应用概率统计 第三十一卷
第四期 2015年8月

Chinese Journal of Applied Probability
and Statistics Vol.31 No.4 Aug. 2015

Covariate-Adjusted Nonparametric Regression for Time Series *

Ma Yunyan

(*School of Mathematics and Statistics Science, Ludong University, Yantai, 264025*)

Kou Guangjie

(*School of Information and Electrical Engineering, Ludong University, Yantai, 264025*)

Abstract

The covariate-adjusted regression model was initially proposed for the situations where both the predictors and the response variables are not directly observed, but are distorted by some common observable covariates. In this paper, we investigate a covariate-adjusted nonparametric regression (CANR) model and consider the proposed model on time series setting. We develop a two-step estimation procedure to estimate the regression function. The asymptotic property of the proposed estimation is investigated under the α-mixing conditions. Both the real data and simulated examples are provided for illustration.

Keywords: α-mixing, covariate-adjusted nonparametric regression, time series.

AMS Subject Classification: Primary 62G08; Secondary 62M10.

§1. Introduction

The covariate-adjusted regression (CAR) model was first proposed by Sentürk and Müller (2005) to analyze the regression relationship between plasma fibrinogen concentration and serum transferrin for haemodialysis patients. Both the response and predictors are not directly observed but are thought to be distorted by unknown functions of a confounding observable covariate in a multiplicative fashion. In order to estimate the coefficients, a two-step estimation procedure was proposed and the consistency of estimators was also established. Subsequently, the asymptotic normality of the estimators was proved by Sentürk and Müller (2006).

*The project was supported by the Doctoral Fund of Ludong University (LY2013001, LY201222), Science and Technology Development Projects of Shandong Province (2012YD01056), Shangdong Province Young and Middle-Aged Scientists Research Awards Fund (BS2013SF029), National Natural Science Foundation of China – Tianyuan Fund for Mathematics (11426126) and Natural Science Foundation of Shandong Province (ZR2014AP007).

Received August 8, 2014. Revised January 27, 2015.

doi: 10.3969/j.issn.1001-4268.2015.04.009

Despite the covariate-adjusted regression is originally designed for independent cross-sectional data, applications to other type of data settings have gained a lot of attention with important practical value in health sciences research, psychology and sociology. For example: Sentürk (2006) proposed a covariate-adjusted varying coefficient regression (CAVCR), where the goal was to target the covariate-adjusted relationship among longitudinal variables; Sentürk and Nguyen (2009) discussed the asymptotic properties of covariate-adjusted regression when the observations were correlated; Nguyen et al. (2008) proposed a covariate-adjusted linear mixed effects model and applied the model to longitudinal data; Nguyen and Sentürk (2009) proposed an estimation method which incorporates the correlation/covariance structure between repeated measurements in the covariate-adjusted regression models for distorted longitudinal data. On the other hand, Ma and Luan (2012) considered a covariate-adjusted regression model on time series setting and studied the asymptotic property of the estimators.

Besides the parametric covariate-adjusted regression models mentioned above, there are many extensions of the CAR model: Cui et al. (2009) proposed a covariate-adjusted nonlinear regression model and applied the model to study the relationship between glomerular filtration rate and serum creatinine. They introduced a nonparametric method to estimate the distorting functions by regressing the predictors and the response on the distorting covariate and obtained the nonlinear least squares estimators for the parameters using the estimated response and predictors. Li et al. (2009) introduced a covariate-adjusted partially linear regression model. Inspired by the work of Cui et al. (2009), they proposed an estimation method. They also obtained the asymptotic normality of the parametric component estimator.

In time series analysis, it can be classified into linear and nonlinear time series models. The most popular class of linear time series models consists of autoregressive (AR) model, autoregressive moving-average (MA) model moving average (ARMA) models, autoregressive integrated moving average (ARIMA) model, and so on. However, in many real time series data, there exists some nonlinear features, include, nonnormality, asymmetric cycles, bimodality, nonlinear relationship between lagged variables, variation of prediction performance over the state-space, time irreversibility, and others (Fan and Yao, 2003). Linear time series models encounter limitation to deal with these data with nonlinear features. Beyond the linear domain, there are many nonlinear forms to be explored. Recently, developments in nonparametric regression techniques provide an alternative to model nonlinear time series (Masry and Fan, 1997; Fan and Yao, 2003). The advantage of the nonparametric regression techniques is that little prior information on model structure is assumed, and it may offer useful insights for further parametric fitting. Furthermore, with increasing computing power in recent years, it has become commonplace to attempt to analyze time series data of unprecedented size and complexity. All these changes have an increasing demand for nonparametric method that can analyze complex data and identify the underlying structure.

In this paper, we introduce a covariate-adjusted nonparametric regression (CANR) model on time series setting. The CANR model is proposed as

$$\begin{cases} Y = m(X) + \epsilon, \\ \widetilde{Y} = \psi(U)Y, \\ \widetilde{X} = \phi(U)X, \end{cases} \tag{1.1}$$

where $m(\cdot)$ is a regression function, Y is a response variable, X is a predictor, and ϵ is an error term satisfying $\mathsf{E}(\epsilon|X) = 0$, $\mathsf{Var}\,(\epsilon|X) = 1$. In this model, X and Y are not directly observed but are distorted by unknown functions of a confounding observable covariate U. $\widetilde{X}$ and $\widetilde{Y}$ are distorted observations, and $\psi(\cdot)$ and $\phi(\cdot)$ are unknown distorting functions. In order to estimate the regression function $m(\cdot)$, a two-step method will be proposed. In step 1, the distorting functions will be estimated using the nonparametric method which were introduced by Cui et al. (2009) and the estimators of the response and the predictors will be obtained. In step 2, a nonparametric method will be employed to estimate the regression function using the estimated response and predictors. We will study the asymptotic convergence property of the estimated regression function and illustrate the proposed procedure by a simulated example.

In financial analysis, the relationships between the spot markets and the futures markets are affected by many factors, such as stock price, stock index futures, interest rate, and so on. In this paper, we will consider the the relationships between the spot markets and the futures markets of China. Since the public listing on 16 April 2010, Shanghai and Shenzhen 300 Stock Index Futures (IF) has exerted important influences on Chinese financial markets. Thus, it is reasonable to take IF as a covariate variable and consider the IF-adjusted nonparametric regression model. In Ma and Luan (2012), a parametric regression method has been used to study this relationship and has showed that there exist nonlinear causality between the spot and the futures markets. In this paper, we will apply the proposed nonparametric regression method to study this relationship and draw the conclusions coinciding with realities.

