Kernel multivariable semiparametric regression model in estimating the level of open unemption in East Java Province

Andi Tenri Ampa1,*, I Nyoman Budiantara2, and Ismaini Zain2

1Department of Statistics, FMIPA, Halu Oleo University, Kendari-Indonesia
2Department of Statistics, FSAD, Sepuluh Nopember Institute of Technology, Surabaya, Indonesia

*anditenriampa910@gmail.com (corresponding author)

Abstract. Semiparametric regression is a combination of parametric regression and nonparametric regression. Parametric regression curve components are approximated by multivariable linear functions, nonparametric regression curve components are approximated by Gaussian Kernel function. The purpose of this study is to obtain an estimate of the shape in semiparametric regression using the Kernel estimator and to model the Open Unemployment Rate (TPT) in East Java Province using the semiparametric regression model. This semiparametric regression model estimates on bandwidth. Semiparametric regression model is obtained by minimizing the Generalized Cross Validation function. The semiparametric regression model used to model TPT case data in East Java Province.

Keywords: Semiparametric Regression, Generalized Cross Validation, Kernel.

1. Introduction

Unemployment is a big problem facing the country. The large number of open unemployment rates has broad social implications because those who do not work have no income. The higher the open unemployment rate, the greater the potential for social vulnerability to cause, for example, crime. Conversely, the lower the open unemployment rate, the more stable social conditions in society will be. It is very appropriate if the government often uses this indicator as a measure of development success.

Some of the variables used and thought to affect the Open Unemployment Rate (TPT) are Labor Force Participation Rate (TPAK), Regency / City Minimum Wage (UMK), Percentage of population aged 15+ who graduated from high school, population density, Gross Regional Domestic Product (PDRB), and investment.

One of the methods in statistics that is often used to analyze the effect of the relationship between the response variable and the variables that affect it is regression analysis (predictor variables). According to the pattern of the relationship between the predictor variable and the response variable, there are 3 (three) types of regression analysis, namely parametric, nonparametric, and semiparametric. If the shape of the regression curve is known, for example following a linear/quadratic/cubic pattern, then a suitable regression approach is linear / quadratic / cubic parametric regression. Nonparametric regression was used for the pattern of relationships with
unknown shapes and unavailability of information related to data patterns. If some data patterns are
known and partially unknown, semiparametric regression is used. Regression curve estimation is one
of the main objectives in regression analysis.

In nonparametric and semiparametric regression, especially for the multivariable case it is possible
to have several types of estimators. Several estimators in nonparametric regression, including Kernel
[13] are used for pattern less data, Spline Smoothing [8] pattern data that does not depend on point
knots, Fourier Series [3], and Spline Truncated [17].

Kernel Estimators are used more frequently in nonparametric regression. This is because it is
simpler [15]. In addition, the Kernel estimator has a convergence speed that is relatively faster than the
Local Polynomial, Fourier Series, or Spline estimator [6]. Another method that has often received
attention is the Fourier Series method. This method is very specific and very well used in the case of
data that has a repeating pattern on the relationship between the response variable and the predictor
variable [1], [3], [11] and [16].

This paper discusses the semiparametric regression estimator used to estimate the regression model
in which the predictor variable has a linear relationship pattern, and the other predictor variables are
unknown. The parametric component will be approximated by the linear function, while the
nonparametric component will be approximated by the Kernel. The estimation method uses Weighted
Leas Square (WLS).

Kernel estimator depends on bandwidth, small bandwidth will produce too rough estimation, while
large bandwidth will produce too smooth estimation, so it cannot estimate the data well. This paper
provides the optimal bandwidth selection method. Furthermore, the estimator obtained is used to
model TPT in East Java Province.

