The Estimating Conditional Heteroscedastic Nonlinear Autoregressive model by Using Smoothing Spline and Penalized Spline Methods

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Songklanakarin Journal of Science and Technology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>SJST-2017-0390.R1</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Original Article</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>01-Feb-2018</td>
</tr>
<tr>
<td>Complete List of Authors:</td>
<td>Araveeporn, Autcha; Statistics;</td>
</tr>
<tr>
<td>Keyword:</td>
<td>Conditional Heteroscedastic Nonlinear Autoregressive model, Smoothing Spline Method, Penalized Spline Method</td>
</tr>
</tbody>
</table>
The Estimating Conditional Heteroscedastic Nonlinear Autoregressive Model by Using Smoothing Spline and Penalized Spline Methods

Autcha Araveeporn¹*

¹Department of Statistics, Faculty of Science,

King Mongkut’s Institute of Technology Ladkrabang, Bangkok 10520, Thailand

*Corresponding Author

E-mail Address: kaautcha@hotmail.com, autcha.ar@kmitl.ac.th

Abstract

We propose smoothing spline (SS) and penalized spline (PS) methods in a class of nonparametric regression method for estimating the unknown functions in a conditional heteroscedastic nonlinear autoregressive (CHNLAR) model. The CHNLAR model consists of trend and heteroscedastic functions in terms of past data at lag 1. The SS and PS methods have been performed to estimate the unknown functions which transformed data to fit the trend and heteroscedastic functions. For simulation study, time series data are generated and hypothesis testing of the bias is used to check the accuracy. The SS and PS methods exhibit a good power estimation in most cases of generated data. For real data, the gold price is then applied by using SS and PS methods based on CHNLAR model. The results show that the SS method performs better than the PS method.
Keywords: Conditional Heteroscedastic Nonlinear Autoregressive model, Smoothing Spline Method; Penalized Spline Method;

1. Introduction

Currently the economics growth is of interest to developing countries. These data are mostly stored in the form of time series data, whether it is a daily, a monthly, a quarterly, a yearly, like the unemployment rate, an economics growth rate, the gold price, the exchange rate. The information is sensitive and rapid fluctuations by the external factors such as natural disasters, wars, epidemics which these factors can’t be controlled and predictable events. For this reason, it is difficult to estimate or predict the precise business information.

Heteroscedastic or volatility means that the economics time series data quickly changes values in time-plot data. Heteroscedastic model is useful to study for estimating or forecasting time series data where the right approach is to solve this problem when the time series are often clear evidence of changing mean and variance.

There are several methods to model heteroscedastic in time series, such as the autoregressive conditional heteroscedastic model (ARCH) by Engle (1982), who was the first to introduce the ARCH model. The mean-corrected asset return is serially uncorrelated value, but condition heteroscedastic are changing over time. Ghosh, Pual and Prajneshu (2010) applied on zero conditional mean residual series to identify time varying volatility in the data set by mixture periodic ARCH model. Bollerslev (1986) extended the Generalized ARCH (GARCH), and assumed that the mean equation can be adequately described by ARMA model. Pual, Ghosh and Prajneshu (2009) carried out to fit autoregressive integrated moving average (ARIMA) and GARCH model for modeling and forecasting of the data. Peng and Yao (2003) showed that the conditional
maximum quasilikelihood estimator suffers from complex limit distributions and slow convergence rates by ARCH and GARCH model with heavy-tailed errors. The nonlinear autoregressive (NLAR) model is developed from the nonlinear regressive model which is introduced by Jones (1978). Gouri´eroux Monfort (1992); Masry and Tjøstheim (1995) have proposed the conditional heteroscedastic nonlinear autoregressive (CHNLAR) model in financial time series. For simplicity, the case is the first- lag of the CHNLAR model were studied to model the foreign exchange rates (Bossaerts, H¨ardle, & Hafner, 1996).

The parametric and nonparametric methods are the choice for estimating regression function between two sets of variables that consist of a vector of predictors and a response variable. A parametric regression model requires an assumption that the form of the underlying regression function. The selection of parametric model depends much on the problem and may be too restrictive in some applications. To overcome the difficulty caused by the restrictive assumption of the parametric form of the regression function, one may remove the restriction that the regression function belongs to a parametric family. This approach leads to so-called nonparametric regression.