The paper is organized as follows. In Section 2, we describe the model in detail and propose the estimator of the regression function. In Section 3, the results on asymptotic property are presented. A simulated example and the applications to real data are given in Section 4. The technical proofs of the main results are presented in Section 5.

§2. Model and Estimations

We write a sample version of the covariate-adjusted nonparametric regression(CANR) model as

$$\begin{cases} Y_i = m(X_i) + \epsilon_i, \\ \widetilde{Y}_i = \psi(U_i) Y_i, \qquad\qquad i = 1, 2, \ldots, n. \\ \widetilde{X}_i = \phi(U_i) X_i, \end{cases} \tag{2.1}$$

In this paper, we consider the model on time series setting. We assume the unobservable data $\{(U_i, X_i, Y_i), i = 1, 2, \ldots, n\}$ is a joint strictly stationary α-mixing sequence, then by property of α-mixing, the observable data $\{(U_i, \widetilde{X}_i, \widetilde{Y}_i), i = 1, 2, \ldots, n\}$ is also a joint strictly stationary α-mixing sequence. The main objective is to estimate the regression function $m(\cdot)$ and to consider the asymptotic property based on the observable data $\{(U_i, \widetilde{X}_i, \widetilde{Y}_i), i = 1, 2, \ldots, n\}$.

To achieve this goal, some basic assumptions are needed as follows:

(A1) The variables ϵ, U, X are mutually independent, and $\mathsf{E}(\epsilon|U,X) = 0$, $\mathsf{Var}\,(\epsilon|U,X) = 1$.

(A2) The functions $\psi(\cdot)$ and $\phi(\cdot)$ are twice continuously differentiable, and satisfy identifiability conditions: $\mathsf{E}\psi(U) = 1$, $\mathsf{E}\phi(U) = 1$ with $\phi(\cdot) > 0$.

(A3) $\{(U_i, X_i, Y_i), i = 1, 2, \ldots, n\}$ is a strictly stationary α-mixing sequence with $\alpha(l) \le cl^{-\beta}$ for some $c > 0$ and $\beta > 5/2$.

(A4) For each fixed $x \in (-\infty, +\infty)$, $m(\cdot)$ satisfies Lipschitz condition, that is, there exist $0 < q \le 1$, $\gamma > 0$ and some neighborhood N of x, such that

$$|m(t) - m(x)| \le \|t - x\|^q, \qquad \forall\, t \in N.$$

To estimate the regression function, a two-step estimation procedure is proposed as follows:

Step 1: Under condition (A2), We assume that the mean distorting effect vanishes in model (2.1). As in Cui et al. (2009),

$$\psi(U) = \frac{\mathsf{E}(\widetilde{Y}|U)}{\mathsf{E}Y}, \qquad \phi(U) = \frac{\mathsf{E}(\widetilde{X}|U)}{\mathsf{E}X}.$$

Then the nonparametric estimators of $\psi(U)$ and $\phi(U)$ are proposed as

$$\widehat{\psi}(u) = \frac{\frac{1}{n}\sum_{i=1}^{n} K_{h_1}(u-U_i)\widetilde{Y}_i}{\frac{1}{n}\sum_{j=1}^{n} K_{h_1}(u-U_j)} \times \frac{1}{\overline{\overline{\widetilde{Y}}}} \triangleq \widehat{g}_{\widetilde{Y}}(u) \times \frac{1}{\overline{\overline{\widetilde{Y}}}}, \tag{2.2}$$

$$\widehat{\phi}(u) = \frac{\frac{1}{n}\sum_{i=1}^{n} K_{h_2}(u-U_i)\widetilde{X}_i}{\frac{1}{n}\sum_{j=1}^{n} K_{h_2}(u-U_j)} \times \frac{1}{\overline{\overline{\widetilde{X}}}} \triangleq \widehat{g}_{\widetilde{X}}(u) \times \frac{1}{\overline{\overline{\widetilde{X}}}}, \tag{2.3}$$

where $K(\cdot)$ is a kernel function, h_1 and h_2 are bandwidths, $K_h(\cdot) = (1/h)K(\cdot/h)$, $\overline{\overline{\widetilde{Y}}} = \sum_{i=1}^{n} \widetilde{Y}_i$, and $\overline{\overline{\widetilde{X}}} = \sum_{i=1}^{n} \widetilde{X}_i$. Consequently, estimators of Y and X can be presented as

$$\widehat{Y}_i = \frac{\widetilde{Y}_i}{\widehat{\psi}(U_i)}, \qquad \widehat{X}_i = \frac{\widetilde{X}_i}{\widehat{\phi}(U_i)}, \tag{2.4}$$

respectively. It follows from (2.1) and (2.4) that

$$\widehat{Y}_i \approx m(\widehat{X}_i) + \sigma(\widehat{X}_i)\epsilon_i, \tag{2.5}$$

where $\sigma^2(x) = \mathsf{Var}\,(\widehat{Y}_i|\widehat{X}_i = x)$ and ϵ_i satisfies

$$\mathsf{E}(\epsilon_i) = 0, \qquad \mathsf{Var}\,(\epsilon_i) = 1.$$

Step 2: In Step 1, we have proposed the estimators $\widehat{X}$ and $\widehat{Y}$ of unobservable variables X and Y. Then the regression function of model (2.1) can be estimated by the approximate formula (2.5), which is a general nonparametric regression model. By the Nadaraya-Watson estimation (Nadaraya, 1964; Watson, 1964), the estimated regression function is proposed as

$$\widehat{m}(x) = \frac{\frac{1}{n}\sum_{i=1}^{n} K_{h_0}(x-\widehat{X}_i)\widehat{Y}_i}{\frac{1}{n}\sum_{j=1}^{n} K_{h_0}(x-\widehat{X}_i)} \triangleq \sum_{i=1}^{n} W_{h_0}(x-\widehat{X}_i)\widehat{Y}_i, \tag{2.6}$$

where h_0 is a bandwidth,

$$W_{h_0}(x-\widehat{X}_i) = \frac{K_{h_0}(x-\widehat{X}_i)}{\sum_{j=1}^{n} K_{h_0}(x-\widehat{X}_i)}.$$

For simplicity, we assume the same kernel function in equations (2.2), (2.3), and (2.4).

