# รายงานการวิจัยฉบับสมบูรณ์

## การวิเคราะห์ตัวแบบพหุสหสัมพันธ์ของสัมประสิทธิ์แบบสุ่มสำหรับข้อมูลอนุกรมเวลา

## The Analysis of Multi Random Coefficient Autoregressive Model on Time Series Data

นางสาวอัชฌา อระวีพร

**ได้รับทุนสนับสนุนงานวิจัยจากเงินรายได้ ประจำปีงบประมาณ 2562**
**คณะวิทยาศาสตร์**
**สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง**

**Research Title :** The Analysis of Multi Random Coefficient Autoregressive Model on Time Series Data

**Researcher : Autcha Araveeporn**

**Department : Statistics**

**Faculty : Science**


## ABSTRACT

The random coefficient autoregressive (RCA) model develops from the autoregressive model and the hierarchical model. The RCA model has considered a constant parameter and coefficient parameter depended on past data. The least squares method is a widely used method by minimizing the sum of squared residuals and differential with respect to the unknown parameter. In this paper, the concept of the least squares method is used to estimate an unknown parameter of the first and the second orders of Random Coefficient Autoregressive (RCA) model or called RCA(1) and RCA(2) models. The efficiency of the two models is to compare by considering the minimum value of mean square error. The RCA(1) and RCA(2) are then applied to a time series data in the form of nonstationary data. The monthly averages of the Stock Exchange of Thailand (SET) index and the daily volume of exchange rate Baht/Dollar are fitted on these models. The prediction of RCA(1) and RCA(2) models is shown that the RCA(l) model outperforms the RCA(2) model, similar to two data sets.

The random coefficient autoregressive (RCA) model develops for estimating parameter of nonparametric regression model that consisted of independent and dependent variables. The smoothing spline, penalized spline, and B-spline methods in a class of smoothing techniques are considered to estimate the unknown parameter. We also compare these methods by fitting nonparametric regression model on simulation data and real data. The nonlinear model is generated in different 2 models. According to the results, it can be found that the smoothing spline, penalized spline, and B-spline methods have a good performance to fit nonlinear data but the penalized spline method shows the minimum mean square errors. For a real data, we use the data from a light detection and ranging (LIDAR) experiment that contained the range distance travelled before the light as an

independent variable and logratio logarithm of the ratio of received light from two laser sources as a dependent variables. From the mean square errors of fitting data, the penalized spline shows the minimum values as the simulated data.

A smoothing spline (SS) method and penalized spline (PS) method are proposed in a class of nonparametric regression method for estimating the unknown functions on 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, the data is generated in terms of time series data and used the hypothesis testing of bias 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 (US Dollars per Troy Ounce) 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.

# CONTENTS

# Chapter 1

Random Coefficient Autoregressive Model

## 1.1 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 time series models fitted to time series data which is applied in term of stationary and nonstationary data such as finance, business, and economics. The modelling of time series models can help to estimate and forecast the values in future time. For nonstationary time series data, the Autoregressive Integrated Moving Average (ARIMA) model can be used to fit model. This model has some problems to over specify the model and estimate the integration parameter. An alternative way to model by volatility is to use the Conditional Heteroscedastic Autoregressive Moving average (CHARMA) model. The Random Coefficient Autoregressive (RCA) model is proposed based on volatility that employed the least squares and maximum likelihood methods to estimate parameters.

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, and penalized splines. These smoothing techniques are generally based on the assumption of homoscedastic variance model which may not be suitable when the data involves high volatility.

## 1.2 Objective of Research

1.2.1 Create the new model and apply for forecasting the economics of time series data.

1.2.2 Study the property of volatility on the economics of time series data.

1.2.3 Study the relationship of random coefficient autoregressive for time series analysis.

1.2.4 Simulation by computer from statistics and mathematics models for forecasting future time series data and investment.

1.2.5 Develop research and apply on real data.

## 1.3 Scope of Research

This study interested to estimate the parameter of random coefficient autoregressive model on lag one and two or called RCA(1) and RCA(2) based on least squares method. The performance these models is mentioned on the minimum value of the Mean Square Error (MSE) using on real time series data such as the Stock Exchange of Thailand (SET) index and the exchange rate of Thai Baht to one U.S. Dollar.

The Stock Exchange of Thailand (SET) is an important index in Thailand that served to promote saving and long-term capital funding for the economic development of the nation. Furthermore, the volatility of exchange rate Baht/Dollar has been related to the speculative motives in the international financial transection. The forecasts of volatility play a crucial role in the financial decision and need for accurate future volatility.

We will consider the application of conditional heteroscedastic nonlinear autoregressive (CHNLAR) model using the smoothing spline (SS) and penalized spline (PS) methods. 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. These data are obtained from http://www.indexmundi.com.

## 1.4 Research Methodology

1.4.1 Study the research about the random coefficient autoregressive model.

1.4.2 Study the research about the analysis data by using random coefficient autoregressive model.

1.4.3 Study R programming for data simulation.

1.4.4 Collect real data from Stock Exchange of Thailand and related data

1.4.5 Use R program for data analysis based on simulation data with random coefficient autoregressive model.

1.4.6 Use R program for data analysis based on real data with random coefficient autoregressive model.

1.4.7 Check accuracy of random coefficient autoregressive model based on simulation and real data.

1.4.8 Conclusion this research.

## 1.5 Benefit of Research

1.5.1 Obtain model for forecasting time series data.

1.5.2 Obtain property of volatility on time series data.

1.5.3 Obtain relationship of variable on time series data.

1.5.4 Obtain simulation data closed real time series data.

# <u>ผลผลิต</u>

| ผลงาน | ระบุรายละเอียดให้ชัดเจนระบุรายละเอียดให้ชัดเจน เช่น ฐานข้อมูลที่จะเผยแพร่/ชื่อวารสาร/ประชุมวิชาการ/ ชื่อผลงาน เป็นต้น | จำนวนที่คาดว่าจะได้ | ปีที่คาดว่าจะสำเร็จ |
|---|---|---|---|
| 1. การเผยแพร่ผลงานทางวิชาการ(Publications) | | | |
| **ระดับนานาชาติ** | | | |
| ■ การประชุม / สัมมนา ระดับนานาชาติ (International Conference) | 2019 2nd International Conference on Mathematics and Statistics (Scopus) | 1 | 2019 |
| ■ วารสาร ระดับนานาชาติ (International Journal) | Statistical Journal of the IAOS (Scopus) | 1 | 2019 |
| | Songklanakarin Journal of Science (Scopus) | 1 | 2019 |
| 2. **การผลิตบัณฑิต** | | | |
| ■ ป.ตรี/โท/เอก | ป.โท | 1 | 2019 |

# Proceedings of

# 2019 2nd International Conference on Mathematics and Statistics

## ICoMS 2019

**Prague, Czech Republic**

**July 8-10, 2019**

**ISBN: 978-1-4503-7168-1**

**The Association for Computing Machinery**
**2 Penn Plaza, Suite 701**
**New York New York 10121-0701**

Published by ACM

**ACM ISBN:** 978-1-4503-7168-1

# Table of Contents

**Proceedings of 2019 2nd International Conference on Mathematics and Statistics (ICoMS 2019)**



- ***Session 1- Mathematical Model Analysis***



- ***Session 2- Statistical Science***

# Comparing the First and the Second Orders of Random Coefficient Autoregressive Model on Time Series Data

Autcha Araveeporn
Department of Statistics,
Faculty of Science, King Mongkut's
Institute of Technology Ladkrabang,
Bangkok 10520, Thailand
kaautcha@hotmail.com

Somsri Banditvilai
Department of Statistics,
Faculty of Science, King Mongkut's
Institute of Technology Ladkrabang,
Bangkok 10520, Thailand
somsri_b2000@yahoo.com

## ABSTRACT
The random coefficient autoregressive (RCA) model develops from the autoregressive model and the hierarchical model. The RCA model has considered a constant parameter and coefficient parameter depended on past data. The least squares method is a widely used method by minimizing the sum of squared residuals and differential with respect to the unknown parameter. In this paper, the concept of the least squares method is used to estimate an unknown parameter of the first and the second orders of Random Coefficient Autoregressive (RCA) model or called RCA(1) and RCA(2) models. The efficiency of the two models is to compare by considering the minimum value of mean square error. The RCA(1) and RCA(2) are then applied to a time series data in the form of nonstationary data. The monthly averages of the Stock Exchange of Thailand (SET) index and the daily volume of exchange rate Baht/Dollar are fitted on these models. The prediction of RCA(1) and RCA(2) models is shown that the RCA(l) model outperforms the RCA(2) model, similar to two data sets.

## CCS Concepts
**Applied Computing → Forecasting**

## Keywords
Least Squares; Mean Square Error; Random Coefficient Autoregressive Model

## 1. INTRODUCTION
In a field of business, most data are collected in the form of time series data that often nonstationarity and volatility because of sudden random changes in the market, especially when data are systematically collected over a long period of time. There are several time series models fitted to time series data which is applied in term of stationary and nonstationary data such as finance, business, and economics. The modelling of time series models can help to estimate and forecast the values in future time.

For nonstationary time series data, the Autoregressive Integrated Moving Average (ARIMA) model can be used to fit model. This model has some problems to over specify the model and estimate the integration parameter. An alternative way to model by volatility is to use the Conditional Heteroscedastic Autoregressive Moving average (CHARMA) model [1]. [2] proposed the Random Coefficient Autoregressive (RCA) model based on volatility that employed the least squares and maximum likelihood methods to estimate parameters. The estimating function technique is equivalent to a weight least squares estimator discussed by [3] and [4]. Presently, it is very popular because it uses past data to help estimate the unknown parameter.

The least squares method is a well-known method that used to determine a parameter of best fit by minimizing the sum of squares. [5] applied least squares criterion to estimate the parameter of the random coefficient dynamic regression model on stationary and nonstationary to the exchange rate of Baht/GBP and Baht/EUR. [6] compared two least-squared random coefficient autoregressive model based on with and without autocorrelated errors and applied two models to the daily volume of Thai gold price.

This study interested to estimate the parameter of RCA(1) and RCA(2) models based on least squares method. The performance these models is mentioned on the minimum value of the Mean Square Error (MSE) using on real time series data such as the Stock Exchange of Thailand (SET) index and the exchange rate of Thai Baht to one U.S. Dollar.

The Stock Exchange of Thailand (SET) is an important index in Thailand that served to promote saving and long-term capital funding for the economic development of the nation. Furthermore, the volatility of exchange rate Baht/Dollar has been related to the speculative motives in the international financial transection. The forecasts of volatility play a crucial role in the financial decision and need for accurate future volatility.

## 2. THE RCA MODEL
The general class of Random Coefficient Autoregressive (RCA) model of order $p$, that is given by

$$x_t = \alpha + \sum_{i=1}^{p} \beta_{ti} x_{t-1} + \varepsilon_t, t = 2,3,\ldots,n. \qquad (1)$$

[7] suggested hierarchical model by setting

$$\underline{\beta}_{ti} = \underline{\mu}_{\beta} + \Omega_{\beta}\, \underline{u}_t,$$

where $\alpha$ is the scalar of constant, $\underline{\beta}_{ti} = (\beta_{t1}, \beta_{t2}, \ldots, \beta_{tp})'$ is a sequence of independent random vectors with mean $\underline{\mu}_{\beta} = (\mu_{t1}, \mu_{t2}, \ldots, \mu_{tp})'$ and covariance matrix $\Omega_{\beta}$. It is


*ICoMS'19*, July 8–10, 2019, Prague, Czech Republic

ACM ISBN 978-1-4503-7168-1/19/07…$15.00
DOI:https://doi.org/10.1145/3343485.3343486

assumed that $\varepsilon_t$ 's are the sequence of iid (independent and identically distributed random variables) from distribution mean zero and unit variance.

In this paper, we focus on the first and the second orders as RCA(1) and RCA(2) model.

The RCA(1) model can written as

$$x_t = \alpha + \beta_{t1} x_{t-1} + \varepsilon_t\,,\ t = 2,3,\ldots,n \qquad (2)$$

$$\beta_{t1} = \mu_\beta + \sigma_\beta v_t,$$

where $x_t$ 's are iid random variables with mean $\mu_\beta$ , and variance $\sigma_\beta^2$, $\varepsilon_t$ 's are iid random variables with mean zero and variance $\sigma_\varepsilon^2$ . The RCA(1) model can be rewritten as

$$x_t = \alpha + \beta_t x_{t-1} + \varepsilon_t = \alpha + \mu_\beta x_{t-1} + u_t\,, \qquad (3)$$

where $u_t = \sigma_\beta v_t x_{t-1} + \varepsilon_t$ , where $v_t$ is a random variable with mean zero and variance one and independent of $\varepsilon_t$ .

The RCA(2) medel can written as

$$x_t = \alpha + \beta_{t1} x_{t-1} + \beta_{t2} x_{t-2} + \varepsilon_t\,,\ t = 3,4,\ldots,n \quad (4),$$

and

$$\beta_{t1} = \mu_{\beta 1} + \sigma_{\beta 1} v_{t1},$$
$$\beta_{t2} = \mu_{\beta 2} + \sigma_{\beta 2} v_{t2}$$

where $x_t$ 's are iid random variables with mean $\mu_{\beta 1}, \mu_{\beta 2}$ , and variance $\sigma_{\beta 1}^2, \sigma_{\beta 2}^2$ , $\varepsilon_t$ 's are iid random variables with mean zero and variance $\sigma_\varepsilon^2$ . The RCA(2) model can be rewritten as

$$\begin{aligned} x_t &= \alpha + \beta_{t1} x_{t-1} + \beta_{t2} x_{t-2} + \varepsilon_t \\ &= \alpha + \mu_{\beta 1} x_{t-1} + \mu_{\beta 2} x_{t-2} + u_t \end{aligned} \qquad (5)$$

where $u_t = \sigma_{\beta 1} v_{t1} x_{t-1} + \sigma_{\beta 2} v_{t2} x_{t-2} + \varepsilon_t$ , and $v_{t1}$, $v_{t2}$ is a random variable with mean zero and variance one and independent of $\varepsilon_t$ .

## 3. LEAST SQUARES ESTIMATION

To estimate the parameter of the RCA model, we propose the process of least squares criterion for estimating the unknown parameter of the first and second orders on RCA model.

### 3.1 RCA(1) Model

The first estimation method, we propose the least squares criterion to estimate parameter $\theta_1 = (\alpha, \mu_{\beta 1})$ by minimizing sum of squared residuals.

Given a sample $x_1, x_2, \ldots, x_n$ , the parameter $\theta_1 = (\alpha, \mu_{\beta 1})$ is to estimate by minimizing $\sum_{t=1}^{n} u_t^2$ with respect to $\theta_1 = (\alpha, \mu_{\beta 1})$ , thus the least square estimate $\hat{\theta}_1 = (\hat{\alpha}, \hat{\mu}_{\beta 1})$ is given by

$$\sum_{t=1}^{n} (u_t)^2 = \sum_{t=1}^{n} \left( x_t - \alpha - \mu_{\beta 1} x_{t-1} \right)^2. \qquad (6)$$

From (6), differential with respect to parameter $\hat{\alpha}$ ,

$$\begin{aligned} \frac{\partial}{\partial \alpha} \sum_{t=1}^{n} (\mu_t)^2 &= \frac{\partial}{\partial \alpha} \sum_{t=1}^{n} \left( x_t - \alpha - \mu_{\beta 1} x_{t-1} \right)^2, \\ &= 0 \end{aligned}$$

then we get

$$\hat{\alpha} = \frac{\sum_{t=1}^{n} x_t}{n} - \mu_{\beta 1} \frac{\sum_{t=2}^{n} x_{t-1}}{n}. \qquad (7)$$

The least square estimate $\hat{\mu}_{\beta 1}$ is obtained by,

$$\begin{aligned} \frac{\partial}{\partial \mu_{\beta 1}} \sum_{t=1}^{n} (u_t)^2 &= \frac{\partial}{\partial \mu_{\beta 1}} \sum_{t=1}^{n} \left( x_t - \alpha - \mu_{\beta 1} x_{t-1} \right)^2 \\ &= 2 \sum_{t=1}^{n} \left( x_t - \alpha - \mu_{\beta 1} x_{t-1} \right) x_{t-1} \\ &= \sum_{t=2}^{n} x_t x_{t-1} - \alpha \sum_{t=2}^{n} x_{t-1} - \mu_{\beta 1} \sum_{t=2}^{n} x_{t-1}^2 \\ &= 0 \end{aligned}$$

,and

$$\hat{\mu}_{\beta 1} = \frac{\sum_{t=2}^{n} x_t x_{t-1} - \alpha \sum_{t=2}^{n} x_{t-1}}{\sum_{t=2}^{n} x_{t-1}^2}. \qquad (8)$$

From (8), let us replace in (7) and the solution of $\hat{\alpha}$ is

$$\alpha = \frac{\sum_{t=2}^{n} x_{t-1}^2 \sum_{t=1}^{n} x_t - \sum_{t=2}^{n} x_t x_{t-1} \sum_{t=2}^{n} x_{t-1}}{\left( n \sum_{t=2}^{n} x_{t-1}^2 - \left( \sum_{t=2}^{n} x_{t-1} \right)^2 \right)}, \qquad (9)$$

or $\hat{\mu}_{\beta 1}$ can be rewritten as

$$\hat{\mu}_{\beta 1} = \frac{n\sum_{t=2}^{n} x_t x_{t-1} - \sum_{t=1}^{n} x_t \sum_{t=2}^{n} x_{t-1}}{\left(n\sum_{t=2}^{n} x_{t-1}^2 - \left(\sum_{t=2}^{n} x_{t-1}\right)^2\right)}. \qquad (10)$$

For RCA(1) model, it can be fitted model as

$$\hat{x}_t = \hat{\alpha} + \mu_{\beta 1} x_{t-1} \quad , t = 2, 3, \ldots n.$$

## 3.2 RCA(2) Model

The RCA(2) model is extended from RCA(1) model by considering the parameter of the second order of RCA model that is approximated $\hat{\theta}_2 = (\hat{\alpha}, \hat{\mu}_{\beta 1}, \mu_{\beta 2})$ and estimated by minimizing sum of squared residuals in the following form :

$$\sum_{t=1}^{n} (u_t)^2 = \sum_{t=1}^{n} \left(x_t - \alpha - \mu_{\beta 1} x_{t-1} - \mu_{\beta 2} x_{t-2}\right)^2. \qquad (11)$$

From (11), differential with respect to parameter $\hat{\alpha}_1$ as

$$\frac{\partial}{\partial \alpha} \sum_{t=1}^{n} (u_t)^2 = \frac{\partial}{\partial \alpha} \sum_{t=1}^{n} \left(x_t - \alpha - \mu_{\beta 1} x_{t-1} - \mu_{\beta 2} x_{t-2}\right)^2$$

$$= \sum_{t=1}^{n} x_t - n\,\alpha - \mu_{\beta 1} \sum_{t=2}^{n} x_{t-1} - \mu_{\beta 2} \sum_{t=3}^{n} x_{t-2}$$

$$= 0$$

, then we obtain

$$\hat{\alpha} = \frac{\sum_{t=1}^{n} x_t}{n} - \mu_{\beta 1} \frac{\sum_{t=2}^{n} x_{t-1}}{n} - \mu_{\beta 2} \frac{\sum_{t=3}^{n} x_{t-2}}{n}. \qquad (12)$$

The least square estimate $\hat{\mu}_{\beta 1}$ is evaluated by

$$\frac{\partial}{\partial \mu_{\beta 1}} \sum_{t=1}^{n} (u_t)^2 = \frac{\partial}{\partial \mu_{\beta 1}} \sum_{t=1}^{n} \left(x_t - \alpha - \mu_{\beta 1} x_{t-1} - \mu_{\beta 2} x_{t-2}\right)^2$$

$$= 2\sum_{t=1}^{n} \left(x_t - \alpha - \mu_{\beta 1} x_{t-1} - \mu_{\beta 2} x_{t-2}\right) x_{t-1}$$

$$= \sum_{t=2}^{n} x_t x_{t-1} - \alpha \sum_{t=2}^{n} x_{t-1} - \mu_{\beta 1} \sum_{t=2}^{n} x_{t-1}^2 - \mu_{\beta 1} \sum_{t=3}^{n} x_{t-2} \sum_{t=2}^{n} x_{t-1}$$

$$= 0$$

, then

$$\hat{\mu}_{\beta 1} = \frac{\sum_{t=2}^{n} x_t x_{t-1} - \alpha \sum_{t=2}^{n} x_{t-1} - \mu_{\beta 2} \sum_{t=3}^{n} x_{t-2} \sum_{t=2}^{n} x_{t-1}}{\sum_{t=2}^{n} x_{t-1}^2}. \qquad (13)$$

From (13), let us replace in (12) and the solution of $\hat{\alpha}$ can be rewritten in the following form :

$$\hat{\mu}_{\beta 1} \sum_{t=2}^{n} x_{t-1}^2 =$$

$$\sum_{t=2}^{n} x_t x_{t-1} - \left( \frac{\sum_{t=1}^{n} x_t}{n} - \hat{\mu}_{\beta 1} \frac{\sum_{t=2}^{n} x_{t-1}}{n} - \mu_{\beta 2} \frac{\sum_{t=3}^{n} x_{t-2}}{n} \right) \sum_{t=2}^{n} x_{t-1} - \hat{\mu}_{\beta 2} \sum_{t=3}^{n} x_{t-2} \sum_{t=2}^{n} x_{t-1}$$

$$\hat{\mu}_{\beta 1} \left( \sum_{t=2}^{n} x_{t-1}^2 - \frac{\left(\sum_{t=2}^{n} x_{t-1}\right)^2}{n} \right) =$$

$$\sum_{t=2}^{n} x_t x_{t-1} - \frac{\sum_{t=1}^{n} x_t \sum_{t=2}^{n} x_{t-1}}{n} + \hat{\mu}_{\beta 2} \left( \frac{\sum_{t=3}^{n} x_{t-2} \sum_{t=2}^{n} x_{t-1}}{n} - \sum_{t=3}^{n} x_{t-2} x_{t-1} \right).$$

Each term of the above equation replaces to $A_1$, $A_2$, and $A_3$.

