BAREKENG: Journal of Mathematics and Its Applications
June 2025 Volume 19 Issue 2 Page 0805–0818
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol19iss2pp0805-0818

MODELING THE INFLUENCE OF CRUDE OIL PRODUCTION AGAINST INDONESIAN SOLAR WHOLESALE PRICE INDEX WITH LEAST SQUARE SPLINE ESTIMATOR APPROACH

Rosidun Nindyo Pratiwi[1*], Nathania Fauziah[2], Bimo Okta Syahputra[3], Ahmad Wahyu Firmanda[4], Dita Amelia[5]

[1,2,3,4,5]*Statistics Study Program, Department of Mathematics, Faculty of Science and Technology, Universitas Airlangga*
Jln. Dr. Ir. H. Soekarno, Mulyorejo, Surabaya, 60115, Indonesia

*Corresponding author's e-mail: * rosidun.nindyo.pratiwi-2022@fst.unair.ac.id*

Article History:

Received: 1st July 2024
Revised: 30th January 2025
Accepted: 28th February 2025
Published: 1st April 2025

Keywords:

Crude Oil Production;
Least Square Spline Estimator;
Nonparametric Regression;
Solar Wholesale Price Index.

ABSTRACT

Solar plays a crucial role in supporting energy sector activities in Indonesia. The fluctuating price of solar is influenced by crude oil production, as crude oil is the main raw material in solar production. The Russia-Ukraine war, which reached its peak in March 2020, also impacted global oil production, given that Russia is one of the largest oil producers and exporters in the world. This study aims to model the effect of crude oil production on the Solar Wholesale Price Index (SWPI) in Indonesia after the Russia-Ukraine war using the Least Squares Spline estimator approach. This approach was chosen because the relationship between the variables is complex and nonlinear, making linear models unsuitable. The results show that the best model is a nonparametric model with three knot points at a polynomial degree of one, which explains 90.26% of the variability in crude oil production relative to the SWPI. The optimal knot points were selected using the Generalized Cross Validation (GCV) method, resulting in a minimum GCV value of 320.9889. Crude oil production was found to have a significant effect on the SWPI and meets the classical assumption tests. However, this study has limitations, as it only considers the effect of crude oil production without including other external factors, such as energy policies or geopolitical influences. Additionally, the model still has limitations in capturing more complex relationship patterns. This study offers an original contribution through the application of the Least Squares Spline estimator approach, which has not been widely used before in analyzing the relationship between crude oil production and SWPI in Indonesia. For future research, it is recommended that the model be expanded by considering more knot points and higher polynomial degrees to capture more complex relationship patterns between these variables.

How to cite this article:

R. N. Pratiwi, N. Fauziah, B. O. Syahputra, A. W. Firmanda and D. Amelia., "MODELING THE INFLUENCE OF CRUDE OIL PRODUCTION AGAINST INDONESIAN SOLAR WHOLESALE PRICE INDEX WITH LEAST SQUARE SPLINE ESTIMATOR APPROACH," *BAREKENG: J. Math. & App.*, vol. 19, iss. 2, pp. 0805-0818, June, 2025.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

Solar has a very important role in supporting Indonesia's energy sector activities as it is widely used as a fuel for transportation, industry, and various other needs [1]. However, diesel prices are volatile and often experience significant fluctuations. The Solar Wholesale Price Index is a key indicator to track changes in the wholesale price of diesel fuel in the market. The price of diesel fuel in Indonesia is always changing as in September 2022 the price of diesel fuel experienced a considerable increase from IDR 5.150 per liter to IDR 6.800 per liter or by 32.4% [2]. The increase in fuel prices in Indonesia is often caused by the rise in global oil prices and an imbalance between domestic demand and supply. This situation contributes to high inflation and public unrest due to the surge in the prices of essential goods [3]. Changes in diesel prices are influenced by global factors, especially crude oil production. As a derivative product of the petroleum refining process, diesel is sensitive to changes in crude oil production in the global energy market.

