

# Fuzzy Varying Coefficient Bilinear Regression of Yield Series

Ting He, Qiujun Lu

College of Science, University of Shanghai for Science and Technology, Shanghai, China  
Email: [hetingfs@163.com](mailto:hetingfs@163.com)

Received 3 June 2015; accepted 26 July 2015; published 29 July 2015

Copyright © 2015 by authors and Scientific Research Publishing Inc.  
This work is licensed under the Creative Commons Attribution International License (CC BY).  
<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

We construct a fuzzy varying coefficient bilinear regression model to deal with the interval financial data and then adopt the least-squares method based on symmetric fuzzy number space. Firstly, we propose a varying coefficient model on the basis of the fuzzy bilinear regression model. Secondly, we develop the least-squares method according to the complete distance between fuzzy numbers to estimate the coefficients and test the adaptability of the proposed model by means of generalized likelihood ratio test with SSE composite index. Finally, mean square errors and mean absolutely errors are employed to evaluate and compare the fitting of fuzzy auto regression, fuzzy bilinear regression and fuzzy varying coefficient bilinear regression models, and also the forecasting of three models. Empirical analysis turns out that the proposed model has good fitting and forecasting accuracy with regard to other regression models for the capital market.

## Keywords

Fuzzy Varying Coefficient Bilinear Regression Model, Fuzzy Financial Assets Yield, Least-Squares Method, Generalized Likelihood Ratio Test, Forecast

---

## 1. Introduction

Researchers usually expect a more reliable estimate of dynamic indicators of the market data through rational design and adjustment with more flexible and applicable models when they study the financial assets price changes. The actual financial data which are often given in the form of interval are not only random but also contain fuzziness. Therefore, taking the innate fuzziness of actual financial data into account, starting with interval financial observed data and combining the features of interval financial series and finally establishing an analytical mode are big problems in analyzing assets price changes.

Since Zadeh [1] put forward the fuzzy set in 1965, the fuzzy set theory has been widely used in social science

research, especially in the economic construction, financial investment, capital market operation and management etc. For interval financial observed data, Li Zhuyu *et al.* [2] defined the stationary of fuzzy financial time series as well as fuzzy financial assets yield series and then constructed a  $p$ th-order fuzzy auto regression (FAR) model, and also estimated the unknown coefficients by using fuzzy linear program (FLP) with satisfying the minimum fuzzy index and real meaning of financial yields. The result of estimation in [2] shows that yields convergence and fluctuation have the same trends; in addition, the model only can apply to centralized fuzzy data. This leads to deviations with the reality. In order to reflect the dynamic changes in financial yields in a period of time, Li Zhuyu *et al.* [4] learned the idea from D'Urso and T. Gastaldi [3] that fluctuations depended on the centers to some extent in a dynamic process, then built the fuzzy bilinear regression (FBR) model with fuzzy financial yields centers and fluctuations respectively, and estimated the unknown coefficients by fuzzy least-squares (FLS) method. Wang Donghua [5] suggested that the method of fuzzy linear program (FLP) was relatively simple so as to get a wide range result and was worse for our application.

Due to the influence of various kinds of social factors, financial asset price changes with non-linear dynamic characteristics, so the traditional method of estimation in the description of the non-linear problems tends to have larger error when modelling. Financial assets yield prediction model in literature [4] still needs to combine the explanatory ability on model, the relationship between the dynamic data, the fitting precision and prediction precision more ideally. Varying coefficient models (also called functional coefficient model) can effectively avoid the problem of the curse of dimensionality because of its obvious flexible model structure, and have a distinct advantage in exploring non-linear dynamic characteristics, reducing model specification errors, describing the features of data as well as forecasting. This article generalizes the fuzzy model in literature [4] to varying coefficient model, which is called fuzzy varying coefficient bilinear regression (FVCBR) model, by the correspondence between financial yields and the symmetric numbers used to depict its fuzziness and use fuzzy least-squares (FLS) method to deduce the estimator. Additionally, we test the adaptability of the proposed model by means of generalized likelihood ratio test. Mean square errors and mean absolutely errors are employed to evaluate and compare the fitting of fuzzy auto regression (FAR), fuzzy bilinear regression (FBR) and fuzzy varying coefficient bilinear regression (FVCBR) models, and also the forecasting of three models in and out of the sample period.

## 2. The Distance of Fuzzy Number Space

In order to introduce the distance of fuzzy number space, we first need to give the introduction to the symmetric numbers and its operational properties.

Definition 1 [6]. A fuzzy number  $\tilde{A}$  is a fuzzy set of the real line  $R$  with satisfying the following conditions:

- 1) there exists an  $x_0 \in R$  such that  $A(x_0) = 1$ ;
- 2) for any  $\alpha \in (0, 1]$ , the  $\alpha$ -level set  $A_\alpha = \{x | A(x) \geq \alpha\}$  is an interval number.

The set of all the fuzzy numbers is denoted by  $\tilde{R}$ .

Definition 2 [6]. A fuzzy number  $\tilde{A}$  with the following membership function:

$$A(x) = L\left(\frac{x-a}{\alpha}\right) \quad (1)$$

is called a symmetric fuzzy number and usually denoted by  $(a, \alpha)_L$ , where  $a$  ( $a \in R$ ) and  $\alpha$  ( $\alpha > 0$ ) are the center and spread of  $\tilde{A}$  respectively. Besides,  $L(\cdot)$  is a strictly decreasing function on  $[0, +\infty)$  with  $L(0) = 1$  and  $L(x) = L(-x)$ . When  $\alpha = 0$ , symmetric fuzzy number  $\tilde{A}$  becomes an ordinary real number  $a$  and is denoted by  $(a, 0)_L = a$ .

The set of all the symmetric fuzzy numbers is denoted by  $\tilde{R}_L$ . Symmetric fuzzy number  $(a, \alpha)_L$  is called symmetric triangle fuzzy number if  $L(x) = \max(0, 1 - |x|)$ .

Proposition 1 [6]. Suppose  $(a, \alpha)_L, (b, \beta)_L \in \tilde{R}_L$  and  $k \in R$ , therefore:

- 1)  $k \cdot (a, \alpha) = (ka, |k|\alpha)_L$ ;
- 2)  $(a, \alpha)_L + (b, \beta)_L = (a+b, \alpha+\beta)_L$ .

