Mean Volatility Regressions

Lu Lin*
Feng Li[§]
Lixing Zhu[#]
Wolfgang Karl Härdle[σ]

* - Shandong University Jinan, China
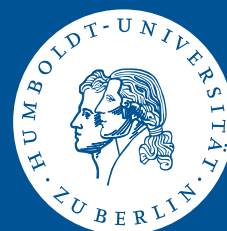§ - Zhengzhou Institute of Aeronautical Industry Management, China
- Hong Kong Baptist University, China
σ - Humboldt-Universität zu Berlin, Germany

BERLIN

ECONOMIC RISK

SFB 649

Mean Volatility Regressions

Lu Lin · Feng Li · Lixing Zhu · Wolfgang Karl Härdle

Abstract Motivated by increment process modeling for two correlated random and non-random systems from a discrete-time asset pricing with both risk free asset and risky security, we propose a class of semiparametric regressions for a combination of a non-random and a random system. Unlike classical regressions, mean regression functions in the new model contain variance components and the model variables are related to latent variables, for which certain economic interpretation can be made. The motivating example explains why the GARCH-M of which the mean function contains a variance component cannot cover the newly proposed models. Further, we show that statistical inference for the increment process cannot be simply dealt with by a two-step procedure working separately on the two involved systems although the increment process is a weighted sum of the two systems. We further investigate the asymptotic behaviors of estimation by using sophisticated nonparametric smoothing. Monte Carlo simulations are conducted to examine finite-sample performance, and a real dataset published in Almanac of

L. Lin
School of Mathematics, Shandong University, Jinan, China
E-mail: linlu@sdu.edu.cn

F. Li
Zhengzhou Institute of Aeronautical Industry Management, Zhengzhou, China
E-mail: lifengsta@gmail.com

L. Zhu
Department of Mathematics, Hong Kong Baptist University, Hong Kong, China
E-mail: lzhu@hkbu.edu.hk

W. K. Härdle
CASE-Center for Applied Statistics & Economics, Humboldt-Universität zu Berlin, Germany and CDA - Centre Complex Data Analysis. National Central University, Taiwan.
E-mail: haerdle@wiwi.hu-berlin.de

China's Finance and Banking (2004 and 2005) is analyzed for illustration about the increment process of wealth in financial market of China from 2003 to 2004.

Keywords Non-random systems · Random systems · Semiparametric regression · Variance built-in Mean

   JEL classification: C00; C14; J01; J31

1 Introduction

The standard regression paradigm separates mean and variance effects in the sense that the scale of the stochastic error does not enter the conditional mean effect given the explanatory variable X. This is also true for most of the time series models of autoregressive type with a few exceptions like e.g. the $GARCH - M$ model, see Tsay(2002). Discrete versions of many continuous time dynamics though naturally give motivation for considering dependence of the mean function on the conditional variance (or standard deviation) given X. Stochastic differential equations, as studied by Karoui and Peng (1997) in their discretized form may be seen as special cases of

$$\mathsf{E}\{\bar{Y}(t)|X(t)\} = f[\theta(t), Y\{X(t)\}, Z\{X(t)\}]\Delta_t^{1/2}, \qquad (1.1)$$

where f is a given regression function and Δ_t is the time period. In (1.1), the parameter vector $\theta(t)$ is time-dependent, $Y\{X(t)\}$, $\bar{Y}(t)$ and $Z\{X(t)\}$ depend on an observable variable function $X(t)$. Further, $Y\{X(t)\}$ and $\bar{Y}(t)$ are observable, while $Z\{X(t)\}$ is an unobservable random variable satisfying

$$Z^2\{X(t)\} = \text{Var}\{\bar{Y}(t)|X(t)\}. \qquad (1.2)$$

Here the basic difference to classical regressions (including the classical time series models) is that the mean $\mathsf{E}\{\bar{Y}(t)|X(t)\}$ of (1.1) contains a standard deviation $Z(x)$ (or variance $Z^2(x)$) and depends on the latent variable $X(t)$. The main difference between diffusion processes (or forward stochastic differential equations) and forward-backward stochastic differential equations will be discussed in the next section.

   Briefly speaking, the dynamics of $\bar{Y}(t)$ is determined by two systems, one is non-random (e.g. a money market instrument or bond) and the other has a random disturbance (e.g. a stock). Allocating randomly the asset to the two systems results in our new model. The new model can describe an increment process (e.g. wealth increment) and may be applied to different scientific fields such as mathematical finance. We adopt a two-step procedure with estimating parameters in the two systems separately and then working on the model about increment process of interest. This technique leads us to investigating statistical properties such as the distribution of the increment process and confidence interval construction for parameters; for details see the next section. The proposed model contains well known models, such as linear, nonlinear and varying coefficient models, as its special cases.

   The GARCH-M is also of a structure in which the mean function contains the variance function. We will give a brief explanation about the differences between our modeling and the GARCH-M in Section 2.1 below.

The paper is organized in the following way. In Section 2 we first briefly describe the motivation from a problem arising in asset pricing with one risk-free asset and one risky security.Consequently, a new model and its extensions are proposed in Section 2. For the model with varying coefficients, some estimation methods are introduced and the asymptotic properties are investigated in Section 3. Section 4 contains the estimation for nonlinear models. We separate these two cases, although the estimation in Section 3 can be a special case of that in Section 4, because of its importance of the model under linear structure. Monte Carlo studies are given in Section 5 for illustration and an empirical application is investigated in Section 6. The proofs of the theorems are presented in the Appendix.

2 Modeling

In this section, we start with two examples to motivate the modeling with varying coefficients under linear structure, and then extend them to the ones with nonlinear structure.

2.1 Models under linear structure

Consider the varying coefficient regression:

$$\bar{Y}(t_i) = [a(t_i) + b(t_i)Y(X(t_i)) + c(t_i)Z\{X(t_i)\}]\Delta_i^{1/2} + Z\{X(t_i)\}\varepsilon(t_i) \qquad (2.1)$$

for $i = 1, \cdots, n-1$, where $\Delta_i = t_{i+1} - t_i$ is the time period, $a(t)$, $b(t)$ and $c(t)$ are time-dependent parameters, $X(t_i)$ and $Y\{X(t_i)\}$ are random variables observed at t_i, and $\bar{Y}(t_i) = [Y(X(t_{i+1})) - Y\{X(t_i)\}]/\Delta_i^{1/2}$, $Z\{X(t_i)\}$ is an unobservable random variable valued at t_i, and $\{\varepsilon(t_i)\}$ is a sequence of unobservable independent standard normal random variables independent of $X(t_i)$. A special case is that all the parameters $a(\cdot)$, $b(\cdot)$, and $c(\cdot)$ are constants.

A very simple example is geometric Brownian motion where the stock dynamics $X(t)$ is:

$$dX(t)/X(t) = \mu dt + \sigma dW(t)$$

Discretising at t_i with steps Δ_i as defined above yields:

$$X(t_{i+1}) - X(t_i) = \mu X(t_i)\Delta_i + \sigma X(t_i)\Delta_i^{(1/2)}\varepsilon(t_i).$$

A further motivating example for (2.1) is as follows. For discrete-time asset pricing, one may think of two basic securities: the risk free asset (e.g., money market instrument or bond) and the risky asset (e.g., stock). The risk free asset is with price per unit $P_0(t)$ governed by

$$P_0(t_{i+1}) - P_0(t_i) = r(t_i)P_0(t_i)(t_{i+1} - t_i), \qquad (2.2)$$

where $r(t)$ is the short rate. The risky asset is traded and the price process $X(t)$ of the stock is modeled (similar to the example above) by a linear equation as:

$$X(t_{i+1}) - X(t_i) = \mu(t_i)X(t_i)(t_{i+1} - t_i) + \sigma(t_i)X(t_i)(t_{i+1} - t_i)^{1/2}\varepsilon(t_i), \qquad (2.3)$$

where $\mu(t)$ and $\sigma(t) > 0$ are unknown functions. Let $\pi_0(t_i)$ and $\pi_1(t_i)$ be the amounts of wealth invested respectively in bond and stock at time t_i, and $n_0(t_i)$ and $n_1(t_i)$ be the numbers respectively of bond and stock. Further, denote the portfolio of bond and stock as $Y\{X(t_i)\} = n_0(t_i)P_0(t_i) + n_1(t_i)X(t_i)$. Let $Z\{X(t_i)\} = \sigma(t_i)\pi_1(t_i) = \sigma(t_i)n_1(t_i)X(t_i)$. Assume that the wealth process $Y\{X(t_i)\}$ is self-financing (Karatzas and Shreve 1998). Then the average wealth process $\bar{Y}(t)$ satisfies

$$\bar{Y}(t_i) = \left[r(t_i)Y\{X(t_i)\} + \frac{\mu(t_i) - r(t_i)}{\sigma(t_i)}Z\{X(t_i)\} \right]\Delta_i^{1/2} + Z\{X(t_i)\}\varepsilon(t_i). \quad (2.4)$$

The proof for (2.4) is given in the Appendix. Note that $Z\{X(t)\}$ is in general unobservable in the financial market because $\sigma(t)$ is an unknown function and $\pi_1(t)$ is unobservable. However $Z\{X(t)\}$ is certainly a function of interest because it is related to the amount $\pi_1(t)$ of risky investment.