2. Methodology
To achieve the research objectives, the following steps were made.

1. Obtaining Estimator of Kernel in Multivariable Semiparametric Regression.
2. Selecting Multivariable Kernel Estimator Bandwidth Parameters in Semiparametric
   Regression.
3. Application of Kernel Estimator Multivariable in Semiparametric Regression in East Java
   Open Unemployment Rate Data.

2.1. Data Structure
In this study, the data structure used was data with response variables and predictor variables
consisting of parametric and nonparametric components (Table 1).

| Observes | Response | Predictor |
|----------|----------|-----------|
|          | y        | x_1       | x_2       | x_3       | x_4       | x_5       | x_6       |
| 1        | y_1      | x_1(1)    | x_2(1)    | x_3(1)    | x_4(1)    | x_5(1)    | x_6(1)    |
| 2        | y_2      | x_1(2)    | x_2(2)    | x_3(2)    | x_4(2)    | x_5(2)    | x_6(2)    |
| ...      | ...      | ...       | ...       | ...       | ...       | ...       | ...       |
| 38       | y_{38}   | x_1(38)   | x_2(38)   | x_3(38)   | x_4(38)   | x_5(38)   | x_6(38)   |

Table 1. Research Data Structure
2.2. Operational Definition of Variables

The variables in this study consist of 1 (one) response variable, namely the open unemployment rate and 5 (five) predictor variables, namely: District/City Minimum Wage, Percentage of Population 15+ who passed SMA/SMK, Population Density and Investment, GRDP. Furthermore, the operational definitions of the variables used in the study are given, these are as follows.

1. Open Unemployment Rate (TPT) = y
   TPT is the percentage of the total unemployed against the total workforce.
   \[ \text{TPT} = \frac{\text{number of open unemployed}}{\text{number of labor force}} \times 100\% \]

2. Labor Force Participation Rate = \( x_1 \)
   The ratio between the labor force and the total working age population. This indicator can be written using the formula below:
   \[ \text{TPAK} = \frac{\text{number of labor force}}{\text{population 15 years and over}} \times 100\% \]

3. Regional Minimum Wage (UMR) = \( x_2 \)
   UMR is the amount of wages paid in each region based on regional regulations set by each region. UMR is a minimum standard in a district / city that is used by entrepreneurs or business people to pay wages to employees or laborers in their work environment (Minister of Manpower and Transmigration Regulation No.7 of 2015)

4. Percentage of population 15+ who graduated from SMA / SMK = \( x_3 \)
   The percentage of the population aged 15 years and over who graduated from SMA / SMK is the population 15 years and over who graduated from SMA / SMK divided by the number of populations aged 15 years and over based on the results of the National Socio-Economic Survey. The percentage of population 15+ graduating from SMA/SMK is:
   \[ \frac{\text{number of people aged 15 + who graduated from SMA / SMK}}{\text{number of people aged 15 years and over"}} \]

5. Population density= \( x_4 \)
   Number of inhabitants per square kilometer. The population is divided widely in an area.

6. Infestation = \( x_5 \)
   The number of realizations of domestic investment and foreign model investment as well as in each district / city is sourced from BPS data.

7. Gross Regional Domestic Product (GRDP) = \( x_6 \)
   Gross Regional Domestic Product based on market prices is the total gross value-added arising from all economic sectors in a region. Added value is the value added from a combination of production factors and raw materials in the production process. The calculation of added value is the value of production (output) minus intermediate costs. The gross value added here includes the components of factor income (wages and salaries, interest, land rent and profits), depreciation and net indirect taxes. So, by adding up the gross added value from each sector and adding up the gross added value from all these sectors, the Gross Regional Domestic Product will be obtained based on market prices.
3. Result and Discussion

3.1 Semiparametric Regression

Semiparametric regression is a regression that contains a parametric component and a nonparametric component. Semiparametric regression is more suitable for data where some of the predictor components have a certain pattern of relationships and some are not.

