รายงานการวิจัยฉบับสมบูรณ์

การเปรียบเทียบการประมาณค่าสำหรับการถดถอยแบบพารามิเตอร์ แบบไม่ใช้พารามิเตอร์ และแบบผสม

The Comparison for estimation of Parametric Regression, Non-Parametric Regression, and Semiparametric Regression

นางสาวอัชฌา อระวีพร

ได้รับทุนสนับสนุนงานวิจัยจากเงินรายได้ ประจำปีงบประมาณ 2557

คณะวิทยาศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

รายงานการวิจัยฉบับสมบูรณ์

การเปรียบเทียบการประมาณค่าสำหรับการถดถอยแบบพารามิเตอร์ แบบไม่ใช้พารามิเตอร์ และแบบผสม

The Comparison for estimation of Parametric Regression, Non-Parametric Regression, and Semiparametric Regression

นางสาวอัชฌา อระวีพร

ได้รับทุนสนับสนุนงานวิจัยจากเงินรายได้ ประจำปีงบประมาณ 2557

คณะวิทยาศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

Research Title : The Comparison of estimation of Parametric Regression model, Nonparametric Regression model, and Semiparametric Regression model in Case of Two Explanatory Variables

Research : Autcha Araveeporn

Department : Statistics

Faculty : Science

ABSTRACT

In this paper, we proposed the parametric regression model, nonparametric regression model, and semiparametric regression model to estimate parameter in these models. The penalized spline method based on nonparametric regression method is used in term of nonparametric regression model, and semiparametric regression model. The minimum of Mean Square Error (MSE) is a criterion for choosing optimality model.

Here, we simulated the response variable and two explanatory variables that correlated a non-linear data based on uniform distribution. The real data can be applied for these models by setting the response variable as the gold price (US Dollars per Troy Ounce) and two explanatory variables as the crude oil price (Dollars per Barrel) and the number of month. The estimated values of nonparametric regression model is a good performance in simulated data and real data.

Keywords: nonparametric regression model, parametric regression model, penalized spline method, semiparametric regression model

CONTENTS

1. Introduction

Regression Analysis or parametric regression analysis are a commonly used method for the investigation of relationship between variables, and obtaining the coefficient of regression function. The simplest of regression function consists of a response variable (y) and a single explanatory variable (x). The multiple regression function is extended from the simple regression function to include additional explanatory variables. To obtain useful the regression analysis, the assumption is investigated before data analysis such as; the variables x and y describes a linear relationship, the variables x and y are normal distribution, the observations of variables are independent, and two explanatory variables (x) occur correlated variation called multicollinearity.

Nonparametric regression method relaxes the assumption of regression analysis. The gain of nonparametric regression method is to estimate the smoothing function which is produces a smoother directly, rather than to estimate regression coefficients. The smoother is a tool for summarizing the trend of a response variable as a function of one or more explanatory variables. The simple nonparametric regression model is often called scatterplot smoothing which is used one explanatory variable. The nonparametric regression method presents more than one explanatory variables called the nonparametric additive model.

There are many approaches to estimate nonparametric regression model, e.g., a local polynomial regression method ([1] and [2]), regression splines method ([3] and [4]), smoothing splines method ([5] and [6]), and penelized splines method ([7]). Moreover, the nonparametric regression model has been made to develop for time series data which may have a nonlinear relationship. Robinson [8] suggested the nonparametric estimation in the context of time series data.

The method of nonparametric regression has a long history in the smoothing method. Wahba [5] defined the natural polynomial spline model that used to measure the roughness of curve. Green and Silverman [6] emphasized the simple case of the natural polynomial spline so-called the natural cubic spline. The smoothing spline method is considered a least square criteria to fit the natural cubic spline. Stone [9] examined the consistency properties of nonparametric regression estimators in local polynomial regression. Fan [10], [11] demonstrated the desirable mean square error properties, as well as establishing the local linear regression based on kernel regression. In the local polynomial regression method, the local neighborhoods are specified by a bandwidth but Eubank [3], [4] introduced the regression spline that the local neighborhoods are specified by a group of locations. Penalized spline method has developed from regression spline and smoothing spline, which is used of fitting and flexible choice of knots and smoothing parameter in nonparametric regression model. Ruppert, Wand, and Carroll [7] described penalized spline method based on reduced-knot truncated power function basis with penalties on the untransformed coefficients, fitted as a mixed model, and motivated as a simple low-rank smoothing spline.

Current paper studies the estimation of parametric regression model, nonparametric regression model, and semiparametric regression model from a response variable and two explanatory variables. The parametric regression model is used the least square method to estimate a coefficient of regression function. The nonparametric regression is concerned the penalized spline method to estimate the parameter of smoothing function. The semiparametric regression is a mixed model between parametric regression model and nonparametric regression model based on penalized spline method.

This paper is organized as follows: Section 2 describes the parameter estimation of parametric regression model, nonparametric regression model, and semiparametric regression model. Section 3 shows the process and the results of simulated data and discuss the results. We apply our proposed models to real data in Section 4. Finally, in Section 5, we present the conclusions.

2. Parameter Estimation

2.1 Parametric Regression Model

The parametric regression model consists of a response variable and a single explanatory variable. In this case, we study two explanatory variables denoted by the $(x_{1t}, x_{2t}), \quad t = 1, 2, ..., n$, the regression parameters $\underset{\sim}{\theta} = (\beta_0, \beta_1, \beta_2)^T$, the regression estimators $\hat{\underset{\sim}{\theta}} = (b_0, b_1, b_2)^T$, and the errors $\underset{\sim}{\varepsilon} = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_n)^T$. The parametric regression model can be written as

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon_t, \quad t = 1, 2, ..., n. \tag{1}$$

The assumptions of the parametric regression model in (1) are contained error variables then the mean $E(\varepsilon_t) = 0$, and variance $V(\varepsilon_t) = \sigma^2$. By the error term, ε_t and ε_k are not correlated. The sample regression model is obtained as follows

$$\hat{y}_t = b_0 + b_1 x_{1t} + b_2 x_{2t}, \quad t = 1, 2, ..., n.. \tag{2}$$

Estimator $\hat{\underset{\sim}{\theta}} = (b_0, b_1, b_2)^T$ is estimated by the Least Square Error (LSE) that specified to minimize the Sum of Square Error (SSE) as

$$SSE = \sum_{t=1}^{n} (y_t - b_0 - b_1 x_{1t} - b_2 x_{2t})^2 . \tag{3}$$

It is convenient to use matrices to approximate the regression estimators by solving the normal equation for $\hat{\underset{\sim}{\theta}}$:

$$X^T X \hat{\underset{\sim}{\theta}} = X^T \underset{\sim}{y}, \tag{4}$$

where $\underset{\sim}{y}$ is an $n \times 1$ vector, X is an $n \times 3$ matrix, and $\hat{\underset{\sim}{\theta}}$ is a 3×1 vector, given by

$$\underset{\sim}{y} = \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix}, \qquad X = \begin{bmatrix} 1 & x_{11} & x_{21} \\ 1 & x_{12} & x_{22} \\ \cdots & \cdots & \cdots \\ 1 & x_{1n} & x_{2n} \end{bmatrix}, \qquad \hat{\underset{\sim}{\theta}} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix}.$$