Let

$$A_1 = \frac{\sum_{t=3}^{n} x_{t-2} \sum_{t=2}^{n} x_{t-1}}{n} - \sum_{t=3}^{n} x_{t-2} x_{t-1},$$

$$A_2 = \sum_{t=2}^{n} x_{t-1}^2 - \frac{\left(\sum_{t=2}^{n} x_{t-1}\right)^2}{n},$$

$$A_3 = \sum_{t=2}^{n} x_t x_{t-1} - \frac{\sum_{t=1}^{n} x_t \sum_{t=2}^{n} x_{t-1}}{n}.$$

So it is easy to write as

$$\hat{\mu}_{\beta 1} A_2 - A_3 = \mu_{\beta 2} A_1. \qquad (14)$$

The estimator $\hat{\mu}_{\beta 2}$ is used the process as the $\hat{\mu}_{\beta 1}$ by differential with respect to parameter $\mu_{\beta 2}$ as

$$\frac{\partial}{\partial \mu_{\beta 2}} \sum_{t=1}^{n} (u_t)^2 = \frac{\partial}{\partial \mu_{\beta 1}} \sum_{t=1}^{n} \left(x_t - \alpha - \mu_{\beta 1} x_{t-1} - \mu_{\beta 2} x_{t-2}\right)^2$$

$$= 2\sum_{t=1}^{n} \left(x_t - \alpha - \mu_{\beta 1} x_{t-1} - \mu_{\beta 2} x_{t-2}\right) x_{t-2}$$

$$= \sum_{t=3}^{n} x_t x_{t-2} - \alpha \sum_{t=3}^{n} x_{t-2} - \mu_{\beta 1} \sum_{t=3}^{n} x_{t-1} x_{t-2} - \mu_{\beta 2} \sum_{t=3}^{n} x_{t-2}^2$$

$$= 0$$

, and

$$\hat{\mu}_{\beta 2} = \frac{\sum_{t=3}^{n} x_t x_{t-2} - \hat{\alpha}\sum_{t=3}^{n} x_{t-2} - \hat{\mu}_{\beta 1}\sum_{t=3}^{n} x_{t-2}\sum_{t=2}^{n} x_{t-1}}{\sum_{t=3}^{n} x_{t-2}^2}. \quad (15)$$

From (15), let us replace in (12) and the solution of $\hat{\alpha}$ can be rewritten in the following form :

$$\hat{\mu}_{\beta 2}\sum_{t=3}^{n} x_{t-2}^2 =$$

$$\sum_{t=3}^{n} x_t x_{t-2} - \left( \frac{\sum_{t=1}^{n} x_t}{n} - \beta_1 \frac{\sum_{t=2}^{n} x_{t-1}}{n} - \beta_2 \frac{\sum_{t=3}^{n} x_{t-2}}{n} \right) \sum_{t=3}^{n} x_{t-2} - \hat{\mu}_{\beta 1}\sum_{t=3}^{n} x_{t-2}x_{t-1}$$

$$\hat{\mu}_{\beta 2}\left( \sum_{t=3}^{n} x_{t-2}^2 - \frac{\left(\sum_{t=3}^{n} x_{t-2}\right)^2}{n} \right) =$$

$$\sum_{t=3}^{n} x_t x_{t-2} - \frac{\sum_{t=1}^{n} x_t \sum_{t=3}^{n} x_{t-2}}{n} + \hat{\mu}_{\beta 1}\left( \frac{\sum_{t=2}^{n} x_{t-1}\sum_{t=3}^{n} x_{t-2}}{n} - \sum_{t=3}^{n} x_{t-2}x_{t-1} \right).$$

Let

$$A_4 = \sum_{t=3}^{n} x_{t-2}^2 - \frac{\left(\sum_{t=3}^{n} x_{t-2}\right)^2}{n},$$

$$A_5 = \sum_{t=2}^{n} x_t x_{t-1} - \frac{\sum_{t=1}^{n} x_t \sum_{t=3}^{n} x_{t-2}}{n}$$

, it can be rewritten as

$$\hat{\mu}_{\beta 2} A_4 - A_5 = \mu_{\beta 1} A_1. \quad (17)$$

From (14) and (17), it can be solved to estimate $\hat{\mu}_{\beta 2}$ by

$$\frac{\hat{\mu}_{\beta 1} A_2 - A_3}{A_1} = \hat{\mu}_{\beta 2} \quad (18)$$

$$\frac{\hat{\mu}_{\beta 2} A_4 - A_5}{A_1} = \hat{\mu}_{\beta 1} \quad (19)$$

$$\left( \frac{\hat{\mu}_{\beta 2} A_4 - A_5}{A_1} \right) A_2 - A_3 = \hat{\mu}_{\beta 2} A_1 \quad (20)$$

$$\frac{\hat{\mu}_{\beta 2} A_4 A_2}{A_1} - \frac{A_5 A_2}{A_1} - A_3 = \hat{\mu}_{\beta 2} A_1 \quad (21)$$

$$\hat{\mu}_{\beta 2}\left( \frac{A_4 A_2}{A_1} - A_1 \right) = \frac{A_5 A_2}{A_1} + A_3 \quad (22)$$

$$\hat{\mu}_{\beta 2} = \frac{\frac{A_5 A_2}{A_1} + A_3}{\frac{A_4 A_2}{A_1} - A_1} \quad (23)$$

The estimate $\hat{\mu}_{\beta 1}$ is computed by

$$\left( \frac{\hat{\mu}_{\beta 1} A_2 - A_3}{A_1} \right) A_4 - A_5 = \hat{\mu}_{\beta 1} A_1 \quad (24)$$

$$\frac{\hat{\mu}_{\beta 1} A_4 A_2}{A_1} - \frac{A_4 A_3}{A_1} - A_5 = \hat{\mu}_{\beta 1} A_1 \quad (25)$$

$$\hat{\mu}_{\beta 1}\left( \frac{A_4 A_2}{A_1} - A_1 \right) = \frac{A_4 A_3}{A_1} + A_5 \quad (26)$$

$$\hat{\mu}_{\beta 1} = \frac{\frac{A_4 A_3}{A_1} + A_5}{\frac{A_4 A_2}{A_1} - A_1}. \quad (27)$$

For RCA(2) model, it can be fitted model as

$$\hat{x}_t = \hat{\alpha} + \mu_{\beta 1} x_{t-1} + \hat{\mu}_{\beta 2} x_{t-2}, t = 3, 4, \ldots n.$$

# 4. APPLICATION IN REAL DATA

In this section, we consider the application of RCA(1) and RCA(2) models using the least squares method that developed in the previous section. The mean square error is a criterion to show the performance of these models that defined by

$$MSE = \frac{\sum_{t=2}^{n} (x_t - \hat{x}_t)^2}{n-1},$$

where $x_t$ denotes the real data, and $\hat{x}_t$ denotes the fitted values.

<u>Case 1</u> The monthly average of the Stock Exchange of Thailand (SET) index is applied for RCA(1) and RCA(2) models in term of real time series data. The SET index is started for trading on April 30, 1975, and we have fitted data from 1975 to 2016 for estimating a parameter of RCA(1) and RCA(2) models, then we forecast data from 2017-2018. This data is collected from http://www.set.or.th/th/market/market/statistics.html which is shown the time series plot in Figure 1.

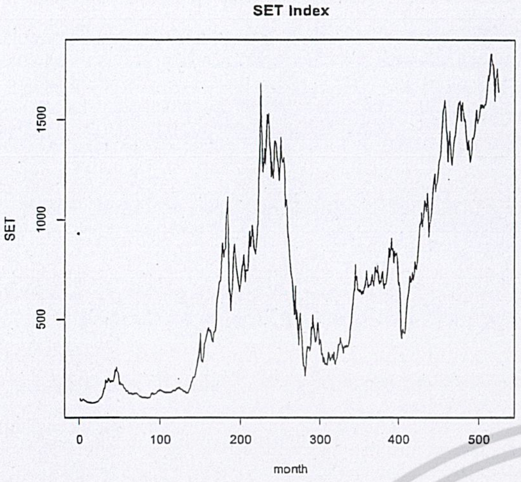



**Figure 1. The time series plot for the SET index**

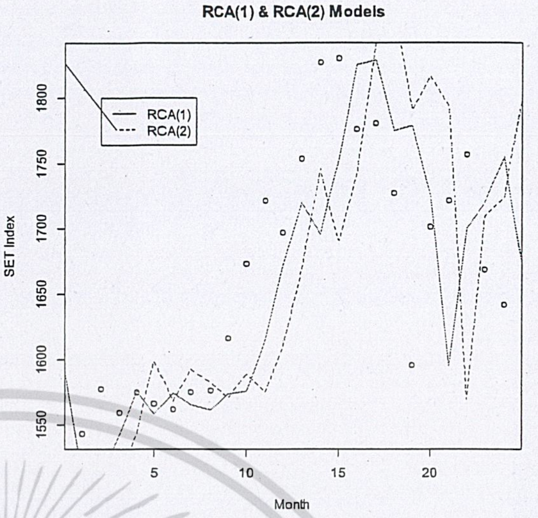



**Figure 3. The plot for the SET index and forecasting of RCA(1) and RCA(2) models**

<u>Case 2</u> We consider the exchange rate of Thai Baht to one U.S. Dollar of 1,426 records of the daily volume of the exchange rate from January, 2 of 2013 to December, 2 of 2018. We performed the RCA(1) and RCA(2) models to forecast on December 2 of 2018 to January, 2 of 2019. This data is collected from http://www.bot.or.th/thai/exchangerate which is shown the time series plot in Figure 2.

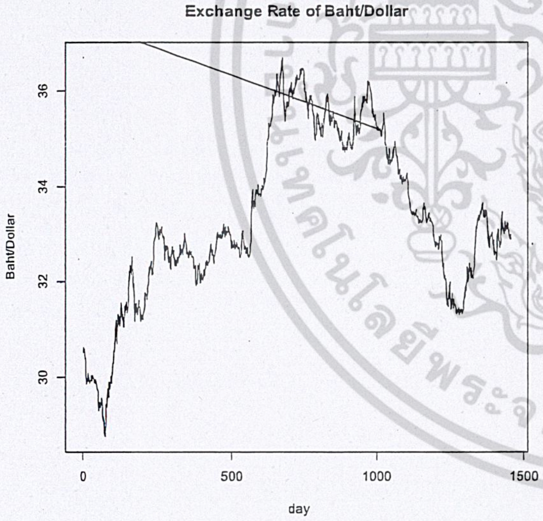



**Figure 2. The time series plot for the exchange rate of Baht/Dollar**

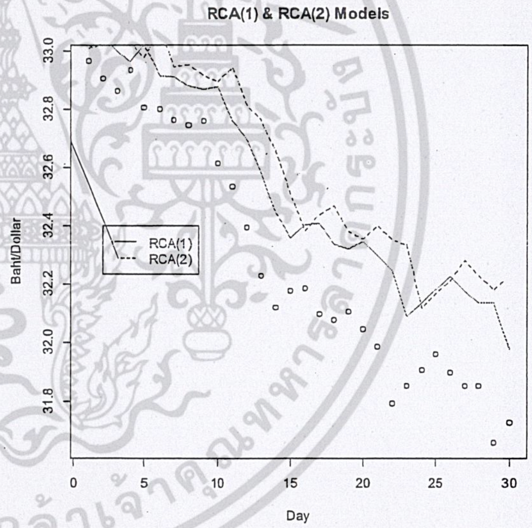



**Figure 4. The plot for the exchange rate of Baht/Dollar and forecasting of RCA(1) and RCA(2) models**

Based on fitting an RCA(1) and RCA(2) models to the fitted SET index volume data and exchange rate of Baht/Dollar. We showed the forecasting values of two data sets in Figures 3 and 4. Figure 3 gives the plot of the SET index and the dashed line of RCA(1) and the dotted line of RCA(2). Figure 4 presents the plot of the exchange rate of Baht/Dollar and the dashed line of RCA(1) and the dotted line of RCA(2).

Comparing with Figures 3- 4, it is seen that the forecasting values of RCA(1) are fairly close to the actually observed series on two data sets. Therefore, we should be more convinced by the MSE of RCA(1) given 3010.227, but the MSE of RCA(2) is shown by 9405.540 based on the SET index. For exchange rate Baht/Dollar, the MSE of RCA(1) as 0.1092 is less than the RCA(2) as 0.1396.

## 5. CONCLUSION

In this paper, we studied the least squares method of the first and the second orders in RCA or called RCA(1) and RCA(2) models. We showed the process to estimate the unknown parameter of these models. By using the differential with respect to a parameter, we also considered the sum squares error which includes the

RCA(1) and RCA(2) models. Through a time series data, we interest the two data sets such as the SET index in term of monthly volume and the exchange rate of Baht/Dollar in a class of daily volume. For fitting model, we are also interested in the power of estimating by the mean square error. We can see that RCA(1) model outperforms the RCA(2) model. We would recommend fitting the RCA(1) model on time series data by least squares method where nonstationary data is expected.

## 6. ACKNOWLEDGMENTS

This work was financially supported by the grant research fund of the Faculty of Science, King Mongkut's Institute of Technology Ladkrabang.

## 7. REFERENCES

[1] Tsay, R. 1987. Conditional heteroscedastic time series models. *Journal of the American Statistical Association*. 82, 590-604. DOI = 10.2307/2289470 https://www.jstor.org/stable/2289470

[2] Nicholls, D.F. and Quinn, B.G. 1982. *Random coefficient autoregressive models: An introduction*, Springer- Verlag Inc, New York.

[3] Hwang, S.Y., and Basawa, U.V. 1998. Parameter estimation for generalized random coefficient autoregressive process. *Journal of Statistics Planning and Inference*. 68, 323-337. https://doi.org/10.1016/S0378-3758(97)00147-X

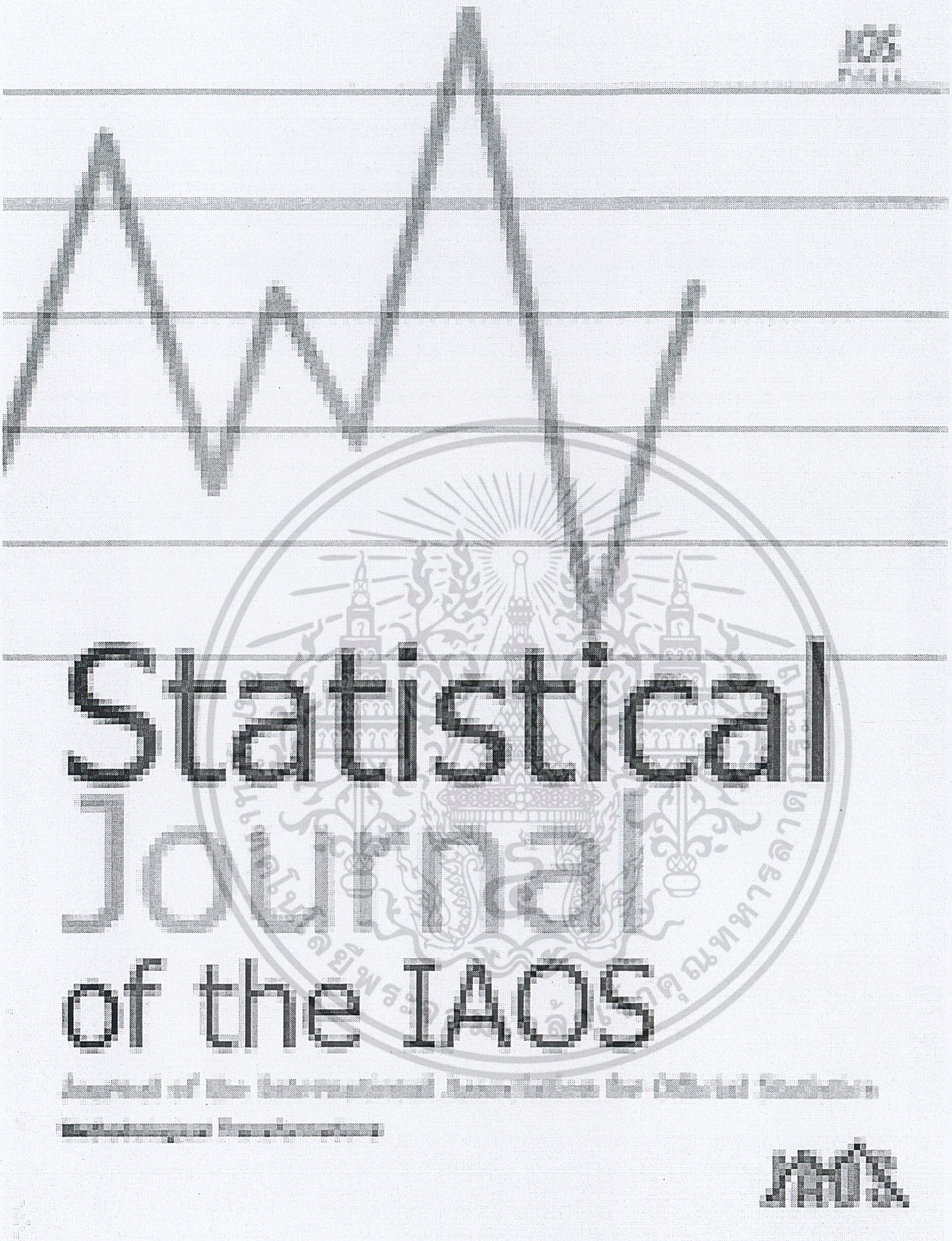
[4] Chandra, S.A. and Taniguchi, M. 2001. Estimating function for nonlinear time series data. *Annals of the Institute of Statistics Mathematic*. 53(1), 125-136. https://link.springer.com/article/10.1023/A:1017924722711

[5] Araveeporn, A. 2012. The least-squares criterion of the random coefficients dynamic regression model. *Journal of Statistical Theory and Practice*. 6(2), 315-333. https://dx.doi.org/10.1080/15598608.2012.673891

[6] Araveeporn, A. 2013. A comparison of two least-squared random coefficient autoregressive models: with and without autocorrelated errors. *International Journal of Advanced Statistics and Probability*. 1(3), 151-162. www.sciencepubco.com/index.php/IJASP

[7] Wang, D., and Ghosh, S.K. 2008. Bayesian estimation and unit root tests for random coefficient autoregressive models. *Model Assisted Statistics and Applications*. 3, 4, 281-295.