Model (2.4) is of course a special case of the dynamics described in (2.1). When the data $X(t_i)$ and $Y(t_i)$ are available, we can infer the parameter functions $a(t)$, $b(t)$ and $c(t)$, and the variable $Z^2\{X(t)\}$, and then we can predict future increments $\bar{Y}(t)$ and risky investments $Z\{X(t)\}$.

A very natural concern is that, to fully use the information provided by observed data, we would first separately estimate the parameters in (2.2) and (2.3), and then define the estimators of the associated parameters in model (2.1) or (2.4), and investigate the statistical inference. We still use the first example to explain its disadvantages. For point estimation, it is possible because the point estimation does not involve a random relationship between the asset allocation of $n_0(t_i)$ and of $n_1(t_i)$. However, this relation does have impact on statistical inference so that such a two-step estimation does not very much help for inferring model (2.1) or (2.4). First, note that in our case, the model of interest is total asset or increment, but in asset pricing, the numbers $n_0(t_i)$ and $n_1(t_i)$ respectively of risk free and risky assets are highly related. By model (2.2) and (2.3) only, one has difficulty to study some statistical properties, such as the distribution of average increment $\bar{Y}(t_i)$ of wealth. It also causes the difficulty for, say, confidence region construction for the parameters in (2.1) or (2.4), when we do it only through the confidence regions for the parameters in models (2.2) and (2.3). This is because, firstly, correlation between the data in models (2.2) and (2.3) can influence the confidence regions (including the coverage probability and shape of the confidence regions) for the parameters in model (2.1) or (2.4); Secondly, the confidence regions for the parameters in models (2.2) and (2.3) can not be directly transformed into the confidence regions for the associated parameters in model (2.1) or (2.4). Therefore, an indirect investigation in terms of model (2.1) or (2.4) is necessary.

When the normality condition on $\varepsilon(t)$ is not supposed to hold, (2.1) becomes linear semiparametric regression model

$$\begin{cases} \mathsf{E}(\bar{Y}(t)|X(t)) = [a(t) + b(t)Y\{X(t)\} + c(t)Z\{X(t)\}]\Delta_t^{1/2}, \\ \mathrm{Var}(\bar{Y}(t)|X(t)) = Z^2\{X(t)\} \quad \text{for } t \in [0, T]. \end{cases} \quad (2.5)$$

From (2.1) and (2.5), we can see the main difference from the standard semiparametric regression is as follows. In (2.5), the regression function contains an unknown deviation function $Z(x)$, and the observations $\bar{Y}(t)$ and $Y\{X(t)\}$ depend on an observable variable $X(t)$. In this case, these models are different from the

GARCH-M, a time series model. In the GARCH-M, although the mean term also contains the variance, the variance is a squared sum of white noises and then is independent of model variables (Tsay 2002). In our approach, however, $Z\{X(t)\}$ is dependent on $X(t)$ where both $Z\{X(t)\}$ and $X(t)$ have special meanings such as price process and investment process, respectively. Further our models are different from the classical diffusion processes because (2.4) is determined by two systems (2.2) and (2.3), and then $\bar{Y}(t)$ and $Y\{X(t)\}$ depend on yet another variable $X(t)$. Note that the standard diffusion process does not have this property although drift functions in some diffusion processes also contain variance functions; see for example Barndorff-Nielsen and Shephard (2002), and Fan (2005).

2.2 Extension

Note that the above semiparametric models are of linear structure in the sense that the conditional expectation of $\bar{Y}(t)$ is linearly related to the time-dependent parameters $a(t)$, $b(t)$ and $c(t)$. Such a linearity would need a check for it. However, this is beyond the scope of this paper, we leave it to a further study. By contrast, some diffusion processes are nonlinear in parameters (Ait-Sahalia (1996), Chan *et al.* (1992), Fan *et al.* (2003), and Fan 2005). We can verify that if models (2.2) and (2.3) are nonlinear in parameters, the resulting model (2.1) or (2.4) will be nonlinear accordingly. Then nonlinear extension of model (2.1) or (2.4) is of interest. For flexibility of modeling the above process, a nonlinear semiparametric regression with a built-in standard deviation can be defined as

$$\begin{cases} \mathsf{E}\{\bar{Y}(t)|X(t)\} = f[\theta(t), Y\{X(t)\}, Z\{X(t)\}]\Delta_t^{1/2}, \\ \mathrm{Var}\{\bar{Y}(t)|X(t)\} = Z^2\{X(t)\} \ \ \text{for } t \in [0, T], \end{cases} \tag{2.6}$$

where f is a given function and $\theta(t)$ is an unknown p-dimensional parameter vector depending on time t.

3 Estimation for linear semiparametric regression

We now consider estimation for model (2.5). First, for studying the asymptotic properties of the estimators defined below, we need the (locally) strict stationarity of the sequence of $X(t_i)$ in (2.5). However, it is usually not the case because $X(t_i)$ are generated from (2.3) and this equation produces unstationary $X(t_i)$. To satisfy the stationarity condition, a common way is to make a transform on $X(t_i)$. For example a familiar transformation is $S(t_i) = \log X(t_i)$. The process $S(t_i)$ is then often a (locally) stationary process. Correspondingly, $P_0(t_i)$ in (2.2) should be transformed as $P(t_i) = \log P_0(t_i)$. $S(t_i)$ and $P(t_i)$ can then satisfy an equation similar to (2.1) or (2.5). Therefore, without loss of generality, we assume that $X(t_i)$ are (locally) stationary. This is a main condition assumed in the following theorems.

As is stated in Subsection 2.1, $Z^2(x)$ is in general unobservable. But the second equation of (2.5) shows that $Z^2(x)$ is a conditional expectation of $\bar{Y}(t)$ given $X(t) = x$. Then we can use nonparametric method to estimate $Z^2(x)$ by historical observations $(X(t_i), \bar{Y}(t_i))$ $(i = 1, \cdots, n)$ of $(X(t), \bar{Y}(t))$ (see for example

Fan (2005), Fan *et al* (2003), Cai (2007) and Wand and Jones (1995), Simonoff (1996)). Note that, compared with the deviation term $Z(x)$, the expectation term $[a(t) + b(t)Y(t) + c(t)Z\{X(t)\}]\Delta_t^{1/2}$ is negligible if Δ_t is small enough. Then, a simple estimator of the variance is defined as

$$\hat{Z}^2(x_0) = \frac{\sum_{i=1}^{n-2} \bar{Y}^2(t_i) K_h\{X(t_i) - x_0\}}{\sum_{i=1}^{n-2} K_h\{X(t_i) - x_0\}}, \tag{3.1}$$

where $K_h(x) = K(x/h)/h$, $K(x)$ is a kernel function satisfying the regularity conditions given below and h is the bandwidth depending on n. We use this estimation mainly for ease of exposition, other smoothing methods, such as local polynomial(Fan and Gijbels 1996), may also be used.

Let $\mathscr{X} = \{X(t_1), \cdots, X(t_n)\}$ and $\bar{\mathscr{Y}} = \{\bar{Y}(t_1), \cdots, \bar{Y}(t_n)\}$. Without loss of generality, suppose that $0 \leq t_1 \leq t_2 \leq \cdots \leq t_n \leq T$. We need the following conditions: $\Delta_{t_i} = t_{i+1} - t_i = \mathcal{O}(T/n)$ for $i = 1, \cdots, n-2$, the kernel function $K(u)$ and the sequence $\{X(t_1), \cdots, X(t_n)\}$ respectively satisfy the following conditions:

C1) $K(u)$ is symmetric with respect to $u = 0$, has a bounded derivative and support $(-1, 1)$, and

$$\int_{-1}^{1} K(u)du = 1, \ \sigma_K^2 = \int_{-1}^{1} u^2 K(u)du \neq 0,$$
$$\int_{-1}^{1} |u|^j K^k(u)du < \infty \ \ for \ j \leq k = 1, 2. \tag{3.2}$$

C2) $X(t_1), \cdots, X(t_n)$ are ρ-mixing dependent; the ρ-mixing coefficients $\rho(l)$ satisfy $\rho(l) \to 0$ as $l \to \infty$, where

$$\rho(l) = \sup_{X_i, X_{i+l} \in \mathscr{X}} \frac{|\mathsf{E}(X_{i+l}X_i) - \mathsf{E}(X_{i+l})\,\mathsf{E}(X_i)|}{(\,Var(X_{i+l})\ Var(X_i))^{1/2}} \tag{3.3}$$

with $X_i = X(t_i)$.