Crude oil, also known as petroleum, plays a crucial role in various sectors around the world. According to International Energy Statistics over the past six years, the United States has been the world's largest producer of crude oil. By 2023, the contribution from the United States, Russia, and Saudi Arabia will reach 40% of total global oil production, which is equivalent to 32.8 million barrels per day [4]. Crude oil is a non-renewable natural resource that is very important for every country to support economic development. However, the availability of crude oil is decreasing as time goes by. The high demand for crude oil results in shortages which cause price fluctuations [5].

Crude oil has an impact on diesel prices. Changes in crude oil production in the global energy market directly affect the cost of diesel production. Crude oil production cuts by OPEC (Organization of the Petroleum Exporting Countries) have helped stabilize the market, which in turn has caused oil prices to recover from a precipitous drop below $20 per barrel to a range of $60-$70 per barrel in 2021, affecting diesel prices [6]. However, in addition to crude oil production, diesel prices are also affected by other factors such as fluctuations in crude oil prices in the global market, government policies regarding subsidies, and political instability in major oil producers [7].

According to previous research on Parametric Estimation of Spline Regression Models, it shows that the results of the analysis using the Least Square Spline Estimator tryout scores affect the UNAS of SMKN 1 Samarinda students by 80.12% [8]. In addition, research on the Implications of Limiting Oil Production Quotas by the Organization Of The Petroleum Exporting Countries (OPEC) Joint Technical Committee Until April 2020 on the Oil Industry in Indonesia, shows that crude oil production has a significant effect on changes in diesel prices [9]. However, this research is limited to OPEC production. Thus, in this study the authors used global data to gain a broader understanding of the effect of crude oil production on the diesel fuel wholesale price index.

The advantage of this research method compared to previous similar studies lies in the use of nonparametric regression, especially the Least Square Spline Estimator approach. This method allows control of the smoothness of data that changes patterns at certain points by using knots and orders, without requiring assumptions about the form of the relationship between the variables used [10]. In previous studies, the parametric methods that are often used usually assume linearity or a predetermined form of relationship, which limits the model's ability to capture more complex relationship patterns [11]. In contrast, nonparametric regression is more flexible, so it is able to adjust the model to data that is nonlinear or changing. By focusing on clear data patterns, the use of knots and orders in this model allows for better optimization, so that it can provide a deeper and more accurate understanding of the relationship between crude oil production and the Solar Wholesale Price Index (SWPI).

This research is expected to be a reference material in the realization of Sustainable Development Goals (SDGs) point 7, namely regarding clean and affordable energy [12]. It is related because the production of crude oil and diesel in Indonesia fluctuates and the longer the petroleum supply must be depleted, it needs a concrete solution about finding alternative energy that is more affordable and renewable so that it can also reduce the impact of environmental pollution caused by diesel, especially related to air pollution and global warming originating from diesel-fueled vehicle emissions.

2. RESEARCH METHODS

2.1 Data Source

The data used in this study are secondary data obtained from the Central Bureau of Statistics and the Energy Information Administration. The data obtained is the production of crude oil and the wholesale price index of diesel fuel, especially Solar in Indonesia from April 2020 to December 2023 as many as 45 observations.

2.2 Research Variable

This study consists of one independent variable and one response variable. All research variables are presented in Table 1 below:

Table 1. Research Variables

Variable Type	Variable Name	Unit	Type
Independent ($\mathbf{x}$)	Crude Oil Production	Million barrels/day	Continuous
Response ($\mathbf{y}$)	Solar Wholesale Price Index	Index	Continuous

2.3 Analysis Technique

The research steps that will be carried out in modeling the effect of world crude oil production against solar wholesale price index using Rstudio and Minitab are as follows [13].