Let  $\tilde{A}$  and  $\tilde{B}$  be two fuzzy numbers. Xu [7] proposed a formula that defines complete distance between fuzzy numbers generalized by distance between interval numbers:

$$d(\tilde{A}, \tilde{B}) = \left( \int_0^1 f(\lambda) \bar{d}^2(A_\lambda, B_\lambda) d\lambda \right)^{\frac{1}{2}} \quad (2)$$

where  $f(\lambda)$  is an increasing function on  $[0,1]$  satisfying  $f(0) = 0$  and  $\int_0^1 f(\lambda) d\lambda = 1/2$ ,  $A_\lambda = [a_1(\lambda), a_2(\lambda)]$  and  $B_\lambda = [b_1(\lambda), b_2(\lambda)]$  are  $\lambda$ -level sets of  $\tilde{A}$  and  $\tilde{B}$  respectively and

$$\bar{d}^2(A_\lambda, B_\lambda) = [a_1(\lambda) - b_1(\lambda)]^2 + [a_2(\lambda) - b_2(\lambda)]^2.$$

As pointed out in [7], the monotonically increasing function  $f(\lambda)$  emphasizes the contribution of higher values of  $\lambda$  to the distance between  $\tilde{A}$  and  $\tilde{B}$ . Furthermore,  $f(0) = 0$  and  $\int_0^1 f(\lambda) d\lambda = 1/2$  ensure  $d(\tilde{A}, \tilde{B})$  is generalized by ordinary distance. In fact, we usually let  $f(\lambda) = \lambda$ .

When  $\tilde{A} = (a, s_A)_L$  and  $\tilde{B} = (b, s_B)_L$  are two symmetric fuzzy numbers, the  $\lambda$ -level sets of  $\tilde{A}$  and  $\tilde{B}$  are respectively  $A_\lambda = [a - L^{-1}(\lambda)s_A, a + L^{-1}(\lambda)s_A]$  and  $B_\lambda = [b - L^{-1}(\lambda)s_B, b + L^{-1}(\lambda)s_B]$ , therefore  $\bar{d}^2(A_\lambda, B_\lambda) = 2(a-b)^2 + 2(L^{-1}(\lambda))^2 (s_A - s_B)^2$  and the distance (2) becomes

$$d(\tilde{A}, \tilde{B}) = \left( (a-b)^2 + \Delta(L, f)(s_A - s_B)^2 \right)^{\frac{1}{2}} \quad (3)$$

where  $\Delta(L, f) = 2 \int_0^1 f(\lambda) (L^{-1}(\lambda))^2 d\lambda$ .

Proposition 2<sup>1</sup>. If  $\tilde{A}, \tilde{B}, \tilde{C} \in \tilde{R}_L$  and  $k \in R$ , therefore:

- 1)  $d(k\tilde{A}, k\tilde{B}) = |k|d(\tilde{A}, \tilde{B})$ ;
- 2)  $d(\tilde{A}, \tilde{B}) = d(\tilde{A} + \tilde{C}, \tilde{B} + \tilde{C})$ .

Proposition 3<sup>2</sup>.  $(\tilde{R}_L, d)$  is separable distance space.

### 3. The Estimation and Test of Model

Suppose that  $\{[x_t^l, x_t^u], t = 1, 2, \dots, n\}$  is a given interval series, in which  $x_t^l$  and  $x_t^u$  are respectively the lowest and highest prices of financial products. Fuzzy financial time series  $\{\tilde{x}_t = (x_t, s_t)_L, t = 1, 2, \dots, n\}$  is the fuzzy depiction of interval series  $\{[x_t^l, x_t^u], t = 1, 2, \dots, n\}$ , where  $x_t = 0.5(x_t^l + x_t^u)$  represents the average price of financial product on  $t$ -th day and the series of centers  $\{x_t\}$  depicts the trend of convergence of financial product price;  $s_t = 0.5(x_t^u - x_t^l)$  is the radius of fluctuations in price and the series of spreads  $\{s_t\}$  depicts the uncertainty of  $\{\tilde{x}_t\}$  on  $t$ -th day, namely the magnitude of nonrandomfluctuations. Series  $\{\tilde{r}_t\}$  is called fuzzy financial yields series and denoted by  $\tilde{r}_t = \nabla \ln \tilde{x}_t$ . Furthermore,  $\{\tilde{r}_t\}$  can also be put in another form of symmetric fuzzy numbers series  $\{\tilde{r}_t\} = \{(c_t, u_t)_L, t = 2, \dots, n\}$ , then the series of centers reflects the convergence being the following formula:

$$\{c_t\} = \left\{ \frac{1}{2} \left[ \ln \frac{x_t^u}{x_{t-1}^l} + \ln \frac{x_t^l}{x_{t-1}^u} \right], t = 2, \dots, n \right\} \quad (4)$$

the series of spreads reflects the magnitude of nonrandomfluctuations and is expressed as following:

$$\{u_t\} = \left\{ \frac{1}{2} \left[ \ln \frac{x_t^u}{x_{t-1}^l} - \ln \frac{x_t^l}{x_{t-1}^u} \right], t = 2, \dots, n \right\} \quad (5)$$

<sup>1</sup>It is proved by proposition 1.

<sup>2</sup>It is proved by the density of the set of rational number in the set of real numbers.

### 3.1. Fuzzy Auto Regression (FAR) Model

If  $\{\tilde{r}_t\} = \{(c_t, u_t)_L, t = 2, \dots, n\}$  is a conditional stationary series, a  $p$ -th order fuzzy auto regression model [2] can be established through  $\tilde{r}_t$  and  $\tilde{r}_{t-1}, \dots, \tilde{r}_{t-p}$ :

$$\tilde{r}_t = A_1 \tilde{r}_{t-1} + A_2 \tilde{r}_{t-2} + \dots + A_p \tilde{r}_{t-p} + \tilde{\varepsilon}_t, t = p+1, \dots, n \quad (6)$$

where  $A_1, A_2, \dots, A_p \in R$  are auto regressive parameters and  $\tilde{\varepsilon}_t$  are the fuzzy errors on  $t$ -th day.

The  $p$ -th order fuzzy auto regression model has two obvious limitations, one of those is that the fuzzy auto regression as (6) will finally leads to the same trends of centers and spreads, another is that formula (6) can be equivalent to the combination of auto regression models built by two ordinary time series with centers and spreads respectively. However, only centralized data can omit the constant terms in auto regression models with traditional time series while in the common case of data, the constant terms usually cannot be ignored.