Condition C1) is standard and condition C2) is commonly used for weakly dependent process; see for example Rosenblatt (1956, 1970), Kolmogorov and Rozanov (1960), Bradley and Bryc (1985), Lu and Lin (1997). Denote $J_K^j = \int_{-1}^{1} u^j K^2(u)du$ for $j = 0, 1, 2$. Furthermore, some common stochastic processes satisfy condition C2) as shown by Bradley (2005). We have the following theorem.

THEOREM 1 *In addition to conditions C1) and C2), assume that in model (2.5), $X(t_i) \in (x_0 - h, x_0 + h)$ have a common probability density $p(x)$; both the functions $p(x)$ and $Z(x)$ have two continuous derivatives; $\{X(t_i) : X(t_i) \in (x_0 - h, x_0 + h)\}$ is a stationary Markov process; the ρ-mixing coefficients satisfy $\rho(l) = \rho^l$ for $0 < \rho < 1$; and $a(t)$, $b(t)$ and $c(t)$ are bounded on $[0, T]$. Then as $nh \to \infty$, $nh^5 \to 0$ and $nh\Delta_t^2 \to 0$,*

$$\sqrt{nh}\{\hat{Z}^2(x_0) - Z^2(x_0)\} \overset{\mathcal{L}}{\longrightarrow} N(0, J_K^0 Z^4(x_0)/p(x_0)).$$

The proof of Theorem 1 is presented in the Appendix. The asymptotic normality given in this theorem is similar to that of classical nonparametric estimation. Here we only need the conditions locally in the neighbor x_0 $(x_0 - h, x_0 + h)$ because the kernel estimation only utilizes the data around x_0. Also, from the proof we can see the ρ-mixing condition can be replaced by $p_l(u, v) - p(u)p(v) = \mathcal{O}(\rho^l)$, where $p_l(u, v)$ is the joint probability density of $\{X(t_i), X(t_{i+l})\}_{i=1}^{n-2}$ and $0 < \rho < 1$.

Furthermore, the proof of this theorem indicates that when the expectation term $[a(t) + b(t)Y(t) + c(t)Z\{X(t)\}]\Delta_t^{1/2}$ is large (when Δ_t is large), the conventional

estimator defined in (3.1) may be no longer consistent. Thus, we define a bias corrected estimator for $Z^2(x_0)$ in terms of the difference between $\bar{Y}(t_{i+1})$ and $\bar{Y}(t_i)$ for successive time points t_i and t_{i+1} as

$$\tilde{Z}^2(x_0) = \frac{\sum_{i=1}^{n-2} \{\bar{Y}(t_{i+1}) - \bar{Y}(t_i)\}^2 K_h\{X(t_i) - x_0\}}{2\sum_{i=1}^{n-2} K_h\{X(t_i) - x_0\}}. \tag{3.4}$$

The above estimator is a new one designed for the model that has an unspecified mean function. The following theorem states the asymptotical normality.

THEOREM 2 *In addition to the conditions of Theorem 1, the conditional probability density $p_l(v|u)$ of $X(t_{i+l})$ given $X(t_i)$ is free of i for $X(t_{i+l}), X(t_i) \in (x_0 - h, x_0 + h)$, and has two continuous derivatives. Then as $nh \to \infty$, $nh^5 \to 0$ and $nh\Delta_t^4 \to 0$,*

$$\sqrt{nh}\{\tilde{Z}^2(x_0) - Z^2(x_0)\} \xrightarrow{\mathcal{L}} N(0, J_K^0 Z^4(x_0)/p(x_0)).$$

The proof is also postponed to the Appendix.

As is shown in the above two theorems, we can estimate the nonparametric function $Z(x)$. Furthermore, according to (2.2) and the commonly used estimation for forward stochastic differential equations, we can also estimate $\sigma(t)$ whose estimator is denoted by $\hat{\sigma}(t)$. From the relation between $Z\{X(t)\}$, $n_1(t)$ and $\pi_1(t)$, we can see that both $n_1(t)$ and $\pi_1(t)$ can be estimated respectively with

$$\hat{n}_1(t) = \frac{\hat{Z}\{X(t)\}}{X(t)\hat{\sigma}(t)} \quad \text{or} \quad \tilde{n}_1(t) = \frac{\tilde{Z}\{X(t)\}}{X(t)\hat{\sigma}(t)}, \tag{3.5}$$

$$\hat{\pi}_1(t) = \frac{\hat{Z}\{X(t)\}}{\hat{\sigma}(t)} \quad \text{or} \quad \tilde{\pi}_1(t) = \frac{\tilde{Z}\{X(t)\}}{\hat{\sigma}(t)}. \tag{3.6}$$

These estimators show that although in some situations, such as in a natural environment, $n_1(t)$ (the number of groups of a species defined in Section 2) cannot be directly observed, it can be estimated.

We now turn to estimating $\beta(t) = (a(t), b(t), c(t))^\tau$. In the remainder of this section, for the convenience of representation, we assume that $\Delta_i = \Delta$ for all i. First, we plug the estimator $\hat{Z}$ or $\tilde{Z}$ into the first equation of (2.5). Then by the local polynomial smoothing (Fan and Gijbels 1996), we can obtain an estimator of $\beta(t)$ as

$$\hat{\beta}(t) = \Delta^{-1/2}(\hat{D}^\tau W \hat{D})^{-1}\hat{D}^\tau W\bar{Y}, \tag{3.7}$$

where $W = \text{diag}\{K_h(t_1 - t), \cdots, K_h(t_n - t)\}$, $\bar{Y} = (\bar{Y}_1, \cdots, \bar{Y}_n)^\tau$ and

$$\hat{D} = \begin{pmatrix} 1 & Y_1 & \hat{Z}_1 \\ \cdots & \cdots & \cdots \\ 1 & Y_n & \hat{Z}_n \end{pmatrix}$$

with $Y_i = Y\{X(t_i)\}$, $\bar{Y}_i = \bar{Y}(t_i)$ and $\hat{Z}_i = \hat{Z}\{X(t_i)\}$ or $\hat{Z}_i = \tilde{Z}\{X(t_i)\}$.

Different from the case for estimating $Z\{X(t)\}$, in which we use state variable $X(t)$ to construct kernel estimators (3.1) and (3.4), we consider time t as a covariate to construct nonparametric estimator of $\beta(t)$ given in (3.7). In the following theorem the local condition on $X(t_i)$ is determined by time t rather than by state variable x.

THEOREM 3 *In addition to the conditions of Theorems 1 and 2, assume that $\beta(t)$ has two continuous derivatives and the matrix $\Omega(t)$ defined below is positive definite. Then, as $nh \to \infty$, $nh^5 \to 0$ and $nh\Delta \to \infty$, we have*

$$\sqrt{nh\Delta}\{\hat{\beta}(t) - \beta(t)\} \xrightarrow{\mathcal{L}} N(0, J_K^0 Z_t^2 \Omega^{-1}(t)),$$

where $\Omega(t) = \begin{pmatrix} 1 & \mathsf{E}(Y_t) & \mathsf{E}(Z_t) \\ \mathsf{E}(Y_t) & \mathsf{E}(Y_t^2) & \mathsf{E}(Y_t Z_t) \\ \mathsf{E}(Z_t) & \mathsf{E}(Y_t Z_t) & \mathsf{E}(Z_t^2) \end{pmatrix}$ *with* $X(t) = X_t$, $Y\{X(t)\} = Y_t$ *and* $Z\{X(t)\} = Z_t$.

Note that unlike the estimator of $Z(x)$ and nonparametric estimator in classical nonparametric regression, which can achieve the standard convergence rate, the convergence rate of $\hat{\beta}(t)$ depends on Δ. Usually, $\Delta \to 0$ and then the estimator $\hat{\beta}(t)$ has a slow convergence rate. The reason why this happens is that the mean term is higher order infinitesimal than the variance term. This argument will be further verified by the simulations in Section 6. On the other hand, since $X(t_1), \cdots, X(t_n)$ are supposed to be locally and identically distributed, the matrix $\Omega(t)$ given in the above theorem is dependent on t.