Typically, the nonparametric regression methods are based on a smoothing technique which produces a smoother. A smoother is a tool for summarizing the trend of a response variable as a function of one or more predictor variables. The single predictor case is called scatterplot smoothing that can be used to enhance the visual appearance of the scatterplot of response versus predictor variable. There are many smoothing techniques, e.g., smoothing splines (Wahba, 1990; Green & Silverman, 1994), and penalized splines (Ruppert, Wand, & Carroll, 2003). These smoothing
techniques are generally based on the assumption of homoscedastic variance model
which may not be suitable when the data involves high volatility.

For these various reasons, we are interesting to extend NLAR model to
CHNLAR model for approximating the heteroscedastic values by adjusting the past
value. The smoothing spline and penalized spline methods are applied to estimate trend
and heteroscedastic values for prediction based on the simulated data, and real data.

2. CHNLAR model

Some nonlinear time series models focus on various volatility forms such as
ARCH model, GARCH model, threshold autoregressive model, and nonparametric
autoregressive model. The nonparametric autoregressive conditional heteroscedastic
(NARCH) model (Fan & Yao, 2003) adopted the nonparametric and nonlinear time
series model or called conditional heteroscedastic nonlinear autoregressive (CHNLAR)
model. It can be written as

\[ y_t = \mu(y_{t-1}, \ldots, y_{t-p}) + \sigma(y_{t-1}, \ldots, y_{t-p}) \varepsilon_t, \]

and \( \sigma(\cdot) \) is an constant called nonparametric autoregressive (NAR) model or called
nonlinear autoregressive (NLAR) model.

In this case, we employ the first-order of conditional heteroscedastic nonlinear
autoregressive (CHNLAR) model following

\[ y_t = \mu(y_{t-1}) + \sigma(y_{t-1}) \varepsilon_t, \quad t = 2,3,\ldots,n, \]  

(1)

where \( y_t, t = 2,3,\ldots,n \) are observed dependent variables, \( y_{t-1}, t = 2,3,\ldots,n \) are the past
of dependent variables at lag 1, \( \mu(y_{t-1}) \) are the trend function of CHNLAR model,
\( \sigma(y_{t-1}) \) are the heteroscedastic function of CHNLAR model, and \( \varepsilon_t, \ t = 2,3,\ldots,n \) denote the random variable of error terms with mean zero and variance one.

3. Nonparametric Regression Method

The popular nonparametric regression methods include smoothing spline method and penalized spline method. The concept of these methods is to let the estimated data interpolate the most suitable form the fitting function by smoothing parameter.

In this section we study the nonlinear autoregressive (NLAR) model following:

The NLAR model is written as

\[
y_t = \mu(y_{t-1}) + \varepsilon_t, \quad t = 2,3,\ldots,n,
\]

where \( y_t, t = 2,3,\ldots,n \) are observed dependent variables, \( y_{t-1}, t = 2,3,\ldots,n \) are the past of dependent variables at lag 1, \( \mu(y_{t-1}) \) are the trend function of nonlinear autoregressive model, and \( \varepsilon_t, t = 2,3,\ldots,n \) denote the random variable of error terms with mean zero and variance one.

3.1 Smoothing Spline (SS) Method

The smoothing spline was studied by Wahba (1990) that the smoothing spline estimator is estimated by the natural polynomial spline \( S_\lambda^{(k)}(\mu) \) which is depended on the smoothing parameter \( (\lambda) \) following:

\[
S_\lambda^{(k)}(\mu) = \sum_{i=1}^{n} \left( y_i - \mu(x_i) \right)^2 + \lambda \int_{a}^{b} \left\{ \mu^{(m)}(x) \right\}^2 dx, \quad (3)
\]
where $K$ is the number of knots on trend function with domain $[a, b]$, $m$ is the $m$th derivative of $\mu(x)$, $y_t$ is the dependent variables, and $\mu(x)$ is trend function in a class of nonparametric regression function with independent variables.

Green and Silverman (1994) emphasized $m = 2$ so-called the natural cubic spline to fit the nonparametric regression function by minimizing

$$S_2^{(k)}(\mu) = \sum_{t=1}^{n} (y_t - \mu(x_t))^2 + \lambda \int_{a}^{b} [\mu''(x_t)]^2 dx_t. \quad (4)$$

In this case, we propose the NLAR model via smoothing spline method, and the natural cubic spline can be written as

$$S_2^{(k)}(\mu) = \sum_{t=2}^{n} (y_t - \mu(y_{t-1}))^2 + \lambda \int_{a}^{b} [\mu''(y_{t-1})]^2 dy_t. \quad (5)$$

The natural cubic spline is given the value and second derivatives at each knots $y_t$ as

$$\mu = \mu(y_{t-1}),$$

$$\gamma = \mu''(y_{t-1}), \quad t = 2, 3, \ldots, n.$$  

Let $\mu$ be the vector $(\mu_1, \ldots, \mu_{n-1})^T$ and let $\gamma$ be the vector $(\gamma_1, \ldots, \gamma_{n-1})^T$.