### 3.2. Fuzzy Bilinear Regression (FBR) Model

The fuzzy bilinear regression (FBR) model in [4] set up by the centers and spreads of fuzzy financial yields series respectively solves those two problems fundamentally.

$$\begin{cases} c_t = \alpha_0 + \alpha_1 c_{t-1} + \dots + \alpha_p c_{t-p} + \varepsilon_t \\ u_t = \beta_0 + \beta_1 u_{t-1} + \dots + \beta_q u_{t-q} + \gamma c_t + e_t \end{cases}, t = k+1, \dots, n \quad (7)$$

where  $k = \max(p, q)$ ,  $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \dots, \alpha_p)^T$  and  $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_q, \gamma)^T$  are unknown coefficient vectors and  $\varepsilon_t$  and  $e_t$  are error terms of the series of centers and spreads respectively on  $t$ -th day.

The model (7) respectively describes the auto regression relationship between the convergence of fuzzy financial yields on  $t$ -th day and the  $p$ -th order lagging value of yields, as well as the auto regression relationship between fluctuations and its  $p$ -th order lagging value. Meanwhile, the model also expresses the interdependent relationship between the fluctuations and the convergence of fuzzy financial yields. When  $p = q$ ,  $\alpha_0 = \beta_0 = 0$ ,  $\alpha_i = \beta_i$  ( $1 \leq i \leq p$ ) and  $\gamma = 0$ , the model (7) will turn into the model (6) in form. Therefore, on one hand, the model (7) is formally generalized by the model (6), and on other hand, the model (7) can improve the explanatory ability to fuzzy financial yields.

### 3.3. Fuzzy Varying Coefficient Bilinear Regression (FVCBR) Model

In the FAR model and FBR model, the regression coefficients being constants suggests that explanatory variables impact on explained variables constantly during the sample period. While in the study of financial assets yields prediction, the yields is a time series and the level of influence of various factors in different intervals will change. Consequently, linear models with constant coefficients are unfit for prediction and the varying coefficient models should put to use. The form of FVCBR model shows below:

$$\begin{cases} c_t = \alpha_0(t) + \alpha_1(t)c_{t-1} + \dots + \alpha_p(t)c_{t-p} + \varepsilon_t \\ u_t = \beta_0(t) + \beta_1(t)u_{t-1} + \dots + \beta_q(t)u_{t-q} + \gamma(t)c_t + e_t \end{cases}, t = k+1, \dots, n \quad (8)$$

In which the model coefficients are the function of  $t$ .

#### 3.3.1. Coefficients Estimate of Fuzzy Varying Coefficient Bilinear Regression (FVCBR) Model

Considering the number of unknown coefficients of varying coefficient model will multiply as the sample size gets large, thus fitting models with traditional methods of estimation are not appropriate. Because of the above-mentioned disadvantages, the restricted weighted least-squares estimation is much used at present. Generally speaking, suppose that  $t_0$  is a given point in the domain of variable  $t$ , the existing observations  $\{(c_t, u_t)_L\}$  all provide information to the model coefficients  $\alpha_0(t_0), \alpha_1(t_0), \dots, \alpha_p(t_0)$ ,  $\beta_0(t_0), \beta_1(t_0), \dots, \beta_q(t_0)$ ,  $\gamma(t_0)$  around  $t_0$ , but different observations work different effects. The importance is shown by assigning each observation a weight, and the  $t$ -th weight corresponds to the  $t$ -th observation  $(c_t, u_t)_L$ .

The kernel estimation is used to estimate the unknown coefficients  $\alpha_0(t), \alpha_1(t), \dots, \alpha_p(t)$ ,  $\beta_0(t), \beta_1(t), \beta_2(t), \dots, \beta_q(t)$ ,  $\gamma(t)$ . On the basis of distance (3) and the principle of the kernel smoothing in statistics, we formulate the following restricted weighted least-square problem. That is, the objective function is

$$\begin{aligned}
& \Psi(\alpha_0(t_0), \alpha_1(t_0), \dots, \alpha_p(t_0), \beta_0(t_0), \dots, \beta_q(t_0), \gamma(t_0)) \\
&= \sum_{t=k+1}^n d^2 \left( (c_t, u_t)_L, \left( \left( \alpha_0(t_0) + \sum_{i=1}^p \alpha_i(t_0) c_{t-i} \right), \left( \beta_0(t_0) + \sum_{j=1}^q \beta_j(t_0) u_{t-j} + \gamma(t_0) c_t \right) \right) \right)_L K_h(t-t_0) \\
&= \sum_{t=k+1}^n \left( \left( c_t - \alpha_0(t_0) - \sum_{i=1}^p \alpha_i(t_0) c_{t-i} \right)^2 + \Delta(L, f) \left( u_t - \beta_0(t_0) - \sum_{j=1}^q \beta_j(t_0) u_{t-j} - \gamma(t_0) c_t \right)^2 \right) K_h(t-t_0) \\
&= \sum_{t=k+1}^n \left( c_t - \alpha_0(t_0) - \sum_{i=1}^p \alpha_i(t_0) c_{t-i} \right)^2 K_h(t-t_0) + \Delta(L, f) \sum_{t=k+1}^n \left( u_t - \beta_0(t_0) - \sum_{j=1}^q \beta_j(t_0) u_{t-j} - \gamma(t_0) c_t \right)^2 K_h(t-t_0)
\end{aligned}$$

And it is minimized with respect to  $\alpha_0(t_0), \alpha_1(t_0), \dots, \alpha_p(t_0), \beta_0(t_0), \beta_1(t_0), \dots, \beta_q(t_0), \gamma(t_0)$  where  $K_h(\cdot) = K(\cdot/h)/h$ , with  $K(\cdot)$  being a given kernel function and  $h$  being the smoothing parameter.