Particularly, when β is time-independent, similar to (3.7), the estimator is simplified to be

$$\hat{\beta} = \Delta^{-1/2}(\hat{D}^\tau \hat{D})^{-1}\hat{D}^\tau \bar{Y}. \tag{3.8}$$

By a similar argument used in Su and Lin (2009), we can prove that, under some regularity conditions,

$$\sqrt{n\Delta}(\hat{\beta} - \beta) \xrightarrow{\mathcal{L}} N(0, J_K^0 \mathsf{E}(Z^2)\Omega^{-1}), \tag{3.9}$$

where $\Omega = \begin{pmatrix} 1 & \mathsf{E}(Y) & \mathsf{E}(Z) \\ \mathsf{E}(Y) & \mathsf{E}(Y^2) & \mathsf{E}(YZ) \\ \mathsf{E}(Z) & \mathsf{E}(YZ) & \mathsf{E}(Z^2) \end{pmatrix}$.

4 Estimation for nonlinear semiparametric regression

We now consider the estimation for model (2.6). Because the estimator of $Z(x_0)$ is free of the first equation of (2.6), it can be estimated by the same method as that given in (3.1) or (3.4), and then the resulting estimator has the same asymptotic properties as those in Theorems 1 and 2. Thus, we only focus on the estimation for the parameter vector $\theta(t)$ in this section.

We can plug the estimator of $Z(x_0)$ into the first equation of (2.6) and then use the local polynomial smoother (Fan and Gijbels 1996) to obtain an estimator of $\theta(t)$. For example, an estimator of $\theta(t)$, denoted by $\hat{\theta}(t)$, can be obtained by minimizing

$$\hat{Q}(\theta) = \frac{1}{n}\sum_{i=1}^{n-2}\{\bar{Y}_i - \hat{f}_i(\theta)\Delta_i^{1/2}\}^2 K_h(t_i - t) \tag{4.1}$$

over θ, where $\hat{f}_i(\theta) = f(t_i, \theta, X_i, Y_i, \hat{Z}_i)$. Under some regularity conditions, $\hat{\theta}$ can also be obtained by solving the following equation

$$\hat{L}(\theta) = \hat{Q}'(\theta) = -\frac{1}{n}\sum_{i=1}^{n-2}\{\bar{Y}_i - \hat{f}_i(\theta)\Delta_i^{1/2}\}\hat{f}_i'(\theta)\Delta_i^{1/2}K_h(t_i - t) = 0,$$

where $Q'(\theta)$ denotes the derivative of $Q(\theta)$. In the remainder of this section, for convenience of representation, we also assume that $\Delta_i = \Delta$ for all i. The following lemma is a preliminary for the asymptotic result of $\hat{\theta}$.

LEMMA 1 *For nonlinear model (2.6), when the functions $\theta(t)$ and $f(\theta)$ respectively have two and three continuous derivatives, $\mathsf{E}\{f'(\theta(t))f'(\theta(t))^\tau\}$ is positive definite, the conditions on $X(t_i)$ in Theorem 3 hold, we have*

$$\sup_{t\in[0,1]} |\hat{\theta}(t) - \theta(t)| = \mathcal{O}_p(n^{-\varsigma}),$$

where the positive number ς satisfies $n^\varsigma[h^2 + (\log(nh)/nh)^{1/2}] \to 0$.

The proof of the lemma is delayed to the Appendix. The following theorem follows.

THEOREM 4 *Under the conditions of Lemma 1, if $nh^5 \to 0$ and $nh\Delta \to \infty$, then*

$$\sqrt{nh\Delta}\{\hat{\theta}(t) - \theta(t)\} \xrightarrow{\mathcal{L}} N\Big(0, J_K^0 Z_t^2 \{\mathsf{E}[f'\{\theta(t)\}f'\{\theta(t)\}^\tau)]\}^{-1}\Big).$$

The proof of the theorem is also in the Appendix. When the regression function $f\{\theta(t)\}$ is linear in $\theta(t)$, the above theorem is reduced to Theorem 3. As is shown above, $X(t_i), i = 1, \cdots, n$, are supposed to be locally and identically distributed, we then write the matrix $\mathsf{E}[f'\{\theta(t)\}f'\{\theta(t)\}^\tau)]$ as a function of t.

When θ is time-independent, the estimator $\hat{\theta}$ can be defined as the minimizer of

$$\hat{Q}(\theta) = \frac{1}{n} \sum_{i=1}^{n-2} \{\bar{Y}_i - \hat{f}_i(\theta)\Delta_i^{1/2}\}^2$$

over θ. Similar to (3.9), we have that, under some regularity conditions,

$$\sqrt{n\Delta}(\hat{\theta} - \theta) \xrightarrow{\mathcal{L}} N\Big(0, J_K^0 \mathsf{E}(Z^2)\{\mathsf{E}[f'(\theta)f'(\theta)^\tau]\}^{-1}\Big).$$

5 Monte Carlo studies

In this subsection, the models for simulations are chosen to satisfy the stationary condition. The simulations are based on 500 repetitions with the size $n = 300$. We use the mean and mean squared error (MSE) of estimators to evaluate their performance.

Example 1. The main goal of this example is to examine the behavior of estimators of the diffusion term and the drift term in linear semiparametric models. Consider the common riskless asset model

$$P_0(t_{i+1}) - P_0(t_i) = rP_0(t_i)(t_{i+1} - t_i) \tag{5.1}$$

with the Cox-Ingersoll-Ross (CIR) model

$$X(t_{i+1}) - X(t_i) = \kappa\{\theta - X(t_i)\}(t_{i+1} - t_i) + \sigma X^{1/2}(t_i)(t_{i+1} - t_i)^{1/2}\varepsilon(t_i), \tag{5.2}$$

where θ, κ and σ are the time-independent parameters. Note that model (5.1) is the same as (2.2) but model (5.2) differs from (2.3). Here model (5.2) does not

describe the stock price but an interest rate term structure, in which the spot rate $X(t)$ moves around its long-run equilibrium level θ at speed κ. When $2\kappa\theta \geq \sigma^2$, this process is shown to be positive and stationary. By the same notations as used in Section 2, we have

$$
\begin{aligned}
\bar{Y}(t_i) &= \left[n_1\kappa\theta + rY(t_i) - \frac{\kappa+r}{n_1\sigma^2}Z^2\{X(t_i)\}\right]\Delta^{1/2} + Z\{X(t_i)\}\varepsilon(t_i) \\
&\triangleq [a + bY(t_i) + cZ^2\{X(t_i)\}]\Delta^{1/2} + Z\{X(t_i)\}\varepsilon(t_i)
\end{aligned}
\tag{5.3}
$$

and $Z\{X(t_i)\} = n_1\sigma X^{1/2}(t_i)$.

Let $\kappa = 0.214, \theta = 0.086, \sigma = 0.078, r = 0.05, n_0 = 10, n_1 = 10, \Delta = 1/4, X(t_0) = 0.08, P(t_0) = 0.1$. In the estimation, we use the Gaussian kernel $K(t) = \frac{1}{\sqrt{2\pi}}exp(-\frac{t^2}{2})$.

Table 1 reports the simulation results for the estimators of the coefficients $a = n_1\kappa, b = r$ and $c = -\frac{\kappa+r}{n_1\sigma^2}$. The results show that the estimator of b works well, but those for a and c have large bias, and especially for c, the MSE is pretty large. The curves however are fitted well. To be precise, the estimated curves of drift and diffusion are close to the true ones. Figure 1 presents the estimated curves by one simulation.

<center>(Table 1 is about here)</center>
<center>(Figure 1 is about here)</center>

Example 2. As is shown in the previous sections, the asymptotic normality depends on the condition about Δ, but the rates required for estimating the drift and the diffusion are different. Now we examine the behavior of the estimator with different choices of Δ. The model is identical to that in Example 1 except choosing $\Delta = 4/10$, which is larger than that in Example 1. Table 2 reports the simulation results. We see that when Δ is large, the resulting estimators work better than those obtained with small Δ. By contrast, comparing Figure 1 with Figure 2, we see that $Z\{X(t)\}$ is estimated more accurately when Δ is small. On the other hand, the drift term is estimated more accurately when Δ is large. These findings coincide with our theoretical results of course.

<center>(Table 2 is about here)</center>
<center>(Figure 2 is about here)</center>

Example 3. In this example θ is time-dependent as $\theta(t) = 0.086 + \frac{1}{4}(\frac{t}{n\Delta} - 0.5)^2$, the other parameters are fixed at the values as in Example 1. Therefore, $a = a(t) = 0.184 + 0.535(\frac{t}{n\Delta} - 0.5)^2$. The simulations show that the estimator $\hat{a}(t)$ works very well. Figure 3 reports the simulation results about estimated curves and confidence bands respectively with bandwidth $h = 10\Delta$ and $h = 20\Delta$ by one simulation. We may conclude that the estimation of $a(t)$ is not very sensitive to the bandwidth selection.