§3. Asymptotic Results

In this section, we consider the asymptotic property of the estimator given in (2.6) on time series setting. Theorem 3.1 and Theorem 3.2 prove the asymptotic convergence and the asymptotic normality of the estimated regression function $\widehat{m}(\cdot)$. For simplicity, we denote the bandwidth by h throughout this section. The proofs are given in Section 5. Same as in Section 2, we denote

$$g_{\widetilde{X}}(U)=\mathsf{E}(\widetilde{X}|U),\qquad g_{\widetilde{Y}}(U)=\mathsf{E}(\widetilde{Y}|U),$$
$$\widehat{g}_{\widetilde{X}}(u)=\frac{\frac{1}{n}\sum\limits_{i=1}^{n}K_{h_1}(u-U_i)\widetilde{X}_i}{\frac{1}{n}\sum\limits_{j=1}^{n}K_{h_1}(u-U_j)},\qquad \widehat{g}_{\widetilde{Y}}(u)=\frac{\frac{1}{n}\sum\limits_{i=1}^{n}K_{h_2}(u-U_i)\widetilde{Y}_i}{\frac{1}{n}\sum\limits_{j=1}^{n}K_{h_2}(u-U_j)}.$$

The following conditions are needed for the results.

(C1) The kernel $K(\cdot)$ is a bounded function with a bounded support, satisfying Lipschitz condition.

(C2) For some $s>2$ and some intervals $[a,b]$ and $[A,B]$, $\mathsf{E}|X|^s<\infty$, $\mathsf{E}|Y|^s<\infty$ and

$$\sup_{u\in[a,b]}\int|x|^s f_1(u,x)\mathrm{d}x<\infty,$$
$$\sup_{u\in[a,b]}\int|y|^s f_2(u,y)\mathrm{d}y<\infty,$$
$$\sup_{x\in[A,B]}\int|y|^s f_3(x,y)\mathrm{d}y<\infty,$$

where f_1,f_2,f_3 denote the joint density of (U,X), (U,Y) and (X,Y), respectively.

(C3) For some $\delta>0$, $h\to0$, $n^{1-2s^{-1}-2\delta}h\to\infty$, as $n\to\infty$, where s is the same as in (C2).

(C4) For some A_1,A_2,A_3, and for any $l\geq1$,

$$f_{U_0,U_l|\widetilde{X}_0,\widetilde{X}_l}(u_0,u_l|\widetilde{x}_0,\widetilde{x}_l)\leq A_1<\infty,$$
$$f_{U_0,U_l|\widetilde{Y}_0,\widetilde{Y}_l}(u_0,u_l|\widetilde{y}_0,\widetilde{y}_l)\leq A_2<\infty,$$
$$f_{X_0,X_l|\widetilde{Y}_0,\widetilde{Y}_l}(x_0,x_l|y_0,y_l)\leq A_3<\infty.$$

(C5) For some $\delta>2$, $a>1-2/\delta$ and some B_1,B_2,B_3, assume $\sum\limits_{l}l^a[\alpha(l)]^{1-2/\delta}<\infty$, $\mathsf{E}|X_0|^\delta<\infty$, $\mathsf{E}|Y_0|^\delta<\infty$,

$$f_{U_0|\widetilde{X}_0}(u|\widetilde{x})\leq B_1<\infty,\qquad f_{U_0|\widetilde{Y}_0}(u|\widetilde{y})\leq B_2<\infty,\qquad f_{X_0|Y_0}(x|y)\leq B_3<\infty.$$

The following Lemma 3.1 gives some asymptotic results of $\widehat{X}$, $\widehat{Y}$ and $W_h(x-\widehat{X})$, which will be used in the proof of Theorem 3.1 and Theorem 3.2.

Lemma 3.1 Under the conditions (A3) and (C1)-(C5), the following asymptotic representations hold:

$$\begin{aligned}
&\widehat{X_i}-X_i=O_P\big(h+(nh/\log(1/h))^{-1/2}\big),\\
&\widehat{Y}_i-Y_i=O_P\big(h+(nh/\log(1/h))^{-1/2}\big),\\
&W_h(x-\widehat{X_i})-W_h(x-X_i)=o_P(1).
\end{aligned}$$

In the following theorem, we present the asymptotic convergence of the estimated regression function $\widehat{m}(x)$, and give the convergence rate.

Theorem 3.1 Let $\widehat{m}(x)$ be defined by (2.6). If conditions (A1)-(A3) and (C1)-(C5) are satisfied, then

$$\sup_{x\in[A,B]}|\widehat{m}(x)-m(x)|=O_P\big(h+(nh/\log(1/h))^{-1/2}\big).$$

Theorem 3.2 Let $\widehat{m}(x)$ be defined by (2.6). If conditions (A1)-(A4) and (C1)-(C5) are satisfied, then

$$\sqrt{nh}(\widehat{m}(x)-m(x)-b_n)\xrightarrow{D}N(0,\Sigma^2),$$

where

$$\Sigma^2=\frac{\sigma^2(x)\displaystyle\int_{-\infty}^{+\infty}K^2(t)\mathrm{d}t}{f(x)\Big(\displaystyle\int_{-\infty}^{+\infty}K(t)\mathrm{d}t\Big)^2},$$

$f(\cdot)$ is the density function of X, $b_n=O(h^q)$, q is the same as in condition (A4).

§4. Numerical Studies

4.1 Simulated Example

In this subsection, we investigate the finite-sample behavior of the proposed estimation. The underlying unobserved nonparametric regression model is

$$Y=\sin(\pi X)+e,$$

where X is simulated from an AR(1) model $X_i=0.9X_{i-1}+\varepsilon_i$ with ε_i generated from $N(0,0.3)$. The confounding covariate U is simulated from an AR(1) model $U_i=0.8U_{i-1}+$

ε_{i0} with ε_{i0} generated from $N(0, 0.5)$. The error term e is generated from $N(0, 0.03)$. The true regression function $m(x)$ is set as $\sin(\pi x)$. The distorting functions are chosen as

$$\psi(U) = \frac{U+3}{3.0008}, \qquad \phi(U) = \frac{(U+2)^2}{4.0030},$$

satisfying the identifiability conditions (A2). The simulation is repeated 10000 times. Figure 1 displays the plots of underlying variables X and Y and their estimators $\widehat{X}$, $\widehat{Y}$ for $n = 300$, respectively. By (2.6), we can obtain the estimated function of $m(\cdot)$ in CANR model. Figure 2 displays the curves of estimated function $\widehat{m}(\cdot)$ (circle line) and true function $m(\cdot)$ (solid line) for $n = 1000$ by 10000 simulations. In order to analyze the influence of covariate U, we also estimate $m(\cdot)$ in general nonparametric regression (GNR) model without considering the covariate U. The performance of estimator $\widehat{m}(\cdot)$ is assessed via the average squared error (ASE), defined by

$$\mathrm{ASE}(\widehat{m}) = \frac{1}{n}\sum_{k=1}^{n}\{\widehat{m}(x_k) - m(x_k)\}^2,$$

with $\widehat{m}$ denotes estimator either in CANR model or in GNR model. The mean and standard deviation of ASE through 10000 simulations are displayed in Table 1. It is easy to see that the covariate-adjusted estimation is better than the estimation without considering the covariate U.