The coefficient of parameter with parametric regression model is

$$\hat{\underset{\sim}{\theta}} = \left(X^T X\right)^{-1} X^T \underset{\sim}{y}. \tag{5}$$

2.2 Nonparametric Regression Model

The nonparametric regression method base on a smoothing technique which is produced a smoothing function as a smoother . A smoother is a tool for summarizing the trend of a response variable as a function of one or more explanatory variables.

We mention a penalized spline method for data analysis in the class of nonparametric regression model. Penalized spline method is quite similar to smoothing spline method especially more flexible choice of the spline model, the basis function, and the penalty function.

The simple nonparametric regression model consists of single explanatory variable and response variable, but in this case we focus two explanatory variables called nonparametric additive model which is written as

$$y_t = f(x_{1t}) + f(x_{2t}) + \varepsilon_t, \qquad t = 1,\ldots,n, \tag{6}$$

where y_t is a response variable, $f(x_{1t})$ is a smooth function of x_1, $f(x_{2t})$ is a smooth function of x_2, and ε_t is a error term.

Penalized spline smoother is estimated using truncated power ([12]), and the penalized spline regression model is rewritten as

$$y_t = \mu(x_{1t}, x_{2t}, \underset{\sim}{\theta}) + \varepsilon_t, \; t = 1,\ldots,n, \tag{7}$$

where

$$\mu(x_{1t}, x_{2t}, \underset{\sim}{\theta}) = \sum_{j=1}^{m-1} \alpha_j^{x_1} x_{1t}^j + \sum_{k=1}^{K_{x1}} \beta_k \mid x_{1t} - \tau_k \mid^{2m-1} + \sum_{j=1}^{m-1} \delta_j^{x_2} x_{2t}^j + \sum_{k=1}^{K_{x2}} \gamma_k \mid x_{2t} - \upsilon_k \mid^{2m-1}, t = 1, 2, ..., n, \quad (8)$$

with $\underset{\sim}{\beta} \equiv \left[\beta_1, ..., \beta_{K_{x1}}\right]^T \sim N\left(0, \sigma_\beta^2\, \Omega_{x1}^{-1/2} (\Omega_{x1}^{-1/2})^T\right), \quad \Omega_{x1} \equiv \left[\left|x_{1t} - \tau_k\right|_{1 \le k \le K_{x1}}^{2m-1}\right]_{1 \le t \le n}$,

and $\underset{\sim}{\gamma} \equiv \left[\gamma_1, ..., \gamma_{K_{x2}}\right]^T \sim N\left(0, \sigma_\gamma^2\, \Omega_{x2}^{-1/2} (\Omega_{x2}^{-1/2})^T\right), \quad \Omega_{x2} \equiv \left[\left|x_{2t} - \upsilon_k\right|_{1 \le k \le K_{x2}}^{2m-1}\right]_{1 \le t \le n}$.

The penalized spline regression model are specified a group of locations in a range of interval $[a,b]$, where $a < \tau_1 < .... < \tau_{K_{x1}} < b$ and $a < \upsilon_1 < .... < \upsilon_{K_{x2}} < b$ introduced by Eubank [3], [4] . These locations are known as knots, and $\tau_k, k = 1, 2, ..., K_{x1}$ and $\upsilon_k, k = 1, 2, ..., K_{x2}$ are called interior knots. We will focus on the low-rank thin-plate splines (m=2) which tend to fit with the non-linear data. The low-rank thin-plate spline is presented by

$$\mu(x_{1t}, x_{2t}, \underset{\sim}{\theta}) = \sum_{j=1}^{2-1} \alpha_j x_1^j + \sum_{k=1}^{K_{x1}} \beta_k \mid x_{1t} - \tau_k \mid^3 + \sum_{j=1}^{2-1} \delta_j x_2^j + \sum_{k=1}^{K_{x2}} \gamma_k \mid x_{2t} - \upsilon_k \mid^3, \quad (9)$$

where $\underset{\sim}{\theta} = (\alpha_1, \beta_1, ..., \beta_{K_{x1}}, \delta_1, \gamma_1, ..., \gamma_{K_{x2}})^T$ is the vector of penalized spline regression model, and $\tau_1 < .... < \tau_{K_{x1}}$ and $\upsilon_1 < .... < \upsilon_{K_{x2}}$ are fixed knots. Following Ruppert [13], we consider a number of knots that is large enough to ensure the desired flexibility, and (τ_k, υ_k) are the sample quartile of x_1 and x_2 corresponding to probability $k / (K_{x1} + 1)$ and $k / (K_{x2} + 1)$. To avoid overfitting, we minimize

$$\sum_{t=1}^{n} \{y_t - \mu(x_{1t}, x_{2t}, \underset{\sim}{\theta})\}^2 + \underset{\sim}{\theta}^T D \underset{\sim}{\theta}, \quad (10)$$

$$D=\begin{bmatrix} 0_{2\times 2} & 0_{2\times(K_{x1}+K_{x2})} \\ 0_{(K_{x1}+K_{x2})\times 2} & \Omega_{(K_{x1}+K_{x2})\times(K_{x1}+K_{x2})} \end{bmatrix}, \text{ and } \Omega=\begin{bmatrix} \lambda_{x1}^{2m-1}\left|x_{1t}-\tau_k\right|^3_{1\le k\le K_{x1}} & \lambda_{x2}^{2m-1}\left|x_{2t}-\upsilon_k\right|^3_{1\le k\le K_{x2}} \end{bmatrix}_{1\le t\le n},$$

where $\lambda_{x1}\equiv\dfrac{\sigma_\beta^2}{\sigma_\varepsilon^2}$, $\lambda_{x2}\equiv\dfrac{\sigma_\gamma^2}{\sigma_\varepsilon^2}$ are the smoothing parameter , D is known positive semi-definite penalty matrix. Smoothing parameter can be approximated by restricted maximum likelihood and approximated best linear unbiased prediction ([14]).