## Statistical Journal of the IAOS - Volume 35, issue 2




ISSN 1874-7655 (P)
ISSN 1875-9254 (E)

This is the flagship journal of the International Association for Official Statistics and is expected to be widely circulated and subscribed to by individuals and institutions in all parts of the world.

The main aim of the journal is to support the IAOS mission by publishing articles to promote the understanding and advancement of official statistics and to foster the development of effective and efficient official statistical services on a global basis. Papers are expected to be of wide interest to readers. Such papers may or may not contain strictly original material. All papers are refereed.

The journal has an Editor-in-Chief who is responsible for ensuring that the journal focuses on current and emerging issues and challenges related to the management, production and use of official statistics and related public policy matters. The journal should publish papers of wide interest to both users and producers of official statistics. The journal should encourage papers with a focus on the basic principles for official statistics covering areas such as the importance of applying the best scientific methods, the need for statistical independence, balancing the needs of users with the burden on respondents, the continuing challenges around confidentiality, and the growing need for consistency and coherence across statistical domains and over time and for international comparability.

Authors and referees of all papers should bear in mind that many of the readers of the journal (particularly in the developing world) do not have easy access to libraries or to other journals, and therefore are strongly encouraged to make the papers as self-contained as possible, while of course giving proper bibliographic credit. Papers of more than 30 manuscript pages, including diagrams, tables and references, will normally not be considered for publication.

Show: 10 results per page

☐ Mark all

## ☐ Assessing the quality of life in the European Union: The European Index of Life Satisfaction (EILS) (https://content.iospress.com:443/articles/statistical-journal-of-the-iaos/sji180481)

**Authors:** Maricic, Milica (https://content.iospress.com:443/search?q=author%3A%28%22Maricic%2C+Milica%22%29)

**Article Type:** Research Article

**Abstract:** Quality of Life (QoL) is a multidimensional phenomenon concerned with the overall well-being of a society or an individual. The indicator-based approach to measuring QoL has attracted the attention of both researchers who propose composite indexes and policy creators who eventually use their results in the process of decision making. To overcome the observed disadvantages of the currently devised composite indexes on QoL, herein we propose a composite index which ranks European Union (EU) member states based on the subjective satisfaction indicators measured by Eurostat. The statistical multivariate analysis on which we base our proposed framework are the Principal Component ... Show more

**Keywords:** Quality of life, Composite index, European Union, Composite I-distance Indicator methodology, enhanced Scatter Search

**DOI:** 10.3233/SJI-180481

**Citation:** Statistical Journal of the IAOS (https://content.iospress.com:443/journals/statistical-journal-of-the-iaos), vol. 35, no. 2, pp. 261-267, 2562

## ☐ An estimating parameter of nonparametric regression model based on smoothing techniques (https://content.iospress.com:443/articles/statistical-journal-of-the-iaos/sji180477)

**Authors:** Araveeporn, Autcha (https://content.iospress.com:443/search?q=author%3A%28%22Araveeporn%2C+Autcha%22%29)

**Article Type:** Research Article

**Abstract:** This paper studies the estimating parameter of a nonparametric regression model that consists of the function of independent variables and observation of dependent variables. The smoothing spline, penalized spline, and B-spline methods in a class of smoothing techniques are considered for estimating the unknown parameter on nonparametric regression model. These methods use a smoothing parameter to control the smoothing performance on data set by using a cross-validation method. We also compare these methods by fitting a nonparametric regression model on simulation data and real data. The nonlinear model is a simulation data which is generated in two different models in ... Show more

**Keywords:** B-spline, nonparametric regression, penalized spline, smoothing spline

**DOI:** 10.3233/SJI-180477

**Citation:** Statistical Journal of the IAOS (https://content.iospress.com:443/journals/statistical-journal-of-the-iaos), vol. 35, no. 2, pp. 269-276, 2562

---

## ☐ Influence of technologies on the growth rate of GDP from agriculture: A case study of sustaining economic growth of the agriculture sector in Bihar (https://content.iospress.com:443/articles/statistical-journal-of-the-iaos/sji180436)

**Authors:** Sinha, Jitendra Kumar (https://content.iospress.com:443/search?q=author%3A%28%22Sinha%2C+Jitendra+Kumar%22%29)

**Article Type:** Research Article

**Abstract:** The influence of agricultural technologies on the growth of agricultural value-added based on time series data of Bihar (India) over the period 1990–2016 has been examined in this paper. The technological progress appears to be a major determinant of boosting the potential productivity of land and affecting positively the economic growth. The results indicated that there are significant and certain benefits from the utilization of a system of technological innovations including mechanization, renewed capital stocks, as well as transfer of new knowledge to farmers' and permanent cropping practices. Farming practices involving crop rotation, multi-cropping, and agro forestry are recommended to ... Show more

**Keywords:** Sustainable economic growth, agricultural technology, Cobb-Douglas production function

**DOI:** 10.3233/SJI-180436

**Citation:** Statistical Journal of the IAOS (https://content.iospress.com:443/journals/statistical-journal-of-the-iaos), vol. 35, no. 2, pp. 277-287, 2562

---

## ☐ Integrating the results of a nonresponse follow-up survey into the survey from which its items were selected (https://content.iospress.com:443/articles/statistical-journal-of-the-iaos/sji170406)

**Authors:** Kott, Phillip S. (https://content.iospress.com:443/search?q=author%3A%28%22Kott%2C+Phillip+S.%22%29)

**Article Type:** Research Article

**Abstract:** A nonresponse follow-up (NRFU) survey was conducted for the National Pilot of the Residential Energy Consumption Survey (RECS), an address-based sample survey of potential primary residences in the US enumerated by web and mail. Virtually all unit (i.e., whole-record) nonrespondents to the National Pilot were sent a short mail questionnaire containing 18 key items from the full survey. Here, we first compare two ways of adjusting variables collected on the NRFU for unit nonresponse. In one, only the weights for respondents to the full National Pilot survey were adjusted to compensate for nonresponse using a calibration weighting procedure that assumes ... Show more

**Keywords:** Residential Energy Consumption Survey, calibration weighting, augmented sample, compositing factor

**DOI:** 10.3233/SJI-170406

**Citation:** Statistical Journal of the IAOS (https://content.iospress.com:443/journals/statistical-journal-of-the-iaos), vol. 35, no. 2, pp. 289-297, 2562

## ☐ Determinants of rural household financial literacy: Evidence from south india (https://content.iospress.com:443/articles/statistical-journal-of-the-iaos/sji180438)

**Authors:** Jayanthi, M. (https://content.iospress.com:443/search?q=author%3A%28%22Jayanthi%2C+M.%22%29) | Rau, S.S. (https://content.iospress.com:443/search?q=author%3A%28%22Rau%2C+S.S.%22%29)

**Article Type:** Research Article

**Abstract:** The purpose of this paper is to investigate the study of rural household's financial literacy. Financial literacy is understood as the set of knowledge, attitudes, and behaviors, which has assumed a fundamental role in allowing and enabling people to make responsible decisions as they strive to attain financial wellbeing. In this research, the researcher throws a descriptive research design to explore the relationship between variables. Respondents were obtained from 126 households in rural areas of Tamil Nadu and data was collected using the stratified sampling method. The current study found that education, marital status, type of family and banks relationship ... Show more

**Keywords:** Financial literacy, demographic variables, rural household, Tamil Nadu

**DOI:** 10.3233/SJI-180438

**Citation:** Statistical Journal of the IAOS (https://content.iospress.com:443/journals/statistical-journal-of-the-iaos), vol. 35, no. 2, pp. 299-304, 2562

Statistical Journal of the IAOS 35 (2019) 269–276
DOI 10.3233/SJI-180477
IOS Press


# An estimating parameter of nonparametric regression model based on smoothing techniques

Autcha Araveeporn
*Department of Statistics, Faculty of Science, King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand*
*E-mail: kaautcha@hotmail.com*

**Abstract.** This paper studies the estimating parameter of a nonparametric regression model that consists of the function of independent variables and observation of dependent variables. The smoothing spline, penalized spline, and B-spline methods in a class of smoothing techniques are considered for estimating the unknown parameter on nonparametric regression model. These methods use a smoothing parameter to control the smoothing performance on data set by using a cross-validation method. We also compare these methods by fitting a nonparametric regression model on simulation data and real data. The nonlinear model is a simulation data which is generated in two different models in terms of mathematical function based on statistical distribution. According to the results, the smoothing spline, the penalized spline, and the B-spline methods have a good performance to fit nonlinear data by considering the hypothesis testing of biased estimator. However the penalized spline method shows the minimum mean square errors on two models. As real data, we use the data from a light detection and ranging (LIDAR) experiment that contained the range distance travelled before the light as an independent variable and the logarithm of the ratio of received light from two laser sources as a dependent variable. From the mean square errors of fitting data, the penalized spline again shows the minimum values.

Keywords: B-spline, nonparametric regression, penalized spline, smoothing spline

## 1. Introduction

In statistical modelling, regression analysis is a statistical process for estimating parameters of the relationships between dependent and independent variables in terms of a regression function. However, regression analysis requires an assumption of the underlying regression function to be met. If an inappropriate assumption is used, it is possible to produce misleading results. To overcome this problem, the nonparametric regression is a choice to analyze data when the data are not meeting the assumption of regression analysis. The nonparametric regression is an alternative way for looking at scatter diagram smoothing to depict the relationship between dependent and independent variables. The single independent variable is called scatterplot smoothing it can be used to enhance the visual appearance to help our eyes pick out the trend in the plot.

The smoothing technique is a part of a method to estimate unknown parameters (trend or smoothing estimators) of nonparametric regression models. There are many popular smoothing techniques such as the smoothing spline [1,2], the penalized spline [3], and the B-spline [4]. The estimating parameters of these methods depend on the smoothing parameter which is controlled the trade off between fidelity to the data and roughness of function. Smoothing Spline (SS) is a technique that estimates the natural polynomial spline by minimizing the penalized sum of squares based on a smoothing parameter. The penalized Spline (PS) smoother is approximated by minimizing the truncated power function on a low rank thin-plate spline depended on the smoothing parameter. The concept of the B-spline is similar to the smoothing spline and penalized spline. This requires the piecewise constant B-spline that can be obtained from truncated counterparts by differencing the B-spline function.

In this paper, we consider the nonparametric regression model in Section 2, and use the smoothing spline, penalized spline, and B-spline methods to estimate the unknown parameter of nonparametric regression model in Section 3. In Sections 4 and 5, we show the estimation of these methods for simulation data and real data. The conclusion is presented in Section 6.

## 2. The nonparametric regression model

The nonparametric regression model consists of the cubic spline of piecewise polynomials function based on a function of independent variables ($S(x_t)$), error process ($\varepsilon_t$), and dependent variables ($y_t$) following

$$y_t = S(x_t) + \varepsilon_t, t = 1, 2, 3, ..., n. \tag{1}$$

The error process is assumed to follow the normal distribution with mean zero and variance one.

## 3. Method of smoothing techniques

The following smoothing techniques show the process to estimate parameters based on nonparametric regression model.

### *3.1. Smoothing spline method*

Wahba [1] defined the natural polynomial spline $S(x_t) = S_K^m(x_t)$ as a real-valued function on $[a, b]$ with the aid of $K$ so-called knots $-\infty \leqslant a < x_1 < x_2 < ... < x_K < b \leqslant \infty$. The class of $m$-order splines with domain $[a, b]$ will be denoted by $W^m[a, b]$.

The natural measure associated with the function $f \in W^m[a, b]$ that used to measure the roughness of curve which is called the quadratic penalty function given by

$$\int_b^a \left\{S^{(m)}(x_t)\right\}^2 dx \tag{2}$$

where $S^{(m)}(x_t)$ is the $m$th derivative of $S(x)$ with respect to $x$.

Consider the simple nonparametric regression model, to estimate $\hat{S}(\cdot)$ minimizes $S_K^{(m)}(x_t)$ over the class of function $S(\cdot)$ following

$$\begin{aligned} S_K^{(m)}(x_t) = \min \sum_{t=1}^{n} \{y_t - S(x_t)\}^2 \\ + \lambda \int_a^b \left\{S^{('')}(x_t)\right\}^2 dx \end{aligned} \tag{3}$$

where $\lambda > 0$ denotes a smoothing parameter. In this study, we emphasize $m = 2$ so-called the natural cubic spline which is commonly considered in the statistical literature [2].

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

$$\underset{\sim}{S} = S(x_t) = \beta_0 + \beta_1 S_1(x) + \dots \beta_{n+3} S_{n+3}(x),$$

$$\underset{\sim}{\gamma} = S''(x_t), \quad t = 1, 2, 3, \dots, n.$$

Let $\underset{\sim}{S}$ be the vector $(S_1, \dots, S_{n+3})^{\mathrm{T}}$ and let $\underset{\sim}{\gamma}$ be the vector $(\gamma_1, \dots, \gamma_{n+3})^{\mathrm{T}}$.

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

$$Q = \begin{pmatrix} h_1^{-1} & 0 & \cdots & 0 \\ -h_1^{-1} - h_2^{-1} & h_2^{-1} & \cdots & 0 \\ h_2^{-1} & -h_2^{-1} - h_3^{-1} & \cdots & 0 \\ 0 & h_3^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & h_{n-1}^{-1} \end{pmatrix}_{(n)\times(n-2)},$$

where $h_t = x_{t+1} - x_t$, for $t = 1, 2, \dots, n$, then $Q$ is a $n \times (n-2)$ matrix.

Matrix $R$ is a symmetric $(n-2) \times (n-2)$ matrix with elements below

$$R = \begin{pmatrix} \frac{1}{3}(h_1 + h_3) & \frac{1}{6}h_2 & \cdots & 0 \\ \frac{1}{6}h_2 & \frac{1}{3}(h_2 + h_3) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{3}(h_{n-2} + h_{n-1}) \end{pmatrix}_{(n-2)\times(n-2)}.$$

The matrix $K$ can be decomposed by

$$K = QR^{-1}Q^{\mathrm{T}}. \tag{4}$$

The roughness penalty will satisfy

$$\begin{aligned} \int_a^b \left\{S''(x_t)\right\}^2 dx_t = \gamma^{\mathrm{T}} R \gamma = S^{\mathrm{T}} K S \\ = \Omega_N(i, j). \end{aligned} \tag{5}$$

To illustrate, it can be written in matrix form introduced by [2] as residual sum of squares (RSS)

$$\begin{aligned} RSS &= \sum_{t=2}^{n} \{y_t - S(x_t)\}^2 \\ &= (\underset{\sim}{y} - \underset{\sim}{S})^{\mathrm{T}}(\underset{\sim}{y} - \underset{\sim}{S}), \end{aligned} \tag{6}$$

where $\underset{\sim}{y} = (y_1, \ldots, y_n)^{\mathrm{T}}$ and $S = (S(x_1), \ldots, S(x_n))^{\mathrm{T}}$. Letting $N$ be a matrix with $N(i,j) = S_j(x_i)$ and $\underset{\sim}{S} = N\beta$.

The roughness penalty term $\int S''^2$ as $\Omega_N$ in Eq. (5) to obtain

$$\hat{S}_\lambda(x_t) = (\underset{\sim}{y} - N\beta)^{\mathrm{T}}(\underset{\sim}{y} - X\beta) + \lambda\beta^{\mathrm{T}}\Omega_N\beta \tag{7}$$

It therefore follows that Eq. (7) has a unique minimum, other smoothing spline estimator is obtained by

$$\hat{\beta} = (N^{\mathrm{T}}N + \lambda\Omega_N)^{-1}N^{\mathrm{T}}\underset{\sim}{y},$$

then

$$\hat{S}_\lambda(x_t) = N(N^{\mathrm{T}}N + \lambda\Omega_N)^{-1}N^{\mathrm{T}}. \tag{8}$$

In this paper, we also select the smoothing parameter using the method of generalized cross-validation (GCV) suggested by Wahba [5] and Craven and Wahba [6]. In practice, this step can be implemented by using the function of smooth.spline in the software R.

### 3.2. Penalized spline method

Eubank [7,8] introduced the regression spline that the local neighbourhoods are specified by a group of locations:

$$\tau_0, \tau_1, \tau_2, \ldots, \tau_K, \tau_{K+1} \tag{9}$$

in the range of interval $[a, b]$, where $a = \tau_0 < \tau_1 < \ldots < \tau_K < \tau_{K+1} < b$. These locations are known as knots, and $\tau_r, r = 1, 2, \ldots, K$ are called interior knots.

A regression spline can be constructed using the $k$-th degree truncated power basis or called the B-spline basis with $K$ knots $\tau_1, \tau_2, \ldots, \tau_K$:

$$1, x_t, \ldots, x_t^k, (x_t - \tau_1)_+^k, \ldots, (x_t - \tau_K)_+^k, \tag{10}$$

where $w_+^k$ denotes $k$-th power of the positive part of $w$ where $w_+ = \max(0, w)$. The first $(k+1)$ basis functions of the truncated power basis Eq. (12) are polynomials of degree up to $k$, and the others are all the truncated power functions of degree $k$. A regression spline can be expressed as

$$S(x_t) = \sum_{s=0}^{k} \beta_s x_t^s + \sum_{r=1}^{K} \beta_{k+r}(x_t - \tau_r)_+^k, \tag{11}$$

where $\beta_0, \beta_1, \ldots, \beta_{k+K}$ are the unknown coefficients to be estimated by a suitable loss minimization.

The penalized spline is a method to estimate a unknown smooth function using the truncated power function [9], and the penalized spline can be expressed as

$$S(x_t) = \sum_{j=0}^{m-1} \alpha_j x_s^j + \sum_{k=1}^{K} \beta_k (x_t - \tau_k)^{2m-1}, \tag{12}$$

where $\beta = (\beta_1, \ldots, \beta_K)^{\mathrm{T}} \sim N\left(0, \sigma_\beta^2 \Omega^{-1/2}(\Omega^{1/2})^{\mathrm{T}}\right)$, and the $(l, k)$ th entry of $\Omega$ is $|\tau_1 - \tau_k|^{2m-1}$ and only the coefficient of $|x_t - \tau_k|^{2m-1}$ are penalized so that a reasonably large order $K$ can be used.