The restricted weighted least-squares problem is equivalent to minimizing the following equations:

$$\begin{cases} \sum_{t=k+1}^n \left( c_t - \alpha_0(t_0) - \sum_{i=1}^p \alpha_i(t_0) c_{t-i} \right)^2 K_h(t-t_0) \\ \sum_{t=k+1}^n \left( u_t - \beta_0(t_0) - \sum_{j=1}^q \beta_j(t_0) u_{t-j} - \gamma(t_0) c_t \right)^2 K_h(t-t_0) \end{cases} \quad (9)$$

Let  $\mathbf{1} = (1, 1, \dots, 1)^T$ ,  $\mathbf{c} = (c_{k+1}, c_{k+2}, \dots, c_n)^T$ ,  $\mathbf{u} = (u_{k+1}, u_{k+2}, \dots, u_n)^T$ ,  $\mathbf{c}_{-i} = (c_{k+1-i}, c_{k+2-i}, \dots, c_{n-i})^T$  ( $1 \leq i \leq p$ ),  $\mathbf{u}_{-j} = (u_{k+1-j}, u_{k+2-j}, \dots, u_{n-j})^T$  ( $1 \leq j \leq q$ ),  $\mathbf{X} = (\mathbf{1}, \mathbf{c}_{-1}, \mathbf{c}_{-2}, \dots, \mathbf{c}_{-p})$ ,  $\mathbf{Y} = (\mathbf{1}, \mathbf{u}_{-1}, \mathbf{u}_{-2}, \dots, \mathbf{u}_{-q}, \mathbf{c})$ ,  $\boldsymbol{\alpha}(t_0) = (\alpha_0(t_0), \alpha_1(t_0), \dots, \alpha_p(t_0))^T$ ,  $\boldsymbol{\beta}(t_0) = (\beta_0(t_0), \beta_1(t_0), \dots, \beta_q(t_0), \gamma(t_0))^T$ ,

$\mathbf{W}(t_0) = \text{diag}(K_h(k+1-t_0), K_h(k+2-t_0), \dots, K_h(n-t_0))$ .

We here assume that the inverse matrix of  $\mathbf{X}^T \mathbf{W}(t_0) \mathbf{X}$  and  $\mathbf{Y}^T \mathbf{W}(t_0) \mathbf{Y}$  exist for each  $t_0$ . Then the solution of the weighted least-squares problem (9), that is, the estimation of the vectors  $\boldsymbol{\alpha}(t_0)$  and  $\boldsymbol{\beta}(t_0)$  of the fuzzy coefficients can be obtained using matrix notation as

$$\begin{cases} \hat{\boldsymbol{\alpha}}(t_0) = (\mathbf{X}^T \mathbf{W}(t_0) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}(t_0) \mathbf{c} \\ \hat{\boldsymbol{\beta}}(t_0) = (\mathbf{Y}^T \mathbf{W}(t_0) \mathbf{Y})^{-1} \mathbf{Y}^T \mathbf{W}(t_0) \mathbf{u} \end{cases} \quad (10)$$

From (10) we can see the factor  $\Delta(L, f)$  is independent of the unknown parameters.

If the observed values of explanatory variables are known, we can obtain the fitted values of explained variables at  $t_0$ . Furthermore, performing the above estimation procedure at  $t_0 = k+1, k+2, \dots, n$  respectively, we can obtain the estimation of explained variables during the whole study period.

$$\begin{cases} \hat{c}_t = \hat{\alpha}_0(t) + \hat{\alpha}_1(t) c_{t-1} + \dots + \hat{\alpha}_p(t) c_{t-p} \\ \hat{u}_t = \hat{\beta}_0(t) + \hat{\beta}_1(t) u_{t-1} + \dots + \hat{\beta}_q(t) u_{t-q} + \hat{\gamma}(t) c_t \end{cases}, t = k+1, k+2, \dots, n \quad (11)$$

As in statistical nonparametric regression, two kinds of kernel functions are commonly used and one of them is Gaussian kernel:

$$K(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right) \quad (12)$$

and the other is Beta kernel [8]:

$$K_\gamma(t) = \frac{1}{\text{Beta}(1/2, \gamma+1)} (1-t^2)^\gamma I(|t| \leq 1) \quad (13)$$

Here, we use the distance (3) used to fuzzify the cross-validation procedure [8] in statistics for selecting the optimal value of the smoothing parameter  $h$ , that is, let

$$\begin{aligned}
CV(h) &= \sum_{t=k+1}^n d^2 \left( (c_t, u_t)_L, (\hat{c}_{(-t)}(h), \hat{u}_{(-t)}(h))_L \right) \\
&= \sum_{t=k+1}^n \left( c_t - \hat{\alpha}_0^{(-t)}(t, h) - \sum_{i=1}^p \hat{\alpha}_i^{(-t)}(t, h) c_{t-i} \right)^2 \\
&\quad + \Delta(L, f) \sum_{t=k+1}^n \left( u_t - \hat{\beta}_0^{(-t)}(t, h) - \sum_{j=1}^q \hat{\beta}_j^{(-t)}(t, h) u_{t-j} - \hat{\gamma}^{(-t)}(t, h) c_{(-t)}(h) \right)^2
\end{aligned} \tag{14}$$

where  $\hat{\alpha}_i^{(-t)}(t, h)$  ( $0 \leq i \leq p$ ),  $\hat{\beta}_j^{(-t)}(t, h)$  ( $0 \leq j \leq q$ ) and  $\hat{\gamma}^{(-t)}(t, h)$  are the resulting estimates of the centers and spreads of the fuzzy coefficients under  $h$  through deleting the  $t$ -th observation and computing the estimates according to the restricted weighted least-squares described above. Thus,

$$\begin{cases} \hat{c}_{(-t)}(h) = \hat{\alpha}_0^{(-t)}(t, h) + \sum_{i=1}^p \hat{\alpha}_i^{(-t)}(t, h) c_{t-i} \\ \hat{u}_{(-t)}(h) = \hat{\beta}_0^{(-t)}(t, h) + \sum_{j=1}^q \hat{\beta}_j^{(-t)}(t, h) u_{t-j} + \hat{\gamma}^{(-t)}(t, h) c_{(-t)}(h) \end{cases}, t = k+1, k+2, \dots, n \tag{15}$$

Then, select  $h_0$  as the optimal value of the smoothing parameter such that  $CV(h_0) = \min_{h>0} CV(h)$ .

### 3.3.2. The Test of the FVCBR Model

As point out in [4], FBR model can be used to analyze the fuzzy financial yields series, but whether the analysis model with constant coefficients are enough to embody the dynamic changes of yields is a question worth thinking about, that is, whether the effects of explanatory variables on explained variables will significantly change or nor as the time goes by? For that reason, we need to test the constant coefficients hypothesis, and

$$H_0 : \alpha(t) = \hat{\alpha}, \beta(t) = \hat{\beta}$$

where  $\hat{\alpha}$  and  $\hat{\beta}$  are estimates of coefficients of constant coefficients model.