Just as with the linear model, we can generalize penalized spline in general linear mixed model ([15]) is

$$\underset{\sim}{y}=X\underset{\sim}{\alpha}+Z_{Kt}\underset{\sim}{\beta}+\underset{\sim}{\varepsilon}, \tag{11}$$

where $\underset{\sim}{y}$ is an $n\times 1$ vector, X is an $n\times 2$ matrix, $\underset{\sim}{\alpha}$ is a 2×1 vector, $\underset{\sim}{\beta}$ is a $(K_{x1}+K_{x2})\times 1$ vector, Z_{Kt} is an $n\times(K_{x1}+K_{x2})$ matrix, $\underset{\sim}{\varepsilon}$ is a $n\times 1$ vector, given by

$$\underset{\sim}{y}=\begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix}, \quad X=\begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ \cdots & \cdots \\ x_{n1} & x_{n2} \end{bmatrix}, \quad \underset{\sim}{\alpha}=\begin{bmatrix} \alpha_1 \\ \delta_1 \end{bmatrix}, \underset{\sim}{\beta}=\begin{bmatrix} \beta_1 \\ \vdots \\ \beta_{K_{x1}} \\ \gamma_1 \\ \vdots \\ \gamma_{K_{x2}} \end{bmatrix},$$

$$Z_{Kt}=\begin{bmatrix} \left|x_{11}-\tau_1\right|^3 & \cdots & \left|x_{11}-\tau_{K_{x1}}\right|^3 & \left|x_{21}-\upsilon_1\right|^3 & \cdots & \left|x_{21}-\upsilon_{K_{x2}}\right|^3 \\ \left|x_{12}-\tau_1\right|^3 & \cdots & \left|x_{12}-\tau_{K_{x1}}\right|^3 & \left|x_{22}-\upsilon_1\right|^3 & \cdots & \left|x_{22}-\upsilon_{K_{x2}}\right|^3 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \left|x_{1n}-\tau_1\right|^3 & \cdots & \left|x_{1n}-\tau_{K_{x1}}\right|^3 & \left|x_{2n}-\upsilon_1\right|^3 & \cdots & \left|x_{2n}-\upsilon_{K_{x2}}\right|^3 \end{bmatrix} \text{ and } \underset{\sim}{\varepsilon}=\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdots \\ \varepsilon_n \end{bmatrix}.$$

This class of penalized spline smoothers $\hat{\mu}(x_1, x_2, \hat{\underset{\sim}{\theta}})$ may also be expressed as

$$\hat{\mu}(x_1, x_2, \hat{\underset{\sim}{\theta}}) = C(C^T C + D)^{-1} C^T \underset{\sim}{y} \ , \tag{12}$$

where

$$C = \begin{bmatrix} x_{1t} & x_{2t} & \left|x_{1t} - \tau_k\right|^3_{1\le k\le K_{x1}} & \left|x_{2t} - \upsilon_k\right|^3_{1\le k\le K_{x2}} \end{bmatrix}_{1\le t\le n} ,$$

$$, \ D = \begin{bmatrix} 0_{2\times 2} & 0_{2\times(K_{x1}+K_{x2})} \\ 0_{(K_{x1}+K_{x2})\times 2} & \Omega_{(K_{x1}+K_{x2})\times(K_{x1}+K_{x2})} \end{bmatrix},$$

and $\Omega = \begin{bmatrix} \lambda_{x1}^3 \left|x_{1t} - \tau_k\right|^3_{1\le k\le K_{x1}} & \lambda_{x2}^3 \left|x_{2t} - \upsilon_k\right|^3_{1\le k\le K_{x2}} \end{bmatrix}_{1\le t\le n}$.

The penalized spline smoothers $\hat{\underset{\sim}{\theta}}$ are computed by

$$\hat{\underset{\sim}{\theta}} = (C^T C + D)^{-1} C^T \underset{\sim}{y} . \tag{13}$$

2.3 <u>Semiparametric Regression Model</u>

The semiparametric regression is a mixed model between parametric regression model and nonparametric regression model. The mixed model of penalized spline method allows the two explanatory variables and a response variable on the penalized spline regression model which is written as

$$y_t = \beta_0 + \beta_1 x_{1t} + f(x_{2t}) + \varepsilon_t , \quad t = 1,\ldots,n, \tag{14}$$

where y_t is a response variable, (β_0 , β_1) are coefficient of parametric regression model, $f(x_{2t})$ is a smooth function of x_2 , and ε_t is a error term.

The penalized spline regression model through the mixed model is

$$y_t = \mu(x_{1t}, x_{2t}, \underset{\sim}{\theta}) + \varepsilon_t \text{, t = 1,…,n,} \tag{15}$$

where

$$\mu(x_{1t}, x_{2t}, \underset{\sim}{\theta}) = \beta_0 + \beta_1 x_{1t} + \sum_{j=1}^{m-1} \delta_j^{x_2} x_{2t}^j + \sum_{k=1}^{K_{x2}} \gamma_k \mid x_{2t} - \upsilon_k \mid^{2m-1}, t = 1,\ldots,n, \tag{16}$$

with $\underset{\sim}{\gamma} \equiv \left[\gamma_1,\ldots,\gamma_{K_{x2}}\right]^T \sim N\left(0, \sigma_\gamma^2\, \Omega_{x2}^{-1/2} (\Omega_{x2}^{-1/2})^T\right)$, $\Omega_{x2} \equiv \left[\left|x_{2t} - \upsilon_k\right|_{1\le k\le K_{x2}}^{2m-1}\right]_{1\le t\le n}$.

