$. Furthermore, let $\widehat{\beta}^d(t)=(\widehat{\beta}_1^d(t),\cdots,\widehat{\beta}_p^d(t))^\tau$. The $\widehat{\beta}_j^d(t)$ is piecewise-constant approximation to $\widehat{\beta}_j(t)$ at resolution 2^{-m}. The reason that $\widehat{\beta}_j(t)$ is approximated by $\widehat{\beta}_j^d(t)$ is that the variance of $\widehat{\beta}_j(t)$ is stable at the dyadic points but fails to converge at non-dyadic points. The detail discussion see Antoniadis et al. (1994).

Theorem 3.2 Suppose that Assumption A1-A6 hold and $\mathsf{E}(\|x\|^4)<\infty$, where $\|\cdot\|$ denotes Euclidean norm. Let $\upsilon^*=\min(3/2,\upsilon,\gamma+1/2)-\epsilon$ and $\epsilon=0$ for $\upsilon\neq 3/2$, $\epsilon>0$ for $\upsilon=3/2$. If $n2^{-m}\to\infty$ and $n2^{-2m\upsilon^*}\to 0$, then

$$\sqrt{n2^{-m}}(\widehat{\beta}^d(t)-\beta(t))\stackrel{D}{\longrightarrow}\mathrm{N}(0,[\mathsf{E}(xx^\tau)]^{-1}\sigma^2k(t)\omega_0^2), \tag{3.2}$$

where $\omega_0^2=\sum\limits_{k\in Z}\phi^2(k)$ and Z is the set of integers.

Remark 2 According to Wu et al. (1998), continuous second derivative of $\beta_j(\cdot)$ are required to achieve the asymptotic normality. In our result, however, it suffices for $\beta_j(\cdot)$'s to belong Sobolev space with order $\upsilon>1/2$. This highlights the power of wavelet procedures.

Remark 3 To use the above asymptotic normality result to obtain confidence intervals for $\beta_j(t)$ at given t. One needs to consistently estimate $\mathsf{E}(xx^\tau)$ and the noise variance σ^2. $\mathsf{E}(xx^\tau)$ can be estimated by $(1/n)\sum\limits_{i=1}^n x_i x_i^\tau$ and σ^2 can be estimated as follows:

$$\widehat{\sigma}^2=\frac{1}{n}\sum_{i=1}^n(y_i-\widehat{y}_i)^2,$$

where $\widehat{y}_i=x_i^\tau\widehat{\beta}(t_i)$.

§4. Proofs of the Main Results

In order to prove the main results we first present several lemmas.

Lemma 4.1 Suppose that Assumption A4 holds, Then we have

(a) $E_0(t,s)\le c_k/(1+|t-s|)^k$ and $E_m(t,s)\le 2^m c_k/(1+2^m|t-s|)^k$, where k is a positive integer and c_k is a constant depending on k only.

(b) $\sup\limits_{0\le t,s\le 1}|E_m(t,s)|=O(2^m)$.

(c) $\sup_{0\le t\le 1}\int_0^1 E_m(t,s)\mathrm{d}s \le c$, where c is a positive constant.

The proofs can be found in Antoniadis et.al (1994).

Lemma 4.2 Suppose that Assumption A4 and A5 hold and $r(t)$ satisfies Assumption A2 and A3. Then

$$\sup_{0\le t\le 1}\Big|r(t)-\sum_{i=1}^{n} r(t_i)\int_{A_i} E_m(t,s)\mathrm{d}s\Big| = O(n^{-\gamma})+O(\eta_m),$$

where

$$\eta_m = \begin{cases} (1/2^m)^{\gamma-1/2} & \text{if } 1/2<\upsilon<3/2,\\ \sqrt{m}/2^m & \text{if } \upsilon=3/2,\\ 1/2^m & \text{if } \upsilon>3/2. \end{cases}$$

Lemma 4.2 follows easily from Theorem 3.2 in Antoniadis et al. (1994).