Given paired data \((x_1, x_2, \ldots, x_{pi}, t_1, t_2, \ldots, t_{qi}, y_i)\) following a semiparametric regression model:

\[
y_i = \mu(x_{i1}, x_{i2}, \ldots, x_{ipi}, t_{i1}, t_{i2}, \ldots, t_{iqi}, y_i) + \varepsilon_i, \quad i = 1, 2, \ldots, n,
\]

\(\mu\) is an additive regression, \(\varepsilon\) curves a random error that is assumed to be independent and identical in normal distribution with zero mean and variance \(\sigma^2\).

\[
y_i = \mu(x_{i1}, x_{i2}, \ldots, x_{ipi}, t_{i1}, t_{i2}, \ldots, t_{iqi}, y_i) + \varepsilon_i, \quad i = 1, 2, \ldots, n
\]

(1)
The semiparametric model in equation (1) can be written in the following matrix form:

\[
\begin{bmatrix}
y_1 \\
y_2 \\
y_3 \\
\vdots \\
y_n
\end{bmatrix} = 
\begin{bmatrix}
1 & x_{11} & x_{21} & \cdots & x_{ipi} \\
1 & x_{12} & x_{22} & \cdots & x_{ipi} \\
1 & x_{13} & x_{23} & \cdots & x_{ipi} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & x_{in} & x_{2n} & \cdots & x_{pm}
\end{bmatrix}
\begin{bmatrix}
\beta_0 \\
\beta_1 \\
\beta_2 \\
\vdots \\
\beta_p
\end{bmatrix} + 
\begin{bmatrix}
\sum_{k=1}^{q} g_k(t_{k1}) \\
\sum_{k=1}^{q} g_k(t_{k2}) \\
\sum_{k=1}^{q} g_k(t_{k3}) \\
\vdots \\
\sum_{k=1}^{q} g_k(t_{kn})
\end{bmatrix}
\begin{bmatrix}
\varepsilon_1 \\
\varepsilon_2 \\
\varepsilon_3 \\
\vdots \\
\varepsilon_n
\end{bmatrix}
\]

which \(\sum_{j=1}^{p} m_j(x_{ji}), \quad i = 1, 2, \ldots, n\) is a parametric component and \(\sum_{k=1}^{q} g_k(t_{ki})\) is a nonparametric component.

Which for \(i = 1, 2, \ldots, n\), obtained, the equation becomes:

\[
y_1 = \sum_{j=1}^{p} m_j(x_{ji}) + \sum_{k=1}^{q} g_k(t_{k1}) + \varepsilon_1, \quad \ldots, \quad y_n = \sum_{j=1}^{p} m_j(x_{jn}) + \sum_{k=1}^{q} g_k(t_{kn}) + \varepsilon_n
\]

Next can be written:

\[
y_1 = m_1(x_{i1}) + \cdots + m_p(x_{ipi}) + g_1(t_{i1}) + \cdots + g_q(t_{iqi}) + \varepsilon_1
\]

\[
y_2 = m_1(x_{i2}) + \cdots + m_p(x_{ipi}) + g_1(t_{i2}) + \cdots + g_q(t_{iqi}) + \varepsilon_2
\]

\[
\vdots
\]

\[
y_n = m_1(x_{in}) + \cdots + m_p(x_{mpi}) + g_1(t_{in}) + \cdots + g_q(t_{mqi}) + \varepsilon_n
\]

As a result, the regression model in equation (1) can be written as follows:
In the form of a matrix it can be presented as follows:

\[
\begin{bmatrix}
  y_1 \\
  y_2 \\
  \vdots \\
  y_n
\end{bmatrix} =
\begin{bmatrix}
  \sum_{j=1}^{p} m_j(x_{j1}) & \sum_{j=1}^{q} g_k(t_{k1}) & \varepsilon_1 \\
  \sum_{j=1}^{p} m_j(x_{j2}) & \sum_{j=1}^{q} g_k(t_{k2}) & \varepsilon_2 \\
  \vdots & \vdots & \vdots \\
  \sum_{j=1}^{p} m_j(x_{jn}) & \sum_{j=1}^{q} g_k(t_{kn}) & \varepsilon_n
\end{bmatrix}
\]