In this case, we focus $m = 2$, as the natural cubic spline, or called low-rank thin-plate spline which present of $S(\cdot)$ as

$$S(x_t) = \alpha_0 + \alpha_1 x_t + \sum_{k=1}^{K} \beta_k |x_t - \tau_k|^3, \tag{13}$$

where $\theta = (\alpha_0, \alpha_1, \beta_1, \ldots, \beta_K)^{\mathrm{T}}$ is the vector of regression coefficients, and $\tau_1 < \tau_2 < \ldots < \tau_K$ are fixed knots. The number of knots, $K$ can be selected using a cross-validation method or information theoretic methods (e.g., BIC or AIC).

This class of penalized spline smoothers $(\hat{S}(\cdot))$ may also be expressed as

$$\hat{S}(x_t) = C(C^{\mathrm{T}}C + \lambda^3 D)^{-1}C^{\mathrm{T}}\underset{\sim}{y}, \tag{14}$$

where

$$C = \left[1\; x_t \; |x_t - \tau_k|^3_{1 \leqslant k \leqslant K}\right]_{1 \leqslant t \leqslant n},$$

$$D = \begin{bmatrix} 0_{2\times 2} & 0_{2\times K} \\ 0_{K\times 2} & (\Omega_K^{1/2})^{\mathrm{T}}\Omega_K^{1/2} \end{bmatrix},$$

and $\lambda = \sigma_\beta^2/\sigma_\varepsilon^2$ is a smoothing parameter. The penalized spline smoothers are estimated by using the SemiPar package in the software R.

### 3.3. B-spline method

B-splines are very interesting as a basic function for univariate independent variable of nonparametric regression function. De Boor [10] gave an algorithm to compute B-spline of lower degree on piece wise polynomials function.

The $m^{\text{th}}$ degree of B-spline function are evaluated from $(m-1)^{\text{th}}$ degree as

$$\begin{aligned} B_j^m(x) &= \frac{x_t - \tau_j}{\tau_{j+m-1} - \tau_j} B_j^{m-1} \\ &\quad + \frac{\tau_{j+m} - x_t}{\tau_{j+m} - \tau_{j+1}} B_{j+1}^{m-1}, \end{aligned} \tag{15}$$

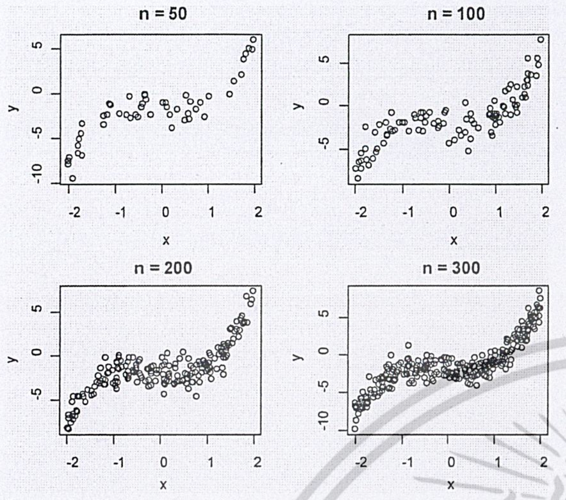



Fig. 1. The scatter plot of dependent and independent variables on model 1.

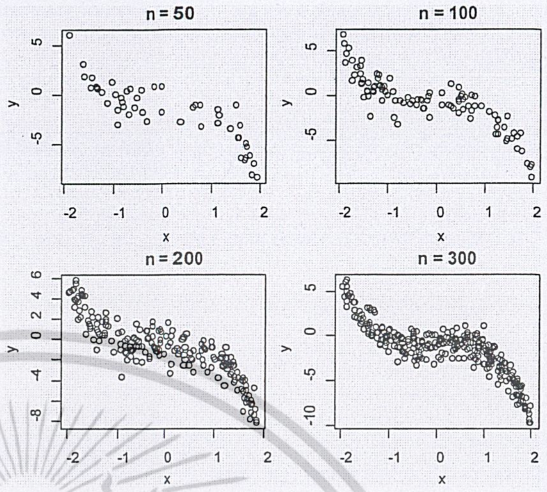



Fig. 2. The scatter plot of dependent and independent variables on model 2.

where basis of order $m$ with knots $\{B_i^m | i = M - m + 1, ..., M + K\}$, and auxiliary knots $\tau_j$. B-splines base on non-zero over domain spanned by at most $M + 1$ knots. In this case, we focus the $m = 4$ or called the cubic B-spline with $K$ knots has basis expansion as

$$S(x_t) = \sum_{j=1}^{K+4} B_j^4(x_t)\beta_j.$$

The nonparametric regression model can written in form of B-splines as

$$y_t = \sum_{j=1}^{K+4} B_j^4(x_t)\beta_j + \varepsilon_t, t = 1, 2, 3, ..., n. \quad (16)$$

In matrix form, B-splines can be written in form a linear model

$$B = \begin{bmatrix} B_1^4(x_1) \dots B_{K+4}^4(x_1) \\ \vdots \qquad\qquad \vdots \\ B_1^4(x_n) \dots B_{K+4}^4(x_n) \end{bmatrix}, \underset{\sim}{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix},$$

$$\text{and } \underset{\sim}{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

The B-splines estimators are approximated by least square problems as

$$\hat{\underset{\sim}{\beta}} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_{K+4} \end{bmatrix} = (B^{\mathrm{T}}B)^{-1}B^{\mathrm{T}}\underset{\sim}{y}.$$

The B-spline and penalties are studied by Eilers and Marx [4] that advocate the use of the equally spaced knots, instead of the order statistics of the independent variable. The B-spline coefficients can be estimated as

$$\hat{\underset{\sim}{\beta}} = (B^{\mathrm{T}}B + \lambda D^{\mathrm{T}}D)^{-1}B^{\mathrm{T}}\underset{\sim}{y}, \quad (17)$$

where $D$ is a banded matric which correspond to the difference penalty and denote by

$$D = \begin{bmatrix} 0 & 0 & 0\dots0 & 0 \\ -1 & 1 & 0\dots0 & 0 \\ 0 & -1 & 1\dots0 & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & 0\dots-1 & 1 \end{bmatrix}.$$

The fitting cubic B-splines are $\hat{S}(x_t) = \sum_{j=1}^{K+4} B_j^4 (x_t)\hat{\beta}_j$. The smoothing parameter $\lambda$ choosing by minimizing the ordinary function or the generalized cross-validation function.

## 4. Simulation study

The nonlinear data of this study is simulated in two models for estimating the performance of smoothing techniques based on independent variables which are considered in the class of uniform distribution. These models in the process of construction a curve on mathematical function, that show the best fit to a series of data points. Figures 1 and 2 show the scatter plot of $x_t$ and $y_t$ on models 1 and 2 with 50, 100, 200, 300 sample sizes.

Table 1
The summary statistics of simulation studies with model 1 based on smoothing spline (SS), penalized spline (PS), and B-spline (BS)

| Sample sizes | Methods | Mean | S.D. | LCI | UCI | $t$-statistic | $p$-values |
|---|---|---|---|---|---|---|---|
| $n = 50$ | SS | −0.3344 | 6.5444 | −0.9095 | 0.2405 | −1.1428 | 0.2537 |
| | PS | −0.3484 | 6.7454 | −0.9411 | 0.2442 | −1.1552 | 0.2468 |
| | BS | 0.3666 | 6.6935 | −0.2214 | 0.9548 | 1.2249 | 0.2212 |
| $n = 100$ | SS | 0.0719 | 4.2680 | −0.3031 | 0.4469 | 0.3767 | 0.7065 |
| | PS | 0.0837 | 4.2029 | −0.2855 | 0.4530 | 0.4454 | 0.6562 |
| | BS | −0.0895 | 4.2152 | −0.4599 | 0.2807 | −0.4751 | 0.6349 |
| $n = 200$ | SS | −0.0954 | 2.8932 | −0.3504 | 0.1595 | −0.7353 | 0.4625 |
| | PS | −0.0921 | 2.9123 | −0.3480 | 0.1637 | −0.7072 | 0.4798 |
| | BS | 0.0149 | 3.2685 | −0.2299 | 0.2598 | 0.1200 | 0.9045 |
| $n = 300$ | SS | −0.1574 | 4.2673 | −0.5347 | 0.2197 | −0.8201 | 0.4125 |
| | PS | −0.1441 | 4.1982 | −0.5130 | 0.2247 | −0.7676 | 0.4431 |
| | BS | 0.1475 | 1.8514 | −0.2786 | 0.5738 | 0.6802 | 0.4966 |

Table 2
The summary statistics of simulation studies with model 2 based on smoothing spline (SS), penalized spline (PS), and B-spline (BS)

| Sample sizes | Methods | Mean | S.D. | LCI | UCI | $t$-statistic | $p$-values |
|---|---|---|---|---|---|---|---|
| $n = 50$ | SS | −0.0037 | 24.2710 | −2.1363 | 2.1288 | −0.0034 | 0.9972 |
| | PS | −0.2582 | 23.8736 | −2.3559 | 1.9893 | −0.2419 | 0.8089 |
| | BS | −0.3230 | 27.1906 | −2.7121 | 2.0660 | −0.2656 | 0.7906 |
| $n = 100$ | SS | −0.3966 | 15.8261 | −1.7871 | 0.9939 | −0.5603 | 0.5755 |
| | PS | −0.2489 | 16.4557 | −1.6947 | 1.1969 | −0.3382 | 0.7353 |
| | BS | 0.5461 | 16.2381 | −0.8806 | 1.9729 | 0.7520 | 0.4524 |
| $n = 200$ | SS | −0.3704 | 8.1031 | −1.0838 | 0.3430 | −1.0201 | 0.3082 |
| | PS | −0.3447 | 7.8716 | −1.0363 | 0.3468 | −0.9792 | 0.3279 |
| | BS | 0.4035 | 8.5116 | −0.3443 | 1.1514 | 1.0602 | 0.2896 |
| $n = 300$ | SS | 0.0411 | 7.3081 | −0.6055 | 0.6878 | 0.1250 | 0.9005 |
| | PS | 0.0408 | 7.0664 | −0.5800 | 0.6617 | 0.1291 | 0.8973 |
| | BS | 0.0344 | 7.2205 | −0.5999 | 0.6688 | 0.1066 | 0.9151 |

Model 1

$$S(x_t) = (x_t^3) - \cos(x_t) - \exp\left\{\frac{x_t}{1+|x_t|}\right\},$$

$$x_t \sim Uniform(-2,2), t = 1,2,3,...,n$$

$$y_t = S(x_t) + \varepsilon_t, \varepsilon_t \sim Normal(0,1),$$

$$t = 1,2,3,...,n$$

Model 2

$$S(x_t) = \sin(x_t) - \exp\left\{\frac{x_t}{1+|x_t|}\right\} - x_t^3,$$

$$x_t \sim Uniform(-2,2), t = 1,2,3,...,n$$

$$y_t = S(x_t) + \varepsilon_t, \varepsilon_t \sim Normal(0,1),$$

$$t = 1,2,3,...,n$$

The next step, the estimates of $S(x_t)$ or called $\hat{S}(x_t)$ are approximated from smoothing spline (SS), penalized spline (PS), and B-spline (BS) that used to compute the bias and MSE of $S(x_t)$ following

$$S_{bias} = \sum_{t=1}^{n}\left(\frac{\hat{S}(x_t) - S(x_t)}{S(x_t)}\right),$$

$$MSE = \frac{\sum_{t=1}^{n}\left[\hat{S}(x_t) - S(x_t)\right]^2}{n}.$$

The data are generated and repeated for fitting the model 500 times. A $t$-statistic is adopted to test that the mean of bias is equal the zero or called unbiased estimator. Tables 1 and 2 present the various summary statistics for the smoothing estimator obtained from three methods. The third and the fourth columns of these tables represent the sample mean and standard deviation of biases. The sample mean for the lower and upper bounds of the 95% confidence interval are given in the next two columns. The last two columns of these tables list the $t$-statistic, and $p$-values for hypothesis testing ($H_0 : \mu_{S_{bias}} = 0$ versus $H_0 : \mu_{S_{bias}} \neq 0$) that means when reject $H_0 : \mu_{S_{bias}} = 0$ the estimator SS, PS, and BS with bias. The histogram of the bias estimator of SS, PS, and BS in model 1 are presented in Figs 3–5, and model 2 are presented in Figs 6–8.

From Tables 1 and 2, by observing the $p$-values, the SS, PS, and BS provide asymptotically unbiased estimates for estimating parameter of $S(x_t)$ nearly for all sample sizes of two models. From the $p$-values for the two tables it is seen that are seen that the SS, PS, and

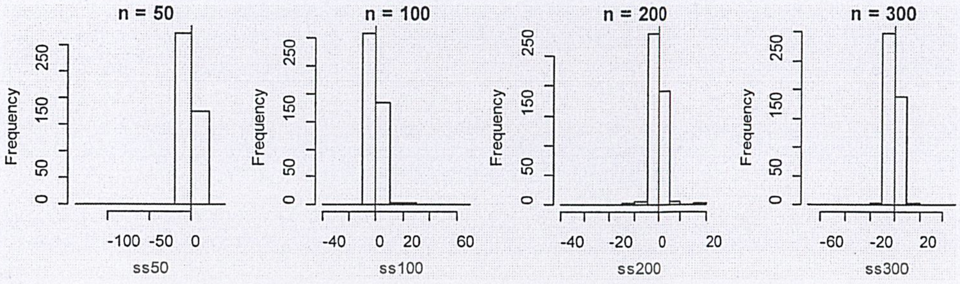



Fig. 3. Histogram of bias for fitting data of smoothing spline method with model 1.

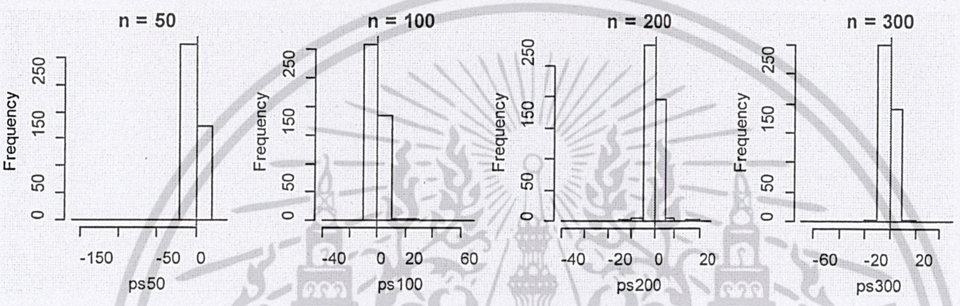



Fig. 4. Histogram of bias for fitting data of penalized spline method with model 1.

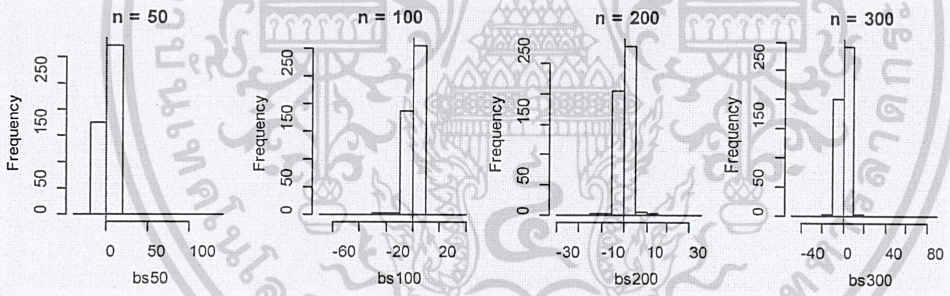



Fig. 5. Histogram of bias for fitting data of B-spline method with model 1.

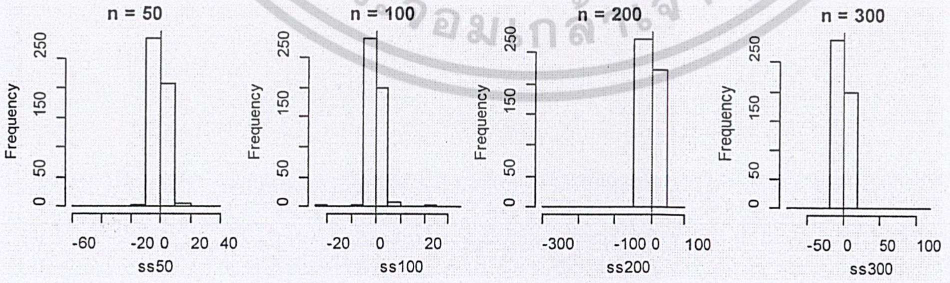



Fig. 6. Histogram of bias for fitting data of smoothing spline method with model 2.

BS of smoothing method have a good performance to fit data in a class of nonlinear data. From the histogram it is apparent that a standard deviation of relative biases increase with increasing sample sizes, so it makes the leptokurtic distribution. The average of MSE can answer the final question which smoothing method is

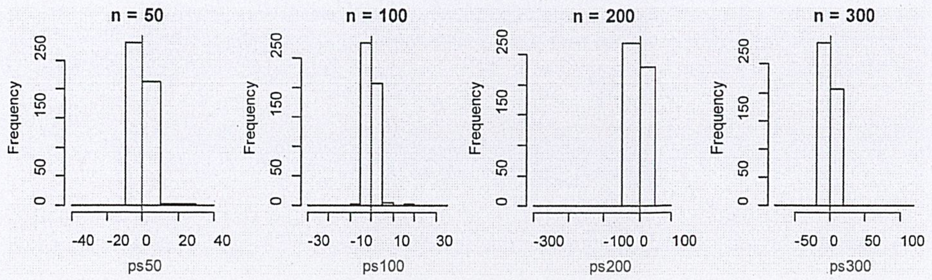



Fig. 7. Histogram of bias for fitting data of penalized spline method with model 2.

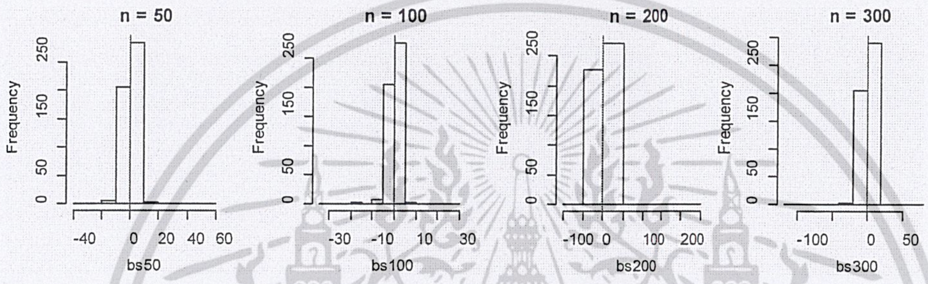



Fig. 8. Histogram of bias for fitting data of B-spline method with model 2.

Table 3
The average MSE of simulation studies with 3 models based on smoothing spline (SS), penalized spline (PS), and B-spline (BS)

| Sample sizes | Methods | Model 1 | Model2 |
|---|---|---|---|
| $n = 50$ | SS | 0.8225 | 0.8348 |
| | PS | **0.7712** | **0.7846** |
| | BS | 0.9174 | 0.9364 |
| $n = 100$ | SS | 0.9087 | 0.9029 |
| | PS | **0.8810** | **0.8719** |
| | BS | 0.9632 | 0.9508 |
| $n = 200$ | SS | 0.9528 | 0.9521 |
| | PS | **0.9387** | **0.9439** |
| | BS | 0.9784 | 0.9840 |
| $n = 300$ | SS | 0.9666 | 0.9648 |
| | PS | **0.9640** | **0.9615** |
| | BS | 0.9899 | 0.9877 |

the best estimator. Table 3 shows the average MSE for fitting 500 times on two models, and it can be seen that the PS method shows the minimum of average MSE for all sample sizes and models.