A group of locations in a range of interval $[a,b]$ is $a < \upsilon_1 < \ldots. < \upsilon_{K_{x2}} < b$, and the interior knots are $\upsilon_k, k = 1,2,\ldots,K_{x2}$. The low-rank thin-plate spline (m=2) is written by

$$\mu(x_{1t}, x_{2t}, \underset{\sim}{\theta}) = \beta_0 + \beta_1 x_{1t} + \sum_{j=1}^{2-1} \delta_j x_2^j + \sum_{k=1}^{K_{x2}} \gamma_k \mid x_{2t} - \upsilon_k \mid^3, \tag{17}$$

where $\underset{\sim}{\theta} = (\beta_0, \beta_1, \delta_1, \gamma_1,\ldots,\gamma_{K_{x2}})^T$ is the vector of penalized spline regression model, and $\upsilon_1 < \ldots. < \upsilon_{K_{x2}}$ are fixed knots. We minimize

$$\sum_{t=1}^{n} \left\{y_t - \mu(x_{1t}, x_{2t}, \underset{\sim}{\theta})\right\}^2 + \underset{\sim}{\theta}^T D \underset{\sim}{\theta}, \tag{18}$$

$$D = \begin{bmatrix} 0_{3\times3} & 0_{3\times K_{x2}} \\ 0_{K_{x2}\times3} & \Omega_{K_{x2}\times K_{x2}} \end{bmatrix}, \text{ and } \Omega = \left[\lambda_{x2}^{2m-1} \left|x_{2t} - \upsilon_k\right|_{1\le k\le K_{x2}}^3\right]_{1\le t\le n},$$

where λ_{x2} is the smoothing parameter computed by $\lambda_{x2} \equiv \dfrac{\sigma_\gamma^2}{\sigma_\varepsilon^2}$ and D is known positive semi-definite penalty matrix.

We can generalize penalized spline in the matrix form as

$$\underset{\sim}{y} = X\underset{\sim}{\alpha} + Z_{Kt}\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}, \tag{19}$$

where $\underset{\sim}{y}$ is an $n\times 1$ vector, X is an $n\times 3$ matrix, $\underset{\sim}{\alpha}$ is a 3×1 vector, $\underset{\sim}{\beta}$ is a $K_{x2}\times 1$ vector, Z_{Kt} is an $n\times K_{x2}$ matrix, $\underset{\sim}{\varepsilon}$ is a $n\times 1$ vector, given by

$$\underset{\sim}{y} = \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix}, \qquad X = \begin{bmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ \cdots & \cdots & \cdots \\ 1 & x_{n1} & x_{n2} \end{bmatrix}, \qquad \underset{\sim}{\alpha} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \delta_1 \end{bmatrix}, \ \underset{\sim}{\beta} = \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_{K_{x2}} \end{bmatrix}.$$

$$Z_{Kt} = \begin{bmatrix} |x_{21}-\upsilon_1|^3 & \cdots & |x_{21}-\upsilon_{K_{x2}}|^3 \\ |x_{22}-\upsilon_1|^3 & \cdots & |x_{22}-\upsilon_{K_{x2}}|^3 \\ \cdots & \cdots & \cdots \\ |x_{2n}-\upsilon_1|^3 & \cdots & |x_{2n}-\upsilon_{K_{x2}}|^3 \end{bmatrix}, \text{ and } \underset{\sim}{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdots \\ \varepsilon_n \end{bmatrix}.$$

The fitted values of penalized spline smoothers $\hat{\mu}(x_1, x_2, \hat{\theta})$ may also be expressed as

$$\hat{\mu}(x_{1t}, x_{2t}, \underset{\sim}{\hat{\theta}}) = C(C^T C + D)^{-1} C^T \underset{\sim}{y}, \tag{20}$$

where

$$C = \begin{bmatrix} 1 & x_{1t} & x_{2t} & |x_{2t}-\upsilon_k|^3_{1\le k\le K_{x2}} \end{bmatrix}_{1\le t\le n}, \quad D = \begin{bmatrix} 0_{3\times 3} & 0_{3\times K_{x2}} \\ 0_{K_{x2}\times 3} & \Omega_{K_{x2}\times K_{x2}} \end{bmatrix},$$

and $\Omega = \begin{bmatrix} \lambda_{x2}^3 |x_{2t}-\upsilon_k|^3_{1\le k\le K_{x2}} \end{bmatrix}_{1\le t\le n}$.

The coefficient of parameter with semiparametric regression model is

$$\underset{\sim}{\hat{\theta}} = (C^T C + D)^{-1} C^T \underset{\sim}{y}. \tag{21}$$

3. Simulation Study

In this section, we display the process and the results of a simulation experiment that we conducted in order to compare the performance of parametric regression model, nonparametric regression model, and semiparametric regression model. To simulate data, we generated data in the class of a response variable (y) and two explanatory variables (x). The response variable is obtained from

$$y_t = f(x_{1t}) + x_{2t} + \varepsilon_t,\quad t = 1,\ldots,n \tag{22}$$

where $f(x_{1t}) = x_{2t}^3 - (x_{2t}^2\cos(x_{2t})) - \exp\left\{\frac{x_{2t}}{1+|x_{2t}|}\right\}$, $x_{2t} \sim Uniform(-r,r)$, and $\varepsilon_t \sim Normal(0,1)$. The uniform distribution of x_{2t} tried several values for r = 1,3,5, and 7. The simulated data consist of 25, 50, 100, and 300 sample sizes, and repeated for model fitting 500 times . Figure 1 illustrates an example of simulated data with 50 sample sizes (n=50) in graph of scatter plots that show the relationship between response variable and two explanatory variables.