<center>(Figure 3 is about here)</center>

6 An empirical application

We now illustrate the proposed technology in a real data situation. To investigate the increment process of wealth in the financial market of China, we fit data

published in the Almanac of China's Finance and Banking (2004 and 2005). The dataset consists of the total amount of deposit and the total amount of market capitalization of stock reported daily from 2003 to 2004. To get daily data of deposit, we approximate them by interpolation. Two models are used to fit the data, one is with constant coefficients and the other has varying-coefficients.

We first consider the following linear model

$$\begin{cases} \mathsf{E}\{\bar{Y}(t_i)|X(t_i)\} = [bY\{X(t_i)\} + cZ\{X(t_i)\}]\Delta_t^{1/2}, \\ \mathrm{Var}\{\bar{Y}(t_i)|X(t_i)\} = Z^2\{X(t_i)\} \quad \text{for } t_i = 1, 2, \cdots, 483, \end{cases} \quad (6.1)$$

where parameters b and c are time-independent, $X(t_i), Y(t_i), \bar{Y}(t_i)$ and $Z\{X(t_i)\}$ are defined as those in Section 2. To estimate $Z\{X(t_i)\}$, the bandwidth is chosen to be $\mathrm{Std}\{X(t)\}n^{-1/5}$ and the length of interval is chosen as $\Delta_{t_i} = 1$. After estimating $Z\{X(t_i)\}$, b and c by the methods proposed in the previous sections, we get the fitted values of wealth increment; that is $\hat{\bar{Y}}(t_i) = [\hat{b}Y(t_i) + \hat{c}\hat{Z}\{X(t_i)\}]\Delta_{t_i}^{1/2}$ (hundred million). Figure 4 reports the fitted values $\hat{\bar{Y}}(t_i)$ and real values $\bar{Y}(t_i)$. It shows that the fitted values are in the centric curve of data cloud. Then our model and method work well in this case. Note that $\hat{Z}\{X(t_i)\}$ changes from time to time largely and the estimation value $\hat{c} = 1.471$ is also large relative to $\hat{b} = -0.004$. It implies that the built-in variance (or deviation) in the regression function is not negligible.

(Figure 4 is about here)

We now turn to considering the varying-coefficient linear model as

$$\begin{cases} \mathsf{E}\{\bar{Y}(t_i)|X(t_i)\} = [b(t_i)Y\{X(t_i)\} + c(t_i)Z\{X(t_i)\}]\Delta_t^{1/2}, \\ \mathrm{Var}\{\bar{Y}(t_i)|X(t_i)\} = Z^2\{X(t_i)\} \quad \text{for } t_i = 1, 2, \cdots, 483. \end{cases} \quad (6.2)$$

For comparison, the bandwidth h for estimating $Z\{X(t)\}$ and length of interval are chosen to be the same as those for model (6.1). Figure 5 reports the fitted values of wealth increment for different h and Table 3 lists the corresponding expectation values of squared residuals (MSR). From these figures and Table 3 we can see that, when h is large, the fitted values are smooth but the mean of squared residuals is large. Thus we should choose a suitable h to balance the smoothness and fitting. On the other hand, the means of squared residuals obtained by time-dependent model (6.2) are uniformly smaller than those obtained by the time-independent model (6.1). Then for this dataset, a time-dependent model seems more suitable than a time-independent model.

(Table 3 is about here)
(Figure 5 is about here)

Appendix:Proofs

Proof of (2.4) The wealth process $Y\{X(t_i)\}$ satisfies

$$Y\{X(t_i)\} = n_0(t_i)P_0(t_i) + n_1(t_i)X(t_i).$$

By the existing literature, see for example Karatzas and Shreve (1998), the self-financing implies

$$\Delta n_0(t_i)P_0(t_{i+1}) + \Delta n_1(t_i)X(t_{i+1}) = 0,$$

where $\Delta n_0(t_i) = n_0(t_{i+1}) - n_0(t_i)$ and $\Delta n_1(t_i) = n_1(t_{i+1}) - n_1(t_i)$. By the representation of $Z\{X(t_i)\} = \sigma(t_i)n_1(t_i)X(t_i)$ and the self-financing equation, we see that if $P_0(t_i)$ and $X(t_i)$ in the above wealth equation are replaced by those in (2.2) and (2.3), respectively, then the equation (2.4) follows.

Proof of Theorem 1 Without loss of generality, we assume $\Delta_{t_i} = \Delta$ for all i. Denote

$$\bar{Y}\{X(t_i)\} = \bar{Y}_i, Z\{X(t_i)\} = Z_i,$$
$$K_h = K_h\{X(t) - x_0\}, K_{h,i} = K_h\{X(t_i) - x_0\}, K_i = K\{X(t_i) - x_0\}.$$

Because $X_i, i = 1, \cdots, n$, are identically distributed, and $Z(x)$ and $p(x)$ have two continuous derivatives, we have

$$\mathsf{E}(\frac{1}{n}\sum_{i=1}^{n-2}\bar{Y}_i^2 K_{h,i}) = \mathsf{E}[\bar{Y}^2\{X(t)\}K_h]$$
$$= \mathsf{E}[\mathsf{E}(\bar{Y}^2\{X(t)\}K_h|\mathscr{X})]$$
$$= \frac{1}{h}\int Z^2(u)K\{(u-x_0)/h\}p(u)du + \mathcal{O}_p(\Delta)$$
$$= Z^2(x_0)p(x_0) + \frac{1}{2}h^2\{Z^2(x_0)p(x_0)\}''\sigma_K^2 + \mathcal{O}_p(\Delta) + \mathcal{O}_p(h^2).$$

Similarly,

$$\mathsf{E}(\frac{1}{n}\sum_{i=1}^{n-2}K_{h,i}) = \frac{1}{h}\int K\{(u-x_0)/h\}p(u)du$$
$$= p(x_0) + \frac{1}{2}h^2 p''(x_0)\sigma_K^2 + \mathcal{O}_p(h^2).$$

Furthermore, $\frac{1}{n}\sum_{i=1}^{n-2}K_{h,i}$ is a kernel estimator of the density $p(x_0)$. By the uniform weak consistency of kernel estimator with mixing dependent variables (Peligrad 1992; Kim and Cox 1996), we have

$$\sup_x |\frac{1}{n}\sum_{i=1}^{n-2}K_{h,i} - p(x)| = \mathcal{O}_p[h^2 + \{\log(nh)/nh\}^{1/2}].$$

Then

$$\hat{Z}^2(x_0) - Z^2(x_0)$$
$$= \frac{\sum_{i=1}^{n-2}\bar{Y}_i^2 K_{h,i} - \sum_{i=1}^{n-2}\mathsf{E}(\bar{Y}_i^2 K_{h,i})}{\sum_{i=1}^{n-2}K_{h,i}} + \frac{\sum_{i=1}^{n-2}\mathsf{E}(\bar{Y}_i^2 K_{h,i}) - Z^2(x_0)\sum_{i=1}^{n-2}K_{h,i}}{\sum_{i=1}^{n-2}K_{h,i}}$$
$$= \frac{\sum_{i=1}^{n-2}\bar{Y}_i^2 K_{h,i} - \sum_{i=1}^{n-2}\mathsf{E}(\bar{Y}_i^2 K_{h,i})}{np(x_0)}\{1 + \mathcal{O}_p(1)\} \qquad (A.1)$$
$$+ \frac{\frac{1}{2}h^2(Z^2(x_0)p(x_0) - p(x_0))''\sigma_K^2 + \mathcal{O}_p(h^2) + \mathcal{O}_p(\Delta)}{p(x_0)}\{1 + \mathcal{O}_p(1)\}.$$