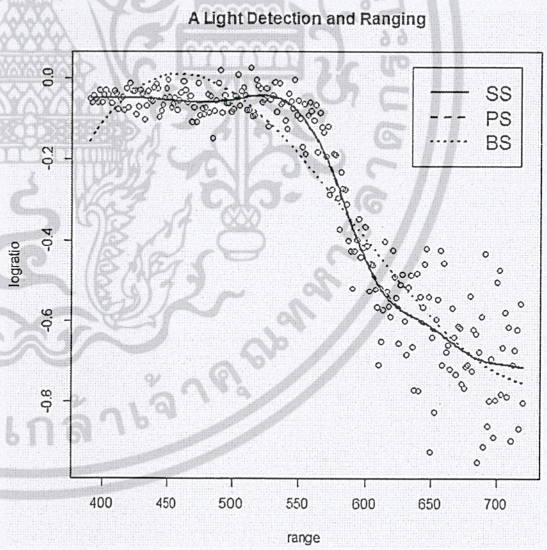



Fig. 9. The plot of LIDAR data frame and model fitting of SS, PS, and BS methods.

## 5. Application of real data

In this section, we consider the application of smoothing method based on SS, PS, and BS methods that we developed in the previous section. As the real data, we use the data frame which consists of 221 observations from a light detection and ranging (LIDAR) experiment. This data frame contains the range distance travelled before the light is reflected back to its source and logarithm of the ratio of received light from two laser sources as shown in the plot in Fig. 9.

After fitting the model, the estimating values play on a plot of light detection of ranging. It can be seen that the SS and PS interpolate in mass data more than the

BS method that followed the MSE values such as SS = 0.006016, PS = 0.006010, and BS = 0.009288. The minimum of MSE is the PS which is closed the SS as the result on Table 3.

## 6. Conclusion

In this section, we used the smoothing techniques of SS, PS, and BS methods based on nonparametric regression models. Through a Monte Carlo simulation study, we evaluated the smoothing estimator of SS, PS, and BS methods. For hypothesis testing based on the $p$-value, the fitting values supported the null value, and showed that the smoothing estimators work reasonably well for all methods, but the PS shows the minimum of average MSE.

## Acknowledgments

This work was supported by Faculty of Science Fund, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand.

## References

[1] Wahba G. Spline Models for Observational Data, SIAM: Philadelphia; 1990.
[2] Green PJ, Silverman BW. Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach, Chapman and Hall: London; 1994.
[3] Ruppert D, Wand MP, Carroll RJ. Semiparametric Regression, Cambridge University Press: New Your; 2003.
[4] Eilers PHC, Marx BD. Flexible Smoothing with B-splies and Penalties, Statistical Science 11(2) (1996), 89–102.
[5] Wahba G. A survey of some smoothing problems and the method of generalized cross-validation for solving them, In Proceeding of the Conference on the Application of Statistic, 1976, pp. 507–523.
[6] Craven P, Wahba G. Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation, Numerische Mathematik 31 (1979), 377–403.
[7] Eubank RL. Spline Smoothing and Nonparametric Regression, Marcel Dekker: New York; 1988.
[8] Eubank RL. Nonparametric Regression and Spline Smoothing, Marcel Dekker: New York; 1999.
[9] Ruppert D, Carroll RJ. Spatial-adaptive penalties for spline fitting Australian and New Zealand Journal of Statistics 42 (2000), 205–224.
[10] De Boor C. A Practical Guide to Splines, Springer: Berlin; 1978.

Songklanakarin J. Sci. Technol.
41 (4), 813-821, Jul. – Aug. 2019

http://www.sjst.psu.ac.th

*Original Article*

# Estimating conditional heteroscedastic nonlinear autoregressive model by using smoothing spline and penalized spline methods

Autcha Araveeporn*

*Department of Statistics, Faculty of Science, King Mongkut's Institute of Technology Ladkrabang, Lat Krabang, Bangkok, 10520 Thailand*

Received: 30 September 2017; Revised: 1 February 2018; Accepted: 1 April 2018

## Abstract

We propose smoothing spline (SS) and penalized spline (PS) methods in a class of nonparametric regression methods for estimating the unknown functions in a conditional heteroscedastic nonlinear autoregressive (CHNLAR) model. The CHNLAR model consists of a trend and heteroscedastic functions in terms of past data at lag 1. The SS and PS methods were tested in estimating the unknown functions used to transform data so that it fits the trend and the heteroscedastic functions. In a simulation study, time series data were generated and hypothesis testing of the bias was used to check the accuracy. The SS and PS methods exhibit a good power estimation in most cases of generated data. As real data, gold price was modeled by using SS and PS methods in the CHNLAR model. The results show that the SS method performed better than the PS method.

**Keywords:** conditional heteroscedastic nonlinear autoregressive model, smoothing spline method, penalized spline method

## 1. Introduction

Currently the economic growth is of interest to developing countries. These data are mostly stored in the form of time series data, whether it is daily, monthly, quarterly, or yearly; typical examples are unemployment rate, economic growth rate, gold price, and currency exchange rates. These indicators are sensitive with rapid fluctuations caused by external factors, such as natural disasters, wars, and epidemics, that are not controllable or predictable. Because of such perturbations, it is difficult to make accurate economic forecasts.

Heteroscedasticity or volatility means that an economic time series data displays quick changes in its time-trace. A heteroscedastic model is useful to study for estimating or forecasting time series data, and this is the right approach to take when the time series has clear evidence of changing mean and variance.

*Corresponding author
Email address: kaautcha@hotmail.com;
autcha.ar@kmitl.ac.th

There are several methods to model heteroscedastic 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, but heteroscedastically changes over time. Ghosh, Pual and Prajneshu (2010) applied zero conditional mean residual series to identify time varying volatility in a data set, by using a mixture periodic ARCH model. Bollerslev (1986) extended the model type to Generalized ARCH (GARCH), and assumed that the mean equation can be adequatedly described by an ARMA model. Pual, Ghosh and Prajneshu (2009) carried used autoregressive intergrated moving average (ARIMA) and GARCH model for modeling and forecasting. Peng and Yao (2003) showed that the conditional maximum quasilikelihood estimator suffers from complex limit distributions and slow convergence rates in an ARCH and GARCH model with heavy-tailed errors. The nonlinear autoregressive (NLAR) model developed from the nonlinear regressive model was introduced by Jones (1978). Gouri´eroux Monfort (1992); Masry and Tjøstheim (1995) have proposed conditional heteroscedastic nonlinear autoregressive (CHNLAR) models for financial time series. For simplicity, a single timestep lag in the CHNLAR model was studied to model the foreign exchange rates (Bossaerts, H¨ardle, & Hafner, 1996).

The parametric and nonparametric methods are the alternatives when estimating the regression between two sets of variables that consist of a vector of predictors and a response variable. A parametric regression model requires the user giving the form of the underlying regression function. The selection of a parametric model depends much on the problem and may be too restrictive in some applications. To overcome these difficulties, 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. A smoother is an operator that summarizes the trend of a response variable as a function of one or more predictor variables. The single predictor case is called scatterplot smoothing, and 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 (Green & Silverman, 1994; Wahba, 1990), and penalized splines (Ruppert, Wand, & Carroll, 2003). These smoothing techniques are generally based on the assumption of homoscedastic variance, which may not be suitable when the data involve high volatility.

For these various reasons, we are interested in extending the NLAR model to a CHNLAR model, for approximating heteroscedastic values by adjusting the past value. The smoothing spline and penalized spline methods are applied to estimate the trend and the heteroscedastic values in both simulated 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 and is called a conditional heteroscedastic nonlinear autoregressive (CHNLAR) model. It can be written as

$$y_t = \mu\left(y_{t-1},\ldots,y_{t-p}\right)+\sigma\left(y_{t-1},\ldots,y_{t-p}\right)\varepsilon_t,$$

and $\sigma(\cdot)$ is a function called the nonparametric autoregressive (NAR) model or the nonlinear autoregressive (NLAR) model.

In this current study, we employ the first-order conditional heteroscedastic nonlinear autoregressive (CHN LAR) model

$$y_t - \mu\left(y_{t-1}\right)+\sigma\left(y_{t-1}\right)\varepsilon_t,\ \ t-2,3,\ldots,n, \tag{1}$$

where $y_t, t=2,3,\ldots,n$ are observed and depend on $y_{t-1}, t=2,3,\ldots,n$ with lag 1, $\mu\left(y_{t-1}\right)$ is the trend function of CHNLAR model, $\sigma\left(y_{t-1}\right)$ is the heteroscedastic function of CHNLAR model, and $\varepsilon_t$, $t=2,3,\ldots,n$, denotes a random variable in the error term, with mean zero and variance one.

## 3. Nonparametric Regression Method

The popular nonparametric regression methods include smoothing splines and penalized splines. The concept of these methods is to interpolate the data in most suitable form of the fitting function with a smoothing parameter.

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

The NLAR model is written as

$$y_t = \mu\left(y_{t-1}\right)+\varepsilon_t,\ t=2,3,\ldots,n, \tag{2}$$

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

### 3.1 Smoothing spline (SS) method

The smoothing spline was studied by Wahba (1990) and the smoothing spline is a natural polynomial spline $\left(S_\lambda^{(K)}(\mu)\right)$ that depends on the smoothing parameter $(\lambda)$:

$$S_\lambda^{(K)}(\mu)=\sum_{t=1}^{n}\left\{y_t-\mu(x_t)\right\}^2+\lambda\int_a^b\left\{\mu^{(m)}(x_t)\right\}^2 dx_t, \tag{3}$$

where $K$ is the number of knots in the trend function with domain $[a,b]$, superscript $(m)$ indicates the $m$ th derivative of $\mu(x_t)$, $y_t$ is the dependent variable, and $\mu(x_t)$ is the trend function that is a nonparametric regression function of independent variables.

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

$$S_\lambda^{(K)}(\mu)=\sum_{t=1}^{n}\left\{y_t-\mu(x_t)\right\}^2+\lambda\int_a^b\left\{\mu''(x_t)\right\}^2 dx_t. \tag{4}$$

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

$$S_\lambda^{(K)}(\mu)=\sum_{t=2}^{n}\left\{y_t-\mu(y_{t-1})\right\}^2+\lambda\int_a^b\left\{\mu''(y_{t-1})\right\}^2 dy_t. \tag{5}$$

The natural cubic spline has given value and second derivative at each knot $y_t$ namely these are

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

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

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

The natural cubic spline then depends on the two matrices $\boldsymbol{Q}$ and $\boldsymbol{R}$ with

$$\boldsymbol{Q}=\begin{pmatrix} h_1^{-1} & 0 & \cdots & 0 \\ -h_1^{-1}-h_2^{-1} & h_2^{-1} & \cdots & 0 \\ h_2^{-1} & -h_2^{-1}-h_3^{-1} & \cdots & 0 \\ 0 & h_3^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & h_{n-1}^{-1} \end{pmatrix}_{(n-1)\times(n-3)},$$

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

$\boldsymbol{R}$ is a symmetric $(n-3)\times(n-3)$ matrix

$$\boldsymbol{R}=\begin{pmatrix} \frac{1}{3}(h_2+h_3) & \frac{1}{6}h_3 & \cdots & 0 \\ \frac{1}{6}h_3 & \frac{1}{3}(h_3+h_4) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{3}(h_{n-2}+h_{n-1}) \end{pmatrix}_{(n-3)\times(n-3)}.$$

The matrix $\boldsymbol{K}$ is defined by

$$\boldsymbol{K}=\boldsymbol{Q}\boldsymbol{R}^{-1}\boldsymbol{Q}^{\mathrm{T}}. \tag{6}$$

The vectors $\boldsymbol{\mu}$ and $\boldsymbol{\gamma}$ specify a natural cubic spline $\mu(y_t)$ if and only if the condition

$$\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{\mu}=\boldsymbol{R}\,\boldsymbol{\gamma} \tag{7}$$

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

$$\int_a^b \left\{\mu''(y_{t-1})\right\}^2 dy_t = \boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{R}\,\boldsymbol{\gamma}=\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{K}\,\boldsymbol{\mu}. \tag{8}$$

To illustrate, in the matrix form introduced by Green & Silverman (1994)

$$RSS=\sum_{t=2}^{n}\left\{y_t-\mu(y_{t-1})\right\}^2=(\boldsymbol{y}-\boldsymbol{\mu})^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{\mu}), \tag{9}$$

where $\boldsymbol{y}=(y_2,\ldots,y_n)^{\mathrm{T}}$ with $y_t$ corresponding value to $y_{t-1}$ and

$$\boldsymbol{\mu}=\left(\mu(y_1),\ldots,\mu(y_{n-1})\right)^{\mathrm{T}}. \tag{10}$$

The roughness penalty term $\int \mu''^2$ equals $\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{\mu}$ in (8) to obtain

$$\begin{aligned} S_\lambda(\mu) &= (\boldsymbol{y}-\boldsymbol{\mu})^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{\mu})+\lambda\,\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{\mu} \\ &= \boldsymbol{\mu}^{\mathrm{T}}(\boldsymbol{I}+\lambda\boldsymbol{K})\boldsymbol{\mu}-2\boldsymbol{y}^{\mathrm{T}}\boldsymbol{\mu}+\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}, \end{aligned} \tag{11}$$

Since $\lambda\boldsymbol{K}$ is non-negative definite, the matrix $\boldsymbol{I}+\lambda\boldsymbol{K}$ is strictly positive definite. It therefore follows that (11) has a unique minimum, and the smoothing spline estimator is

$$\hat{\mu}_\lambda=(\boldsymbol{I}+\lambda\boldsymbol{K})^{-1}\boldsymbol{y}, \tag{12}$$

where $\boldsymbol{I}$ denotes 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

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

where $\alpha_j$ and $\beta_k$ denote regression coefficients in the truncated power function.

The natural cubic spline with $m=2$ is called the low-rank thin-plate spline and it tends to have very good numerical properties. The low-rank thin-plate spline representation of $\mu(.)$ is

$$\mu(x_t,\boldsymbol{\theta})=\alpha_0+\alpha_1 x_t+\sum_{k=1}^{K}\beta_k\left|x_t-\tau_k\right|^3,\ t=1,2,\ldots,n, \tag{14}$$

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

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

$$\mu(y_{t-1},\boldsymbol{\theta})=\alpha_0+\alpha_1 y_{t-1}+\sum_{k=1}^{K}\beta_k\left|y_{t-1}-\tau_k\right|^3,\ t=2,3,\ldots,n. \tag{15}$$

To avoid overfitting, we minimize

$$\sum_{t=1}^{n}\left\{y_t-\mu(y_{t-1},\boldsymbol{\theta})\right\}^2+\frac{1}{\lambda}\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{\theta}, \tag{16}$$

where $\lambda$ is the smoothing parameter and $\boldsymbol{D}$ is a known positive semi-definite penalty matrix. The thin-plate spline penalty matrix is

$$\boldsymbol{D}=\begin{bmatrix}\mathbf{0}_{2\times 2} & \mathbf{0}_{2\times K}\\ \mathbf{0}_{K\times 2} & \boldsymbol{\Omega}_K\end{bmatrix}, \tag{17}$$

where the $(l,k)$ th entry of $\boldsymbol{\Omega}$ is $|\tau_l-\tau_k|^3$ and penalizes only the coefficient of $|y_{t-1}-\tau_k|^3$.

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

$$\boldsymbol{y}=\boldsymbol{Y\alpha}+\boldsymbol{Z}_K\boldsymbol{\beta}+\boldsymbol{\varepsilon}, \tag{18}$$

where $\boldsymbol{y}=(y_2,\ldots,y_n)^{\mathrm{T}}$, $\boldsymbol{Y}$ be the matrix with the $t-1$ th row $\boldsymbol{Y}_t=(1,y_{t-1})$, $\boldsymbol{Z}_K$ is the matrix with the $t$ th row $\boldsymbol{Z}_{Kt}=\left\{|y_{t-1}-\tau_1|^3,\ldots,|y_{t-1}-\tau_K|^3\right\}$, $\boldsymbol{\alpha}=(\alpha_1,\alpha_2)^{\mathrm{T}}$, $\boldsymbol{\beta}=(\beta_1,\ldots,\beta_K)^{\mathrm{T}}$, and $\boldsymbol{\varepsilon}$ is $N(0,\sigma_\varepsilon^2 I)$. Consider the vector $\boldsymbol{\alpha}$ as fixed parameters and the vector $\boldsymbol{\beta}$ as a set of random parameters with $E(\boldsymbol{\beta})=0$ and $\mathrm{cov}(\boldsymbol{\beta})=\sigma_\beta^2$. This class of penalized spline smoothers $(\hat{\mu}(.))$ may also be expressed as

$$\hat{\boldsymbol{\mu}}=\boldsymbol{C}\left(\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C}+\lambda^3\boldsymbol{D}\right)^{-1}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{y}, \tag{19}$$

where

$$\boldsymbol{C}=\left[1 \quad y_{t-1} \quad |y_{t-1}-\tau_k|^3_{1\le k\le K}\right]_{1\le t\le n},\ \boldsymbol{D}=\begin{bmatrix}\mathbf{0}_{2\times 2} & \mathbf{0}_{2\times K}\\ \mathbf{0}_{K\times 2} & \left(\boldsymbol{\Omega}_K^{1/2}\right)^{\mathrm{T}}\boldsymbol{\Omega}_K^{1/2}\end{bmatrix}, \tag{20}$$

and $\lambda=\sigma_\beta^2/\sigma_\varepsilon^2$ is a smoothing parameter.

## 4. Proposed Trend and Heteroscedastic Estimators

The trend $\mu(y_{t-1})$ and heteroscedasticity $\sigma^2(y_{t-1})$ 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

$$y_t=\mu(y_{t-1})+\delta_t,\ t=2,3,\ldots,n, \tag{21}$$

where $\delta_t=\sigma(y_{t-1})\ \varepsilon_t$. Next, we obtain $\hat{\mu}(y_{t-1})$ from smoothing spline (SS) and penalized spline (PS) where the residuals can be estimated as

$$\hat{\delta}_t=y_t-\hat{\mu}(y_{t-1}) \tag{22}$$

$$\left(\hat{\delta}_t\right)^2=\left(\sigma(y_{t-1})\ \hat{\varepsilon}_t\right)^2. \tag{23}$$

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

$$\log\hat{\delta}_t^2=h(y_{t-1})+\log\hat{\varepsilon}_t^2, \tag{24}$$

$$\log\hat{\delta}_t^2-E\left[\log\hat{\varepsilon}_t^2\right]=h(y_{t-1})+\log\hat{\varepsilon}_t^2-E\left[\log\hat{\varepsilon}_t^2\right]. \tag{25}$$

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

$$\log\hat{\delta}_t^2+1.2704=h(y_{t-1})+\log\hat{\varepsilon}_t^2+1.2704 \tag{26}$$

$$\tilde{y}_t=h(y_{t-1})+\tilde{\varepsilon}_t \tag{27}$$

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

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

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

$$y_t=\mu(y_{t-1})+\exp\left\{\frac{\hat{h}(y_{t-1})}{2}\right\}\hat{\varepsilon}_t \tag{29}$$

$$\exp\left\{-\frac{\hat{h}(y_{t-1})}{2}\right\}y_t=\exp\left\{-\frac{\hat{h}(y_{t-1})}{2}\right\}\mu(y_{t-1})+\hat{\varepsilon}_t \tag{30}$$

$$\breve{y}_t=g(y_{t-1})+\hat{\varepsilon}_t \tag{31}$$

where

$$\breve{y}_t=\exp\left\{-\frac{\hat{h}(y_{t-1})}{2}\right\}y_t,$$

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

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

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

The simulation study to assess the performances of smoothing spline (SS) method and penalized spline (PS) was divided into two parts. The first part is to study in CHNLAR model

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

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 $y_1 \sim$ 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 $\varepsilon_t, t = 2,3,\ldots,n$ in (33) is assumed to follow the normal distribution with mean zero and variance one.