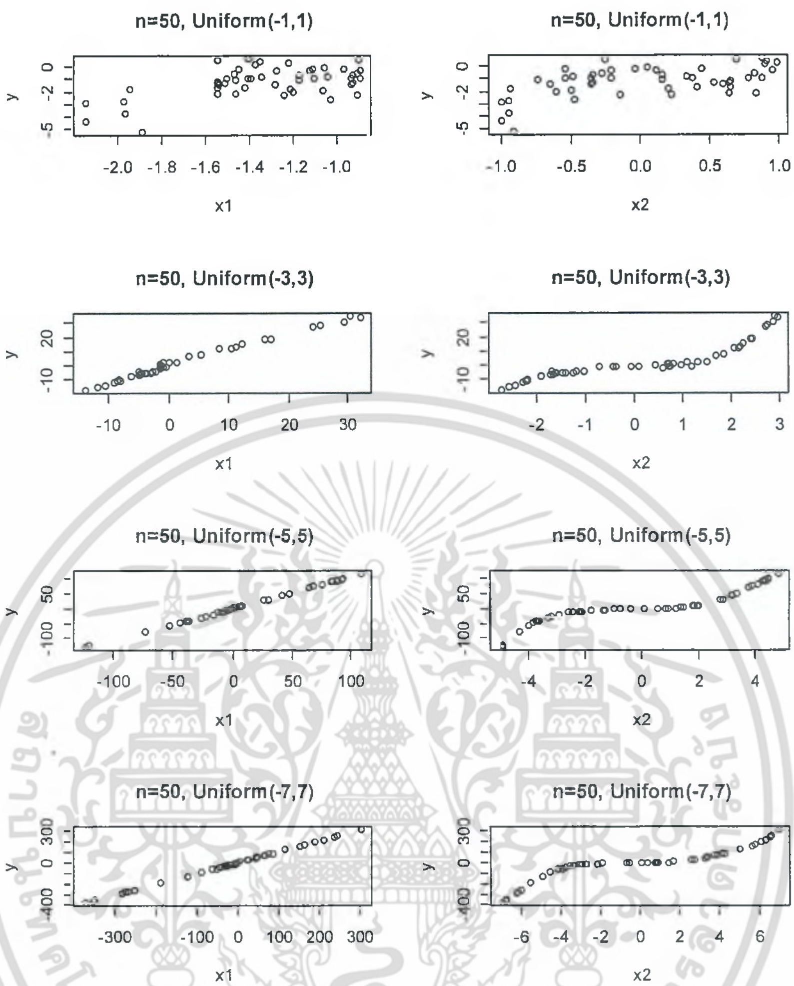

Figure 1. The graph of scatter plots between response variable and two explanatory variables (n=50).

The performance of parameter estimation is compared by Mean Square Errors (MSE) as follows:

$$MSE = \frac{\sum_{t=1}^{n}(y_t - \hat{y})^2}{n},$$

where y_t are the response variable, and $\hat{y}_t$ are the fitted values.

The first and second columns of Table 1 show the various sample sizes and its minimum and maximum values from uniform distribution. The third to fifth columns present the average MSE of parametric regression model (Reg), nonparametric regression model (Non), and semiparametric regression model (Sem).

Table 1. The average MSE of parametric regression model (Reg), nonparametric regression model (Non), and semiparametric regression model (Sem).

Sample sizes	Uniform	Reg	Non	Sem
n=25	-1,1	0.9348	0.8050	0.8482
	-3,3	0.7481	0.5795	0.6067
	-5,5	0.9139	0.8859	0.9139
	-7,7	0.4402	0.2796	0.3512
n=50	-1,1	0.9434	0.8946	0.9114
	-3,3	0.9476	0.8844	0.9033
	-5,5	0.8854	0.8003	0.8609
	-7,7	0.6926	0.6718	0.6855
n=100	-1,1	0.9902	0.9531	0.9680
	-3,3	0.9662	0.9407	0.9512
	-5,5	0.9693	0.9375	0.9502
	-7,7	0.9703	0.9513	0.9606
n=300	-1,1	0.9986	0.9882	0.9924
	-3,3	0.9911	0.9810	0.9847
	-5,5	0.9887	0.9785	0.9832
	-7,7	0.9862	0.9759	0.9796

By observing the average MSE, the results appear that the nonparametric regression model provides the minimum values in all cases. For the Figure 2, the average MSE is decreasing when the range of uniform distribution is increasing especially the sample sizes n = 25, 50, and 100. When the sample size is increasing, the average MSE is increasing shown at Figure 3. However, the average MSE of nonparametric regression model is shown the smallest values since the penalized spline method can be conducted using the truncated power function based on the number of knots which is controlled to trade-off the goodness of fit.

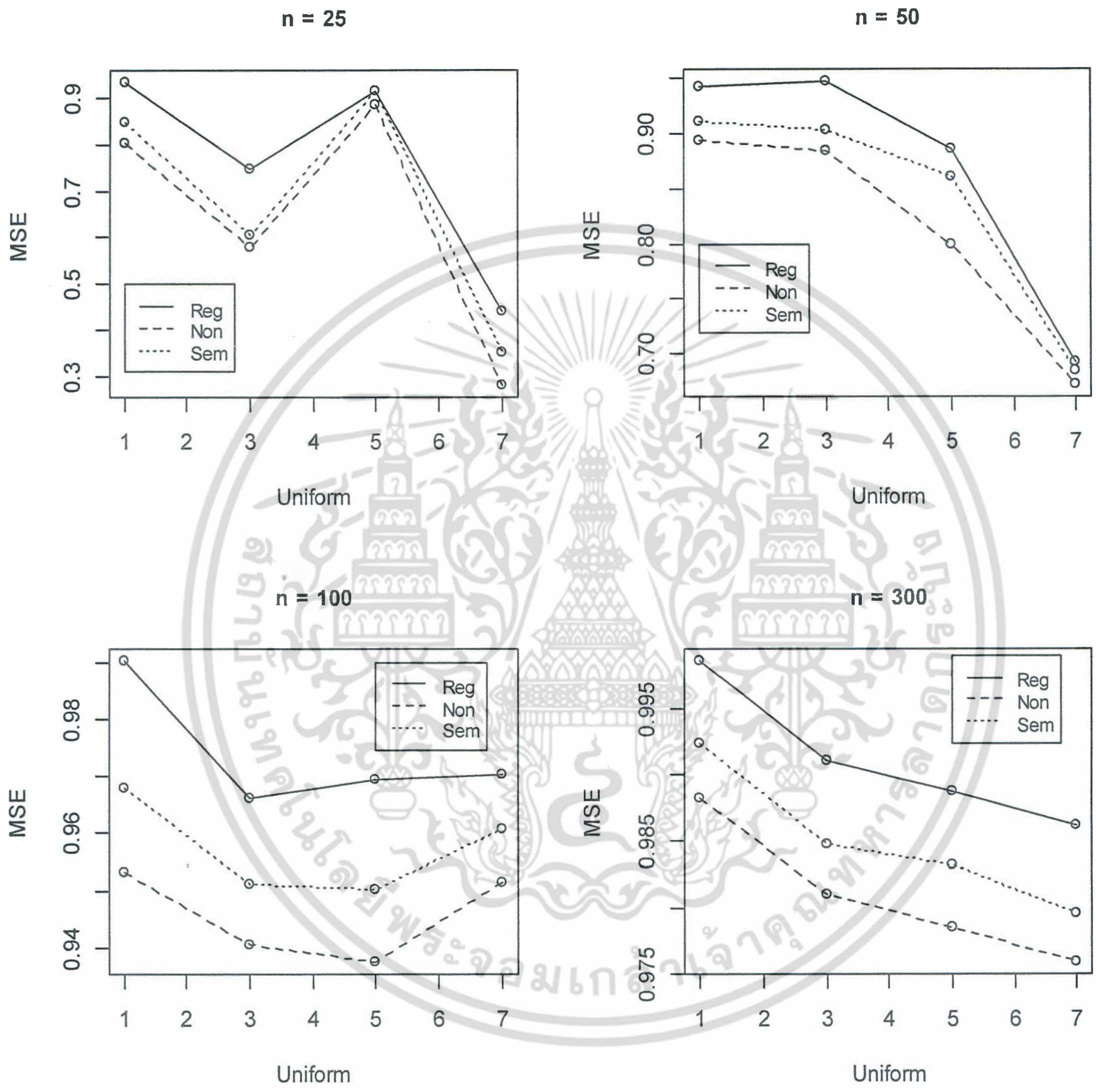

Figure 2. The average MSE depended on the range of uniform distribution of parametric regression model (Reg), nonparametric regression model (Non), and semiparametric regression model (Sem).