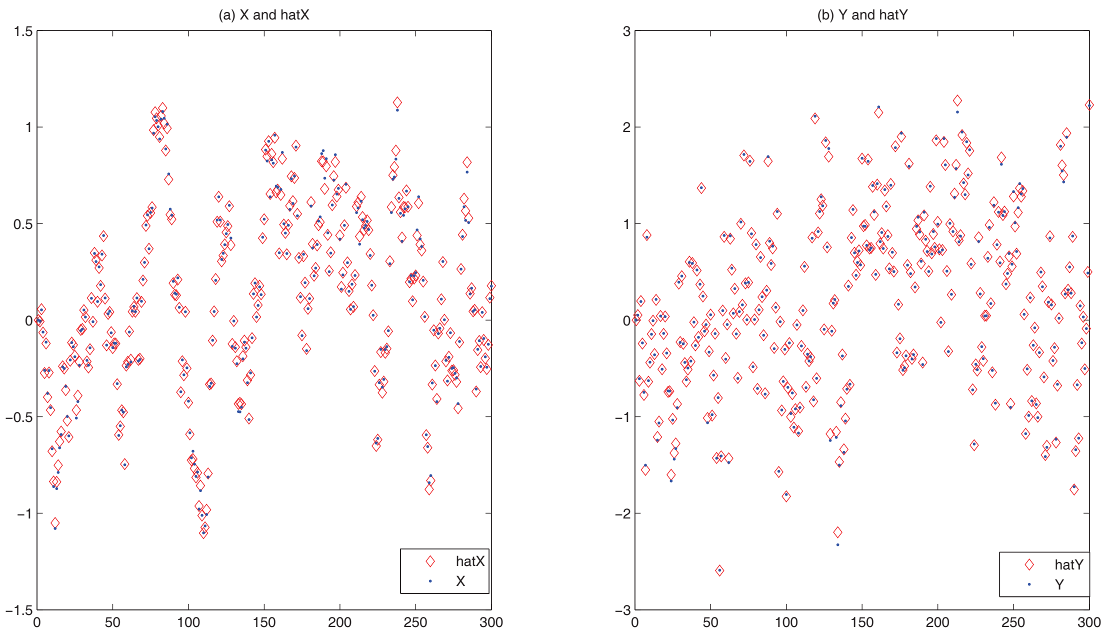

Figure 1 Plots of estimators $\widehat{X}$ and $\widehat{Y}$: (a) X $(\cdot)$ and $\widehat{X}$ $(\diamond)$ (b) Y $(\cdot)$ and $\widehat{Y}$ $(\diamond)$

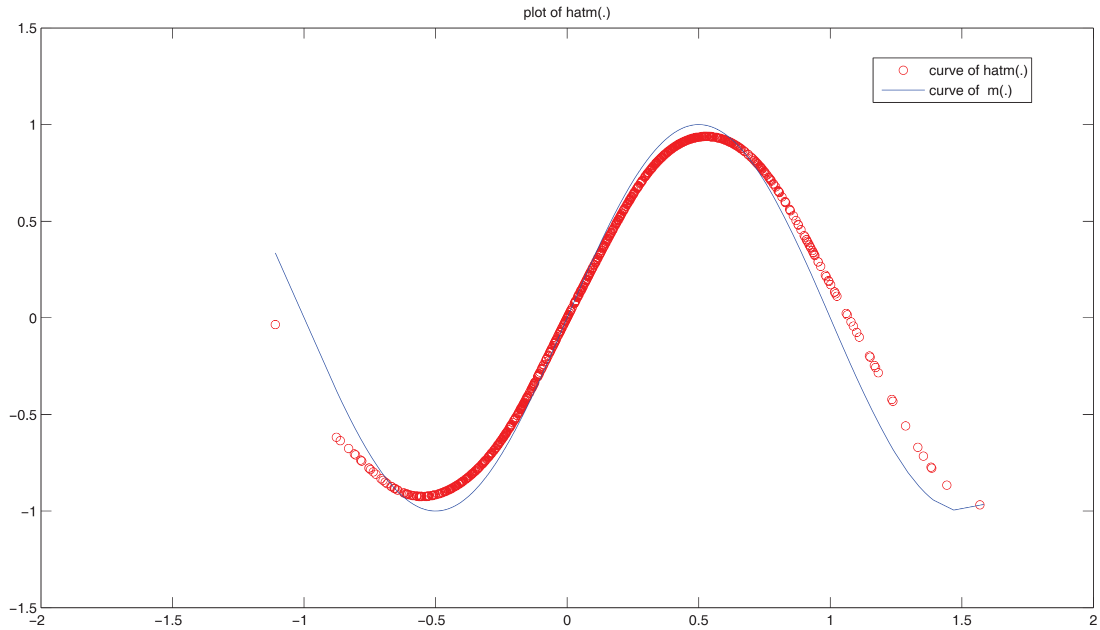

Figure 2 Curves of estimated regression functions $\widehat{m}(\cdot)$ (circle line) and true function $m(\cdot)$ (solid line)

Table 1 Average squared errors (ASE) of $\widehat{m}(\cdot)$

ASE: mean$_{\text{(standard deviation)}}$		
	CANR	GNR
$n = 300$	$0.0089_{(0.0138)}$	$0.0091_{(0.0136)}$
$n = 500$	$0.0061_{(0.0185)}$	$0.0062_{(0.0186)}$
$n = 1000$	$0.0057_{(0.0091)}$	$0.0058_{(0.0091)}$

The simulation is repeated 10000 times for each of sample sizes $n = 300$, 500 and 1000. The ASE of estimator $\widehat{m}(\cdot)$ are reported in Table 1.