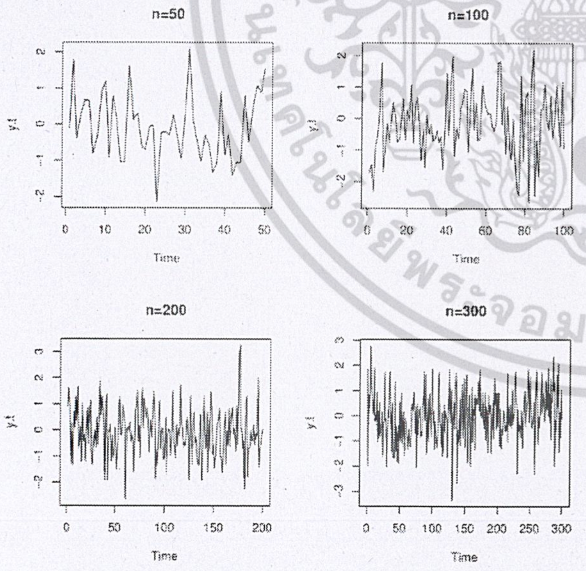



Figure 1. The time series data in CHNLAR model.

In the second part, to estimate $\hat{\mu}(y_{t-1})$ and $\hat{\sigma}(y_{t-1})$, we compute the bias and the Mean Square Error (MSE) of $\mu(.)$ and $\sigma(.)$ by

$$\mu_{bias} = \frac{1}{n}\sum_{t=2}^{n}\frac{\hat{\mu}(y_{t-1}) - \mu(y_{t-1})}{\mu(y_{t-1})},$$

$$\sigma_{bias} = \frac{1}{n}\sum_{t=2}^{n}\frac{\hat{\sigma}(y_{t-1}) - \sigma(y_{t-1})}{\sigma(y_{t-1})},$$

$$MSE(\mu) = \frac{1}{n}\sum_{t=2}^{n}\left(\hat{\mu}(y_{t-1}) - \mu(y_{t-1})\right)^2,$$

$$MSE(\sigma) = \frac{1}{n}\sum_{t=2}^{n}\left(\hat{\sigma}(y_{t-1}) - \sigma(y_{t-1})\right)^2.$$

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 presents the average MSE of SS and PS methods for all sample sizes. The average MSE of $\mu(.)$ and $\sigma(.)$ decreased with sample size. For $\mu(.)$, the average MSE of PS is less than with SS, but the average MSE of PS was larger than with SS for $\sigma(.)$.

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 means of the lower and upper bounds of 95% confidence interval are given in the 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 ≠ 0). If the p-value is less than 0.05, we reject the null hypothesis ($H_0$) that there is difference between the observed values and the fitted values. If the p-value is larger than 0.05, we conclude that we have an unbiased estimator. Based on the p-values, we can claim the following. From Tables 2 and 3, the SS and PS methods provide asymptotically unbiased estimates of $\mu(.)$ and $\sigma(.)$. However $\sigma(.)$ did not get an asymptotically unbiased estimate when the sample sizes was 200 $(n = 200)$ with the SS method.

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

Table 1. The average MSE of SS and PS methods with different sample sizes (500 replications).

| | SS method | | PS method | |
|---|---|---|---|---|
| sample size | $\mu(.)$ | $\sigma(.)$ | $\mu(.)$ | $\sigma(.)$ |
| $n = 50$ | 0.300 | 0.0251 | 0.0249 | 0.1427 |
| $n = 100$ | 0.0105 | 0.0114 | 0.0095 | 0.0622 |
| $n = 200$ | 0.0054 | 0.0061 | 0.0033 | 0.0193 |
| $n = 300$ | 0.00372 | 0.00413 | 0.0020 | 0.0100 |

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

| bias | sample size | mean | s.d. | lci | uci | t-stat | p-value |
|---|---|---|---|---|---|---|---|
| $\mu_{bias}$ | $n = 50$ | 6.507 | 157.106 | -7.296 | 20.3116 | 0.9261 | 0.3548 |
| | $n = 100$ | -1.172 | 23.367 | -3.225 | 0.881 | -1.1216 | 0.2626 |
| | $n = 200$ | -0.2977 | 10.2154 | -1.1953 | 0.5998 | -0.6517 | 0.5149 |
| | $n = 300$ | 3.5053 | 49.927 | -0.8815 | 7.8921 | 1.5699 | 0.1171 |
| $\sigma_{bias}$ | $n = 50$ | -0.0114 | 0.1577 | -0.025 | 0.0024 | -1.6176 | 0.1064 |
| | $n = 100$ | 0.0020 | 0.1072 | -0.0073 | 0.0114 | 0.4283 | 0.6686 |
| | $n = 200$ | -0.0074 | 0.0776 | -0.0142 | -0.0005 | -2.1335 | 0.0333* |
| | $n = 300$ | 0.00184 | 0.0644 | -0.0038 | 0.0075 | 0.6412 | 0.5217 |

* indicates significance at 5% level

Table 3. The simulation of penalized spline method for different sample sizes (500 replications).

| bias | sample size | mean | s.d. | lci | uci | t-stat | p-value |
|---|---|---|---|---|---|---|---|
| $\mu_{bias}$ | $n = 50$ | -4.734 | 79.833 | -11.749 | 2.279 | -1.326 | 0.1854 |
| | $n = 100$ | 1.485 | 47.561 | -2.693 | 5.664 | 0.698 | 0.485 |
| | $n = 200$ | 0.0454 | 32.9606 | -2.8506 | 2.9415 | 0.0308 | 0.9754 |
| | $n = 300$ | -0.5388 | 10.4931 | -1.4608 | 0.3831 | -1.1483 | 0.2514 |
| $\sigma_{bias}$ | $n = 50$ | 0.00429 | 0.172 | -0.010 | 0.019 | 0.556 | 0.578 |
| | $n = 100$ | -0.0001 | 0.1120 | -0.009 | 0.009 | -0.026 | 0.9792 |
| | $n = 200$ | -0.0018 | 0.0790 | -0.0087 | 0.0051 | -0.5170 | 0.6054 |
| | $n = 300$ | 0.0035 | 0.0617 | -0.0018 | 0.0090 | 1.3005 | 0.194 |

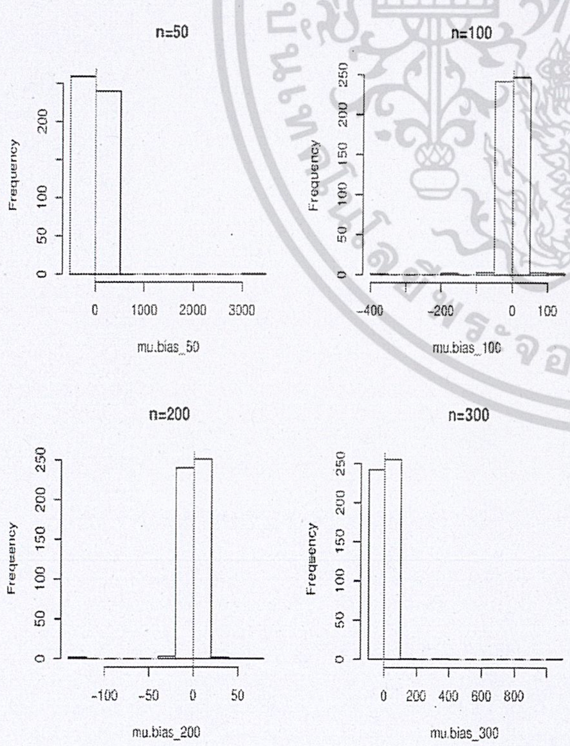



Figure 2. Histogram of bias in $\mu(.)$ with smoothing spline method.

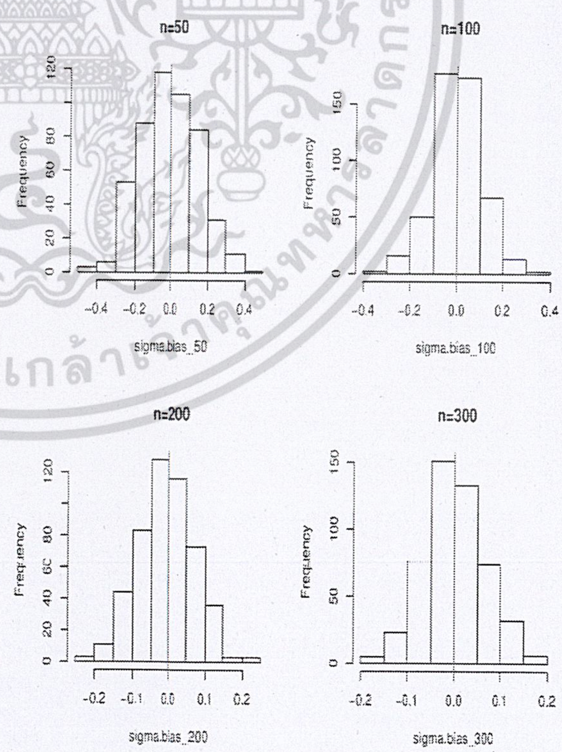



Figure 3. Histogram of bias in $\sigma(.)$ with smoothing spline method.

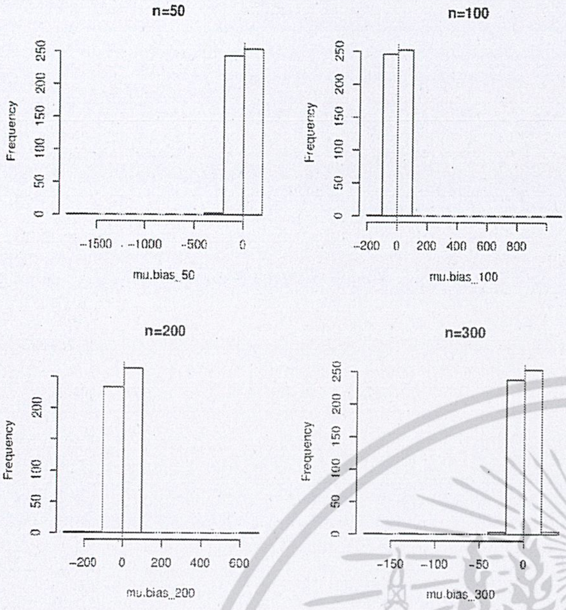



Figure 4. Histogram of bias in $\mu(.)$ with penalized spline method.

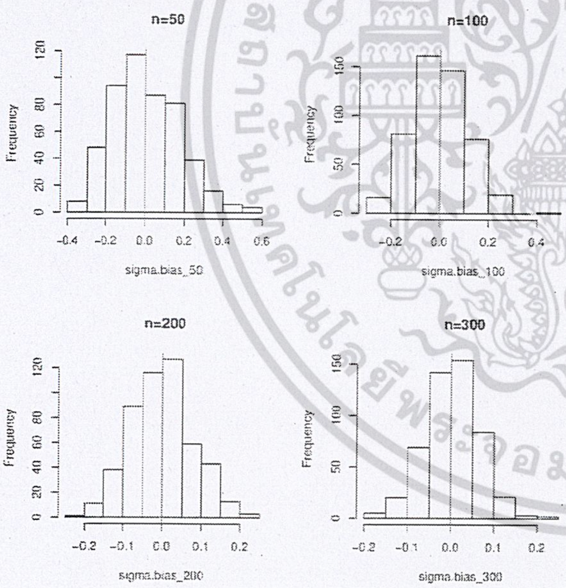



Figure 5. Histogram of bias in $\sigma(.)$ with penalized spline method.

## 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. As a real data set, we use the monthly gold price (US Dollars per Troy Ounce) from January, 1984 to December 2013, which consists of 360 records and is shown in Figure 6. These data were obtained from http://www.index mundi.com.

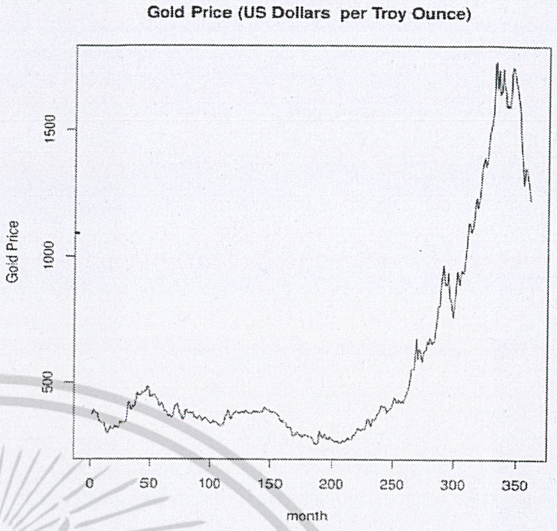



Figure 6. The monthly gold prices from January, 1984 to December, 2013.

The modeling steps were as follows.
At first, we considered the CHNLAR model following

$$y_t = \mu(y_{t-1}) + \sigma(y_{t-1})\ \varepsilon_t,\ t = 2,3,\dots,360, \tag{34}$$

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.

Then we fitted the CHNLAR model to obtain the trend function, $\mu(.)$ and the heteroscedastic function, $\sigma(.)$. We got $\hat{\mu}(y_{t-1}), \hat{\sigma}(y_{t-1}), t = 2,3,\dots,360$ 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

$$\hat{\varepsilon}_t = \frac{y_t - \hat{\mu}(y_{t-1})}{\hat{\sigma}(y_{t-1})},\ t = 2,3,\dots,360, \tag{35}$$

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 $\hat{y}_2,\dots,\hat{y}_n$ using the estimated trend and heteroscedastic based on the CHNLAR model:

$$\hat{y}_t = \hat{\mu}(y_{t-1}) + \hat{\sigma}(y_{t-1})\ \hat{\varepsilon}_t,\ t = 2,3,\dots,360, \tag{36}$$

where the forecast trend $\hat{\mu}(y_{t-1})$ and the forecast heteroscedastic $\hat{\sigma}(y_{t-1})$ from SS and PS methods are presented on Figure 7 and 8.

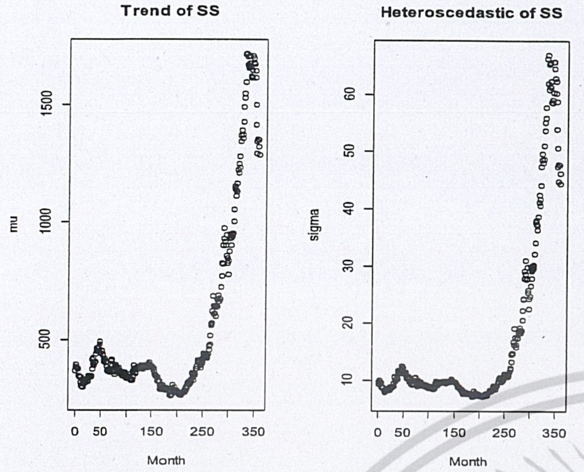



Figure 7. Forecasting trend and heterocsedasticity by smoothing spline (SS) method.

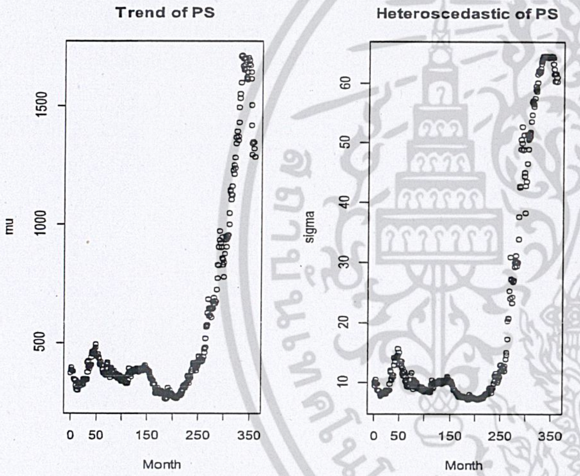



Figure 8. Forecasting trend and heterocsedasticity by penalized spline (PS) method.

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

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

| Estimator | SS method | PS method |
|---|---|---|
| $\mu(.)$ | 564.4387 | 564.4463 |
| | (396.7862) | (396.755) |
| $\sigma(.)$ | 16.92417 | 20.4015 |
| | (15.54389 | (19.66588) |
| MSE | 5.653 | 8.372 |

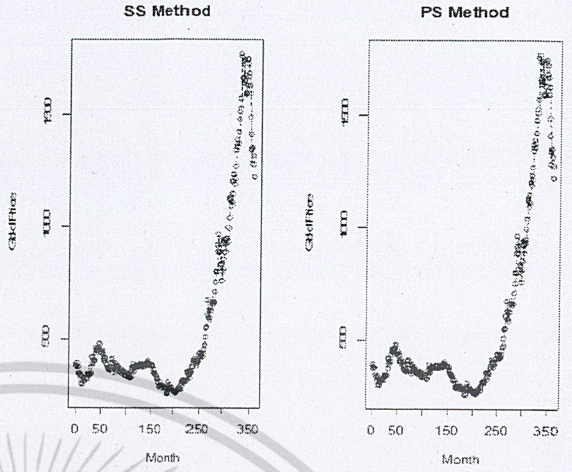



Figure 9. The gold prices with line plot of fitted values by smoothing spline (SS) and penalized spline (PS) methods.

## 7. Conclusions

In this study, we used nonparametric regression methods such as the smoothing spline method and the penalized spline method to estimate a smooth unknown trend and heteroscedasticity in an CHNLAR model. Through a Monte Carlo simulation study, we evaluated the performance of the smoothing spline methods and showed that the trend estimator ($\mu(.)$) and heteroscedasticity estimator ($\sigma(.)$)work reasonably well for most data of all sample sizes, except in one case ($n = 200$) where the heteroscedasticity estimator had a bias. The point volatility estimators approach their corresponding true values as the sample size increases.

In a Monte Carlo study, we showed that the trend estimator of penalized spline works well for all small sample sizes, when the smoothing parameter is high enough, indicating that small sample sizes allow reliable interpolation by these models.

In an application to actual data, we were also interested in comparing the power of estimating values, assessed by considering the Mean Square Error (MSE). The MSEs with smoothing spline were smaller than with penalized spline. However, we consider the mean of the trend and heteroscedastic estimator, and we can see that the means of trend with smoothing spline method were slightly different from the penalized spline method but the variance and MSE of smoothing spline method is smaller than with penalized spline method. It can be concluded that the forecasting performance of CHNLAR depends on heteroscedasticity.

In future work, we intend to study higher order lags in trend and heteroscedasticity of the CHNALR model.

## Acknowledgements

This work was financially supported by the grant research fund of the Faculty of Science, King Mongkut's Institute of Technology Ladkrabang.

## References


Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, *31*, 307–327.

Bossaerts, P., H¨ardle, W., & Hafner, C. (1996). Foreign exchange rates have surprising volatility. In P. M. Robinson (Ed.), *Proceeding in Athens Conference on Applied Probability and Time Series*, *Lecture Notes in Statistics 115* (pp. 55-72), New York, NY: Springer.