(A.1) shows that, $\hat{Z}^2(x_0) - Z^2(x_0)$ is asymptotically equivalent to $\sum_{i=1}^{n-2}\bar{Y}_i^2 K_{h,i} - \sum_{i=1}^{n-2}\mathsf{E}(\bar{Y}_i^2 K_{h,i})$. From Lemma 1 of Politis and Romano (1992) and the relation between the α-mixing condition and the ρ-mixing condition (see, e.g. Theorem 1.1.1 of Lu and Lin 1997), we can see that $\{\bar{Y}_1^2, \cdots, \bar{Y}_n^2\}$ is a ρ-mixing dependent process and the mixing coefficient, denoted by $\rho_Y(l)$, satisfies

$$\sum_{k=1}^{\infty}\rho_Y(2^k) \le C\sum_{k=1}^{\infty}\rho(2^k) = \sum_{k=1}^{\infty}\rho^{2^k} < \infty,$$

where C is a positive constant. By this and Central Limit Theorems for ρ-mixing dependent process (see Ibragimov 1975; or Theorem 6.3 of Fan and Yao 2005; or Theorem 4.0.1 of Lu and Lin 1997), we can obtain that

$$\sqrt{nh}\left\{\frac{1}{n}\sum_{i=1}^{n-2}\bar{Y}_i^2 K_{h,i} - \frac{1}{n}\sum_{i=1}^{n-2}\mathsf{E}(\bar{Y}_i^2 K_{h,i})\right\} \xrightarrow{\mathcal{L}} N(0, D), \qquad (A.2)$$

where D is the limiting variance of $\frac{\sqrt{nh}}{n}\sum_{i=1}^{n-2}\bar{Y}_i^2 K_{h,i}$. The work remaining is to calculate D. By the properties of Markov process and the ρ-mixing coefficients, we have

$$\begin{aligned}
&|\mathsf{E}(\bar{Y}_i^2\bar{Y}_{i+l}^2\{K_{h,i} - \mathsf{E}(K_{h,i})\}\{K_{h,(i+l)} - \mathsf{E}(K_{h,(i+l)})\}|\\
&= |\mathsf{E}[Z_i^2 Z_{i+l}^2\{K_{h,i} - \mathsf{E}(K_{h,i})\}\{K_{h,(i+l)} - \mathsf{E}(K_{h,(i+l)})\}]| + \mathcal{O}(\Delta)\\
&\le \mathsf{E}(Z_i^2 Z_{i+l}^2)\{\operatorname{Var}(K_i)\operatorname{Var}(K_{i+l})\}^{1/2}\rho^l + \mathcal{O}(\Delta)\\
&= C\rho^l + \mathcal{O}(\Delta).
\end{aligned}$$

We can easily verify that, as $n \to \infty$, $\frac{1}{n}\sum_{i=1}^{n-2}\sum_{l=1}^{n-i}\rho^l \to \frac{\rho}{1-\rho}$. Thus

$$\begin{aligned}
D &= \operatorname{Var}(\frac{\sqrt{nh}}{n}\sum_{i=1}^{n-2}\bar{Y}^2 K_{h,i})\\
&= h\,\mathsf{E}[(Z^2\{X(t)\}(K_h - \mathsf{E}(K_h)))^2]\\
&\quad + \frac{2h}{n}\sum_{i=1}^{n-2}\sum_{l=1}^{n-i}\mathsf{E}[\bar{Y}_i^2\bar{Y}_{i+l}^2\{K_{h,i} - \mathsf{E}(K_{h,i})\}\{K_{h,(i+l)} - \mathsf{E}(K_{h,(i+l)})\}]\\
&= h\,\mathsf{E}[Z^4\{X(t)\}(K_h - \mathsf{E}(K_h))^2] + \frac{2h}{n}\mathcal{O}(\sum_{i=1}^{n-2}\sum_{l=1}^{n-i}\rho^l) + \mathcal{O}(h\Delta)\\
&= Z^4(x_0)p(x_0)J_K^0 + \mathcal{O}(1).
\end{aligned} \qquad (A.3)$$

Therefore the result of the theorem follows from $(A.1)$-$(A.3)$. $\qquad\square$

Proof of Theorem 2 The proof is similar to that of Theorem 1 and thus is omitted here. $\qquad\square$

As Theorem 3 is a special case of Theorem 4, we omit the proof details.

Proof of Lemma 1 Let $\theta^0(t)$ be the true value of the parameter in model (2.6) and ν be p-dimensional vector satisfying $\|\nu\| = 1$. It is sufficient to show that, for any given $\eta > 0$, the following holds:

$$P\left\{\inf_{t\in[0,T]}\inf_{\|\nu\|=1}\hat{Q}\{\theta^0(t) + n^{-\varsigma}\nu\} > \hat{Q}\{\theta^0(t)\}\right\} \ge 1 - \eta, \qquad (A.4)$$

where $\hat{Q}$ is defined in (4.1). This is because it implies that with probability at least $1 - \eta$ that $\sup_{t\in[0,T]}\|\hat{\theta}(t) - \theta^0(t)\| \le n^{-\varsigma}$. Hence there exists a minimizer such that $\sup_{t\in[0,T]}\|\hat{\theta}(t) - \theta^0(t)\| = \mathcal{O}_p(n^{-\varsigma})$.

Again by the uniform weak consistency of kernel estimator for mixing dependent variables, we have

$$\begin{aligned}
\sup_x |\hat{Z}^2(x) - Z^2(x)| &= \mathcal{O}_p\{h^2 + (\log(nh)/nh)^{1/2}\},\\
\sup_x |\tilde{Z}^2(x) - Z^2(x)| &= \mathcal{O}_p\{h^2 + (\log(nh)/nh)^{1/2}\}.
\end{aligned}$$

Let $f_i(\theta) = f(t_i, \theta, X_i, Y_i, Z_i)$ and

$$L(\theta) = -\frac{1}{n}\sum_{i=1}^{n-2}\{\bar{Y}_i - f_i(\theta)\Delta^{1/2}\}f_i'(\theta)\Delta^{1/2}K_h(t_i - t).$$

By similar methods and Taylor expansion, there exists a point θ^* between $\theta^0(t) + n^{-\varsigma}\nu$ and $\theta^0(t)$, such that, uniformly for all x and t, noting that f_i contains $Z(x)$,

$$
\begin{aligned}
&\hat{Q}\{\theta^0(t) + n^{-\varsigma}\nu\} - \hat{Q}(\theta^0(t)) \\
&= -n^{-\varsigma}\hat{L}(\theta^0(t)) + \frac{\mathcal{O}_p(n^{-2\varsigma})}{2n} \sum_{i=1}^{n-2} \hat{f}_i'(\theta^*)\hat{f}_i'(\theta^*)^\tau \Delta K_h(t_i - t) \\
&\quad - \frac{\mathcal{O}_p(n^{-2\varsigma})}{2n} \sum_{i=1}^{n-2} (\bar{Y}_i - \hat{f}_i(\theta^*)\Delta^{1/2})\hat{f}_i''(\theta^*)\Delta^{1/2}K_h(t_i - t) \\
&= -n^{-\varsigma}L(\theta^0(t)) + \frac{\mathcal{O}_p(n^{-2\varsigma})}{2n} \sum_{i=1}^{n-2} f_i'(\theta^*)f_i'(\theta^*)^\tau \Delta K_h(t_i - t)(1 + \mathcal{O}_p(1)) \\
&\quad - \frac{\mathcal{O}_p(n^{-2\varsigma})}{2n} \sum_{i=1}^{n-2} (\bar{Y}_i - f_i(\theta^*)\Delta^{1/2})f_i''(\theta^*)\Delta^{1/2}K_h(t_i - t)(1 + \mathcal{O}_p(1)) \\
&= -n^{-\varsigma}L(\theta^0(t)) + \frac{\mathcal{O}_p(n^{-2\varsigma})}{2n} \sum_{i=1}^{n-2} f_i'(\theta^0(t))f_i'(\theta^0(t))^\tau \Delta K_h(t_i - t)(1 + \mathcal{O}_p(1)) \\
&\quad - \frac{\mathcal{O}_p(n^{-2\varsigma})}{2n} \sum_{i=1}^{n-2} (\bar{Y}_i - f_i(\theta^0(t))\Delta^{1/2})f_i''(\theta^0(t))\Delta^{1/2}K_h(t_i - t)(1 + \mathcal{O}_p(1)).
\end{aligned}
\qquad (A.5)
$$

We now compare the convergence rates of the three terms. Again by the uniform consistency of kernel estimator, we have that, uniformly for all t,

$$
\begin{aligned}
&n^{-\varsigma}L\{\theta^0(t)\} = \mathcal{O}_p\left(\Delta n^{-\varsigma}\{h^2 + (\log(nh)/nh)^{1/2}\}\right), \\
&\frac{n^{-2\varsigma}}{2n} \sum_{i=1}^{n-2} f_i'\{\theta^0(t)\}f_i'\{\theta^0(t)\}^\tau \Delta K_h(t_i - t) = \mathcal{O}_p(\Delta n^{-2\varsigma}), \\
&\frac{n^{-2\varsigma}}{2n} \sum_{i=1}^{n-2} [\bar{Y}_i - f_i\{\theta^0(t)\}\Delta^{1/2}]f_i''\{\theta^0(t)\}\Delta^{1/2}K_h(t_i - t) \\
&= \mathcal{O}_p\left[\Delta n^{-2\varsigma}\{h^2 + (\log(nh)/nh)^{1/2}\}\right].
\end{aligned}
$$