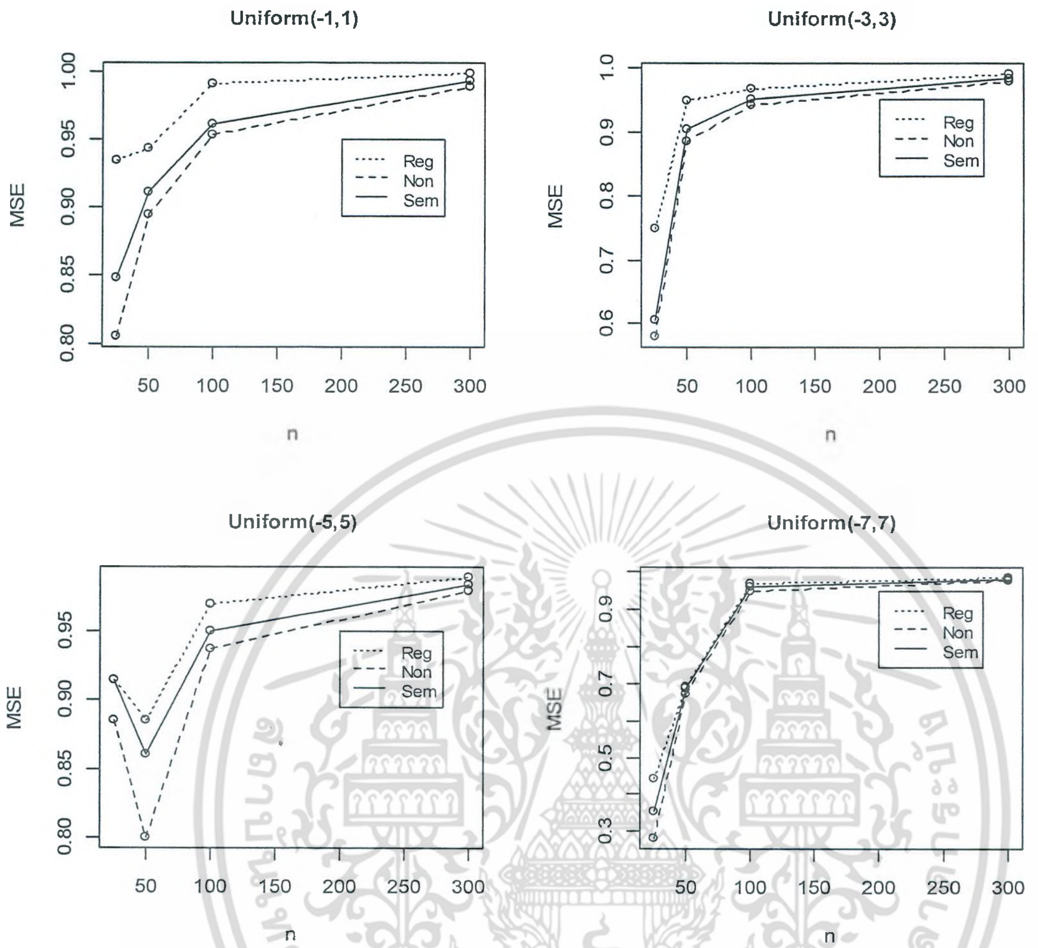

Figure 3. The average MSE depended on sample sizes of parametric regression model (Reg), nonparametric regression model (Non), and semiparametric regression model (Sem).

5. The Application in Real Data

In this section, we apply the model described in Section 2 to analyze with the real data. The gold price (US Dollars per Troy Ounce) is denoted the response variable and two explanatory variables are defined by the crude oil price (US Dollars per Barrel) and the number of month. These data consisted of 312 records of monthly volume from January 1988 to December 2013 that can be founded at *www.cmegroup.com* and *www.eia.gov.* The estimated parameters are obtain from the data analysis of parametric regression model, nonparametric regression model, and Semiparametric regression model given in Table 2 .

Table 2. The parameter estimation of parametric regression model, nonparametric regression model, and semiparametric regression model

	Reg	Non	Sem
Parameter Estimation	$\hat{\underset{\sim}{\theta}} = \begin{bmatrix} b_0 = 92.4411 \\ b_1 = 11.8667 \\ b_2 = -0.8904 \end{bmatrix}$	$f(x_1), \quad \lambda_{x1} = 23.25, K_{x1} = 35$ $f(x_2), \quad \lambda_{x2} = 352.6, K_{x2} = 2$	$\hat{\underset{\sim}{\theta}} = \begin{bmatrix} b_0 = 145 \\ b_1 = 11.89 \end{bmatrix}$ $f(x_2), \quad \lambda_{x2} = 375.5, K_{x2} = 2$
MSE	45393.51	22462.54	45047.07

From Table 2, it is apparent that MSE by the nonparametric regression model is the smallest values. Therefore, it should be noted that the nonparametric regression model performs better than the parametric regression model and semiparametric regression model since the nonparametric regression model contains the two smoothing function which can be interpolated more than the other models.

Figure 4 shown the fitted values from 3 models and the gold price is the bottom panel. It follows from the Figure 4 that the nonparametric regression model can be close the real values more than two models.