The condition of natural cubic spline depends on two matrices $Q$ and $R$ below

$$Q = \begin{pmatrix}
h_1^{-1} & 0 & \cdots & 0 \\
-h_1^{-1} - h_2^{-1} & h_2^{-1} & \cdots & 0 \\
h_2^{-1} & -h_2^{-1} - h_3^{-1} & \cdots & 0 \\
0 & h_3^{-1} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & h_{n-1}^{-1} - h_n^{-1}
\end{pmatrix}$$

where $h_t = y_t - y_{t-1}$, for $t = 2, \ldots, n-1$, then $Q$ is an $(n-1) \times (n-3)$ matrix.

$R$ is a symmetric $(n-3) \times (n-3)$ matrix with elements below
The matrix $K$ is defined here

$$K = QR^{-1}Q^{T}. \quad (6)$$

The vector $\mu$ and $\gamma$ specify a natural cubic spline $\mu(y_i)$ if and only if the condition

$$Q^{T}\mu = R\gamma \quad (7)$$

is satisfied. If (7) is satisfied the roughness penalty will satisfy

$$\int_{a}^{b}[\mu'(y_{i-1})]^{2}dy_i = y^{T}R\gamma = \mu^{T}K\mu. \quad (8)$$

To illustrate, it can be written in matrix form introduced by Green & Silverman (1994) as

$$\text{RSS} = \sum_{i=2}^{n}(y_{i} - \mu(y_{i-1}))^{2} = (y - \mu)^{T}(y - \mu), \quad (9)$$

where $y=(y_{2},\ldots,y_{n})^{T}$ with $y_{i}$ corresponding value to $y_{i-1}$ and

$$\mu=(\mu(y_{1}),\ldots,\mu(y_{n-1}))^{T}. \quad (10)$$

The roughness penalty term $\int \mu^{2}$ as $\mu^{T}K\mu$ in (8) to obtain

$$S_{\lambda} (\mu) = (y - \mu)^{T}(y - \mu) + \lambda \mu^{T}K\mu$$

$$= \mu^{T}(I + \lambda K)\mu - 2y^{T}\mu + y^{T}y, \quad (11)$$

since $\lambda K$ is non-negative definite, the matrix $I + \lambda K$ is strictly positive definite. It therefore follows that (11) has a unique minimum, other smoothing spline estimator is obtained by
\( \hat{\mu}_k = (I + \lambda K)^{-1} y, \) 

(12)

where \( I \) denote the \( n \)-dimensional identity matrix.

### 3.2 Penalized Spline (PS) Method

Penalized spline smoother is estimated using the truncated power function (Ruppert & Carroll, 2000), and the penalized spline model is written as

\[
\mu(x_t) = \sum_{j=0}^{m-1} \alpha_j x_t^j + \sum_{k=1}^{K} \beta_k (x_t - \tau_k)^{2m-1}, \quad t = 1, 2, \ldots, n, \tag{13}
\]

where \( \alpha_j \) and \( \beta_k \) denote regression coefficient of the truncated power function.

The natural cubic spline is denoted \( m = 2 \) or called low-rank thin-plate spline which tend to have very good numerical properties. The low-rank thin-plate spline representation of \( \mu(\cdot) \) is

\[
\mu(x_t; \theta) = \alpha_0 + \alpha_1 x_t + \sum_{k=1}^{K} \beta_k |x_t - \tau_k|^3, \quad t = 1, 2, \ldots, n, \tag{14}
\]

where \( \theta = (\alpha_0, \alpha_1, \beta_1, \ldots, \beta_K)^T \) is the vector of regression coefficients, and \( \tau_1 < \tau_2 < \ldots < \tau_K \) are fixed knots.