1. Describe the diesel wholesale price index and its influencing factor, namely crude oil production, with the following steps:
   a. Analyzing the size of the data based on mean, standard deviation, and range of each variable used.
   b. Creating a scatterplot between the response variable, namely solar wholesale price index and the predictor variable, namely crude oil production, to determine the pattern formed.
2. Modeling the solar wholesale price index based on crude oil production using the nonparametric regression method of least square spline estimator with the following steps:
   a. Using a combination of knot points to be used based on the plotting results of each predictor variable and response variable
   b. Selecting the optimal knot point based on the minimum Generalized Cross Validation (GCV) value
   c. Estimating the model parameters based on the optimal knot points
   d. Calculating the coefficient of determination to evaluate the goodness-of-fit of the model
   e. Evaluating the performance of the nonparametric model using the mean squared error (MSE) or other nonparametric goodness-of-fit criteria to assess the adequacy of the model
   f. Validating the residuals of the nonparametric regression model by analyzing the residual distribution and independence using diagnostic tools such as residual plots and normality tests, as these can still be applied to nonparametric models for assessing assumptions
   g. Formulating the best model equation
3. Analyze and interpret the nonparametric least square spline regression model of crude oil production on diesel fuel wholesale price index with the following steps.
   a. Plot the comparison between the model estimation results and the actual values
   b. Interpret the model formed based on the parameter values
   c. Draw conclusions and suggestions

2.3.1 Least Square Spline Estimator

Least squares is one of the most used methods for estimating regression model parameters. Since the principle of least squares is to minimize the error generated by the model, the regression model is expected to explain the data well. Given that the error generated by the regression model for each observation can include negative and positive values of , then to avoid adding negative values, find the sum of the squares of the error values [14].

The spline method, as one of the nonparametric regression approaches, is highly effective in capturing nonlinear and complex data patterns, as the spline function combines different polynomial segments over specific intervals to form a smooth curve [15]. In the least square spline estimator, there are polynomial cuts on different segments that are joined together at certain knots. The regression function $g(x_i)$ at order p and knot points $\tau_1, \tau_2, \dots, \tau_k$ can be expressed as follows:

$$g(x) = \sum_{j=0}^{p+k} \beta_j \varphi_j(x) \tag{1}$$

Where $\boldsymbol{\beta} = \left(\beta_o, \beta_1, \dots, \beta_{p+k}\right)^T$ is the parameter vector and φ_j can be defined as follows:

$$f(x) = \begin{cases} x^j, & for\ 0 \le j \le p \\ (x - \tau_{j-p})_+^p, & for\ p+1 \le j \le p+k \end{cases} \tag{2}$$

With

$$(x - \tau_{j-p})_+^p = \begin{cases} (x - \tau_{j-p})^p & for\ x \ge \tau_{j-p} \\ 0 & for\ x < \tau_{j-p} \end{cases} \tag{3}$$

If given $\lambda = (\tau_1, \tau_2, \dots, \tau_k)$ then the estimator of g with order p and the number of knots is k, namely $g_\lambda(x)$ can be expressed as follows:

$$g_\lambda(x) = \sum_{j=0}^{p+k} \beta_{\lambda j}\, \varphi_j(x) \tag{4}$$

The function expressed in **Equation (4)** is called a spline function. A least square spline is a polynomial cut with different polynomial segments joined together at knot points $\tau_1, \tau_2, \dots, \tau_k$ while still ensuring continuity.

Furthermore, if n paired samples are taken, the spline function in **Equation (4)** can be written as follows:

$$\left.\begin{aligned} g(x_1) &= \beta_0(x_1)^0 + \beta_1(x_1)^1 + \cdots + \beta_p(x_1)^p + \beta_{(p+1)}(x_1 - \tau_1)_+^p + \cdots + \beta_{(p+k)}(x_1 - \tau_k)_+^p \\ g(x_2) &= \beta_0(x_2)^0 + \beta_1(x_2)^1 + \cdots + \beta_p(x_2)^p + \beta_{(p+1)}(x_2 - \tau_1)_+^p + \cdots + \beta_{(p+k)}(x_2 - \tau_k)_+^p \\ &\ \vdots \\ g(x_n) &= \beta_0(x_n)^0 + \beta_1(x_n)^1 + \cdots + \beta_p(x_n)^p + \beta_{(p+1)}(x_n - \tau_1)_+^p + \cdots + \beta_{(p+k)}(x_n - \tau_k)_+^p \end{aligned}\right\} \tag{5}$$