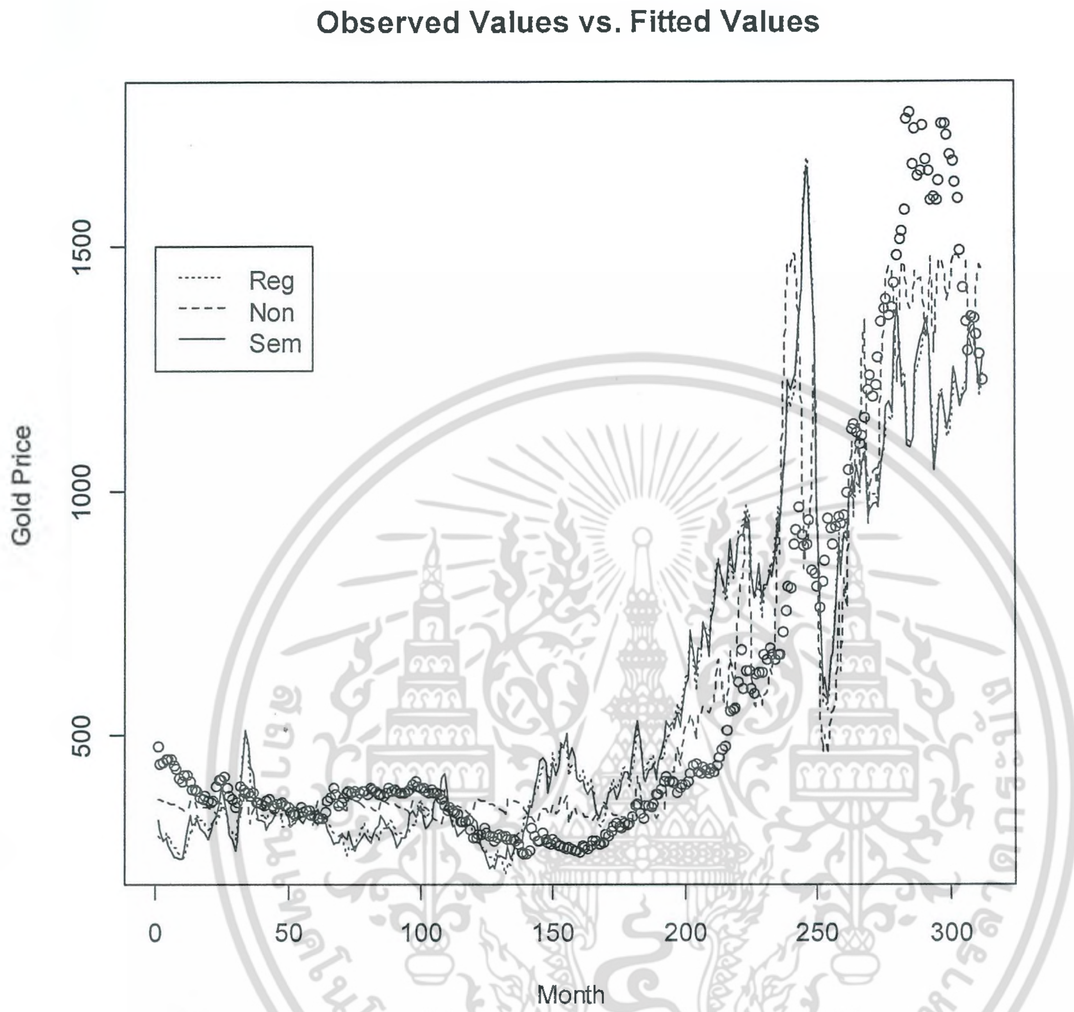

Figure 4. The scatter plot of the gold price and fitted values of parametric regression model, nonparametric regression model, and semiparametric regression model

6. Conclusions

In this paper, we have to compare parametric regression model, nonparametric regression model, and semiparametric regression model for estimating parameter. The nonparametric regression model and semiparametric regression model in term of smoothing function focus the penalized spline method. For simulation study, the nonparametric regression model performs better than two models based method in term of minimizing MSE. For real data, we are also interested the financial data to compare the performance of three models by considering the MSE. The results is similar the simulation that the nonparametric regression is a reasonable working more than parametric regression model and semiparametric regression model.

In other situation, a parametric regression model may not be available to use. To overcome the difficulty caused by restrictive assumption of the parametric regression model, one may remove the restriction that the penalized spline method belongs to nonparametric regression model.

Reference

[1] Wand M.P. and Jones M.C. *Kernel Smoothing,* Chapman and Hall, London, 1995.

[2] Fan J. and Gijbels, I. *Local polynomial modelling and its applications,* Chapman and Hall, London, 1996.

[3] Eubank R.L. *Spline Smoothing and Nonparametric Regression,* Marcel Dekker, New York, 1988.

[4] Eubank R.L. *Nonparametric Regression and Spline Smoothing,* Marcel Dekker, New York, 1999.

[5] Wahba G. *Spline Models for Observational Data,* SIAM, Philadelphia, PA, 1990.

[6] Green P.J. and Silverman B.W. *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach,* Chapman and Hall, London, 1994.

[7] Ruppert D., Wand, M.P. and Carroll R.J. *Semiparametric Regression,* Cambridge University Press, 2003.

[8] Robinson P.M. Non-parametric estimation for time series models, *Journal of Time Series Analysis*, 1983; **4**, 185-208.

[9] Stone C.G. Consistent nonparametric regression, *The Annals of Statistics*, 1977; **5**, 595-620.

[10] Fan J. Design-adaptive nonparametric regression, *Journal of American Statistics Association,* 1992; **87**, 998-1004.

[11] Fan J. Local linear regression smoothers and their minimax efficiency, *The Annals of Statistics*, 1993; **21**, 196-226.

[12] Ruppert D. and Carroll R.J. Spatial-adaptive penalties for spline fitting, *Australian and New Zealand Journal of Statistics,* 2000; **42**, 205-224.

[13] Ruppert D. Selecting the number of knots for penalized splines, *Journal of computational and Graphical Statistics,* 2002; **11**, 735-757.

[14] Robinson D. A. That BLUP is a good thing: the estimation of random effects, *Statistical Science,* 1991; **6**, 15-51.

[15] Brumback B.A., Ruppert D. and Wand M. P. Comment on Shively, Kohn, and Wood, *Journal of the American Statistical Association,* 1999; **94**, 794-797.