Lemma 4.3 Let $\{V_i, i=1,\cdots,n\}$ be a sequence of independent random variables with mean zero and finite $(2+\delta)$th moments, and $\{a_{ij}, i,j=1,\cdots,n\}$ a set of positive numbers such that $\max_{i,j}|a_{ij}|\le n^{-p_1}$ for some $0\le p_1\le 1$ and $\sum_i a_{ij}=O(n^{p_2})$ for some $p_2\ge \max\{0, 2/(2+\delta)-p_1\}$. Then

$$\max_{1\le j\le n}\Big|\sum_i a_{ij}V_i\Big| = O(n^{-(p_1-p_2)/2}\log n) \qquad \text{a.s..}$$

The proof Lemma 4.3 can be found in Härdle et al. (2000) and Zhou and You (2004).

Proof of Theorem 3.1 From (2.1), we have

$$\begin{aligned} &\widehat{\beta}(t)-\beta(t)\\ =\ &(X^\tau W(t)X)^{-1}X^\tau W(t)Y-\beta(t)\\ =\ &\Big[\sum_{i=1}^{n} x_i x_i^\tau w_i(t)\Big]^{-1}\sum_{i=1}^{n} x_i w_i(t)(x_i^\tau\beta(t_i)-x_i^\tau\beta(t))+\Big[\sum_{i=1}^{n} x_i x_i^\tau w_i(t)\Big]^{-1}\sum_{i=1}^{n} x_i w_i(t)\epsilon_i, \quad (4.1)\end{aligned}$$

where $w_i(t)=\int_{A_i}E_m(t,s)\mathrm{d}s$. Let $a_{ij}=w_i(t)$, $j=1,2,\cdots,n$. From Lemma 4.1, we have

$$\max_{i,j}|a_{ij}| = \max_i\Big|\int_{A_i}E_m(t,s)\mathrm{d}s\Big| = O(2^m/n)=O(n^{-2/3})=O(n^{-p_1})$$

and

$$\sum_i a_{ij}=\sum_i w_i(t)=\int_0^1 E_m(t,s)\mathrm{d}s = O(1)=O(n^{p_2}),$$

where $p_1=2/3$ and $p_2=0$. (If $p_1=2/3$ and $p_2=0$, it suffices that δ in Lemma 4.3 is greater than or equal to 1. This can be satisfied by Condition A.1). Hence, by Lemma

4.2 and 4.3, we have

$$\begin{aligned}\sum_{i=1}^{n} x_i x_i^\tau w_i(t) &= \sum_{i=1}^{n}(x_i x_i^\tau - \mathsf{E}(x_1 x_1^\tau))w_i(t) + \Big(\sum_{i=1}^{n} w_i(t) - 1\Big)\mathsf{E}(x_1 x_1^\tau) + \mathsf{E}(x_1 x_1^\tau) \\ &= O(n^{-1/3}\log n) + O(n^{-\gamma}) + O(\eta_m) + \mathsf{E}(x_1 x_1^\tau) \qquad \text{a.s.,} \qquad (4.2)\end{aligned}$$

where $\sum_{i=1}^{n}(x_i x_i^\tau - \mathsf{E}(x_1 x_1^\tau))w_i(t) = O(n^{-1/3}\log n)$, which means that each element of $\sum_{i=1}^{n}(x_i x_i^\tau - \mathsf{E}(x_1 x_1^\tau))w_i(t)$ is $O(n^{-1/3}\log n)$. Similarly, by Lemma 4.2, we have

$$\begin{aligned}&\sum_{i=1}^{n}(x_i w_i(t) x_i^\tau)(\beta(t_i) - \beta(t)) \\ =\ & \sum_{i=1}^{n}(x_i x_i^\tau - \mathsf{E}(x_1 x_1^\tau))(\beta(t_i) - \beta(t))w_i(t) + \sum_{i=1}^{n}\mathsf{E}(x_1 x_1^\tau)(\beta(t_i) - \beta(t))w_i(t) \\ =\ & O(n^{-1/3}\log n) + O(n^{-\gamma}) + O(\eta_m) \qquad \text{a.s..} \qquad (4.3)\end{aligned}$$