Here we use the generalized likelihood ratio (GLR) test [9]. Let

$$\begin{aligned}
RSS_0 &= \sum_{t=k+1}^n \left( c_t - \hat{\alpha}_0 - \sum_{i=1}^p \hat{\alpha}_i c_{t-i} \right)^2 + \sum_{t=k+1}^n \left( u_t - \hat{\beta}_0 - \sum_{j=1}^q \hat{\beta}_j u_{t-j} - \hat{\gamma} c_t \right)^2 \\
&= (\mathbf{c} - \mathbf{X}\hat{\alpha})^T (\mathbf{c} - \mathbf{X}\hat{\alpha}) + (\mathbf{u} - \mathbf{Y}\hat{\beta})^T (\mathbf{u} - \mathbf{Y}\hat{\beta})
\end{aligned} \tag{16}$$

$$\begin{aligned}
RSS_1 &= RSS_1^{(1)} + RSS_1^{(2)} \\
&= \sum_{t=k+1}^n \left( c_t - \hat{\alpha}_0(t) - \sum_{i=1}^p \hat{\alpha}_i(t) c_{t-i} \right)^2 + \sum_{t=k+1}^n \left( u_t - \hat{\beta}_0(t) - \sum_{j=1}^q \hat{\beta}_j(t) u_{t-j} - \hat{\gamma}(t) c_t \right)^2 \\
&= (\mathbf{c} - \mathbf{X}\hat{\alpha}(t))^T (\mathbf{c} - \mathbf{X}\hat{\alpha}(t)) + (\mathbf{u} - \mathbf{Y}\hat{\beta}(t))^T (\mathbf{u} - \mathbf{Y}\hat{\beta}(t))
\end{aligned} \tag{17}$$

Here  $RSS_0$  and  $RSS_1$  are residual sum of squares of null hypothesis and the whole space respectively. The GLR statistic is

$$T = \frac{n-k}{2} \ln \frac{RSS_0}{RSS_1} \tag{18}$$

Then, the asymptotic distribution of statistic can be generalized by the method of Bootstrap. Specific steps are as follows:

Step 1. Let  $\hat{\sigma}_1^2 = \frac{1}{n-k} RSS_1^{(1)}$ , and a series of random numbers obeyed  $N(0, \hat{\sigma}_1^2)$  will be generated, that is,  $\tilde{\varepsilon}_t \sim N(0, \hat{\sigma}_1^2)$  where  $t = k+1, \dots, n$ , and let  $\tilde{\mathbf{c}} = \mathbf{X}\hat{\alpha} + \tilde{\varepsilon}$ ;

Let  $\hat{\sigma}_2^2 = \frac{1}{n-k} RSS_1^{(2)}$ , and a series of random numbers obeyed  $N(0, \hat{\sigma}_2^2)$  will be generated, that is,

$\tilde{\epsilon}_t \sim N(0, \hat{\sigma}_2^2)$  where  $t = k+1, \dots, n$ , and let  $\tilde{\mathbf{u}} = \mathbf{Y}\hat{\boldsymbol{\beta}} + \tilde{\boldsymbol{\epsilon}}$ ;

Step 2. Use the sample data to construct GLR statistics  $\tilde{T}$ ;

Step 3. Repeat Step 1 and Step 2  $m$  times and then get  $m$  GLR statistics  $\tilde{T}_1, \tilde{T}_2, \dots, \tilde{T}_m$ ;

Step 4. The asymptotic distribution of GLR statistic  $T$  of null hypothesis is expressed by the empirical distribution of  $\tilde{T}_1, \tilde{T}_2, \dots, \tilde{T}_m$ , that is

$$\hat{F}_m(t) = \frac{1}{m} \sum_{i=1}^m I\{\tilde{T}_i \leq t\} \quad (19)$$

## 4. Empirical Analysis

Now, we will fit and forecast the fuzzy financial yields with the FVCBR model. The database consists of 119 observations of the SSE Composite index from July 17, 2014 to January 9, 2015. Assume that

$\{[x_t^l, x_t^u], t = 1, 2, \dots, 119\}$  is a series of interval observations of the SSE Composite index, in which  $x_t^l$  and  $x_t^u$  represent the maximum and minimum of the SSE Composite index at time  $t$  respectively, as shown in

**Figure 1**. In order to obtain the yields series  $\{\tilde{r}_t\}, t = 2, 3, \dots, 119$ , as shown in **Figure 2**, we deal with the fuzzy series  $\{\tilde{x}_t\}$  by logarithmic transformation and then first order difference. Further, we divide the data  $\{(c_t, u_t)_L, t = 2, \dots, 119\}$  into two parts. One is the fitted samples  $\{(c_t, u_t)_L, t = 2, \dots, 104\}$  and the other is the forecast samples  $\{(c_t, u_t)_L, t = 105, \dots, 119\}$ , and we will test the empirical results and prediction abilities of models in and out of the sample period.

### 4.1. The Varying Coefficient Bilinear Regression Model of Fuzzy Financial Yields Series

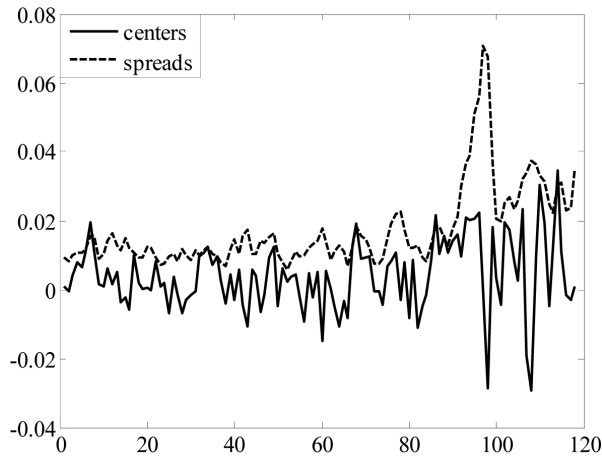
In **Figure 1** we can see that the observations of the SSE Composite index  $\{[x_t^l, x_t^u], t = 1, 2, \dots, 119\}$  show an obvious increase with the time. Then, we adopt the method of run test to test the series of centers  $\{x_t\}$  of  $\{[x_t^l, x_t^u], t = 1, 2, \dots, 119\}$  and the results shown in the column 2 of **Table 1** rejecting the null hypothesis with significant level 5% suggests that the sequence of  $\{x_t\}$  is non-stationary, in other words, the sequence of the SSE Composite index  $\{\tilde{x}_t\}$  is non-stationary.