Furthermore, the regression curve \( \sum_{j=1}^{p} m_j(x_{ji}) \) in the equation (1) is a parametric component which is approximated by a linear function in the form:

\[
m_j(x_{ji}) = \beta_{0j} + \beta_{1j}x_{ji}, \quad i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, p
\]

**Kernel.** Parametric components so that the equation (2) can be written as:

\[
\begin{bmatrix}
  m(x_1) \\
  m(x_2) \\
  \vdots \\
  m(x_n)
\end{bmatrix} =
\begin{bmatrix}
  \sum_{j=1}^{p} m_j(x_{j1}) & \sum_{j=1}^{q} \left( \beta_{0j} + \beta_{1j}x_{j1} \right) \\
  \sum_{j=1}^{p} m_j(x_{j2}) & \sum_{j=1}^{q} \left( \beta_{0j} + \beta_{1j}x_{j2} \right) \\
  \vdots & \vdots \\
  \sum_{j=1}^{p} m_j(x_{jn}) & \sum_{j=1}^{q} \left( \beta_{0j} + \beta_{1j}x_{jn} \right)
\end{bmatrix}
\]

\[
= \begin{bmatrix}
  \beta_{01} + \beta_{11}x_{11} + \beta_{02} + \beta_{12}x_{21} + \ldots + \beta_{0p} + \beta_{1p}x_{p1} \\
  \beta_{01} + \beta_{11}x_{12} + \beta_{02} + \beta_{12}x_{22} + \ldots + \beta_{0p} + \beta_{1p}x_{p2} \\
  \vdots \\
  \beta_{01} + \beta_{11}x_{1n} + \beta_{02} + \beta_{12}x_{2n} + \ldots + \beta_{0p} + \beta_{1p}x_{pn}
\end{bmatrix}
\]

So that equation (3) can be written in matrix form, as follows:

\[
\begin{bmatrix}
  \beta_0^* \\
  \beta_1^* \\
  \vdots \\
  \beta_n^*
\end{bmatrix} = \begin{bmatrix}
  \beta_{01} + \beta_{02} + \ldots + \beta_{0p} \\
  \beta_{11} + \beta_{12} + \ldots + \beta_{1p} \\
  \vdots \\
  \beta_{n1} + \beta_{n2} + \ldots + \beta_{np}
\end{bmatrix}
\]
\[ m_j = X\beta \]

Where:

\[
X = \begin{bmatrix}
1 & x_{11} & x_{21} & \cdots & x_{p1} \\
1 & x_{12} & x_{22} & \cdots & x_{p2} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & x_{1n} & x_{2n} & \cdots & x_{pn}
\end{bmatrix}
\] dan \[ \beta = \begin{bmatrix} \beta_0 & \beta_1 & \beta_2 & \cdots & \beta_p \end{bmatrix}^T \]

For regression curve \( \sum_{k=1}^{q} g_k(t_{k_i}) \) is a nonparametric component of the Kernel from (1),

\[ g_k(t_{k_i}), \; k = 1, 2, \ldots, q \]

approached with the Kernel function, then with the Nadaraya Watson approach, the estimate \( g \) is

\[
g(t) = n^{-1} \sum_{i=1}^{n} \frac{1}{\varphi} K \left( \frac{t-t_i}{\varphi} \right) y_i = n^{-1} \sum_{i=1}^{n} \frac{1}{\varphi} K \left( \frac{t-t_{k_i}}{\varphi} \right) y_i + n^{-1} \sum_{i=1}^{n} \frac{1}{\varphi} K \left( \frac{t-t_{k_i}}{\varphi} \right) y_{k_i} + \ldots + n^{-1} \sum_{i=1}^{n} \frac{1}{\varphi} K \left( \frac{t-t_{n_i}}{\varphi} \right) y_{n_i}
\]

then to find an estimator \( \sum_{k=1}^{q} g_k(t_{k_i}) \) in equation (1), first looking for an estimate:

\[ g_k(t_{k_i}) \] with \( k = k_1 \) up to \( k = k_n \) for \( g_k(t_{k_i}) \).