Moreover, Lemma 4.3 also lead to

$$\sum_{i=1}^{n} x_i w_i \epsilon_i = O(n^{-1/3}\log n) \qquad \text{a.s..} \qquad (4.4)$$

From (4.1)–(4.4),

$$\widehat{\beta}(t) - \beta(t) = O(n^{-1/3}\log n) \qquad \text{a.s..}$$

This completes the proof. □

Proof of Theorem 3.2 From the wavelet estimator (2.1) of $\beta(t)$, it is easy to see

$$\widehat{\beta}(t) - \beta(t) = \Big(\sum_{i=1}^{n} x_i x_i^\tau w_i(t)\Big)^{-1}\sum_{i=1}^{n} x_i w_i(t)(y_i - x_i^\tau \beta(t)). \qquad (4.5)$$

By Theorem 3.2 and 3.3 in Antoniadis et al. (1994), we have

$$\sum_{i=1}^{n} x_i x_i^\tau w_i(t) = \mathsf{E}(x_1 x_1^\tau)(1 + o_p(1)). \qquad (4.6)$$

Let

$$\begin{aligned}\sum_{i=1}^{n} x_i w_i(t)(y_i - x_i^\tau \beta(t)) &= \sum_{i=1}^{n} x_i w_i(t) x_i^\tau(\beta(t_i) - \beta(t)) + \sum_{i=1}^{n} x_i w_i(t)\epsilon_i \\ &= K_n(t) + L_n(t), \qquad (4.7)\end{aligned}$$

where $K_n(t) = \sum_{i=1}^{n} x_i w_i(t) x_i^\tau(\beta(t_i) - \beta(t))$ and $L_n(t) = \sum_{i=1}^{n} x_i w_i(t)\epsilon_i$. By Lemma 4.2, we have

$$\begin{aligned}\mathsf{E}K_n(t) &= \sum_{i=1}^{n}\mathsf{E}(x_i x_i^\tau)(\beta(t_i) - \beta(t))w_i(t) \\ &= \mathsf{E}(x_1 x_1^\tau)(O(n^{-\gamma}) + O(\eta_m)). \qquad (4.8)\end{aligned}$$

Let $K_{nl}(t)$ denote the lth element of $K_n(t)$. The variance of $K_{nl}(t)$ can be written as

$$\begin{aligned}
&\mathsf{Var}\Big(\sum_{i=1}^{n} x_{il}w_i(t)x_i^{\tau}(\beta(t_i)-\beta(t))\Big)\\
=&\sum_{i=1}^{n} w_i^2(t)(\beta(t_i)-\beta(t))^{\tau}\mathsf{Cov}\,(x_ix_{il},x_i^{\tau}x_{il})(\beta(t_i)-\beta(t))\\
\le&\Big(\sum_{i=1}^{n} w_i^3(t)\Big)^{1/2}\Big\{\sum_{i=1}^{n} w_i(t)[(\beta(t_i)-\beta(t))^{\tau}C_{il}(\beta(t_i)-\beta(t))]^2\Big\}^{1/2}, \qquad (4.9)
\end{aligned}$$

where $C_{il}=\mathsf{Cov}\,(x_ix_{il},x_i^{\tau}x_{il})$. Note that

$$\sum_{i=1}^{n} w_i^3(t)=\sum_{i=1}^{n}\Big[\int_{A_i}E_m(t,s)\mathrm{d}s\Big]^3=\sum_{i=1}^{n}E_m^3(t,u_i)(s_i-s_{i-1})^3,$$

where $u_i\in A_i$. The number of terms contributing to the above sum is order $O(2^{-m})$. Hence, using the bound $\sup\limits_{t,s}E_m^3(t,s)\le 2^{3m}$,

$$\sum_{i=1}^{n} w_i^3(t)\le O(2^{2m}n^{-2}). \qquad (4.10)$$

By Lemma 4.2, we have

$$\sum_{i=1}^{n} w_i(t)[(\beta(t_i)-\beta(t))^{\tau}C_{il}(\beta(t_i)-\beta(t))]^2=O(n^{-\gamma})+O(\eta_m). \qquad (4.11)$$