These three rates indicate that the second term is slower convergent to zero than the other two terms do. Thus, the first and third terms can be asymptotically negligible, and then $\hat{Q}(\theta^0(t) + n^{-\varsigma}\nu) - \hat{Q}(\theta^0(t))$ is asymptotically equivalent to

$$
\frac{\mathcal{O}_p(n^{-2\varsigma})}{2n} \sum_{i=1}^{n-2} f_i'\{\theta^0(t)\}f_i'\{\theta^0(t)\}^\tau \Delta K_h(t_i - t).
$$

Note that the above is positive uniformly for $t \in [0, T]$. We then get that for any $\eta > 0$ with probability $1 - \eta$, for large n, $\hat{Q}(\theta^0(t) + n^{-\varsigma}\nu) > \hat{Q}(\theta^0(t))$ uniformly for $t \in [0, T]$. This completes the proof of $(A.6)$. $\square$

Proof of Theorem 4 Note that $\hat{L}(\hat{\theta}(t)) = 0$ and $\hat{L}(\theta(t)) = L(\theta(t))(1 + \mathcal{O}_p(1))$. Then by Lemma 1 and Taylor expansion, we have that there exists a point θ^*

between $\hat{\theta}$ and θ such that, uniformly for all x and t,

$$0 = L(\hat{\theta})\{1 + \mathcal{O}_p(1)\}$$

$$= L(\theta)(1 + \mathcal{O}_p(1)) + \frac{1}{n} \sum_{i=1}^{n-2} f_i'(\theta^*) f_i'(\theta^*)^\tau \Delta K_h(t_i - t)\{\hat{\theta}(t) - \theta(t)\}\{1 + \mathcal{O}_p(1)\}$$

$$- \frac{1}{n} \sum_{i=1}^{n-2} \{\bar{Y}_i - f_i(\theta^*)\Delta^{1-\delta}\} f_i''(\theta^*)\Delta^{1/2} K_h(t_i - t)\{\hat{\theta}(t) - \theta(t)\}\{1 + \mathcal{O}_p(1)\}$$

$$= L(\theta)(1 + \mathcal{O}_p(1)) + \frac{1}{n} \sum_{i=1}^{n-2} f_i'(\theta) f_i'(\theta)^\tau \Delta K_h(t_i - t)\{\hat{\theta}(t) - \theta(t)\}\{1 + \mathcal{O}_p(1)\}$$

$$- \frac{1}{n} \sum_{i=1}^{n-2} \{\bar{Y}_i - f_i(\theta)\Delta^{1/2}\} f_i''(\theta)\Delta^{1/2} K_h(t_i - t)\{\hat{\theta}(t) - \theta(t)\}\{1 + \mathcal{O}_p(1)\}. \qquad (A.6)$$

As is shown in the proof of Lemma 1, the first and third terms on the right hand side of (A.6) can be asymptotically negligible uniformly for all x and t. By the uniform weak consistency of kernel estimator,

$$\sup_{t \in [0,1]} \left| \frac{1}{n} \sum_{i=1}^{n-2} f_i'(\theta) f_i'(\theta)^\tau K_h(t_i - t) - \mathsf{E}[f'(\theta(t)) f(\theta(t))^\tau)] \right|$$
$$= \mathcal{O}_p(h^2 + (\log(nh)/nh)^{1/2}). \qquad (A.7)$$

Furthermore, by the same argument in the proof of Theorem 1, we have that $\Delta^{-1/2}\sqrt{nh}\, L(\theta)$ is asymptotically normal. Hence (A.6) and (A.7) result in that

$$\Delta^{1/2}\sqrt{nh}(\hat{\theta}(t) - \theta(t))$$
$$= \left\{ \frac{1}{n} \sum_{i=1}^{n-2} f_i'(\theta) f_i'(\theta)^\tau K_h(t_i - t) \right\}^{-1} \Delta^{-1/2}\sqrt{nh}\, L(\theta) + \mathcal{O}_p(1)$$
$$\xrightarrow{\mathcal{L}} N\left(0, J_K^0 Z_t^2 \{\mathsf{E}[f'(\theta(t)) f(\theta(t))^\tau)]\}^{-1}\right),$$

as desired. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \Box$

References

1. Äit-Sahalia, Y. (1996). Testing continuous-time models of the spot interest rate. *Review of Financial Studies*, **9**, 385-426.
2. Bardorff-Nielsen, O. E. and Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *J. Royal Statistical Society Series B*, **64**, 253-280.
3. Bismut, J. M. (1973). Conjugate convex functions in optimal stochastic control. *J. Math. Ann. Appl.*, **44**, 384-404.
4. Bradley, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Prob. Surveys*, **2**, 107-144.
5. Bradley, R. C. and Bryc, W. (1985). Multilinear forms and measures of dependence between random variables. *J. Multivariate Anal.*, **16**, 335-367.
6. Cai, Z. (2007). Trending Time-varying coefficient time series model with serially correlated errors. *Journal of Econometrics*, **136**, 163-188.
7. Chan, K. C., Karolyi, A. G, Longstaff, F. A. and Sanders, A. B. (1992). An empirical comparison of alternative models of the short-time interest rate. *Journal of Finance*, **47**, 1209-1227.

8. Fan, J. (2005). A selective overview of nonparametric methods in financial econometrics (with discussion). *Statist. Sci.*, **20**, 317-357.

9. Fan, J., Jiang, J., Zhang, C. and Zhou, Z. (2003). Time-dependent diffusion models for term structure dynamics. *Statistica Sinica*, **12**, 965-992.

10. Hart, J. D. (1997). *Nonparametric Smoothing and Lack-of-fit Tests*. Springer-Verlag New York, Inc.

11. Ibragimov, I. A. (1975). A note on the central limit theorem for dependent random variables. *Probab. Theory Appl.* **20**, 134-139.

12. Karatzas, L. and Shreve, S. (1998). *Methods of Mathematical Finance*. Springer, New York.

13. Karoui, N. E. and Peng, S. (1997). Backward stochastic differential equation in finance. *Mathematical Finance*, **7**, 1-71.

14. Kim, T. T. and Cox, D. D. (1996). Uniform strong consistency of kernel estimator under dependence. *Statist. and Prob. Letter*, **26**, 179-185.

15. Kolmogorov, A. N. and Rozanov, U. A. (1960). On the strong mixing conditions of a stationary Gaussian process. *Probab. Theory Appl.*, **2**, 222-227.

16. Lu, C. R. and Lin, Z. Y. (1997). *Limit theories for mixing dependent variables*. Science Press, Beijing.

17. Pardoux, E. and Peng, S. (1990). Adapted solution of backward stochastic differential equation. *Systems control Lett.*, **14**, 55-61.

18. Peligrad, M (1992). Properties of uniform consistency of the kernel estimators of density and regression functions under dependence assumptions. *Stochastics*, **40**, 147 - 168.

19. Politis N. and Romano P. (1992) A general resampling scheme for triangular arrays of $\alpha-$mixing random variables with application to the problem of spectral density estimation. *Ann. statist*, **20**, 1985-2007.

20. Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci.*, **42**, 43-47.

21. Rosenblatt, M. (1970). Density estimates and Markov sequences, in *Nonparametric Techniques in Statistical Inference*, ed. M. Puri, London: Cambridge University Press, pp. 199-210.

22. Simonoff, J. S. (1996). *Smoothing Methods in Statistics*. Springer, New York.

23. Su, Y. X. and Lin, L. (2009). Semi-parametric estimation for Forward-Backward Stochastic Differential Equations. *Communication in Statistics-Theory and Methods*, **38**, 1759-1775.