In this case, we focus the NLAR model based on penalized spline method, then the natural cubic spline can be written as

\[
\mu(y_{i-1}; \theta) = \alpha_0 + \alpha_1 y_{i-1} + \sum_{k=1}^{K} \beta_k |y_{i-1} - \tau_k|^3, \quad t = 2, 3, \ldots, n. \tag{15}
\]

To avoid overfitting, we minimize

\[
\sum_{i=1}^{n} (y_i - \mu(y_{i-1}; \theta))^2 + \frac{1}{2} \theta^T D \theta, \tag{16}
\]

where \( \lambda \) is the smoothing parameter and \( D \) is a known positive semi-definite penalty matrix. The thin-plate spline penalty matrix is
\[ D = \begin{bmatrix} 0_{2 \times 2} & 0_{2 \times K} \\ 0_{K \times 2} & \Omega_K \end{bmatrix} , \]  \hspace{1cm} (17)

where the \((i,k)\)th entry of \(\Omega\) is \(\left| r_i - r_k \right|^3\) and penalized only coefficient of \(\left| y_{i-1} - r_k \right|^3\).

Just as with the linear model, we can generalize penalized spline in general linear mixed model (Brumback, Ruppert, & Wand, 1999) as

\[ y = Y\alpha + Z_k\beta + \epsilon , \hspace{1cm} (18) \]

where \(y=(y_2,\ldots,y_n)^T, \ Y\) be the matrix with the \(t-1\)th row \(Y_i=(1,y_{i-1})\), \(Z_k\) be the matrix with the \(t\)th row \(Z_{ki}=[\left| y_{i-1} - r_1 \right|^3,\ldots,\left| y_{i-1} - r_k \right|^3]\), \(\alpha=(\alpha_1,\alpha_2)^T, \beta=(\beta_1,\ldots,\beta_K)^T\), and \(\epsilon\) is \(N(0,\sigma^2 I)\). Consider the vector \(\alpha\) as a fixed parameters and the vector \(\beta\) as a set of random parameters with \(E(\beta)=0\) and \(\text{cov}(\beta)=\sigma_\beta^2 I\). This class of penalized spline smoothers \((\hat{\mu}(\cdot))\) may also be expressed as

\[ \hat{\mu} = C (C^T C + \lambda I)^{-1} C^T y , \hspace{1cm} (19) \]

where

\[ C = \begin{bmatrix} 1 & y_{i-1} & \left| y_{i-1} - r_1 \right|^3 & \cdots & \left| y_{i-1} - r_k \right|^3 \end{bmatrix}_{1 \leq i \leq n} , \]
\[ D = \begin{bmatrix} 0_{2 \times 2} & 0_{2 \times K} \\ 0_{K \times 2} & \Omega_K^{1/2} \Omega_K^{1/2} \end{bmatrix} , \hspace{1cm} (20) \]

and \(\lambda = \sigma_\beta^2 / \sigma_\epsilon^2\) is a smoothing parameter.

4. Proposed Trend and Heteroscedastic Estimators

The trend \(\mu(y_{t-1})\) and heteroscedastic \(\sigma^2(y_{t-1})\) can also be considered in CHNLAR model. As an initial step, we start by estimating the trend \(\mu(y_{t-1})\) using the concept of NLAR model written as

\[ y_t = \mu(y_{t-1}) + \delta_t , \hspace{1cm} t = 2,3,\ldots,n , \hspace{1cm} (21) \]
where $\delta_i = \sigma(y_{i-1}) \varepsilon_i$. Next, we obtain $\hat{\mu}(y_{i-1})$ from smoothing spline (SS) and penalized spline (PS) where the residuals can be estimated as

$$\hat{\delta}_i = y_i - \hat{\mu}(y_{i-1})$$  \hspace{1cm} (22)

$$\left(\hat{\delta}_i\right)^2 = \left(\sigma(y_{i-1}) \hat{\varepsilon}_i\right)^2.$$  \hspace{1cm} (23)

We transform $\sigma(y_{i-1}) = \exp\left\{\frac{h(y_{i-1})}{2}\right\}$, and take log with residuals in (23)

$$\log \hat{\delta}_i^2 = h(y_{i-1}) + \log \hat{\varepsilon}_i^2,$$  \hspace{1cm} (24)

$$\log \hat{\delta}_i^2 - E[\log \hat{\varepsilon}_i^2] = h(y_{i-1}) + \log \hat{\varepsilon}_i^2 - E[\log \hat{\varepsilon}_i^2].$$  \hspace{1cm} (25)