The spline function in **Equation (5)** can be written in matrix notation as follows:

$$\begin{pmatrix} g(x_1) \\ g(x_2) \\ \vdots \\ g(x_n) \end{pmatrix} = \begin{pmatrix} 1 & x_1^1 & x_1^2 & \cdots & x_1^p & (x_1 - \tau_1)_+^p & \cdots & (x_1 - \tau_k)_+^p \\ 1 & x_2^1 & x_2^2 & \cdots & x_2^p & (x_2 - \tau_1)_+^p & \cdots & (x_2 - \tau_k)_+^p \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_n^1 & x_n^2 & \cdots & x_n^p & (x_n - \tau_1)_+^p & \cdots & (x_n - \tau_k)_+^p \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{(p+k)} \end{pmatrix} \tag{6}$$

So obtained $\boldsymbol{g}(\boldsymbol{x}) = \mathbf{X}_\lambda \boldsymbol{\beta}_\lambda$, with :

$$\boldsymbol{X}_\lambda = \begin{pmatrix} 1 & x_1^1 & x_1^2 & \cdots & x_1^p & (x_1-\tau_1)_+^p & \dots & (x_1-\tau_k)_+^p \\ 1 & x_2^1 & x_2^2 & \cdots & x_2^p & (x_2-\tau_1)_+^p & \cdots & (x_2-\tau_k)_+^p \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n^1 & x_n^2 & \cdots & x_n^p & (x_n-\tau_1)_+^p & \cdots & (x_n-\tau_k)_+^p \end{pmatrix} \tag{7}$$

and

$$\boldsymbol{\beta}_\lambda = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{(p+k)} \end{pmatrix} \tag{8}$$

Therefore, the estimator for $\boldsymbol{g}(\boldsymbol{x})$, namely $\widehat{\boldsymbol{g}}_\lambda(\boldsymbol{x})$ can be written as follows:

$$\widehat{\boldsymbol{g}}_\lambda(\boldsymbol{x}) = \mathbf{X}_\lambda \widehat{\boldsymbol{\beta}}_\lambda \tag{9}$$

The estimator for $\boldsymbol{\beta}_\lambda$ namely $\widehat{\boldsymbol{\beta}}_\lambda$, is obtained using the least square method by determining the value of $\widehat{\boldsymbol{\beta}}_\lambda$ that minimizes the sum of squared errors as follows:

$$\hat{\boldsymbol{\varepsilon}}^{\boldsymbol{T}}\hat{\boldsymbol{\varepsilon}} = \left(\boldsymbol{y} - \mathbf{X}_\lambda\widehat{\boldsymbol{\beta}}_\lambda\right)^{\boldsymbol{T}}\left(\boldsymbol{y} - \mathbf{X}_\lambda\widehat{\boldsymbol{\beta}}_\lambda\right) \tag{10}$$

From this is obtained:

$$\widehat{\boldsymbol{\beta}}_\lambda = \left(\mathbf{X}_\lambda^{\boldsymbol{T}}\mathbf{X}_\lambda\right)^{-\mathbf{1}}\mathbf{X}_\lambda^{\boldsymbol{T}}\mathbf{y} \tag{11}$$

Based on **Equation (9)** and **Equation (11)**, we obtain:

$$\widehat{\boldsymbol{g}}_\lambda(\boldsymbol{x}) = \mathbf{X}_\lambda\left(\boldsymbol{X}_\lambda^{\boldsymbol{T}}\boldsymbol{X}_\lambda\right)^{-\mathbf{1}}\mathbf{X}_\lambda^{\boldsymbol{T}}\mathbf{y} \tag{12}$$

Equation (12) can be written as follows:

$$\widehat{\boldsymbol{g}}_\lambda(\mathbf{x}) = \boldsymbol{A}(\lambda)\boldsymbol{y} \tag{13}$$