As can be seen in **Figure 2**, the sequence of yields  $\{\tilde{r}_t\}$  has been detrended fluctuation basically and ranges about from  $-0.04$  to  $0.04$ . Then, we also employ the method of run test to the centers  $\{c_t\}$  of  $\{\tilde{r}_t\}$ , and the results shown in the column 3 of **Table 1** cannot reject the null hypothesis with significant level 5%, that is, the sequence  $\{c_t\}$  is stationary so that we can infer  $\{\tilde{r}_t\}$  is conditional stationary.



**Figure 1.** The series of interval observations of the SSE Composite index SSE Composite index (2014.7.17-2015.01.09) (2014.7.17-2015.01.09).





**Figure 2.** The fuzzy yields series of the SSE Composite index SSE Composite index (2014.7.17-2015.01.09) (2014.7.17-2015.01.09).

**Table 1.** The fuzzy time series of the SSE Composite index/the run test of centers of fuzzy yields series.

	Centers of fuzzy time series	Centers of fuzzy yields series
Test values	2353.99300	0.00394
The number of samples that are less than test values	59	59
The number of samples that are equal or greater than test values	60	59
Simple size	119	118
Number of runs	4	51
Z	-10.403	-1.664
P value	0.000	0.096

Note: the test values are the averages of each sequence.

We can establish the FVCBR model for the conditional stationary financial yields  $\{\tilde{r}_t\} = \{(c_t, u_t)_L\}$ . Here we will not consider the order determination of the model and only modelling the fuzzy yields sequence of the SSE Composite index  $\{\tilde{r}_t\} = \{(c_t, u_t)_L\}$  with FVCBR (1,1) and finally discuss the imitative effect.

$$\begin{cases} c_t = \alpha_0(t) + \alpha_1(t)c_{t-1} + \varepsilon_t \\ u_t = \beta_0(t) + \beta_1(t)u_{t-1} + \gamma(t)c_t + e_t \end{cases}, t = 2, 3, \dots, 104 \quad (20)$$

We select Epanechnikov kernel, Briweight kernel, Triweight kernel and Gaussian kernel to analysis respectively, and in turn denote above kernel functions as  $K_1, K_2, K_3$  and  $K_4$ . For each kernel function, we use the cross-validation procedure to find the right  $h$ .

Suppose that  $h$  is selected from 1 to 20 and steps by 0.2. **Figure 3(a)** illustrates the CV values computed by each of the kernel functions change with  $h$ . From **Figure 3(a)** we can see that the CV values of all kernel functions decrease first and then increase with the increase of  $h$ . Therefore, under the assumption of that  $h$  is chosen from 1 to 20, the optimum  $h$  for  $K_4$  is 2 and are 4 for  $K_1, K_2, K_3$  respectively. In order to find more better  $h$ , we further let  $h$  steps by 0.05 from 1 to 5, and the relationship between CV values and  $h$  is shown in **Figure 3(b)**. So we can obtain the right  $h$  for every kernel functions are respectively 3.18, 3.82, 4.36 and 1.82.

For every kernel function, we select the optimal bandwidth and then obtain the estimates of regression coefficients of model (20) at any point in time. **Figure 4** shows the estimate values of regression coefficients of different kernel functions and also indicates that the regression coefficients vary similarly with the time although the different kernel functions, that is, different kernel function has little effect on the estimates values of coefficients. Based on this result, we instead considering Gaussian kernel as the kernel function in this article. As we



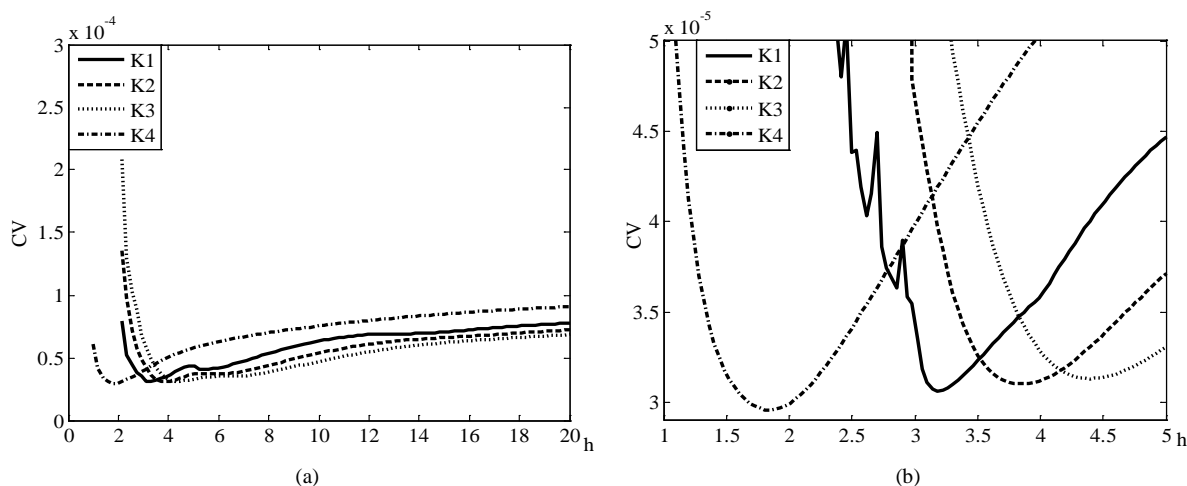


Figure 3. The changes of CV values of every kernel function with  $h$  increasing (a)  $h$  is from 1 to 20; (b)  $h$  is from 1 to 5.