4.2 Applications to Financial Data

In financial analysis, the relationships between the spot markets and the futures markets have been paid close attention by researchers. Most traditional studies focus on the linear Granger causality between the spot markets and the futures markets. In real markets, however, their relationships are influenced by many factors, such as stock price, stock index futures, interest rate, and so on. These imply that there should exist more complicated relationships between the spot and futures markets. Recently, empirical evidences have also shown that nonlinear causality relationship indeed exists in the spot markets and futures markets, but the traditional linear methodology can't discriminate nonlinear causality very well, so as not to describe their relationships clearly. Several studies have employed nonlinear method to describe nonlinear causality relationship between the spot and futures return (Abhyankar, 1996; Chen and Lin, 2004; Silvapulle and Moosa, 1999). Take Chinese financial markets for example, since the public listing of Shanghai and Shenzhen 300 Stock Index Futures (IF) on 16 April 2010, it has produced great impact on the spot markets and the futures markets. A recent study also revealed that there exists nonlinear causality between spot and future prices of Copper market (Dai and Ding, 2010). To illustrate our methods, we consider the relationship between the Copper Spot Price (CSP) (response) and the Copper Futures Price (CFP) (predictor). The underlying nonparametric regression model would be

$$\mathrm{CSP} = m(\mathrm{CFP}) + e, \tag{4.1}$$

where e is an error term. As discussed above, Shanghai and Shenzhen 300 Stock Index Futures (IF), as a stock index futures, affects the relationship between CSP and CFP. Then we choose IF as the covariate U and consider the IF-adjusted estimation method given in Section 2. We denote the distorted CSP and CFP by $\widetilde{\mathrm{CSP}}$ and $\widetilde{\mathrm{CFP}}$, respectively. The distorting equations are expressed as

$$\widetilde{\mathrm{CSP}} = \psi(\mathrm{IF})\mathrm{CSP}, \qquad \widetilde{\mathrm{CFP}} = \phi(\mathrm{IF})\mathrm{CFP}. \tag{4.2}$$

By (2.4), we obtain the estimators of CSP and CFP, which are denoted by $\widehat{\mathrm{CSP}}$ and $\widehat{\mathrm{CFP}}$, respectively. The time series plots of CSP, CFP, $\widehat{\mathrm{CSP}}$ and $\widehat{\mathrm{CFP}}$ are displayed in Figure 4. The estimators of the distorting functions $\psi(\cdot)$ and $\phi(\cdot)$ are displayed in Figure 5. By (2.6), we obtain the estimator of the regression function $m(\cdot)$. In order to illustrate our method, we also give the non-adjusted estimator of regression function without considering the covariate IF. We present the estimators derived by the two methods in Figure 6. From the four figures, it can be seen that the estimators are adjusted when IF are too big or too small, and they also provide the evidence that there exists nonlinear relationship between CSP and CFP and this relationship is affected by IF. The result coincides well with empirical evidence, and then it is helpful to investors to analyze the investment environment clearly and to make the relatively correct investment decision.

We choose each trading day's closing price of CSP, CFP, and IF over the period between 16 April 2010 and 31 December 2010 with 218 observations. In order to ensure the stationarity, we pretreat the raw data and use the first-order difference of the natural logarithm. Figure 3(a) gives the time series plots of CSP (dotted line) and CFP (solid line). Figure 3(b) gives the time series plots of IF (solid line). (Data sources: Wind Information).

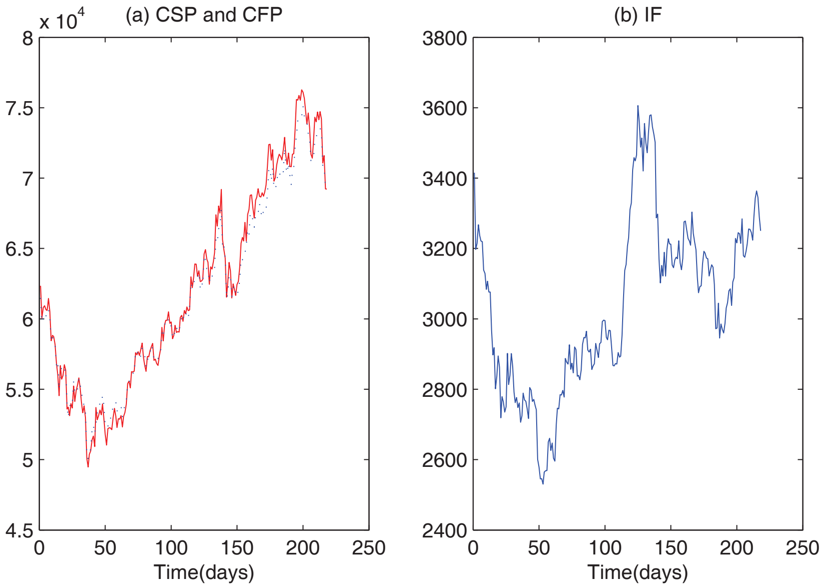

Figure 3 Time series plots of real data: (a) Plots of CSP (solid line) and CFP (dotted line) (b) Plot of IF

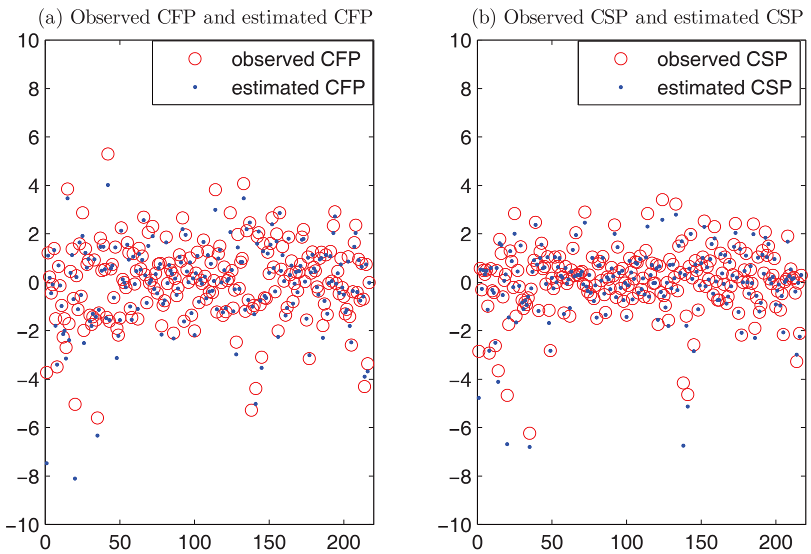

Figure 4 Plots of observed and estimated data: (a) Observed CFP (circle) and estimated CFP (dot) (b) Observed CSP (circle) and estimated CSP (dot)