Brumback, B., Ruppert, D., & Wand, M. P. (1999). Comment on variable selection and function estimation in additive nonparametric regression using data-based prior by shively. *Journal of American Statistics Association*, *94*, 794–797.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, *50*, 987–1007.

Fan, J., & Yao, Q. (2003). *Nonlinear time series nonparametric and parametric methods*. New York, NY: Springer.

Ghosh, H., Pual, R. K., & Prajneshu. (2010). Nonlinear time series modeling and forecasting periodic and ARCH effects. *Journal of the Indian Society of Agricultural Statistics*, *60*(1), 85-92.

Gouri´eroux, C., & Monfort, A. (1992). Qualitative threshold ARCH models. *Journal of Econometrics*, *52*, 159–199.

Green, P. J., & Silverman, B. W. (1994). *Nonparametric regression and generalized linear models: A roughness penalty approach*. London, England: Chapman and Hall.

Jones, D. (1978). Nonlinear regressive processes. *Proceedings of Royal Society London, Series A*, *360*, 71–95.

Marsy, E., & Tjøstheim, D. (1995). Non-parametric estimation and identification of ARCH nonlinear time series: Strong convergence and asymptotic normality. *Econometric Theory*, *11*, 258–289.

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

Pual, R. K., Ghosh, H., & Prajneshu. (2009). GARCH nonlinear time series analysis for modeling and forecasting of India's volatile spices export data. *Journal of the Indian Society of Agricultural Statistics*, *63*(2), 123-131.

Ruppert, D., & Carroll, R. J. (2000). Spatial-adaptive penalties for spline fitting. *Australian and New Zealand Journal of Statistics*, *42*, 205–224.

Ruppert, D., Wand, M. P., & Carroll. R. J. (2003). *Semiparametric regression*. Cambridge, England: Cambridge University Press.

Wahba, G. (1990). *Spline models for observational data*. Philadelphia, PA: SIAM.

Songklanakarin J. Sci. Technol.
41 (4), 813-821, Jul. – Aug. 2019

http://www.sjst.psu.ac.th

*Original Article*

# Estimating conditional heteroscedastic nonlinear autoregressive model by using smoothing spline and penalized spline methods

Autcha Araveeporn*

*Department of Statistics, Faculty of Science, King Mongkut's Institute of Technology Ladkrabang, Lat Krabang, Bangkok, 10520 Thailand*

Received: 30 September 2017; Revised: 1 February 2018; Accepted: 1 April 2018

**Abstract**

We propose smoothing spline (SS) and penalized spline (PS) methods in a class of nonparametric regression methods for estimating the unknown functions in a conditional heteroscedastic nonlinear autoregressive (CHNLAR) model. The CHNLAR model consists of a trend and heteroscedastic functions in terms of past data at lag 1. The SS and PS methods were tested in estimating the unknown functions used to transform data so that it fits the trend and the heteroscedastic functions. In a simulation study, time series data were generated and hypothesis testing of the bias was used to check the accuracy. The SS and PS methods exhibit a good power estimation in most cases of generated data. As real data, gold price was modeled by using SS and PS methods in the CHNLAR model. The results show that the SS method performed better than the PS method.

**Keywords:** conditional heteroscedastic nonlinear autoregressive model, smoothing spline method, penalized spline method

## 1. Introduction

Currently the economic growth is of interest to developing countries. These data are mostly stored in the form of time series data, whether it is daily, monthly, quarterly, or yearly; typical examples are unemployment rate, economic growth rate, gold price, and currency exchange rates. These indicators are sensitive with rapid fluctuations caused by external factors, such as natural disasters, wars, and epidemics, that are not controllable or predictable. Because of such perturbations, it is difficult to make accurate economic forecasts.

Heteroscedasticity or volatility means that an economic time series data displays quick changes in its time-trace. A heteroscedastic model is useful to study for estimating or forecasting time series data, and this is the right approach to take when the time series has clear evidence of changing mean and variance.

*Corresponding author
Email address: kaautcha@hotmail.com; autcha.ar@kmitl.ac.th

There are several methods to model heteroscedastic 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, but heteroscedastically changes over time. Ghosh, Pual and Prajneshu (2010) applied zero conditional mean residual series to identify time varying volatility in a data set, by using a mixture periodic ARCH model. Bollerslev (1986) extended the model type to Generalized ARCH (GARCH), and assumed that the mean equation can be adequatedly described by an ARMA model. Pual, Ghosh and Prajneshu (2009) carried used autoregressive intergrated moving average (ARIMA) and GARCH model for modeling and forecasting. Peng and Yao (2003) showed that the conditional maximum quasilikelihood estimator suffers from complex limit distributions and slow convergence rates in an ARCH and GARCH model with heavy-tailed errors. The nonlinear autoregressive (NLAR) model developed from the nonlinear regressive model was introduced by Jones (1978). Gouri´eroux Monfort (1992); Masry and Tjøstheim (1995) have proposed conditional heteroscedastic nonlinear autoregressive (CHNLAR) models for financial time series. For simplicity, a single timestep lag in the CHNLAR model was studied to model the foreign exchange rates (Bossaerts, H¨ardle, & Hafner, 1996).

The parametric and nonparametric methods are the alternatives when estimating the regression between two sets of variables that consist of a vector of predictors and a response variable. A parametric regression model requires the user giving the form of the underlying regression function. The selection of a parametric model depends much on the problem and may be too restrictive in some applications. To overcome these difficulties, 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. A smoother is an operator that summarizes the trend of a response variable as a function of one or more predictor variables. The single predictor case is called scatterplot smoothing, and 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 (Green & Silverman, 1994; Wahba, 1990), and penalized splines (Ruppert, Wand, & Carroll, 2003). These smoothing techniques are generally based on the assumption of homoscedastic variance, which may not be suitable when the data involve high volatility.

For these various reasons, we are interested in extending the NLAR model to a CHNLAR model, for approximating heteroscedastic values by adjusting the past value. The smoothing spline and penalized spline methods are applied to estimate the trend and the heteroscedastic values in both simulated 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 and is called a conditional heteroscedastic nonlinear autoregressive (CHNLAR) model. It can be written as

$$y_t = \mu\left(y_{t-1},\dots,y_{t-p}\right) + \sigma\left(y_{t-1},\dots,y_{t-p}\right)\varepsilon_t,$$

and $\sigma(\bullet)$ is a function called the nonparametric autoregressive (NAR) model or the nonlinear autoregressive (NLAR) model.

In this current study, we employ the first-order conditional heteroscedastic nonlinear autoregressive (CHN LAR) model

$$y_t = \mu\left(y_{t-1}\right) + \sigma\left(y_{t-1}\right)\varepsilon_t,\ \ t = 2,3,\dots,n, \qquad (1)$$

where $y_t, t = 2,3,\dots,n$ are observed and depend on $y_{t-1}, t = 2,3,\dots,n$ with lag 1, $\mu\left(y_{t-1}\right)$ is the trend function of CHNLAR model, $\sigma\left(y_{t-1}\right)$ is the heteroscedastic function of CHNLAR model, and $\varepsilon_t$, $t = 2,3,\dots,n$, denotes a random variable in the error term, with mean zero and variance one.

## 3. Nonparametric Regression Method

The popular nonparametric regression methods include smoothing splines and penalized splines. The concept of these methods is to interpolate the data in most suitable form of the fitting function with a smoothing parameter.

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

The NLAR model is written as

$$y_t = \mu\left(y_{t-1}\right) + \varepsilon_t,\ t = 2,3,\dots,n, \qquad (2)$$

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

### 3.1 Smoothing spline (SS) method

The smoothing spline was studied by Wahba (1990) and the smoothing spline is a natural polynomial spline $\left(S_\lambda^{(K)}(\mu)\right)$ that depends on the smoothing parameter $(\lambda)$ :

$$S_\lambda^{(K)}(\mu) = \sum_{t=1}^{n}\left\{y_t - \mu(x_t)\right\}^2 + \lambda\int_a^b\left\{\mu^{(m)}(x_t)\right\}^2 dx_t, \quad (3)$$

where $K$ is the number of knots in the trend function with domain $[a,b]$, superscript $(m)$ indicates the $m$ th derivative of $\mu(x_t)$, $y_t$ is the dependent variable, and $\mu(x_t)$ is the trend function that is a nonparametric regression function of independent variables.

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

$$S_\lambda^{(K)}(\mu) = \sum_{t=1}^{n}\left\{y_t - \mu(x_t)\right\}^2 + \lambda\int_a^b\left\{\mu''(x_t)\right\}^2 dx_t. \quad (4)$$

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

$$S_\lambda^{(K)}(\mu) = \sum_{t=2}^{n}\left\{y_t - \mu(y_{t-1})\right\}^2 + \lambda\int_a^b\left\{\mu''(y_{t-1})\right\}^2 dy_t. \quad (5)$$

The natural cubic spline has given value and second derivative at each knot $y_t$ namely these are

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

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

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

The natural cubic spline then depends on the two matrices $\boldsymbol{Q}$ and $\boldsymbol{R}$ with

$$\boldsymbol{Q}=\begin{pmatrix} h_1^{-1} & 0 & \cdots & 0 \\ -h_1^{-1}-h_2^{-1} & h_2^{-1} & \cdots & 0 \\ h_2^{-1} & -h_2^{-1}-h_3^{-1} & \cdots & 0 \\ 0 & h_3^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & h_{n-1}^{-1} \end{pmatrix}_{(n-1)\times(n-3)},$$

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

$\boldsymbol{R}$ is a symmetric $(n-3)\times(n-3)$ matrix

$$\boldsymbol{R}=\begin{pmatrix} \frac{1}{3}(h_2+h_3) & \frac{1}{6}h_3 & \cdots & 0 \\ \frac{1}{6}h_3 & \frac{1}{3}(h_3+h_4) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{3}(h_{n-2}+h_{n-1}) \end{pmatrix}_{(n-3)\times(n-3)}.$$

The matrix $\boldsymbol{K}$ is defined by

$$\boldsymbol{K}=\boldsymbol{Q}\boldsymbol{R}^{-1}\boldsymbol{Q}^{\mathrm{T}}. \tag{6}$$

The vectors $\boldsymbol{\mu}$ and $\boldsymbol{\gamma}$ specify a natural cubic spline $\mu(y_t)$ if and only if the condition

$$\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{\mu}=\boldsymbol{R}\,\boldsymbol{\gamma} \tag{7}$$

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

$$\int_a^b \left\{\mu''(y_{t-1})\right\}^2 dy_t = \boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{R}\,\boldsymbol{\gamma}=\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{K}\,\boldsymbol{\mu}. \tag{8}$$

To illustrate, in the matrix form introduced by Green & Silverman (1994)

$$RSS=\sum_{t=2}^{n}\left\{y_t-\mu(y_{t-1})\right\}^2=(\boldsymbol{y}-\boldsymbol{\mu})^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{\mu}), \tag{9}$$

where $\boldsymbol{y}=\left(y_2,\ldots,y_n\right)^{\mathrm{T}}$ with $y_t$ corresponding value to $y_{t-1}$ and

$$\boldsymbol{\mu}=\left(\mu(y_1),\ldots,\mu(y_{n-1})\right)^{\mathrm{T}}. \tag{10}$$

The roughness penalty term $\int \mu''^2$ equals $\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{\mu}$ in (8) to obtain

$$\begin{aligned} S_\lambda(\mu) &= (\boldsymbol{y}-\boldsymbol{\mu})^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{\mu})+\lambda\,\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{\mu} \\ &= \boldsymbol{\mu}^{\mathrm{T}}(\boldsymbol{I}+\lambda\boldsymbol{K})\boldsymbol{\mu}-2\boldsymbol{y}^{\mathrm{T}}\boldsymbol{\mu}+\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}, \end{aligned} \tag{11}$$

Since $\lambda\boldsymbol{K}$ is non-negative definite, the matrix $\boldsymbol{I}+\lambda\boldsymbol{K}$ is strictly positive definite. It therefore follows that (11) has a unique minimum, and the smoothing spline estimator is

$$\hat{\boldsymbol{\mu}}_\lambda=(\boldsymbol{I}+\lambda\boldsymbol{K})^{-1}\boldsymbol{y}, \tag{12}$$

where $\boldsymbol{I}$ denotes 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

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

where $\alpha_j$ and $\beta_k$ denote regression coefficients in the truncated power function.

The natural cubic spline with $m=2$ is called the low-rank thin-plate spline and it tends to have very good numerical properties. The low-rank thin-plate spline representation of $\mu(.)$ is

$$\mu(x_t,\boldsymbol{\theta})=\alpha_0+\alpha_1 x_t+\sum_{k=1}^{K}\beta_k\left|x_t-\tau_k\right|^3,\ t=1,2,\ldots,n. \tag{14}$$

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

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

$$\mu(y_{t-1},\boldsymbol{\theta})=\alpha_0+\alpha_1 y_{t-1}+\sum_{k=1}^{K}\beta_k\left|y_{t-1}-\tau_k\right|^3,\ t=2,3,\ldots,n. \tag{15}$$

To avoid overfitting, we minimize

$$\sum_{t=1}^{n}\left\{y_t-\mu(y_{t-1},\boldsymbol{\theta})\right\}^2+\frac{1}{\lambda}\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{\theta}, \tag{16}$$

where $\lambda$ is the smoothing parameter and $\boldsymbol{D}$ is a known positive semi-definite penalty matrix. The thin-plate spline penalty matrix is

$$\boldsymbol{D}=\begin{bmatrix}\boldsymbol{0}_{2\times 2} & \boldsymbol{0}_{2\times K}\\ \boldsymbol{0}_{K\times 2} & \boldsymbol{\Omega}_K\end{bmatrix}, \qquad (17)$$

where the $(l,k)$ th entry of $\boldsymbol{\Omega}$ is $|\tau_l-\tau_k|^3$ and penalizes only the coefficient of $|y_{t-1}-\tau_k|^3$.

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

$$\boldsymbol{y}=\boldsymbol{Y\alpha}+\boldsymbol{Z}_K\boldsymbol{\beta}+\boldsymbol{\varepsilon}, \qquad (18)$$

where $\boldsymbol{y}=(y_2,\ldots,y_n)^{\mathrm{T}}$, $Y$ be the matrix with the $t-1$ th row $\boldsymbol{Y}_t=(1,y_{t-1})$, $\boldsymbol{Z}_K$ is the matrix with the $t$ th row $\boldsymbol{Z}_{Kt}=\{|y_{t-1}-\tau_1|^3,\ldots,|y_{t-1}-\tau_K|^3\}.\boldsymbol{\alpha}=(\alpha_1,\alpha_2)^{\mathrm{T}}.\boldsymbol{\beta}=(\beta_1,\ldots,\beta_K)^{\mathrm{T}}$, and $\boldsymbol{\varepsilon}$ is $N(0,\sigma_\varepsilon^2 I)$. Consider the vector $\boldsymbol{\alpha}$ as fixed parameters and the vector $\boldsymbol{\beta}$ as a set of random parameters with $E(\boldsymbol{\beta})=0$ and $\mathrm{cov}(\boldsymbol{\beta})=\sigma_\beta^2$. This class of penalized spline smoothers $(\hat{\mu}(.))$ may also be expressed as

$$\hat{\boldsymbol{\mu}}=\boldsymbol{C}\left(\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C}+\lambda^3\boldsymbol{D}\right)^{-1}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{y}, \qquad (19)$$

where

$$\boldsymbol{C}=\begin{bmatrix}1 & y_{t-1} & |y_{t-1}-\tau_k|^3_{1\le k\le K}\end{bmatrix}_{1\le t\le n}, \boldsymbol{D}=\begin{bmatrix}\boldsymbol{0}_{2\times 2} & \boldsymbol{0}_{2\times K}\\ \boldsymbol{0}_{K\times 2} & (\boldsymbol{\Omega}_K^{1/2})^{\mathrm{T}}\boldsymbol{\Omega}_K^{1/2}\end{bmatrix}, \qquad (20)$$

and $\lambda=\sigma_\beta^2/\sigma_\varepsilon^2$ is a smoothing parameter.

## 4. Proposed Trend and Heteroscedastic Estimators

The trend $\mu(y_{t-1})$ and heteroscedasticity $\sigma^2(y_{t-1})$ 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

$$y_t=\mu(y_{t-1})+\delta_t,\ t=2,3,\ldots,n, \qquad (21)$$

where $\delta_t=\sigma(y_{t-1})\ \varepsilon_t$. Next, we obtain $\hat{\mu}(y_{t-1})$ from smoothing spline (SS) and penalized spline (PS) where the residuals can be estimated as

$$\hat{\delta}_t=y_t-\hat{\mu}(y_{t-1}) \qquad (22)$$

$$(\hat{\delta}_t)^2=(\sigma(y_{t-1})\ \hat{\varepsilon}_t)^2. \qquad (23)$$

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

$$\log\hat{\delta}_t^2=h(y_{t-1})+\log\hat{\varepsilon}_t^2, \qquad (24)$$

$$\log\hat{\delta}_t^2-E\left[\log\hat{\varepsilon}_t^2\right]=h(y_{t-1})+\log\hat{\varepsilon}_t^2-E\left[\log\hat{\varepsilon}_t^2\right]. \qquad (25)$$

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

$$\log\hat{\delta}_t^2+1.2704=h(y_{t-1})+\log\hat{\varepsilon}_t^2+1.2704 \qquad (26)$$

$$\tilde{y}_t=h(y_{t-1})+\tilde{\varepsilon}_t \qquad (27)$$

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

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

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

$$y_t=\mu(y_{t-1})+\exp\left\{\frac{\hat{h}(y_{t-1})}{2}\right\}\hat{\varepsilon}_t \qquad (29)$$

$$\exp\left\{-\frac{\hat{h}(y_{t-1})}{2}\right\}y_t=\exp\left\{-\frac{\hat{h}(y_{t-1})}{2}\right\}\mu(y_{t-1})+\hat{\varepsilon}_t \qquad (30)$$

$$\breve{y}_t=g(y_{t-1})+\hat{\varepsilon}_t \qquad (31)$$

where

$$\breve{y}_t=\exp\left\{-\frac{\hat{h}(y_{t-1})}{2}\right\}y_t,$$

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

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

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

The simulation study to assess the performances of smoothing spline (SS) method and penalized spline (PS) was divided into two parts. The first part is to study in CHNLAR model

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

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 $y_1 \sim$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 $\varepsilon_t, t = 2,3,\ldots,n$ in (33) is assumed to follow the normal distribution with mean zero and variance one.

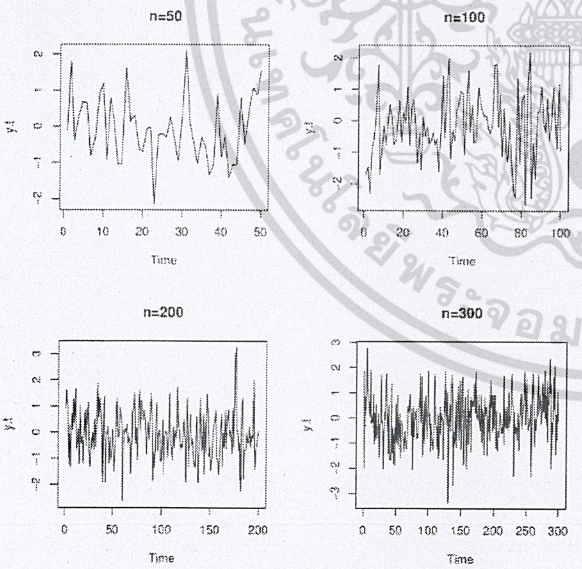



Figure 1. The time series data in CHNLAR model.