If we require $\hat{\varepsilon}_i$ to be normally distributed with mean zero and variance one, then $E[\log \hat{\varepsilon}_i^2] = -1.2704$ on (25) and hence we can apply in SS and PS to obtain

$$\log \hat{\delta}_i^2 + 1.2704 = h(y_{i-1}) + \log \hat{\varepsilon}_i^2 + 1.2704$$ \hspace{1cm} (26)

$$\tilde{y}_i = h(y_{i-1}) + \hat{\varepsilon}_i$$ \hspace{1cm} (27)

where $\tilde{y}_i = \log \hat{\delta}_i^2 + 1.2704$ and $\hat{\varepsilon}_i = \log \hat{\varepsilon}_i^2 + 1.2704$. Next, we get a smooth estimate $\hat{h}(y_{i-1})$ from SS and PS by using (27) and update the heteroscedastic estimate to be

$$\hat{\sigma}(y_{i-1}) = \exp\left\{\frac{\hat{h}(y_{i-1})}{2}\right\}. \hspace{1cm} (28)$$

At the second stage of estimation we update the trend estimate by using the following model

$$y_i = \mu(y_{i-1}) + \exp\left\{\frac{\hat{h}(y_{i-1})}{2}\right\} \hat{\varepsilon}_i$$ \hspace{1cm} (29)

$$\exp\left\{-\frac{\hat{h}(y_{i-1})}{2}\right\} y_i = \exp\left\{-\frac{\hat{h}(y_{i-1})}{2}\right\} \mu(y_{i-1}) + \hat{\varepsilon}_i$$ \hspace{1cm} (30)
\[ \bar{y}_t = g(y_{t-1}) + \hat{\epsilon}_t \] (31)

where \( \bar{y}_t = \exp\left(-\frac{\hat{h}(y_{t-1})}{2}\right) y_t \), and \( g(y_{t-1}) = \exp\left(-\frac{\hat{h}(y_{t-1})}{2}\right) \mu(y_{t-1}) \) on (30). If \( \hat{g}(y_{t-1}) \) is obtained by SS and PS, the second stage estimate of \( \mu(y_{t-1}) \) is given by

\[ \hat{\mu}(y_{t-1}) = \exp\left(-\frac{\hat{h}(y_{t-1})}{2}\right) \hat{g}(y_{t-1}). \] (32)

Finally, the estimates of \( \mu(y_{t-1}) \) and \( \sigma(y_{t-1}) \) converge to \( \hat{\mu}(y_{t-1}) \) and \( \hat{\sigma}(y_{t-1}) \).

5. Simulation Study

In simulation study to estimate the performance of smoothing spline (SS) method and penalized spline (PS) are divided into two parts. The first part is to study in CHNLAR model

\[ y_t = \mu(y_{t-1}) + \sigma(y_{t-1}) \varepsilon_t, \quad t = 2,3,\ldots,n, \] (33)

where \( \mu(y_{t-1}) \) and \( \sigma(y_{t-1}) \) are generated following

\[ \mu(y_{t-1}) = 0.1(y_{t-1}), \]
\[ \sigma(y_{t-1}) = \exp\{0.05 \times y_{t-1}\}, \]

where \( y_t \sim \text{Normal}(0,1) \). In Figure 1, we present \( y_t \) in CHNLAR model at sample sizes \( n = 50, 100, 200, \) and 300. The error process \( \varepsilon_t, t = 2,3,\ldots,n \) from in (33) which is assumed to follow the normal distribution with mean zero and variance one.

Figure 1: The time series data in CHNLAR model

The second part, the estimates of \( \hat{\mu}(y_{t-1}) \) and \( \hat{\sigma}(y_{t-1}) \), we compute the bias and the Mean Square Error (MSE) of \( \mu(.) \) and \( \sigma(.) \) by
\[
\mu_{\text{bias}} = \frac{1}{n} \sum_{i=2}^{n} \frac{\hat{\mu}(y_{i-1}) - \mu(y_{i-1})}{\mu(y_{i-1})}, \\
\sigma_{\text{bias}} = \frac{1}{n} \sum_{i=2}^{n} \frac{\hat{\sigma}(y_{i-1}) - \sigma(y_{i-1})}{\sigma(y_{i-1})}, \\
MSE(\mu) = \frac{1}{n} \sum_{i=2}^{n} (\hat{\mu}(y_{i-1}) - \mu(y_{i-1}))^2, \\
MSE(\sigma) = \frac{1}{n} \sum_{i=2}^{n} (\hat{\sigma}(y_{i-1}) - \sigma(y_{i-1}))^2.
\]

We simulated data with the sample sizes \( n = 50, 100, 200, \) and \( 300, \) and repeated the data generation and model fitting 500 times.