For those who are approximated by the Kernel, the Nadaraya-Watson estimator is:

\[
g_k(t_{k_i}) = n^{-1} \sum_{i=1}^{n} \frac{1}{\varphi_k} K \left( \frac{t_{k_i}-t_{k_i}}{\varphi_k} \right) y_i
\]

Based on equation (5), equation (6) can be written as:

\[ g_k(t_{ki}) = \sum_{i=1}^{n} \frac{1}{\varphi_k} K \left( \frac{t_{k_i}-t_{ki}}{\varphi_k} \right) y_i \]
Furthermore, equation (6) is written as follows:

\[
\begin{bmatrix}
\frac{1}{n} \sum_{i=1}^{n} K \left( \frac{t_{i1} - t_{11}}{\phi_i} \right) y_1 \\
\cdots \\
\frac{1}{n} \sum_{i=1}^{n} K \left( \frac{t_{in} - t_{1n}}{\phi_i} \right) y_n
\end{bmatrix}
\]

Consequently, equation (8), then the estimate for the regression curve

\[
\sum_{k=1}^{q} g_k(t_{ki}) = \sum_{k=1}^{q} P_k(\phi_k) y = P_1(\phi_1) y + P_2(\phi_2) y + \ldots + P_q(\phi_q) y
\]

which \( \Omega_k = P_1(\phi_1) + P_2(\phi_2) + \ldots + P_q(\phi_q) \), \( \phi_k \) is the Kernel \( k \).

Furthermore, to find the parametric component estimator, it can be obtained from minimizing \( \epsilon^T \epsilon \) from the equation:
Based on equations (8) and (9), the estimation for equation (1) is as follows:

\[
\hat{y} = X\hat{\beta} - \hat{g}_k
\]

\[
\hat{y} = X(X^TX)^{-1}X^T(\Omega_k + I)y
\]

So that equation (10) can be written in (11) matrix form, as follows:

\[
\hat{y} = M'y, \quad M' = X(X^TX)^{-1}X^T(2\Omega_k + I)
\]

### 3.2. Selection Optimal Bandwidth

The important thing that plays a role in getting a kernel depends on the bandwidth selection. Bandwidth is a smoothing parameter that controls the smoothness of the estimated curve. One method that is often used in bandwidth selection is Generalized Cross Validation (GCV). When compared with other methods, for example Cross Validation (CV) and the Unimplemented Risk (UBR) method or the Generalized Maximum Likelihood (GML), GCV theoretically has optimal asymptotic properties. The GCV method also has the advantage of not requiring knowledge of population variance and the invariance GCV method for transformation. The GCV method is a development of CV [17].

Selection of smoothing parameters \( \varphi \) the optimum using GCV is defined as follows:

\[
GCV(\lambda) = \frac{MSE(\lambda)}{(n^{-1}trace(I - B(\lambda)))^2}
\]

where:

\[
MSE(\lambda) = n^{-1}y'(I - B(\lambda))(I - B(\lambda))y
\]

and \( B(\lambda) \) is a matrix obtained from the equation \( \hat{y} = B(\lambda)y \). Score \( \lambda \) the optimum is obtained from \( GCV(\lambda) \) the smallest.

### 3.3 Open Unemployment Rate (TPT)

The Open Unemployment Rate (TPT) is defined by the Central Statistics Agency (BPS), the workforce is defined as the working age population (15 years and over / 15+) who are employed, or have a job but are temporarily unemployed and unemployed. Meanwhile, those who are not included in the workforce are the working age population (15 years and over) who are still in school, taking care of the household, or carrying out other activities.