From (4.10) and (4.11), (4.9) is bounded by

$$O(n^{-1}2^m)\{O(\eta_m)+O(n^{-\gamma})\}^{1/2}. \qquad (4.12)$$

By (4.8) and (4.12), we have

$$\mathsf{E}K_{nl}^2(t)=O(n^{-1}2^m)(O(\eta_m)+O(n^{-\gamma}))^{1/2}+(O(\eta_m)+O(n^{-\gamma}))^2.$$

Hence, from $n2^{-2mv^*}\to 0$,

$$\sqrt{n2^{-m}}K_n=o_p(1). \qquad (4.13)$$

Now let's verify the asymptotic normality of $L_n(t^{(m)})$. From the proof of Theorem 3.3 of Antoniadis et al. (1994), we have

$$\mathsf{Cov}\,(L_n(t^{(m)}))=\mathsf{E}x_1x_1^{\tau}\{n^{-1}2^m\sigma^2k(t)(\omega_0^2+o(1))+O(2^m\rho(n))+O(2^{2m}/n^2)\}. \qquad (4.14)$$

Note that $\mathsf{E}L_n(t^{(m)})=0$. By the Cramer-Wold Theorem, to derive the asymptotic normality of $L_n(t^{(m)})$, it suffices to show that for any unit vector $d\in R^p$,

$$d^{\tau}L_n(t^{(m)})/(d^{\tau}\mathsf{Cov}\,(L_n(t^{(m)}))d)^{1/2}\xrightarrow{D}\mathrm{N}(0,1). \qquad (4.15)$$

Note that

$$d^\tau L_n(t^{(m)}) = \sum_{i=1}^n w_i(t^{(m)}) d^\tau x_i \epsilon_i,$$

we shall appeal to a central limit theorem for weighted sums of i.i.d. random variables to obtain (4.15). To complete the proof, we need to check

$$\frac{\max\limits_{1\le i\le n} |w_i^2(t^{(m)})|}{d^\tau \mathsf{Cov}\,(L_n(t^{(m)}))d} \to 0.$$

From (4.14), we have

$$n2^{-m} d^\tau \mathsf{Cov}\,(L_n(t^{(m)}))d = d^\tau \mathsf{E}(x_1 x_1^\tau) d\sigma^2 k(t)\omega_0 + o(1). \tag{4.16}$$

Also using $\max\limits_{1\le i\le n} |w_i(t)|^2 = O(2^{2m}/n^2)$, we have

$$\frac{\max\limits_{1\le i\le n} |w_i^2(t^{(m)})|}{d^\tau \mathsf{Cov}\,(L_n(t^{(m)}))d} = \frac{O(2^{2m}/n^2)}{O(2^m/n)} = o(1).$$

Hence (4.15) holds. That is

$$\sqrt{n2^{-m}} L_n(t^{(m)}) \xrightarrow{D} \mathrm{N}(0, \mathsf{E}(xx^\tau)\sigma^2 k(t)\omega_0^2). \tag{4.17}$$

By(4.5), (4.6), (4.7), (4.13) and (4.17), Theorem 3.2 holds. □

§5. Numerical Simulations

In this section, we use the following examples to illustrate the performance of the wavelet estimators of varying coefficient models.

Example 1 $Y = (-1 + 2\sin(\pi t/60))X_1 + (1 - 2\cos(\pi(t-25)/15))X_2 + \epsilon.$

Example 2 $Y = (-1 + 2\sin(\pi t/2))X_1 + 9\beta(t)X_2 + \epsilon$, where

$$\beta(t) = \begin{cases} 4t^2(3-4t) & \text{if } 0 \le t \le 0.5 \\ (4/3)\cdot t(4t^2 - 10t + 7) - 1.5 & \text{if } 0.5 < t \le 0.75 \\ (16/3)\cdot t(t-1)^2 & \text{if } 0.75 < t \le 1 \end{cases}$$

and $\beta(t)$ is a piecewise polynomial with discontinuity. t follows a uniform distribution on $[0,1]$ and X_1 and X_2 are normally distributed with correlation coefficient $2^{-1/2}$. Further the marginal distributions of X_1 and X_2 are standard normal and ϵ, t and (X_1, X_2) are independent. The random ϵ follows a normal distribution with mean zero and variance σ^2. The σ^2 is chosen so that the signal to noise is about $5:1$, namely