Table 1: The average of MSE of SS and PS methods for different sample (500 replications)

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>MSE(( \mu ))</th>
<th>MSE(( \sigma ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>200</td>
<td></td>
<td></td>
</tr>
<tr>
<td>300</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tables 1 presents an average MSE of SS and PS methods for all sample sizes. That the average MSE of \( \mu(.) \) and \( \sigma(.) \) are reduced with increasing sample sizes. For \( \mu(.) \), the average MSE of PS is less than SS, but the average MSE of PS are larger than the SS at \( \sigma(.) \).

Tables 2 and 3 show various Monte Carlo(MC) summary statistics of the estimates obtained by the SS and PS methods. The third and the fourth columns of these tables represent the MC sample mean and standard deviation of biases. The sample mean for the lower and upper bounds of 95% confidence interval are given in next two
columns. The last two columns of these tables list the t-statistic, and p-value for hypothesis testing ($H_0$: bias = 0 versus $H_1$: bias $\neq$ 0). If the p-value is less than 0.05, we reject the null hypothesis ($H_0$) that there’s difference between the observed values and the fitted values. If the p-value is larger than 0.05, we conclude that there’s unbiased estimator.

Table 2: The simulation of smoothing spline method for different sample (500 replications)

* indicates significance at 5% level

Figure 2: Histogram of Bias for $\mu(.)$ with smoothing spline method

Figure 3: Histogram of Bias for $\sigma(.)$ with smoothing spline method

Table 3: The simulation of penalized spline method for different sample (500 replications)
Figure 4: Histogram of Bias for $\mu(.)$ with penalized spline method

Figure 5: Histogram of Bias for $\sigma(.)$ with penalized spline method

By observing the p-values, the results appear following:

From Tables 2 and 3, the SS and PS methods provide asymptotically unbiased estimates of $\mu(.)$ and $\sigma(.)$ nearly for all two methods. However $\sigma(.)$ is not asymptotically unbiased estimates when the sample sizes is 200 ($n = 200$) on SS method.

The histogram of the biases of all parameter estimates are presented in Figures 2-5. The histogram is apparent that the distribution of $\sigma(.)$ the biases appears to be more normally distributed for all sample sizes.

6 Applications for Real Data

In this section, we will consider the application of CHNLAR model using the smoothing spline (SS) and penalized spline (PS) methods that we developed in the previous chapter. The real data sets, we use the monthly volume of gold price (US Dollars per Troy Ounce) from January, 1984 to December 2013, which consisted of 360 records and shown on Figure 6. These data are obtained from http://www.indexmundi.com.
Figure 6: The time series plot of the monthly gold price values from January, 1984 to December, 2013

The process for predictive future values following:

At the first, we considered the CHNLAR model following

\[ y_t = \mu(y_{t-1}) + \sigma(y_{t-1}) \varepsilon_t, \quad t = 2,3,...,360, \]  

(34)

where \( \varepsilon_t \)'s are independently and identically distributed with mean zero and variance one. In this case, we let \( y_t \) denote the gold price of month \( t \) where \( t = 1 \) represents January of 1984 and \( t = 360 \) represents December of 2013.

The second for parameter estimation, we fitted the CHNLAR model to obtain the trend function, \( \mu(.) \) and the heteroscedastic function, \( \sigma(.) \). We got \( \hat{\mu}(y_{t-1}), \hat{\sigma}(y_{t-1}), \) \( t = 2,3,...,360 \) using SS and PS methods.

Let \( \hat{\mu}(y_{t-1}) \) and \( \hat{\sigma}(y_{t-1}) \) denote the converged estimates of \( \mu(.) \) and \( \sigma(.) \), and let

\[ \hat{\varepsilon}_t = \frac{y_t - \hat{\mu}(y_{t-1})}{\hat{\sigma}(y_{t-1})}, \quad t = 2,3,...,360, \]  

(35)

denote the standardized residuals based on the converged values of \( \hat{\mu}(y_{t-1}) \) and \( \hat{\sigma}(y_{t-1}) \).

Finally, we obtain estimated values of \( \hat{y}_2,...,\hat{y}_n \) using the estimated trend and heteroscedastic based on the CHNLAR model:

\[ \hat{y}_t = \hat{\mu}(y_{t-1}) + \hat{\sigma}(y_{t-1}) \hat{\varepsilon}_t, \quad t = 2,3,...,360, \]  

(36)

where the forecast trend \( \hat{\mu}(y_{t-1}) \) and the forecast heteroscedastic \( \hat{\sigma}(y_{t-1}) \) from SS and PS methods that are presented on Figure 7 and 8.
Figure 7: The scatter plot of forecasting trend and heterocedastic by smoothing spline (SS) method

Figure 8: The scatter plot of forecasting trend and heterocedastic by penalized spline (PS) method

The fitted values from February, 1984 to December, 2013 of the SS and PS methods are shown on Figure 9 and also computed the MSE, mean, and standard deviation of $\mu(\cdot)$ and $\sigma(\cdot)$ that shown in Table 4. From the Table 4, the MSE of PS is larger than SS, but the $\mu(\cdot)$ shown the different slightly values.