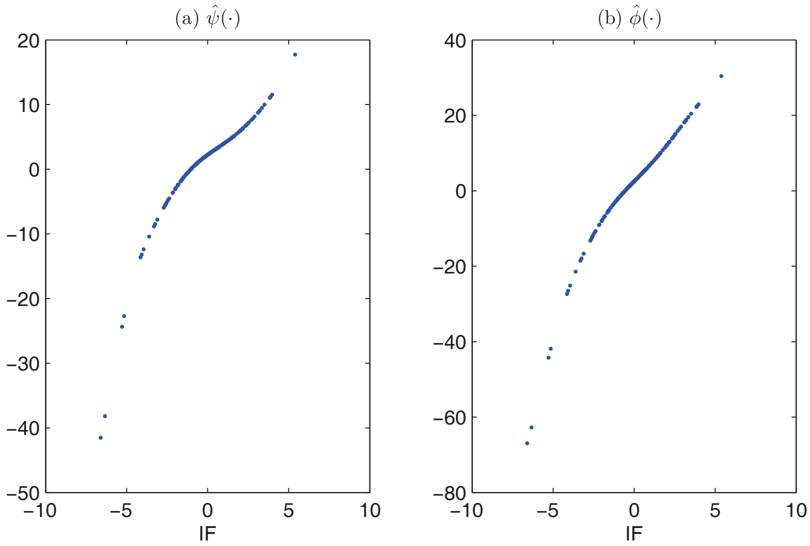

Figure 5　(a) Estimator of the distorting function $\psi(\cdot)$

(b) Estimator of the distorting function $\phi(\cdot)$

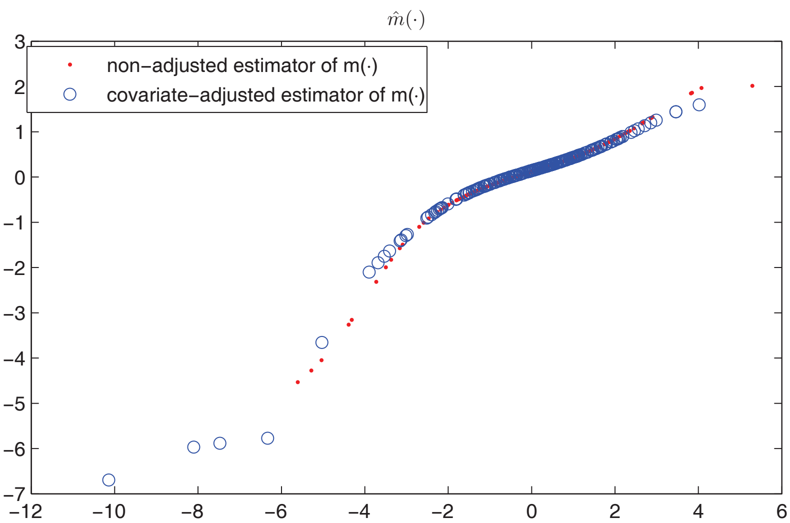

Figure 6　Non-adjusted estimator of $m(\cdot)$ (dotted line) and covariate-adjusted estimator of $m(\cdot)$ (circle line)

§5.　Proof of the Main Results

Proof of Lemma 3.1　By Theorem 6.5 in Fan and Yao (2003), we can get the uniform convergence of the nonparametric estimators $\widehat{g}_{\widetilde{X}}(u)$ and $\widehat{g}_{\widetilde{Y}}(u)$,

$$\sup_{u\in[a,b]}|\widehat{g}_{\widetilde{X}}(u)-g_{\widetilde{X}}(u)|=O_P\big(h+(nh/\log(1/h))^{-1/2}\big),\tag{5.1}$$

$$\sup_{u\in[a,b]} |\widehat{g}_{\widetilde{Y}}(u) - g_{\widetilde{Y}}(u)| = O_P\big(h + (nh/\log(1/h))^{-1/2}\big). \tag{5.2}$$

Firstly, consider $\widehat{X}_i - X_i$. By (2.2), (2.3) and (2.4), we can get the following decomposition,

$$\begin{aligned}\widehat{X}_i - X_i &= X_i\Big(\frac{\phi(U_i)}{\widehat{\phi}(U_i)} - 1\Big)\\ &= X_i\Big(\frac{g_{\widetilde{X}}(U_i)}{\mathsf{E}X}\frac{\overline{\widetilde{X}}}{\widehat{g}_{\widetilde{X}}(U_i)} - 1\Big)\\ &= X_i\Big(\frac{g_{\widetilde{X}}(U_i)}{\widehat{g}_{\widetilde{X}}(U_i)} - 1 + \frac{g_{\widetilde{X}}(U_i)}{\widehat{g}_{\widetilde{X}}(U_i)}\frac{\overline{\widetilde{X}} - \mathsf{E}X}{\mathsf{E}X}\Big).\end{aligned}$$

According to the term $g_{\widetilde{X}}(U_i)/\widehat{g}_{\widetilde{X}}(U_i)$, it can be easily obtained by (5.1) that

$$\frac{g_{\widetilde{X}}(U_i)}{\widehat{g}_{\widetilde{X}}(U_i)} = 1 - \frac{\widehat{g}_{\widetilde{X}}(U_i) - g_{\widetilde{X}}(U_i)}{\widehat{g}_{\widetilde{X}}(U_i)} = O_P\big(h + (nh/\log(1/h))^{-1/2}\big).$$

Besides, applying the law of large numbers to $(\overline{\widetilde{X}} - \mathsf{E}X)/\mathsf{E}X$, it follows that

$$\begin{aligned}\widehat{X}_i - X_i &= O_P\Big(X_i\cdot\frac{\widehat{g}_{\widetilde{X}}(U_i) - g_{\widetilde{X}}(U_i)}{\widehat{g}_{\widetilde{X}}(U_i)}\Big) + O_P\Big(X_i\cdot\frac{\widehat{g}_{\widetilde{X}}(U_i) - g_{\widetilde{X}}(U_i)}{\widehat{g}_{\widetilde{X}}(U_i)}\cdot n^{-1/2}\Big)\\ &= O_P\big(h + (nh/\log(1/h))^{-1/2}\big).\end{aligned}$$

Similarly, it can be shown that $\widehat{Y}_i - Y_i = O_P\big(h + (nh/\log(1/h))^{-1/2}\big)$.