Open unemployment is a workforce that has no jobs at all. TPT itself is defined as the percentage of the total unemployed against the total workforce. Open unemployment consists of:

a) those who are unemployed and looking for work,
b) those who do not have a job and prepare a business,
c) those who do not have a job and do not look for work because they feel it is impossible to get a job,
d) those who already have a job, but haven't started working yet. TPT indicates the large percentage of the workforce that is included in unemployment. The variables for compiling
this indicator are obtained from the National Labor Force Survey. A high TPT indicates that there is a large labor force that is not absorbed in the labor market.

3.4 Selection of Optimal Refining Parameters in Semiparametric Regression

The optimal smoothing parameter is very important in controlling the goodness of fit and smoothness of function. The optimal bandwidth is very important in the Kernel estimator, that is too large or wide will produce a very smooth curve, but not according to the pattern. Furthermore, smoothing parameter have an important role in controlling the goodness of fit and smoothness of function. In this case, the goodness of the estimator Kernel depends very much on the optimal bandwidth.

The GCV method is one of the optimal smoothing parameter selection methods. In this case, smoothing parameter is smoothing parameter between goodness of fit and smoothness of function. Furthermore, the method for selecting optimal smoothing parameters and optimal bandwidth for estimator Kernel in semiparametric regression is derived. The estimator of Kernel in semiparametric regression is presented as follows:

$$\hat{y} = \hat{\mu}_{\beta \varphi}(x, t) = M \hat{y}$$

From equation (12), it is obtained the mean square error (MSE) of Kernel in semiparametric regression, as follows:

$$MSE(\lambda) = n^{-1} \left\| (1 - M^T) y \right\|^2,$$

then the selection of the optimal smoothened can be obtained by finding the minimum value of GCV with the GCV formula given in equation (12), so that it is obtained:

$$GCV(\lambda) = \frac{n^{-1} \left\| (1 - M^T) y \right\|^2}{(n^{-1}trace(1 - M^T))^2}$$

Optimal smoothing parameters, optimal oscillation parameters and optimal bandwidth are obtained at the smallest $GCV(\lambda)$.

3.5 Application of Multivariable Kernel Semiparametric Regression Estimators on the Open Unemployment Rate Data of East Java Province

Selecting the optimal bandwidth parameter based on the minimum GCV value, is given in Table 2.

| Parametric | Kernel Nonparametric | $R^2$ | GCV          |
|-----------|----------------------|-------|--------------|
| $x_1$     | $x_2$                | $x_3$ | $x_4$       | $x_5$ | $x_6$ | 68.34827 | 1.444297e+08 |
| $x_2$     | $x_1$                | $x_3$ | $x_4$       | $x_5$ | $x_6$ | 70.3574  | 1.298639e+08 |
| $x_3$     | $x_1$                | $x_2$ | $x_4$       | $x_5$ | $x_6$ | 68.70495 | 1.450700e+08 |
| $x_4$     | $x_1$                | $x_2$ | $x_3$       | $x_5$ | $x_6$ | 66.05832 | 1.415290e+08 |
| $x_5$     | $x_1$                | $x_2$ | $x_3$       | $x_4$ | $x_6$ | 76.3572  | 1.076903e+08 |
| $x_6$     | $x_1$                | $x_2$ | $x_3$       | $x_4$ | $x_5$ | 76.86773 | 1.082943e+08 |

Then the estimation of parameters of Multivariable Kernel Semiparametric Regression Model is given
\[
y_i = 4219.21326 + 10.10514x_i + \sum_{i=1}^{n} \frac{1}{737863e+04} K \frac{t_i - t_{ij}}{737863e+04} y_i + \sum_{i=1}^{n} \frac{1}{3.892592e+04} K \frac{t_i - t_{ij}}{3.892592e+04} y_i + \sum_{i=1}^{n} \frac{1}{1.616939e+02} K \frac{t_i - t_{ij}}{1.616939e+02} y_i
\]