With

$$\boldsymbol{A}(\lambda) = \mathbf{X}_\lambda\left(\mathbf{X}_\lambda^{\boldsymbol{T}}\mathbf{X}_\lambda\right)^{-\mathbf{1}}\mathbf{X}_\lambda^{\boldsymbol{T}} \tag{14}$$

In general regression, this matrix $\boldsymbol{A}(\lambda)$ is commonly referred to as the Hessian matrix. From **Equation (14)**, we can conclude that$\widehat{\boldsymbol{g}}_\lambda(\mathbf{x})$ is a linear estimator (linear to the observations) for $\boldsymbol{g}(\boldsymbol{x})$.

2.3.2 Model Goodness Measures

In this study, the authors used two measures, namely MSE and the coefficient of determination (R^2). The estimated value of the model, expressed as the average squared error, can be measured by MSE. The coefficient of determination (R^2) is used to determine how much variation can be explained by the model. MSE is expected to have the minimum value possible, while the coefficient of determination (R^2). will be better if the value is closer to 1 **[16]**.

The formula for MSE is as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n} e_i^2 \tag{15}$$

Meanwhile, the formula for the coefficient of determination (R^2) is as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{16}$$

2.3.3 Optimum Knot Point Selection

A knot point is a point of fusion or intersection where there is a change in data behavior [17]. The best nonparametric regression model with the Least Square Spline estimator depends on the optimal knot points. The selection of optimal knot points in the Least Square Spline estimator can use the Generalized Cross Validation (GCV) method. The GCV function for the nonparametric spline regression model is as follows:

$$GCV(\lambda) = \frac{MSE(\lambda)}{\left(\frac{1}{n} tr[\boldsymbol{I} - \boldsymbol{H}(\lambda)]\right)^2} \tag{17}$$

with $\boldsymbol{I}$ is the identity matrix and $\boldsymbol{H}(\lambda)$ is the hessian matrix or Hat Matrix. The formula for MSE is as follows:

$$MSE(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{g}(x_i)\right)^2 \tag{18}$$

A spline regression model has an optimal knot point if it meets the criteria that the value of GCV and MSE is minimum.

2.3.4 Parameter Significance Testing

Model parameter testing is carried out to determine whether the model of a predictor variable significantly affects the response variable. The following is the hypothesis used based on the spline nonparametric regression model.

$\mathrm{H}_0: \beta = 0$ (there is no significant effect of predictor variables on the response variable)
$\mathrm{H}_1: \beta \neq 0$ (there is a significant effect of predictor variables on the response variable)

The test statistic used in this test is the F test statistic which has the following formula.

$$F = \frac{MS_{regression}}{MS_{residual}} \tag{19}$$

With

$$\begin{aligned} MS_{regression} &= \frac{SS_{regression}}{df_{regression}} = \frac{\sum_{i-1}^{n}(\hat{y}_i - \bar{y})^2}{p} \\ MS_{residual} &= \frac{SS_{residual}}{df_{residual}} = \frac{\sum_{i-1}^{n}(\hat{y}_i - \bar{y})^2}{n-p+1} \end{aligned} \tag{20}$$

Where p is the number of parameters in the nonparametric spline regression model, with n is the amount of data.

2.3.5 Residual Assumption Test

Testing the residual assumptions (goodness of fit) of the regression model is carried out to determine whether the resulting residuals have met the assumptions of identical, independent, and normally distributed (IIDN).

1. Identical Residual Assumption Test

One of the assumption tests met in the regression model is that the variance of each error has a constant value or is equal to σ^2. This variance equality is called identical. If the identical assumption is not met, it is called a case of heteroscedasticity. In testing, the detection of heteroscedasticity cases can be done using the Glejser test [18]. The hypothesis used in the Glejser test is as follows.