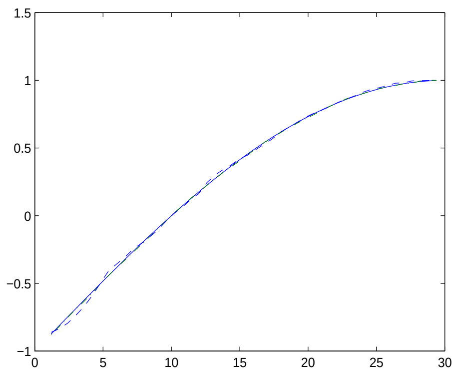
$$\sigma^2 = 0.2\mathsf{Var}\,(m(t, X_1, X_2)) \quad \text{with} \quad m(t, X_1, X_2) = \mathsf{E}(Y|t, X_1, X_2).$$

For each of the above examples, we conducted 100 simulations with sample size $n =$ 100 and 200. The functional coefficients were estimated respectively by the wavelet method and local linear method. The performances of the estimators were evaluated in terms of mean absolute deviation error (MADE), which is defined as

$$\mathrm{MADE}_j = \frac{1}{n}\sum_{i=1}^{n} |\widehat{\beta}_j(t_i) - \beta_j(t_i)|.$$

The tuning parameter m in the wavelet estimation and h in the local linear estimation were chosen by 'leave-one-out' cross-validation procedures. Figure 1 depicts the actual functions and their estimated curves, which attain median performance among 100 simulations with sample size $n = 200$. Table 1 summaries the simulation results with the mean and the standard deviation of MADE_j based on 100 simulations with sample size $n = 100$ and 200.

Example 1: A typical result



Example 2: A typical result

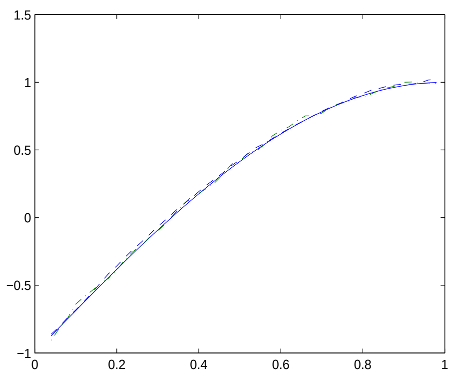


Figure 1 Comparisons of the performance between the wavelet method and local linear method. Solid curve — true functions; dashed curves — estimates based on the wavelet method; dashdot curves — estimates based on the local linear method.

Table 1 The mean (standard deviation) of MADE of the functional coefficient estimators

		Example 1		Example 2	
Size	Method	MADE_1	MADE_2	MADE_1	MADE_2
$n = 100$	Wavelet	0.0449(0.0135)	0.2908(0.0782)	0.1968(0.0613)	0.3234(0.0669)
	Local linear	0.0440(0.0116)	0.2445(0.0674)	0.1999(0.0475)	0.3541(0.0555)
$n = 200$	Wavelet	0.0305(0.0089)	0.1898(0.0465)	0.1502(0.0411)	0.2561(0.0440)
	Local linear	0.0302(0.0081)	0.1752(0.0430)	0.1508(0.0333)	0.2778(0.0322)

From Figure 1, we see that both the wavelet method and local linear method gave fairly good estimates for the functional coefficients in Example 1 and 2. Reduction of the mean of MADE with growing sample size is clearly identified from Table 1. Table 1 also show that, for the irregular functional coefficient like $\beta_2(t)$ in the Example 2, the wavelet method slightly outperforms the local linear method in terms of the observed MADE, but for smoothing functional coefficient the local linear slightly outperforms the wavelet method.

References

[1] Antoniadis, A., Gredoire, G. and Mackeague, W.I., Wavelet methods for curve estimation, *J. Amer. Statist. Assoc.*, **89**(1994), 1340–1353.