Determining the value \( R^2 \) of Multivariable Kernel Semiparametric Regression. \( R^2 = 76.3572 \)

4. Conclusion

The purpose of this study is to obtain an estimate of the shape in semiparametric regression using the Kernel estimator and to model the Open Unemployment Rate (TPT) in East Java Province using the semiparametric regression model. This semiparametric regression model estimates on bandwidth. Semiparametric regression model is obtained by minimizing the Generalized Cross Validation function. The multivariable Kernel semiparametric regression model that is applied to TPT case data in East Java Province produces a model with \( R^2 = 76.3572 \).

In the application of the regression model estimation results, the researcher uses semiparametric regression with 1 variable of parametric component and 5 variables of nonparametric components. On the next occasion, several possible combinations will be made.

Acknowledgement

I would like to thank all those who have helped carry out this research. My special speech goes to my promoter and copromotor.

References

[1] Amato, U., Antoniadis, A., and De-Feis, I. 2002. *Journal of Computational and Applied Mathematics*, 146: 459-479.
[2] Badan Pusat Statistika (BPS), 2017. Ketenagakerjaan: Pengangguran Terbuka. www.bps.go.id diakses pada tanggal 12 Agustus 2019.
[3] Bilodeau, M., 1992, Fourier Smoother and Additive Models, *The Canadian Journal of Statistical*, 3: 257-269.
[4] Budiantara, I N. dan Mulianah, 2007, Pemilihan Bandwidth Optimal dalam Regresi Semiparametrik Kernel dan Aplikasinya, *Sigma*, 10, 2: 159-166.
[5] Chen, H. and Shiau., J.J.H., 1994, Data Driven Efficient Estimators for a Partially Linear Model, *The Annals of Statistics*, 22: 211-237.
[6] Cheng, M.Y., Paige, R.L., Sun, S., and Yan, K., 2010, Variance Reduction for Kernel Estimators in Clustered/Longitudinal Data Analysis, *Journal of Statistical Planning and Inference*, 140: 1389-1397.
[7] Cox, D.D. and O’Sullivan, F., 1996, Penalized Type Estimator for Generalized Nonparametric Regression, 1983, *Journal of Multivariate Analysis*, 56: 185-206.
[8] Drapper, N.R. and Smith, H., 1992, *Applied Regression Analysis 2nd edition*, Marcel Dekker, New York.

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

[10] Frenkel, R. and Ros, J., 2006. *Unemployment next Term and the Real Exchange Rate in Latin America. World Development.* 34, 4: 631-646.

[11] Galtchouk, L. and Pergamenshchikov, S., 2009, “Adaptive Asymptotically Efficient Estimation in Heteroscedastic Nonparametric Regression”, *Journal of the Korean Statistical Society*. 38: 305-322.

[12] Green, P.J. and Silverman, B.W. 1994, *Nonparametric Regression and Generalized Linear Models (a Roughness Penalty Approach)*. Chapman & Hall, London.

[13] Hardle, W., 1994, *Applied Nonparametric Regression*. Humboldt Universität zu Berlin, Berlin.

[14] Nadaraya dan Watson, 1964, *On Estimating Regression. Theori Of Probability and ITS Application*.

[15] Okumura, H. and Naito, K., 2006, *Nonparametric Kernel Regression for Multinominal Data*, *Journal of Statistics Planning and Inference*, 14.

[16] Tripena, A. and Budiantara, I.N., 2009, *Fourier Estimator in Nonparametrik Regression*, *International Conference on Natural Sciences and Applied Natural Sciences*, Ahmad Dahlan University, Yogyakarta.

[17] Wahba, G., 1990, *Spline Models for Observasion Data*, SIAM Pensylvania.