Lastly, because $\widehat{X}$ converges to X a.e., then $\widehat{X}$ converges to X in distribution, and because the kernel function $K(\cdot)$ is continuous, it can be easily obtained that

$$W_h(x-\widehat{X}_i) - W_h(x-X_i) = \frac{K_h(x-\widehat{X}_i)}{\dfrac{1}{n}\sum\limits_{j=1}^n K_h(x-\widehat{X}_j)} - \frac{K_h(x-X_i)}{\dfrac{1}{n}\sum\limits_{j=1}^n K_h(x-X_j)} = o_P(1).$$

This completes the proof of Lemma 3.1. □

Proof of Theorem 3.1

$$\begin{aligned}&\widehat{m}(x) - m(x)\\ =\ &\sum_{i=1}^n W_h(x-\widehat{X}_i)\widehat{Y}_i - m(x)\\ =\ &\sum_{i=1}^n [W_h(x-\widehat{X}_i) - W_h(x-X_i)][\widehat{Y}_i - Y_i] + \sum_{i=1}^n [W_h(x-\widehat{X}_i) - W_h(x-X_i)]Y_i\\ &+\sum_{i=1}^n W_h(x-X_i)[\widehat{Y}_i - Y_i] + \sum_{i=1}^n W_h(x-X_i)Y_i - m(x)\\ =\ &Q_1 + Q_2 + Q_3 + Q_4,\end{aligned} \tag{5.3}$$

where Q_i, $i=1,\ldots,4$ are defined as follows,

$$Q_1=\sum_{i=1}^n[W_h(x-\widehat{X}_i)-W_h(x-X_i)][\widehat{Y}_i-Y_i],$$

$$Q_2=\sum_{i=1}^n[W_h(x-\widehat{X}_i)-W_h(x-X_i)]Y_i,$$

$$Q_3=\sum_{i=1}^n W_h(x-X_i)[\widehat{Y}_i-Y_i],$$

$$Q_4=\sum_{i=1}^n W_h(x-X_i)Y_i-m(x).$$

By Lemma 3.1, the first three items are $o_P(1)$, and the major part Q_4 is

$$O_P\big(h+(nh/\log(1/h))^{-1/2}\big),$$

derived from Theorem 6.5 in Fan and Yao (2003). $\square$

Proof of Theorem 3.2 By Lemma 3.1, the first three items in (5.3) are $o_P(1)$, so it remains to prove the asymptotic normality of the major part Q_4. For simplicity, in the proofs, we denote the bandwidth in (2.6) by h. By

$$W_h(x-X_i)=\frac{K_h(x-X_i)}{\sum\limits_{i=1}^n K_h(x-X_i)},$$

it follows that

$$\begin{aligned}Q_4&=\frac{\sum\limits_{i=1}^n K\Big(\dfrac{x-X_i}{h}\Big)Y_i}{\sum\limits_{i=1}^n K\Big(\dfrac{x-X_i}{h}\Big)}-m(x)\\&=\frac{\sum\limits_{i=1}^n(m(X_i)-m(x))K\Big(\dfrac{x-X_i}{h}\Big)}{\sum\limits_{i=1}^n K\Big(\dfrac{x-X_i}{h}\Big)}+\frac{\sum\limits_{i=1}^n \sigma^2(X_i)K\Big(\dfrac{x-X_i}{h}\Big)\epsilon_i}{\sum\limits_{i=1}^n K\Big(\dfrac{x-X_i}{h}\Big)}\\&\triangleq B_n+R_n.\end{aligned}\tag{5.4}$$

Consequently,

$$\sqrt{nh}(\widehat{m}(x)-m(x)-B_n)=\frac{\sum\limits_{i=1}^n\dfrac{1}{\sqrt{nh}}\sigma(X_i)K\Big(\dfrac{x-X_i}{h}\Big)\epsilon_i}{\dfrac{1}{nh}\sum\limits_{i=1}^n K\Big(\dfrac{x-X_i}{h}\Big)}\triangleq\frac{\sum\limits_{k=1}^n\xi_{nk}}{S_n},\tag{5.5}$$

where $\{\xi_{nk}\}$ is a martingale difference sequence, by Lemma 1.2 in Hu (2002), we have

$$\sum_{k=1}^n\xi_{nk}\xrightarrow{D}N(0,\delta_1^2),$$

where $\delta_1^2=\sigma^2(x)f(x)\int_{-\infty}^{+\infty}K^2(t)\mathrm{d}t$.

For S_n, similar to the proof of Theorem 2.2 in Hu (2002), we have

$$
\begin{aligned}
S_n=&\frac{1}{h}\mathsf{E}K\Big(\frac{x-X_1}{h}\Big)\\
&+O_P\Big(\Big(\frac{1}{nh^2}\Big(\mathsf{Var}\Big(K\Big(\frac{x-X_i}{h}\Big)\Big)+\beta_n\Big(\int_{-\infty}^{+\infty}\Big|K\Big(\frac{x-t}{h}\Big)\Big|f(t)\mathrm{d}t\Big)^2\Big)\Big)^{1/2}\Big)\\
=&\frac{1}{h}\int_{-\infty}^{+\infty}K\Big(\frac{x-t}{h}\Big)f(t)\mathrm{d}t+O\Big(\frac{1}{nh}+\frac{\beta_n}{n}\Big)\xrightarrow{\mathsf{P}}f(x)\int_{-\infty}^{+\infty}K(t)\mathrm{d}t.
\end{aligned}
$$

Then, it follows that

$$
\sqrt{nh}(\widehat{m}(x)-m(x)-B_n)=\frac{\sum\limits_{k=1}^{n}\xi_{nk}}{S_n}\xrightarrow{D}N(0,\Sigma^2),\qquad n\longrightarrow\infty,\tag{5.6}
$$

where

$$
\Sigma^2=\frac{\sigma^2(x)\displaystyle\int_{-\infty}^{+\infty}K^2(t)\mathrm{d}t}{f(x)\Big(\displaystyle\int_{-\infty}^{+\infty}K(t)\mathrm{d}t\Big)^2}.
$$

For B_n,

$$
\frac{\sum\limits_{i=1}^{n}(m(X_i)-m(x))K\Big(\dfrac{x-X_i}{h}\Big)}{\sum\limits_{i=1}^{n}K\Big(\dfrac{x-X_i}{h}\Big)}\triangleq\frac{I_1}{I_2}.\tag{5.7}
$$