Figure 9: The scatter plot of gold price values and line plot of fitted values by smoothing spline (SS) and penalized spline (PS) methods

Table 4: The mean (standard deviation) and Mean Square Error (MSE) of smoothing spline (SS) and penalized spline (PS) methods

7. Conclusion

In this section, we have developed the nonparametric regression method such as the smoothing spline method and the penalized spline method to estimate smooth unknown
trend and heteroscedastic with CHNLAR model. Through a Monte Carlo simulation study, we evaluated the performance of smoothing spline method procedure and showed that the trend estimator \( \hat{\mu}(\cdot) \) and heteroscedastic estimator \( \hat{\sigma}(\cdot) \) work reasonably well for most data of all sample sizes except in one case when heteroscedastic estimator is statistical significant at \( n = 200 \). The point volatility estimators approach their corresponding true values as the sample sizes increase.

Taking a Monte Carlo study into consideration, we shown that the trend estimator of penalized spline works well for all small sample sizes, because the smoothing parameter is the high enough, indicating that the small sample sizes can obtain reliable conclusion from interpolating these models.

For application in actual data, we are also interested in comparing the power of estimating values by considering the Mean Square Error (MSE). The MSEs of a smoothing spline is smaller than a penalized spline method. However, we consider the mean of the trend and heteroscedastic estimator, we can see that the means of trend with smoothing spline method performs slightly different in penalized spline method but the variance and MSE of smoothing spline method is smaller than penalized spline method. It can be concluded that the estimating and forecasting of CHNLAR depend on the heteroscedastic.

As a part of future work, we suggest to study the higher order of lag on trend and heteroscedastic at CHNALR model.

References


Peng, L., & Yao, Q. (2003) Least absolute deviations estimation for ARCH and 
GARCH models. *Biometrika*, 90, 967-975.

modeling and forecasting of India’s volatile spices export data. *Journal of the 


Cambridge University Press.

Table 1: The average of MSE of SS and PS methods for different sample (500 replications)

<table>
<thead>
<tr>
<th>Sample size</th>
<th>SS method</th>
<th>PS method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\mu(.)$</td>
<td>$\sigma(.)$</td>
</tr>
<tr>
<td>$n = 50$</td>
<td>0.300</td>
<td>0.0251</td>
</tr>
<tr>
<td>$n = 100$</td>
<td>0.0105</td>
<td>0.0114</td>
</tr>
<tr>
<td>$n = 200$</td>
<td>0.0054</td>
<td>0.0061</td>
</tr>
<tr>
<td>$n = 300$</td>
<td>0.0037</td>
<td>0.00413</td>
</tr>
</tbody>
</table>

Table 2: The simulation of smoothing spline method for different sample (500 replications)

<table>
<thead>
<tr>
<th>Bias</th>
<th>Sample size</th>
<th>Mean</th>
<th>S.d.</th>
<th>Lci</th>
<th>Uci</th>
<th>T-stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_{bias}$</td>
<td>$n = 50$</td>
<td>6.507</td>
<td>157.106</td>
<td>-7.296</td>
<td>20.3116</td>
<td>0.9261</td>
<td>0.3548</td>
</tr>
<tr>
<td></td>
<td>$n = 100$</td>
<td>-1.172</td>
<td>23.367</td>
<td>-3.225</td>
<td>0.881</td>
<td>-1.1216</td>
<td>0.2626</td>
</tr>
<tr>
<td></td>
<td>$n = 200$</td>
<td>-0.2977</td>
<td>10.2154</td>
<td>-1.1953</td>
<td>0.5998</td>
<td>-0.6517</td>
<td>0.5149</td>
</tr>
<tr>
<td></td>
<td>$n = 300$</td>
<td>3.5053</td>
<td>49.927</td>
<td>-0.8815</td>
<td>7.8921</td>
<td>1.5699</td>
<td>0.1171</td>
</tr>
<tr>
<td>$\sigma_{bias}$</td>
<td>$n = 50$</td>
<td>-0.0114</td>
<td>0.1577</td>
<td>-0.025</td>
<td>0.0024</td>
<td>-1.6176</td>
<td>0.1064</td>
</tr>
<tr>
<td></td>
<td>$n = 100$</td>
<td>0.0020</td>
<td>0.1072</td>
<td>-0.0073</td>
<td>0.0114</td>
<td>0.4283</td>
<td>0.6686</td>
</tr>
<tr>
<td></td>
<td>$n = 200$</td>
<td>-0.0074</td>
<td>0.0776</td>
<td>-0.0142</td>
<td>-0.0005</td>
<td>-2.1335</td>
<td>0.0333$^*$</td>
</tr>
<tr>
<td></td>
<td>$n = 300$</td>
<td>0.00184</td>
<td>0.0644</td>
<td>-0.0038</td>
<td>0.0075</td>
<td>0.6412</td>
<td>0.5217</td>
</tr>
</tbody>
</table>
Table 3: The simulation of penalized spline method for different sample (500 replications)