ประวัติคณะผู้วิจัย

1. ชื่อ : น.ส.อัชฌา อระวีพร

Miss. Autcha Araveeporn

2. สถานภาพปัจจุบัน : อาจารย์ประจำ ภาควิชาสถิติประยุกต์ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหาร ลาดกระบัง

3. หน่วยงานที่สังกัด: ภาควิชาสถิติประยุกต์ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหาร ลาดกระบัง

4. ประวัติการศึกษา :

ระดับการศึกษา	วุฒิการศึกษาที่ได้รับ	สาขาวิชา	ปีที่สำเร็จการศึกษา	สถานศึกษา
ปริญญาเอก	ปรัชญาดุษฎีบัณฑิต (ปร.ด.)	สถิติ	2553	มหาวิทยาลัยธรรมศาสตร์
ปริญญาโท	สถิติศาสตรมหาบัณฑิต (สต.ม.)	สถิติ	2541	จุฬาลงกรณ์มหาวิทยาลัย
ปริญญาตรี	วิทยาศาสตรบัณฑิต (วท.บ.)	สถิติประยุกต์	2539	สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ

5. ผลงาน

5.1 งานวิจัยที่ตีพิมพ์แล้ว

- สมศรี บัณฑิตวิไล และ **อัชฌา อระวีพร.** 2547. *การศึกษาหลักสูตรสถิติประยุกต์ต่อการนำไปประยุกต์ใช้ในงานอาชีพของมหาวิทยาลัยของรัฐในกรุงเทพมหานคร,* การประชุมวิชาการสถิติประยุกต์ภาคเหนือ ครั้งที่ 5 : 423-433.
- สมศรี บัณฑิตวิไล และ **อัชฌา อระวีพร.** 2548. *การศึกษาทัศนคติของนักศึกษาที่มีต่อกองทุนเงินให้กู้ยืมเพื่อการศึกษา ของคณะวิทยาศาสตร์สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง,* วารสารวิทยาศาสตร์ลาดกระบัง, 14(1) : 40-50.
- **A.Araveeporn**. 2011. *Developing Nonparametric Conditional Heteroscedastic Autoregressive Nonlinear Model by Using Maximum Likelihood Method,* Chiang Mai Journal of Science, 38(3) : 331-345.
- **A.Araveeporn.** 2012. *The Least-Squares Criteria of the Random Coefficient Dynamic Regression Model*, Journal of Statistical Theory and Practice, 6(2) : 315-333.
- **อัชฌา อระวีพร**. 2555. *การประมาณค่าดัชนีตลาดหลักทรัพย์แห่งประเทศไทยโดยวิธีการถดถอยเชิงเส้นอย่างง่าย และวิธีการถดถอยแบบไม่ใช้พารามิเตอร์อย่างง่าย*, การประชุมวิชาการสถิติและสถิติประยุกต์ระดับชาติ ครั้งที่ 13 ประจำปี 2555 : 80-87.
- **A.Araveeporn**. 2011. *The Comparison of Bandwidth Selection Methods using Kernel Function,* KMITL Science Technology Journal, 11(2) : 64-78.

- **A.Araveeporn.** 2012. *Bayesian Analysis of Random Coefficient Dynamic AutoRegressive Model*, Thailand Statistician, 10(2) : 199-223.
- A.Araveeporn. 2013. The Maximum Likelihood Method of Random Coefficient Dynamic Regression Model, World Academy of Science, Engineering and Technology, 78, 2055-2060.
- **สุทธิศักดิ์ ป้อมแจ่มศรี และ อัชฌา อระวีพร**. 2556. การเปรียบเทียบประสิทธิภาพการประมาณค่าของตัวแบบการถดถอยในกรณีที่ข้อมูลมีความสัมพันธ์กัน, การประชุมวิชาการสถิติประยุกต์และเทคโนโลยีสารสนเทศระดับชาติ ประจำปี 2556 : 11-21.
- **A.Araveeporn**. 2013. *A Comparison of Two Least-Squared Random Coefficient Autoregressive models: with and without autocorrelated Errors*, International Journal of Advanced Statistics and Probability, 1, 151-162.
- **A.Araveeporn**. 2014. *A Comparison of Smoothing Spline Method and Penalized Spline Regression Method based on Nonparametric Regression Model*, International Science Index, 5, 1232-1236.
- **A.Araveeporn**. 2014. Parameter Estimation of Poisson Distribution by Using Maximum Likelihood, Markov Chain Monte Carlo, and Bayes method, Thammasat International Journal of Science and Technology, 19(3), 1-14.

5.2 บทความที่ตีพิมพ์แล้ว

- **A.Araveeporn**. 2012. *The Estimation of Smoothing Parameter using Smoothing Techniques on Nonparametric Regression*, Silpakorn University Science and Technology Journal, 6(1) : 14-22.
- **อัชฌา อระวีพร**. 2554. *การหาค่าตัวประมาณเบส์ด้วยโปรแกรมวินบั๊ก*, วารสารวิทยาศาสตร์ลาดกระบัง, 20(2) : 45-60.
- **อัชฌา อระวีพร**. 2555. *การวิเคราะห์เบส์จากโปรแกรมวินบั๊กสู่โปรแกรมอาร์*, วารสารวิทยาศาสตร์มหาวิทยาลัยนเรศวร, 9(1) : 30-44.
- **อัชฌา อระวีพร**. 2555. *การประมาณค่าตัวแบบการถดถอยแบบไร้พารามิเตอร์ด้วยชุดคำสั่ง SemiPar ในโปรแกรมอาร์*, วารสารวิทยาศาสตร์ลาดกระบัง, 21(2) : 78-86.

5.3 หัวหน้าโครงการวิจัย

5.3.1 เรื่อง การวิเคราะห์ตัวแบบการถดถอยพลวัตของสัมประสิทธ์แบบสุ่ม ได้รับงบประมาณสนับสนุนจากเงินรายได้คณะวิทยาศาสตร์ ปี 2554 (เสร็จสิ้นแล้ว)

5.3.2 เรื่อง การวิเคราะห์สมการถดถอยแบบไร้พารามิเตอร์ด้วยฟังก์ชันเคอร์เนล ได้รับงบประมาณสนับสนุนจากเงินรายได้คณะวิทยาศาสตร์ ปี 2555 (เสร็จสิ้นแล้ว)

5.3.3 เรื่อง การวิเคราะห์ตัวแบบสหสัมพันธ์ของสัมประสิทธ์แบบสุ่ม เมื่อค่าความคลาดเคลื่อนเกิดสหสัมพันธ์ในตัวเอง ได้รับงบประมาณสนับสนุนจากเงินรายได้คณะวิทยาศาสตร์ ปี 2556 (เสร็จสิ้นแล้ว)

5.3.4 เรื่อง การเปรียบเทียบการประมาณค่าสำหรับการถดถอยแบบพารามิเตอร์ แบบไม่ใช้พารามิเตอร์ และแบบผสม ได้รับงบประมาณสนับสนุนจากเงินรายได้คณะวิทยาศาสตร์ ปี 2557 (อยู่ระหว่างดำเนินการ)

5.3.5 การวิเคราะห์ข้อมูลอนุกรมเวลาด้วยการถดถอยแบบไม่ใช้พารามิเตอร์สำหรับตัวแบบความผันผวนอัตสหสัมพันธ์ไม่เชิงเส้น ได้รับงบประมาณสนับสนุนจากสำนักงานวิจัยแห่งชาติ ปี 2558