For I_1,

$$
\begin{aligned}
I_1=&n\mathsf{E}\Big((m(X_1)-m(x))K\Big(\frac{x-X_1}{h}\Big)\Big)+O_P\Big(\Big(n\mathsf{Var}\Big((m(X_1)-m(x))K\Big(\frac{x-X_1}{h}\Big)\Big)\\
&+2n\beta_n\Big(\int_{-\infty}^{+\infty}\Big|(m(t)-m(x))K\Big(\frac{x-t}{h}\Big)\Big|f(t)\mathrm{d}t\Big)^2\Big)^{1/2}\Big)\\
=&O(nh^{q+1})+O_P\big((nh^{2q+1}+n\beta_nh^{2q+2})^{1/2}\big)\\
=&nh\big(O(h^q)+O_P(h^q(nh)^{-1/2})\big).
\end{aligned}
$$

For I_2,

$$
\begin{aligned}
I_2=&n\mathsf{E}K\Big(\frac{x-X_1}{h}\Big)+O_P\Big(\Big(n\mathsf{Var}\Big(K\Big(\frac{x-X_1}{h}\Big)\Big)+2n\beta_n\Big(\int_{-\infty}^{+\infty}\Big|K\Big(\frac{x-t}{h}\Big)\Big|f(t)\mathrm{d}t\Big)^2\Big)^{1/2}\Big)\\
=&n\mathsf{E}K\Big(\frac{x-X_1}{h}\Big)+O_P\big((nh+2n\beta_nh^2)^{1/2}\big)\\
=&n\mathsf{E}K\Big(\frac{x-X_1}{h}\Big)+O_P\big((nh)^{1/2}\big).
\end{aligned}
$$

Let $b_n = \mathsf{E}\{(m(X_1) - m(x))K[(x - X_1)/h]\}/\mathsf{E}K[(x - X_1)/h]$, then $b_n = O(h^q)$, and

$$B_n = b_n + h^q O_P((nh)^{-1/2}). \tag{5.8}$$

By (5.6) and (5.8), as $n \longrightarrow \infty$, we have

$$\sqrt{nh}(\widehat{m}(x) - m(x) - b_n) \xrightarrow{D} N(0, \Sigma^2),$$

where

$$\Sigma^2 = \frac{\sigma^2(x)\displaystyle\int_{-\infty}^{+\infty} K^2(t)\mathrm{d}t}{f(x)\Big(\displaystyle\int_{-\infty}^{+\infty} K(t)\mathrm{d}t\Big)^2}.$$

This completes the proof of Theorem 3.2. $\square$

References

[1] Abhyankar, A., Does the stock index futures market tend to lead the cash market? New evidence from the FT-SE 100 stock index futures markets, *Working Paper 96-01*, Accountancy and Finance Department, University of Stirling, United Kingdom, 1996.

[2] Chen, A.S. and Lin, J.W., Cointegration and detectable linear and nonlinear causality: analysis using the London Metal Exchange lead contract, *Applied Economics*, **36(11)**(2004), 1157–1167.

[3] Cui, X., Guo, W., Lin, L. and Zhu, L., Covariate-adjusted nonlinear regression, *The Annals of Statistics*, **37(4)**(2009), 1839–1870.

[4] Dai, X. and Ding, L., Nonlinear Granger causality test of spot and future prices: evidence from the copper market, *Journal of Hunan University* (*Social Sciences*), **24(5)**(2010), 42–46. (in Chinese)

[5] Hu, S.H., Asymptotic normality for nonparametric regression function estimate, *Acta Mathematica Sinica, Chinese Series*, **45(3)**(2002), 433–442. (in Chinese)

[6] Fan, J. and Yao, Q., *Nonlinear Time Series: Nonparametric and Parametric Methods*, New York: Springer, 2003.

[7] Fan, J. and Gijbels, I., *Local Polynomial Modelling and Its Applications: Monographs on Statistics and Applied Probability 66*, Chapman & Hall, London, 1996.

[8] Li, F., Lin, L. and Cui, X., Covariate-adjusted partially linear regression models, *Communications in Statistics - Theory and Methods*, **39(6)**(2010), 1054–1074.

[9] Ma, Y. and Luan, Y., Covariate-adjusted regression for time series, *Communications in Statistics - Theory and Methods*, **41(3)**(2012), 422–436.

[10] Nadaraya, E.A., On estimating regression, *Theory of Probability and Its Applications*, **9(1)**(1964), 141–142.

[11] Nguyen, D.V., Sentürk, D. and Carroll, R.J., Covariate-adjusted linear mixed effects model with an application to longitudinal data, *Journal of Nonparametric Statistics*, **20(6)**(2008), 459–481.

[12] Nguyen, D.V. and Sentürk, D., Covariate-adjusted regression for longitudinal data incorporating correlation between repeated measurements, *Australian and New Zealand Journal of Statistics*, **51(3)** (2009), 319–333.

[13] Sentürk, D., Covariate-adjusted varying coefficient models, *Biostatistics*, **7(2)**(2006), 235–251.

[14] Sentürk, D. and Müller, H.G., Covariate-adjusted regression, *Biometrika*, **92(1)**(2005), 75–89.

[15] Sentürk, D. and Müller, H.G., Inference for covariate adjusted regression via varying coefficient models, *The Annals of Statistics*, **34(2)**(2006), 654–679.

[16] Sentürk, D. and Nguyen, D.V., Asymptotic properties of covariate-adjusted regression with correlated errors, *Statistics and Probability Letters*, **79(9)**(2009), 1175–1180.

[17] Silvapulle, P. and Moosa, I.A., The relationship between spot and futures prices: evidence from the crude oil market, *Journal of Futures Markets*, **19(2)**(1999), 175–193.

[18] Masry, E. and Fan, J., Local polynomial estimation of regression functions for mixing processes, *Scandinavian Journal of Statistics*, **24(2)**(1997), 165–179.

[19] Watson, G.S., Smooth regression analysis, *Sankhyā: The Indian Journal of Statistics, Series A*, **26(4)** (1964), 359–372.

时间序列中的协变量调整非参数回归模型

马云艳　　　　寇光杰

(鲁东大学数学与统计科学学院, 烟台, 264025)　(鲁东大学信息与电气工程学院, 烟台, 264025)

在某些场合, 回归模型中的预测变量与响应变量不能被直接观测, 而是受到某个可观测变量的影响, 在这种情况下人们提出了协变量调整模型. 本文在时间序列场合下讨论协变量调整非参数回归模型(CANR), 提出了回归函数的两步估计法, 在α-混合条件下讨论了估计的大样本性质, 最后研究了模型在模拟和实际金融数据中的应用.

关键词:　α-混合, 协变量调整非参数回归, 时间序列.

学科分类号:　O212.7.