<table>
<thead>
<tr>
<th>bias</th>
<th>sample size</th>
<th>mean</th>
<th>s.d.</th>
<th>lci</th>
<th>uci</th>
<th>t-stat</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n = 50</td>
<td>-4.734</td>
<td>79.833</td>
<td>-11.749</td>
<td>2.279</td>
<td>-1.326</td>
<td>0.1854</td>
</tr>
<tr>
<td></td>
<td>n = 100</td>
<td>1.485</td>
<td>47.561</td>
<td>-2.693</td>
<td>5.664</td>
<td>0.698</td>
<td>0.485</td>
</tr>
<tr>
<td></td>
<td>n = 200</td>
<td>0.0454</td>
<td>32.9606</td>
<td>-2.8506</td>
<td>2.9415</td>
<td>0.0308</td>
<td>0.9754</td>
</tr>
<tr>
<td></td>
<td>n = 300</td>
<td>-0.5388</td>
<td>10.4931</td>
<td>-1.4608</td>
<td>0.3831</td>
<td>-0.1483</td>
<td>0.2514</td>
</tr>
<tr>
<td>sigma bias</td>
<td>n = 50</td>
<td>0.00429</td>
<td>0.172</td>
<td>-0.010</td>
<td>0.019</td>
<td>0.556</td>
<td>0.578</td>
</tr>
<tr>
<td></td>
<td>n = 100</td>
<td>-0.0001</td>
<td>0.1120</td>
<td>-0.009</td>
<td>0.009</td>
<td>-0.026</td>
<td>0.9792</td>
</tr>
<tr>
<td></td>
<td>n = 200</td>
<td>-0.0018</td>
<td>0.0790</td>
<td>-0.0087</td>
<td>0.0051</td>
<td>-0.5170</td>
<td>0.6054</td>
</tr>
<tr>
<td></td>
<td>n = 300</td>
<td>0.0035</td>
<td>0.0617</td>
<td>-0.0018</td>
<td>0.0090</td>
<td>1.3005</td>
<td>0.194</td>
</tr>
</tbody>
</table>

Table 4: The mean (standard deviation) and Mean Square Error (MSE) of smoothing spline (SS) and penalized spline (PS) methods

<table>
<thead>
<tr>
<th>Estimator</th>
<th>SS method</th>
<th>PS method</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\mu}(\cdot) )</td>
<td>564.4387</td>
<td>564.4463</td>
</tr>
<tr>
<td></td>
<td>(396.7862)</td>
<td>(396.755)</td>
</tr>
<tr>
<td>( \hat{\sigma}(\cdot) )</td>
<td>16.92417</td>
<td>20.4015</td>
</tr>
<tr>
<td></td>
<td>(15.54389)</td>
<td>(19.66588)</td>
</tr>
<tr>
<td>MSE</td>
<td>5.653</td>
<td>8.372</td>
</tr>
</tbody>
</table>
Figure 1: The time series data in CHNLAR model

Figure 2: Histogram of Bias for $\mu(\cdot)$ with smoothing spline method
Figure 3: Histogram of Bias for $\sigma(.)$ with smoothing spline method

Figure 4: Histogram of Bias for $\mu(.)$ with penalized spline method
Figure 5: Histogram of Bias for $\sigma(.)$ with penalized spline method

Figure 6: The time series plot of the monthly gold price values from January, 1984 to December, 2013
Figure 7: The scatter plot of forecasting trend and heteroscedastic by smoothing spline (SS) method.

Figure 8: The scatter plot of forecasting trend and heteroscedastic by penalized spline (PS) method.
Figure 9: The scatter plot of gold price values and line plot of fitted values by smoothing spline (SS) and penalized spline (PS) methods.