In the second part, to estimate $\hat{\mu}(y_{t-1})$ and $\hat{\sigma}(y_{t-1})$, we compute the bias and the Mean Square Error (MSE) of $\mu(.)$ and $\sigma(.)$ by

$$\mu_{bias} = \frac{1}{n}\sum_{t=2}^{n}\frac{\hat{\mu}(y_{t-1}) - \mu(y_{t-1})}{\mu(y_{t-1})},$$

$$\sigma_{bias} = \frac{1}{n}\sum_{t=2}^{n}\frac{\hat{\sigma}(y_{t-1}) - \sigma(y_{t-1})}{\sigma(y_{t-1})},$$

$$MSE(\mu) = \frac{1}{n}\sum_{t=2}^{n}\left(\hat{\mu}(y_{t-1}) - \mu(y_{t-1})\right)^2,$$

$$MSE(\sigma) = \frac{1}{n}\sum_{t=2}^{n}\left(\hat{\sigma}(y_{t-1}) - \sigma(y_{t-1})\right)^2.$$

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 presents the average MSE of SS and PS methods for all sample sizes. The average MSE of $\mu(.)$ and $\sigma(.)$ decreased with sample size. For $\mu(.)$, the average MSE of PS is less than with SS, but the average MSE of PS was larger than with SS for $\sigma(.)$.

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 means of the lower and upper bounds of 95% confidence interval are given in the 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 is difference between the observed values and the fitted values. If the p-value is larger than 0.05, we conclude that we have an unbiased estimator. Based on the p-values, we can claim the following. From Tables 2 and 3, the SS and PS methods provide asymptotically unbiased estimates of $\mu(.)$ and $\sigma(.)$. However $\sigma(.)$ did not get an asymptotically unbiased estimate when the sample sizes was 200 $(n = 200)$ with the SS method.

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

Table 1. The average MSE of SS and PS methods with different sample sizes (500 replications).

| | SS method | | PS method | |
|---|---|---|---|---|
| sample size | $\mu(.)$ | $\sigma(.)$ | $\mu(.)$ | $\sigma(.)$ |
| $n = 50$ | 0.300 | 0.0251 | 0.0249 | 0.1427 |
| $n = 100$ | 0.0105 | 0.0114 | 0.0095 | 0.0622 |
| $n = 200$ | 0.0054 | 0.0061 | 0.0033 | 0.0193 |
| $n = 300$ | 0.00372 | 0.00413 | 0.0020 | 0.0100 |

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

| bias | sample size | mean | s.d. | lci | uci | t-stat | p-value |
|---|---|---|---|---|---|---|---|
| $\mu_{bias}$ | $n = 50$ | 6.507 | 157.106 | -7.296 | 20.3116 | 0.9261 | 0.3548 |
| | $n = 100$ | -1.172 | 23.367 | -3.225 | 0.881 | -1.1216 | 0.2626 |
| | $n = 200$ | -0.2977 | 10.2154 | -1.1953 | 0.5998 | -0.6517 | 0.5149 |
| | $n = 300$ | 3.5053 | 49.927 | -0.8815 | 7.8921 | 1.5699 | 0.1171 |
| $\sigma_{bias}$ | $n = 50$ | -0.0114 | 0.1577 | -0.025 | 0.0024 | -1.6176 | 0.1064 |
| | $n = 100$ | 0.0020 | 0.1072 | -0.0073 | 0.0114 | 0.4283 | 0.6686 |
| | $n = 200$ | -0.0074 | 0.0776 | -0.0142 | -0.0005 | -2.1335 | 0.0333* |
| | $n = 300$ | 0.00184 | 0.0644 | -0.0038 | 0.0075 | 0.6412 | 0.5217 |

* indicates significance at 5% level

Table 3. The simulation of penalized spline method for different sample sizes (500 replications).

| bias | sample size | mean | s.d. | lci | uci | t-stat | p-value |
|---|---|---|---|---|---|---|---|
| $\mu_{bias}$ | $n = 50$ | -4.734 | 79.833 | -11.749 | 2.279 | -1.326 | 0.1854 |
| | $n = 100$ | 1.485 | 47.561 | -2.693 | 5.664 | 0.698 | 0.485 |
| | $n = 200$ | 0.0454 | 32.9606 | -2.8506 | 2.9415 | 0.0308 | 0.9754 |
| | $n = 300$ | -0.5388 | 10.4931 | -1.4608 | 0.3831 | -1.1483 | 0.2514 |
| $\sigma_{bias}$ | $n = 50$ | 0.00429 | 0.172 | -0.010 | 0.019 | 0.556 | 0.578 |
| | $n = 100$ | -0.0001 | 0.1120 | -0.009 | 0.009 | -0.026 | 0.9792 |
| | $n = 200$ | -0.0018 | 0.0790 | -0.0087 | 0.0051 | -0.5170 | 0.6054 |
| | $n = 300$ | 0.0035 | 0.0617 | -0.0018 | 0.0090 | 1.3005 | 0.194 |

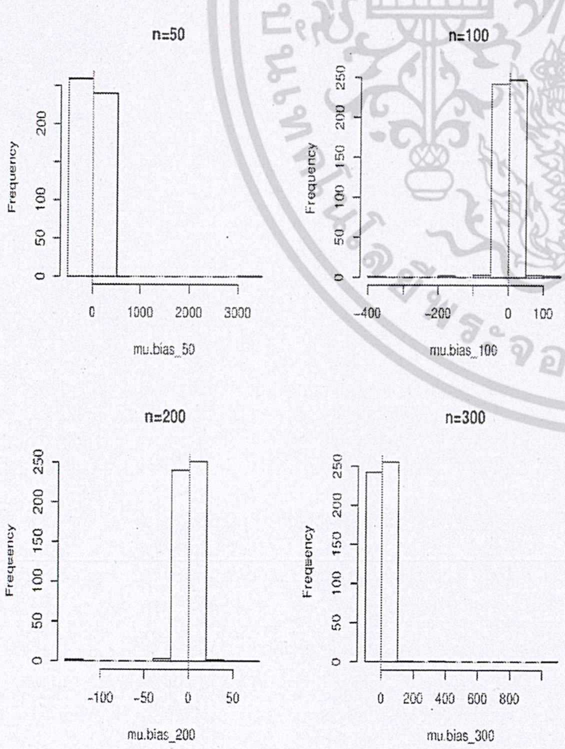



Figure 2. Histogram of bias in $\mu(.)$ with smoothing spline method.

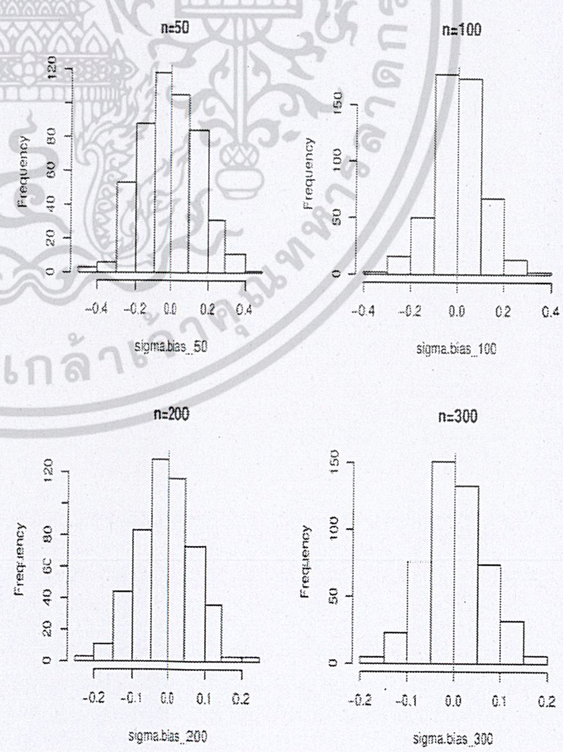



Figure 3. Histogram of bias in $\sigma(.)$ with smoothing spline method.

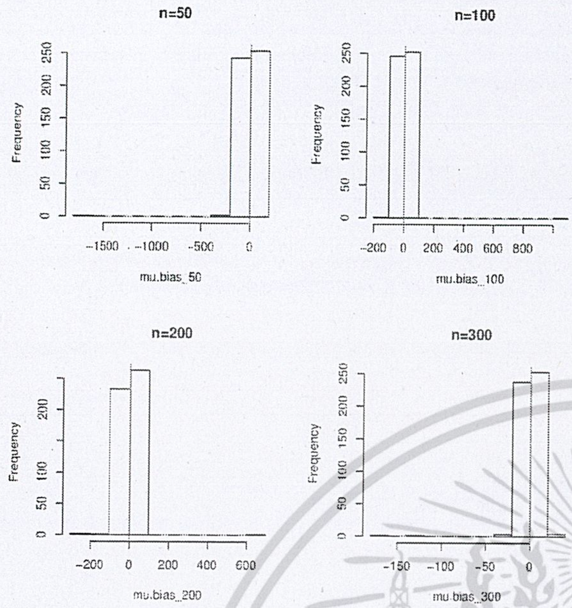



Figure 4. Histogram of bias in $\mu(.)$ with penalized spline method.

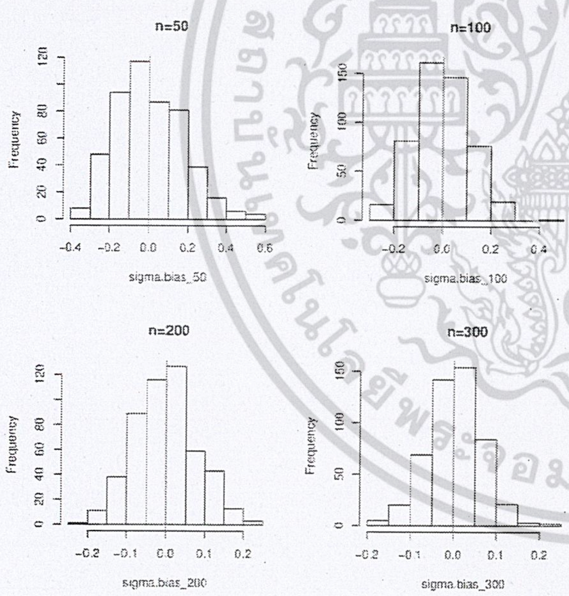



Figure 5. Histogram of bias in $\sigma(.)$ with penalized spline method.

## 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. As a real data set, we use the monthly gold price (US Dollars per Troy Ounce) from January, 1984 to December 2013, which consists of 360 records and is shown in Figure 6. These data were obtained from http://www.index mundi.com.

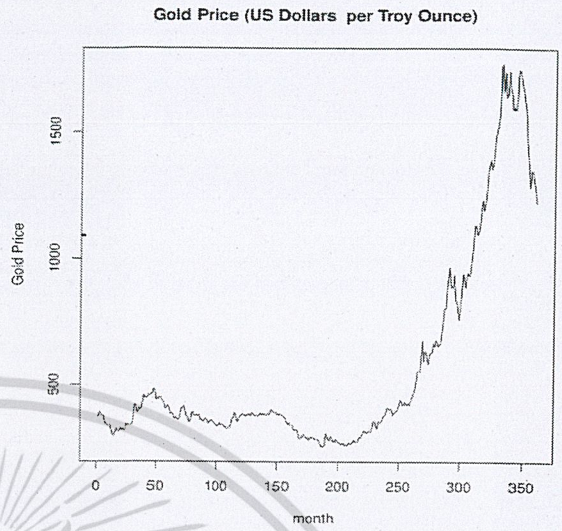



Figure 6. The monthly gold prices from January, 1984 to December, 2013.

The modeling steps were as follows.
At first, we considered the CHNLAR model following

$$y_t = \mu(y_{t-1}) + \sigma(y_{t-1})\ \varepsilon_t,\ t = 2,3,\ldots,360, \tag{34}$$

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.

Then we fitted the CHNLAR model to obtain the trend function, $\mu(.)$ and the heteroscedastic function, $\sigma(.)$. We got $\hat{\mu}(y_{t-1}), \hat{\sigma}(y_{t-1}), t = 2,3,\ldots,360$ 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

$$\hat{\varepsilon}_t = \frac{y_t - \hat{\mu}(y_{t-1})}{\hat{\sigma}(y_{t-1})},\ t = 2,3,\ldots,360, \tag{35}$$

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 $\hat{y}_2,\ldots,\hat{y}_n$ using the estimated trend and heteroscedastic based on the CHNLAR model:

$$\hat{y}_t = \hat{\mu}(y_{t-1}) + \hat{\sigma}(y_{t-1})\ \hat{\varepsilon}_t,\ t = 2,3,\ldots,360, \tag{36}$$

where the forecast trend $\hat{\mu}(y_{t-1})$ and the forecast heteroscedastic $\hat{\sigma}(y_{t-1})$ from SS and PS methods are presented on Figure 7 and 8.

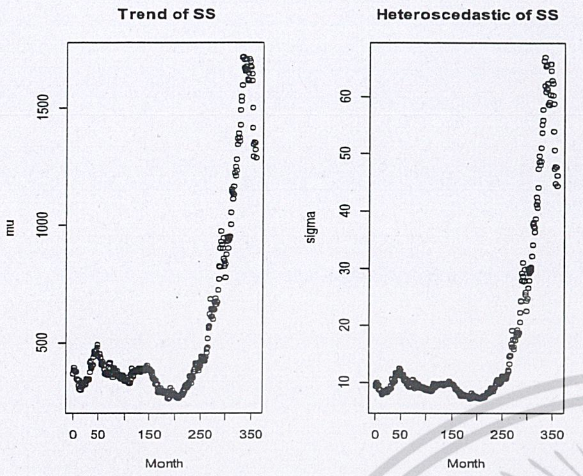



Figure 7. Forecasting trend and heterocsedasticity by smoothing spline (SS) method.

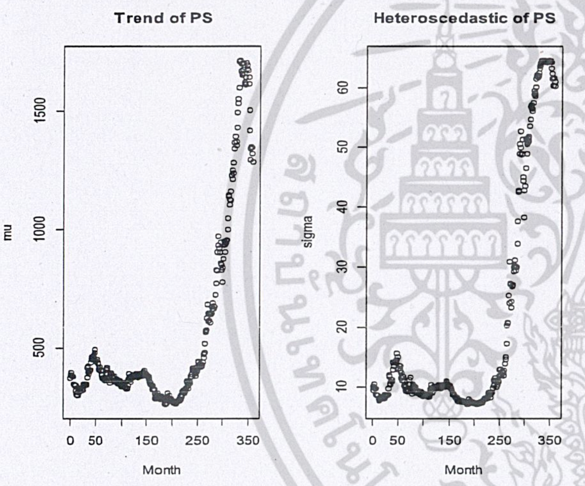



Figure 8. Forecasting trend and heterocsedasticity by penalized spline (PS) method.

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

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

| Estimator | SS method | PS method |
|---|---|---|
| $\mu(.)$ | 564.4387<br>(396.7862) | 564.4463<br>(396.755) |
| $\sigma(.)$ | 16.92417<br>(15.54389 | 20.4015<br>(19.66588) |
| MSE | 5.653 | 8.372 |

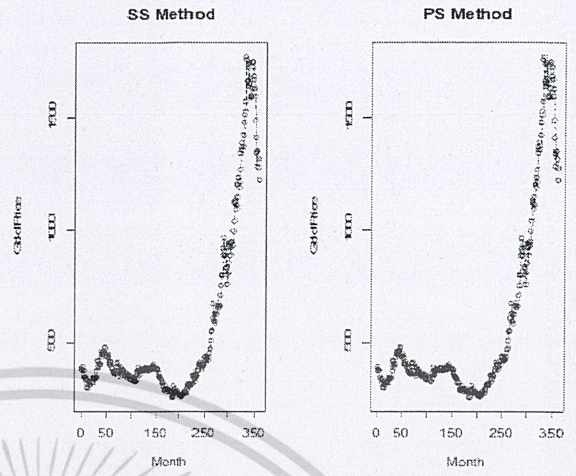



Figure 9. The gold prices with line plot of fitted values by smoothing spline (SS) and penalized spline (PS) methods.

## 7. Conclusions

In this study, we used nonparametric regression methods such as the smoothing spline method and the penalized spline method to estimate a smooth unknown trend and heteroscedasticity in an CHNLAR model. Through a Monte Carlo simulation study, we evaluated the performance of the smoothing spline methods and showed that the trend estimator ( $\mu(.)$ ) and heteroscedasticity estimator ( $\sigma(.)$ )work reasonably well for most data of all sample sizes, except in one case ( $n = 200$ ) where the heteroscedasticity estimator had a bias. The point volatility estimators approach their corresponding true values as the sample size increases.

In a Monte Carlo study, we showed that the trend estimator of penalized spline works well for all small sample sizes, when the smoothing parameter is high enough, indicating that small sample sizes allow reliable interpolation by these models.

In an application to actual data, we were also interested in comparing the power of estimating values, assessed by considering the Mean Square Error (MSE). The MSEs with smoothing spline were smaller than with penalized spline. However, we consider the mean of the trend and heteroscedastic estimator, and we can see that the means of trend with smoothing spline method were slightly different from the penalized spline method but the variance and MSE of smoothing spline method is smaller than with penalized spline method. It can be concluded that the forecasting performance of CHNLAR depends on heteroscedasticity.

In future work, we intend to study higher order lags in trend and heteroscedasticity of the CHNALR model.

## Acknowledgements

This work was financially supported by the grant research fund of the Faculty of Science, King Mongkut's Institute of Technology Ladkrabang.

## References


Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, *31*, 307–327.

Bossaerts, P., H¨ardle, W., & Hafner, C. (1996). Foreign exchange rates have surprising volatility. In P. M. Robinson (Ed.), *Proceeding in Athens Conference on Applied Probability and Time Series*, *Lecture Notes in Statistics 115* (pp. 55-72), New York, NY: Springer.

Brumback, B., Ruppert, D., & Wand, M. P. (1999). Comment on variable selection and function estimation in additive nonparametric regression using data-based prior by shively. *Journal of American Statistics Association*, *94*, 794–797.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, *50*, 987–1007.

Fan, J., & Yao, Q. (2003). *Nonlinear time series nonparametric and parametric methods*. New York, NY: Springer.

Ghosh, H., Pual, R. K., & Prajneshu. (2010). Nonlinear time series modeling and forecasting periodic and ARCH effects. *Journal of the Indian Society of Agricultural Statistics*, *60*(1), 85-92.

Gouri´eroux, C., & Monfort, A. (1992). Qualitative threshold ARCH models. *Journal of Econometrics*, *52*, 159–199.

Green, P. J., & Silverman, B. W. (1994). *Nonparametric regression and generalized linear models: A roughness penalty approach*. London, England: Chapman and Hall.

Jones, D. (1978). Nonlinear regressive processes. *Proceedings of Royal Society London, Series A*, *360*, 71–95.

Marsy, E., & Tjøstheim, D. (1995). Non-parametric estimation and identification of ARCH nonlinear time series: Strong convergence and asymptotic normality. *Econometric Theory*, *11*, 258–289.

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

Pual, R. K., Ghosh, H., & Prajneshu. (2009). GARCH nonlinear time series analysis for modeling and forecasting of India's volatile spices export data. *Journal of the Indian Society of Agricultural Statistics*, *63*(2), 123-131.

Ruppert, D., & Carroll, R. J. (2000). Spatial-adaptive penalties for spline fitting. *Australian and New Zealand Journal of Statistics*, *42*, 205–224.

Ruppert, D., Wand, M. P., & Carroll. R. J. (2003). *Semiparametric regression*. Cambridge, England: Cambridge University Press.

Wahba, G. (1990). *Spline models for observational data*. Philadelphia, PA: SIAM.