$\mathrm{H}_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2 = \sigma^2$

H_1 : At least one $\sigma_i^2 \neq \sigma^2, i = 1,2, \dots, n$

The test statistics used are as follows:

$$F = \frac{\frac{\left[\sum_{i=1}^{n}(|\hat{\varepsilon}_i| - |\bar{\varepsilon}|)^2\right]}{v-1}}{\frac{\left[\sum_{i=1}^{n}(|\hat{\varepsilon}_i| - |\bar{\varepsilon}|)^2\right]}{(n-v)}} \tag{21}$$

where the value of v is the number of parameters of the Glejser model and for the nonparametric least square spline regression model the value of $v = m + r$. The rejection area used is reject H_0 if F is greater that $F_{\alpha(m+r),n-(m+r)-1}$ or $p - value < \alpha$, where the value of m is the Glejser model parameter of the polynomial degree, r is the truncated component parameter and n is the number of observations. The conclusion obtained when rejecting H_0 is that there is at least one $\sigma_i^2 \neq \sigma^2$. This indicates that there is a case of heteroscedasticity so that the identical assumption is not met.

2. Independent Residual Assumption Test

The independent residual assumption is the assumption that there is no correlation between residuals or autocorrelation. One way to detect autocorrelation is to do the Durbin Watson [18]. The following is the hypothesis in the Durbin Watson test.

$\mathrm{H}_0 : \rho = 0$ (residuals are independent or there is no autocorrelation)
$\mathrm{H}_1 : \rho \neq 0$ (residuals are not independent or autocorrelation occurs)

The test statistics used in the Durbin Watson test are as follows:

$$d = \frac{\sum_{i=1}^{n}(\varepsilon_i - \varepsilon_{i-1})^2}{\sum_{i=1}^{n}\varepsilon_t^2} \tag{22}$$

The rejection area is divided into several parts and can be seen in **Table 2** as follows.

Table 2. Durbin-Watson Test Rejection Area

Null Hypothesis	If	Decision
No positive autocorrelation	$0 < d < d_L$	Reject
No positive autocorrelation	$d_L \leq d \leq d_U$	No Decision
No negative autocorrelation	$4 - d_U \leq d \leq 4$	Reject
No negative autocorrelation	$4 - d_U \leq d \leq 4 - d_L$	No Decision
No positive or negative autocorrelation	$d_U \leq d \leq 4 - d_U$	Fail to reject

3. Normal Distributed Residual Assumption Test

The method that can be used in the data normality test is the Kolmogorov-Smirnov test. The working principle of the Kolmogorov-Smirnov test is to compare the cumulative frequency of the theoretical distribution with the cumulative frequency of the empirical distribution [19]. The hypothesis for this test is as follows:

H_0: Residuals are normally distributed
H_1: Residuals are not normally distributed

The test statistic used is as follows:

$$D = Sup|F_n(\varepsilon) - F_0(\varepsilon)| \tag{23}$$

The rejection area for the test is to reject H_0 if $D > D_\alpha$. Where D_α s the critical value for the one-sample Kolmogorov Smirnov test, obtained from the one-sample Kolmogorov Smirnov table. $F_n(\varepsilon)$ is the cumulative probability value based on sample data, $F_0(\varepsilon)$ is the cumulative probability value under H_0. If the normal distribution assumption is not met, several ways of data transformation can be done such as natural logarithm transformation, square root, inverse, and others [18].

3. RESULTS AND DISCUSSION

3.1 Descriptive Analysis of Data

To find out an overview of Crude Oil Production (COP) and Solar Wholesale Price Index (SWPI), a descriptive analysis is carried out which is presented in Table 3 below:

Table 3. Descriptive Analysis Result

Variable	N	Mean	St. Dev	Minimum	Maximum
COP	45	78.728	3.593	70.280	82.920
SWPI	45	141.11	51.60	62.76	211.92

Based on Table 3, the interpretation of descriptive statistics is follows:

1. In the crude oil production variable with 45 data, the average crude oil production is 78.728 million barrels/day and the standard deviation value is 3.593 which is smaller than the average value. This shows that crude oil production has a small diversity. Then, it is known that the minimum production is 70.280 million barrels/day which occurred in May 2020 and the maximum production is 82.90 million barrels/day which occurred in December 2023.
2. In the solar wholesale price index variable with a total of 45 data, the average is 141.11 and the standard deviation value is 51.60 which is smaller than the average value. This shows that SWPI diesel has a small diversity. Then, it is known that the minimum index is 62.76 which occurred in May 2020 and the maximum index is 211.92 which occurred in December 2022.