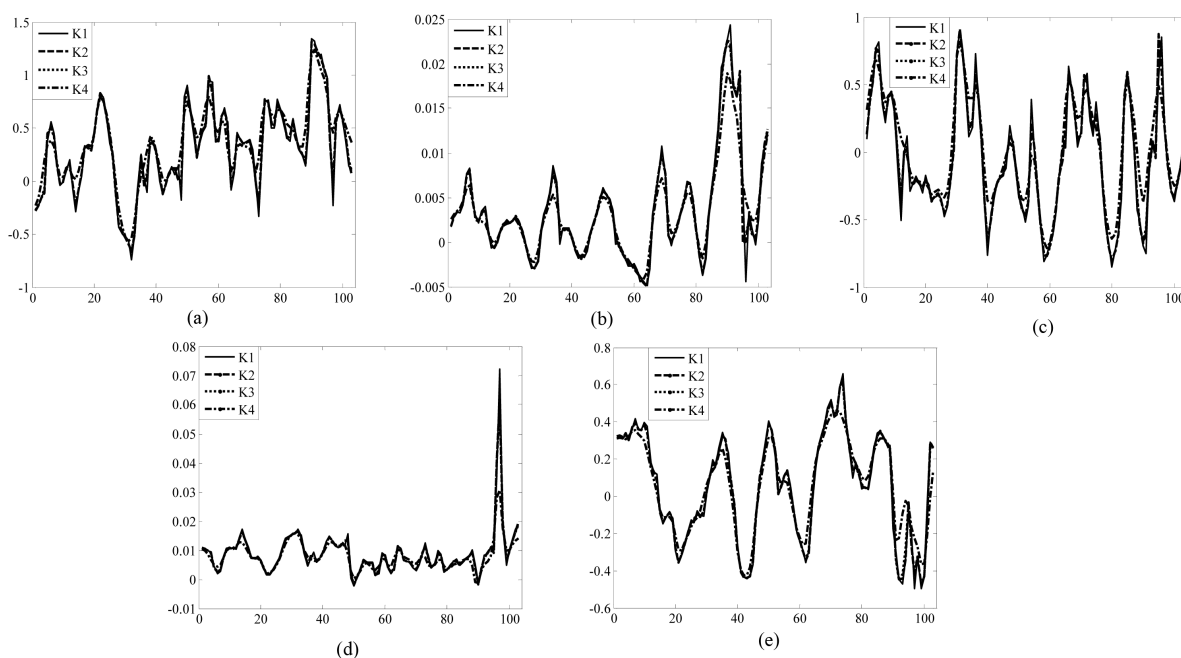


Figure 4. The estimates of regression coefficients of model (20). (a)  $\alpha_0$ ; (b)  $\alpha_1$ ; (c)  $\beta_0$ ; (d)  $\beta_1$ ; (e)  $\gamma$ .

can see from the estimates of model (20) and Figure 4, the spreads of fuzzy yields have positive correlation to its first order lagging values while the relationship between the centers of fuzzy yields and its first order lagging values as well as the relationship between the spreads and current centers remain positive or negative in the dynamic changes.

Next, we will test the hypothesis of the regression coefficients are constants by using Gaussian kernel as the selected kernel function. Based on the data of SSE composite index and formula (18), we can get the generalized likelihood ratio (GLR) statistic  $T = 42.6781$ . When repeat the Step 1 and Step 2 of Bootstrap method for 100 times, that is,  $m = 100$ , the curve of asymptotic distribution of GLR statistic is as shown in Figure 5(a) and at this time the  $p$  value being lower than 0.01 suggests that the result reject the null hypothesis with significant level 0.01. Similarly, when  $m = 1000$ , the curve of asymptotic distribution of GLR statistic is as shown in Figure 5(b), and in this case the  $p$  value being lower than 0.001 illustrates the result reject the null hypothesis with significant level 0.001. In conclusion, we reject the null hypothesis and hold that the regression coefficients

are change with time. As a result, the fuzzy varying coefficient regression model is a better choice.

## 4.2. Forecasting and Evaluation of Simulation

### 4.2.1. Forecasting

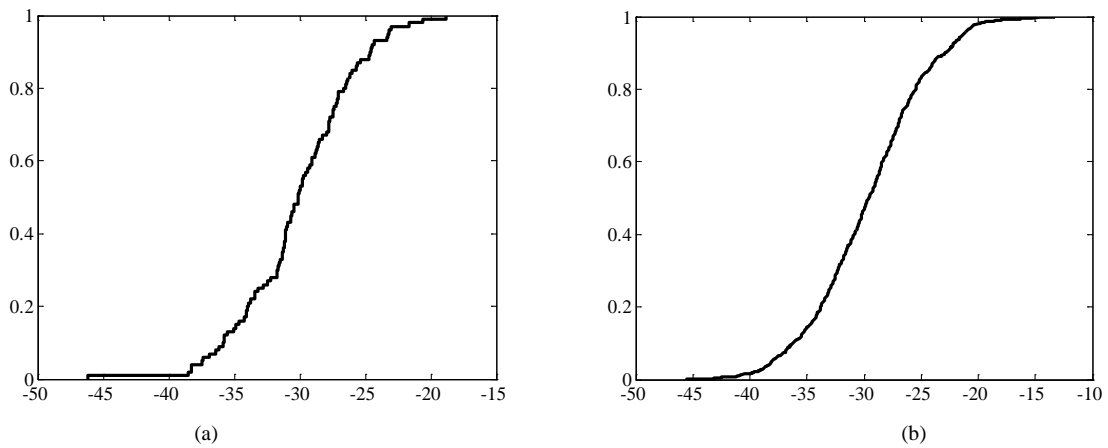
We forecast the real data from 105-th to 119-th with one-step-ahead prediction by using the model (20) and formula (10). The results are as shown in **Table 2**.

### 4.2.2. Evaluation of Simulation of the Model

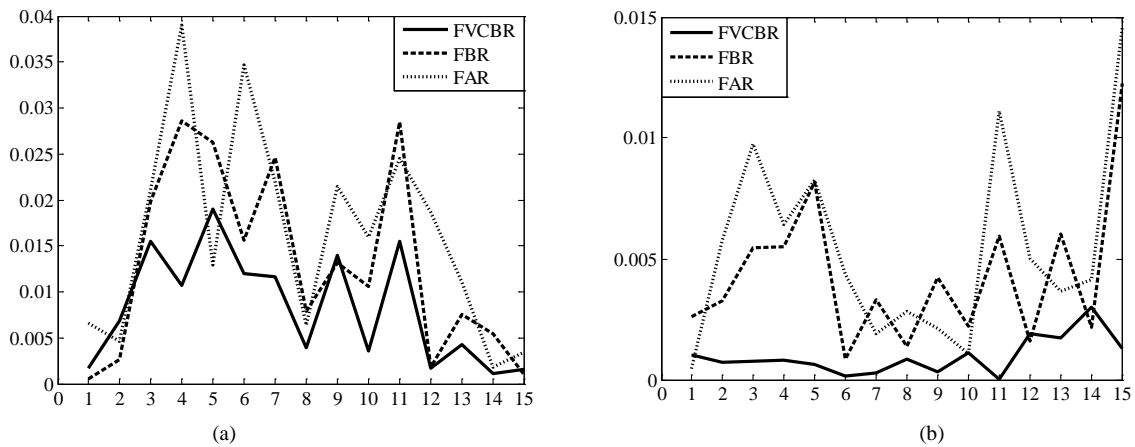
In order to compare to the prediction of FAR model and FBR model, we respectively calculate the absolute error of prediction and show them with curves (as see in **Figure 6**) of FAR model, FBR model and FVCBR model.