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

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

[4] Carroll, R.J., Fan, J., Gijbels, I. and Wand, M.P., Generalized partially linear single-index models, *J. Amer. Statist. Assoc.*, **92**(1997), 477-489.

[5] Chen, R. and Tsay, R.S., Functional-coefficient autoregressive models, *J. Amer. Statist. Assoc.*, **88**(1993), 298–308.

[6] Chiang, C., Rice, J.A. and Wu, C.O., Smoothing spline estimation for varying coefficient models with repeatedly measure dependent variables, *J. Amer. Statist. Assoc.*, **96**(2001), 605–619.

[7] Donoho, D.L. and Johnstone, I.M., Ideal spatial adaption by wavelet shrinkage, *Biometrika*, **81**(1994), 425–455.

[8] Fan, J. and Gijbels, I., *Local Polynomial Modeling and Its Application*, London: Chapman and Hall, 1996.

[9] Friedman, J.H., Multivariate adaptive regression splines (with discussion), *Ann. Statist.*, **19**(1991), 1–141.

[10] Green, P.J. and Silverman, B.W., *Nonparametric Regression and Generalized Linear Models: a Roughness Penalty Approach*, London: Chapman and Hall, 1994.

[11] Gu, C. and Wahba, G., Smoothing spline multiple ANOVA with component-wise Bayesian "confidence intervals", *J. Comput. Graph. Statist.*, **2**(1993), 97–117.

[12] Hall, P. and Patil, P., On wavelet methods for estimating smooth function, *Bernoulli*, **1**(1995), 41–58.

[13] Härdle, W., Liang, H. and Gao, J.T., *Partially Linear Models*, Physica-Verlag, Heidelberg, 2000.

[14] Härdle, W. and Stoker, T.M., Investigating smooth multiple regression by the method of average derivatives, *J. Amer. Statist. Assoc.*, **84**(1989), 986–995.

[15] Hastie, T. and Tibshirani, R., Varying-coefficient model, *J. R. Statist. Soc. Ser. B*, **55**(1993), 757–796.

[16] Hastie, T.J. and Tibshirani, R., *Generalized Additive models*, London: Chapman and Hall, 1990.

[17] Liang, H., Zhu, L.X. and Zhou, Y., Asymptotic efficient estimation based on wavelet of expectation value in a partial linear model, *Commun. Statist. Theorey Meth.*, **28(9)**(1999), 2045–2055.

[18] Lu, Y.Q. and Mao, S.S., Local Asymptotics for B-spline estimators of the varying coefficient model, *Communications in Statistics – Theory and Methods*, **33(5)**(2004), 1119–1138.

[19] Qian, W.M. and Cai, G.X., The strong convergencerate of wavelet estimator in partially models, *Chinese Sci. A*,**29**(1999), 233–240.

[20] Stone, C.J., Hansen, M., Kooperberg, C. and Truong, Y.K., Polynomial splines and their tensor products in extended linear modelling, *Ann. Statist.*, **25**(1997), 1371–1470.

[21] Vidakovic, B., *Statistical Modeling by Wavelet*, John Wiley & Sons Inc., New York, 1999.

[22] Wu, C.O., Chiang, C. and Hoover, D.R., Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data, *J. Amer. Statist. Assoc.*, **93**(1998), 1388–1403.

[23] Zhou, X. and You, J., Wavelet estimation in varying-coefficient partially linear regression models, *Statistics and Probability letters*, **68**(2004), 91–104.

变系数模型的小波估计

卢一强 李志林

(解放军信息工程大学电子技术学院, 郑州, 450004)

变系数模型是近年来文献中经常出现的一种统计模型. 本文主要研究了变系数模型的估计问题, 提出运用小波的方法估计变系数模型中的系数函数, 小波估计的优点是避免了象核估计、光滑样条等传统的变系数模型估计方法对系数函数光滑性的一些严格限制. 并且, 我们还得到了小波估计的收敛速度和渐近正态性. 模拟研究表明变系数模型的小波估计有很好的估计效果.

关键词: 变系数模型, 小波, 最小二乘估计, 渐近正态性, 收敛速度.

学科分类号: O212.7.