24. Tsay, R. S. (2002). *Analysis of Financial Time Series*. John Wiley & Sons, inc.

25. Wand, M. P. and Jones, M. C. (1995). *Kernel Smoothing*. Chapman and Hall, London.

Table 1: The Parameter Estimators For Example 1

Parameter	True value	Mean	MSE
a	0.1840	0.2542	0.0341
b	0.0500	0.0499	0.0000
c	-4.3393	-5.6493	14.0454

Table 2: The Parameter Estimators For Example 2

Parameter	True value	Mean	MSE
a	0.1840	0.1999	0.0210
b	0.0500	0.0500	0.0000
c	-4.3393	-4.2135	6.6967

Table 3: Means Of Squared Residuals For Demonstration Study

Model	(5.4)	(5.5) with $h = 10\Delta$	(5.5) with $h = 15\Delta$
MSR	2.164×10^5	1.887×10^5	1.957×10^5

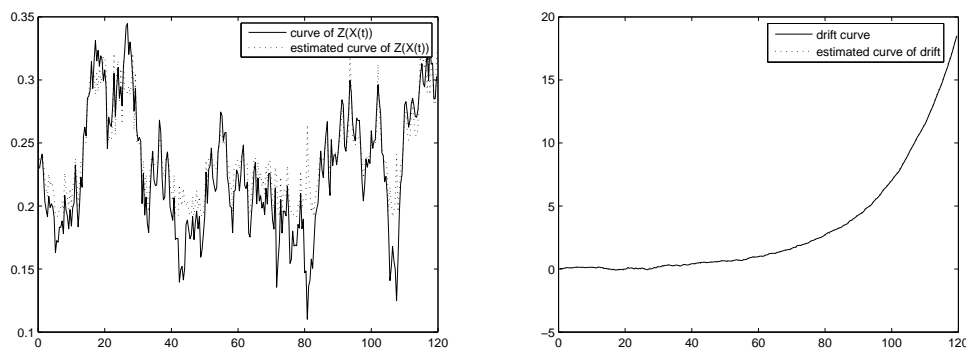| Model	(5.5) with $h = 20\Delta$	(5.5) with $h = 25\Delta$	(5.5) with $h = 40\Delta$
MSR	1.997×10^5	2.021×10^5	2.074×10^5
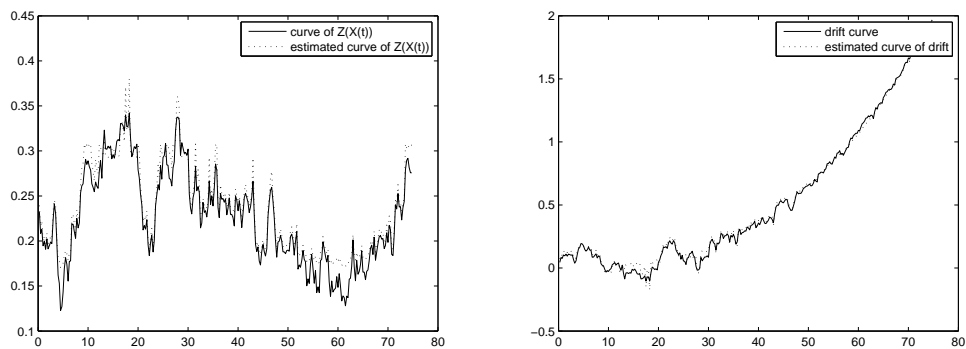
Figure 1. The estimation curves with $\Delta = 1/4$ for model in example 1. The figure on the left presents the true curve of $Z\{X(t)\}$ and estimation curve of $Z\{X(t)\}$, and the figure on the right presents the true curve of drift term and its estimation curve.
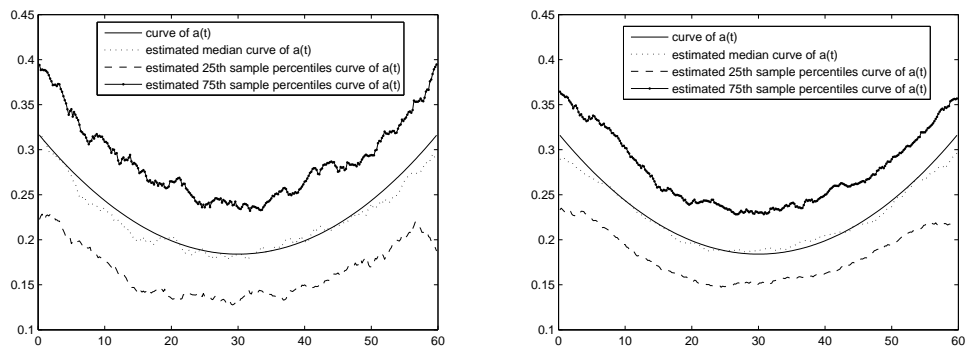

Figure 2. The estimation curves with $\Delta = 4/10$ for model in example 2. The figure on the left presents the true curve of $Z\{X(t)\}$ and estimation curve of $Z\{X(t)\}$, and the figure on the right presents the true curve of drift term and its estimation curve.

Figure 3. The estimated curve and confidence band of $a(t)$ in Example 3. The figures on the left and right present the results with $h = 10\Delta$ and $h = 20\Delta$, respectively.
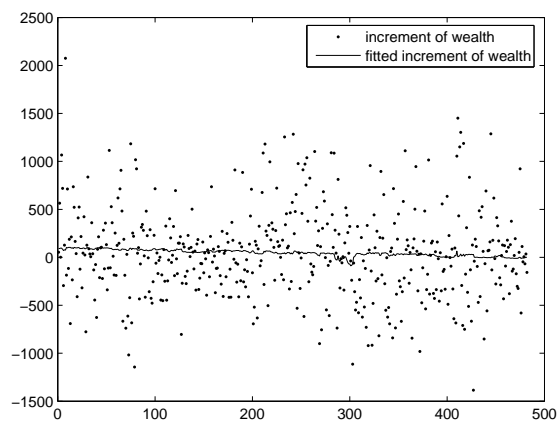
Figure 4. The fitted increment of wealth for model (5.4).
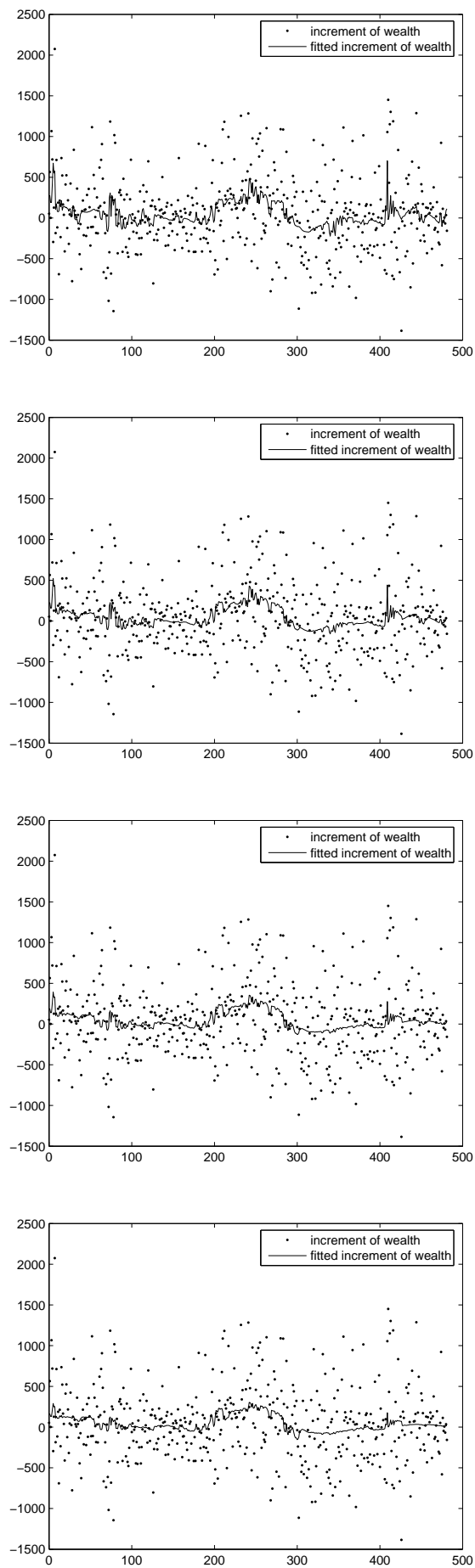
Figure 5. The fitted increment of wealth for model (5.5). The figures on the upper left and right respectively present the results $h = 10\Delta$ and $h = 15\Delta$. The figures on the middle left and right respectively present the results with $h = 20\Delta$ and $h = 25\Delta$.

SFB 649 Discussion Paper Series 2011

For a complete list of Discussion Papers published by the SFB 649,
please visit http://sfb649.wiwi.hu-berlin.de.

001   "Localising temperature risk" by Wolfgang Karl Härdle, Brenda López
       Cabrera, Ostap Okhrin and Weining Wang, January 2011.
002   "A Confidence Corridor for Sparse Longitudinal Data Curves" by
       Shuzhuan Zheng, Lijian Yang and Wolfgang Karl Härdle, January 2011.
003   "Mean Volatility Regressions" by Lu Lin, Feng Li, Lixing Zhu and
       Wolfgang Karl Härdle, January 2011.