As we can see from **Figure 6(a)**, when predict the centers, the absolute errors of FVCBR model are a little greater than FAR model and FBR model only on the 1st, the 2nd, the 5th and the 9th periods; meanwhile, when predict the spreads, the predicted values are all much close to the observed values at every period except the 1st, the 12th and the 14th periods. So, we preliminary infer that the predictions of FVCBR model are more precise than FAR model and FBR model.

For the purpose of judging the predictions and evaluations of simulation of the three models more accurately, we use the mean square error (MSE) and mean absolute error (MAE) that are often used in regression analysis to evaluate the fitting effects and prediction accuracy. Specifically, we calculate the MSE and MAE of financial yields  $\{\tilde{r}_i\}$  in and out of sample period respectively and the results are shown in **Table 3**.



**Figure 5.** The asymptotic distribution of GLR statistic. (a)  $m = 100$ ; (b)  $m = 1000$ .



**Figure 6.** The comparisons of predictions of FAR model, FAR model and FVCBR model. (a) The sequence of centers; (b) The sequence of spreads.

**Table 2.** The results of one-step-ahead prediction of fuzzy varying coefficient bilinear regression model<sup>3</sup>.

Periods	105	106	107	108	109
Observed values	(0.0083, 0.0234)	(0.0025, 0.0259)	(0.0233, 0.0320)	(-0.0190, 0.0339)	(-0.0292, 0.0374)
Predicted values	(0.0065, 0.0245)	(0.0093, 0.0252)	(0.0077, 0.0312)	(-0.0083, 0.0347)	(-0.0101, 0.0367)
Periods	110	111	112	113	114
Observed values	(0.0097, 0.0364)	(0.0302, 0.0332)	(0.0195, 0.0313)	(-0.0047, 0.0247)	(0.0119, 0.0223)
Predicted values	(-0.0024, 0.0363)	(0.0186, 0.0335)	(0.0156, 0.0304)	(0.0092, 0.0244)	(0.0155, 0.0235)
Periods	115	116	117	118	119
Observed values	(0.0347, 0.0303)	(0.0112, 0.0310)	(-0.0015, 0.0230)	(-0.0031, 0.0238)	(0.0007, 0.0350)
Predicted values	(0.0192, 0.0303)	(0.0130, 0.0291)	(0.0028, 0.0247)	(-0.0020, 0.0268)	(-0.0008, 0.0338)

**Table 3.** The errors measurements of fitting and forecasting of SSE composite index.

Model	In the sample period		Out of the sample period	
	MSE	MAE	MSE	MAE
FVCBR model	4.19E-05	0.0061	1.11E-04	0.0098
FBR model	9.46E-05	0.0099	3.15E-04	0.0185
FAR model	1.26E-04	0.0111	4.56E-04	0.0233

From **Table 3**, we can see no matter in and out of the sample period, the MSE and MAE of FVCBR model are all lower than FAR model and FBR model. This result proves that the fitting effects and prediction accuracy of FVCBR model are superior to FAR model and FBR model.

## 5. Conclusion

This article introduces the fuzzy financial yields series to deal with the interval observed samples in financial markets and constructs the fuzzy varying coefficient bilinear regression (FVCBR) model with satisfying the practical significance of financial yields. Besides, based on the complete distance between fuzzy numbers, we develop the fuzzy least squares method to obtain the estimates of the unknown coefficients. Empirical analysis shows that compared with constant coefficient regression model and fuzzy auto regression model, the varying coefficient regression has shown some improvements no matter in fitting effects or prediction. The fuzzy auto regression model has certain limitations in model fitting and forecasting because the auto regression of centers and spreads are considered independently rather than taking the effect that centers have on the spreads into account. Fuzzy varying coefficient bilinear regression explores the problem of financial fuzzy time series more flexibly and applicatively to make the fitted values and predictions be intervals and thus more consistent with the description of real financial market. Finally, this article only discusses the uncertainty of changes of financial assets price, that is, reflects the changes of yields only by fuzzy data but ignores the probability distribution of parameters. So, later research will focus on the coefficient estimates, statistical tests, model fitting and forecasting after introducing the random error, so as to give deciders more perspective to recognize and explain the changes of financial markets.

## References

- [1] Zadeh, L.A. (1965) Fuzzy Sets. *Information and Control*, **8**, 338-353. [http://dx.doi.org/10.1016/S0019-9958\(65\)90241-X](http://dx.doi.org/10.1016/S0019-9958(65)90241-X)
- [2] Li, Z.-Y., Zhang, C. and Wang, T.-J. (2010) Studying on the Interval Financial Times Series and Evaluating on the Forecast. *Journal of Applied Statistics and Management*, **29**, 129-136.

<sup>3</sup>The regression coefficients in one-step-ahead prediction are previous estimates.

- [3] D'Urso, P. and Gastaldi, T. (2000) A Least-Squares Approach to Fuzzy Linear Regression Analysis. *Computational Statistics and Data Analysis*, **34**, 427-440. [http://dx.doi.org/10.1016/S0167-9473\(99\)00109-7](http://dx.doi.org/10.1016/S0167-9473(99)00109-7)
- [4] Li, Z.-Y., Liu, W.-Y. and Wang, T.-J. (2009) Fuzzy Bilinear Regression of Yields Series. *Statistical Research*, **26**, 68-73.
- [5] Wang, H.-D., Guo, S.-C. and Yue, L.-Z. (2014) An Approach to Fuzzy Multiple Linear Regression Mode Based on the Structured Element Theory. *Systems Engineering—Theory & Practice*, **34**, 2628-2636.
- [6] Hu, B.-Q. (2010) Foundations of Fuzzy Theory. Wuhan University Press, Wuhan, 103-114.
- [7] Xu, R.N. (1991) A Linear Regression Model in Fuzzy Environment. *Advances in Modelling Simulation*, **27**, 31-40.
- [8] Fan, J.Q. and Yao, Q.W. (2003) Nonlinear Time Series. Springer, Berlin, 243-245.
- [9] Cai, Z.W., Fan, J.Q. and Yao, Q.W. (2000) Functional-Coefficient Regression Models for Nonlinear Times Series. *Journal of the American Statistical Association*, **95**, 941-956. <http://dx.doi.org/10.1080/01621459.2000.10474284>