3.2 Identification of Variable Relationship Patterns

Identifying the relationship pattern between the independent and response variables is a necessary step to determine the regression model estimation method. The relationship pattern between variables can be visualized using a scatter plot. Regression modeling can be determined through pattern identification, namely using parametric regression if the scatter plot has a quadratic, linear, or other parametric pattern and using nonparametric regression if the scatter plot does not show a certain pattern.

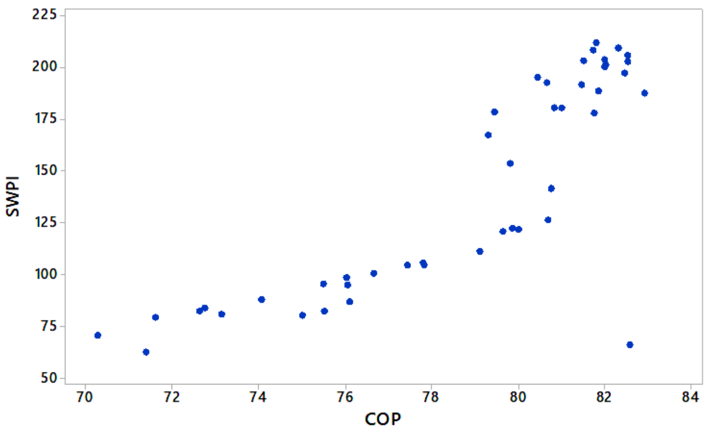

Figure 1. Scatterplot Data

Based on Figure 1, identification of the point distribution pattern on the resulting scatterplot shows that the distribution of crude oil production data against the production index of large-scale trade in diesel fuel, especially Industrial Solar in Indonesia, is not very linear and does not form a certain pattern. Therefore, the model estimation can be solved using nonparametric regression with the least square spline estimator approach because there is a change in behavior at certain intervals.

3.3 Nonparametric Regression Analysis with Least Square Spline Estimator

Nonparametric regression model with the least square spline estimator approach has smoothing parameters in the form of knot points and order. Knot point is a combination point where there is a change in the pattern of data behavior. The best model selection is taken from the model that has the minimum GCV value, smaller the GCV value so that `the knot point chosen will be more optimum. The knot points used in this study are one, two, and three and polynomial orders 1 and 2 which are processed with R software.

The results of nonparametric regression analysis with the least square spline estimator approach of polynomial order 1 and 2 obtained output MSE, GCV, and R^2 values in **Table 4** as follow:

Table 4. Model Calculation with Least Square Spline Estimator

Polynomial Order	Number of Knots	Knot Point	MSE	GCV	R^2
1	1	77.36	688.4504	790.313	0.735524
1	1	78.83	694.0826	796.7785	0.7333603
2	1	80.75	627.1411	755.4793	0.7590766
2	1	81.88	601.4361	724.5141	0.7689515
1	2	77.36 and 82.53	656.0259	790.2751	0.7479802
1	2	82.45 and 82.59	635.3908	765.4173	0.7559074
2	2	78.42 and 82.55	381.9829	483.4471	0.8532569
2	2	82.45 and 82.59	301.688	381.8239	0.8841031
1	**3**	**77.83, 82.53, and 82.59**	**253.6209**	**320.9889**	**0.9025687**
1	3	77.96, 82.53, and 82.68	253.6656	321.0455	0.9025515
2	3	77.83, 82.53, and 82.59	262.0112	348.8314	0.8993454
2	3	77.96, 82.53, and 82.68	262.2236	349.1143	0.8992638

Based on **Table 4**, it can be seen the minimum GCV is generated by 3 knot points at polynomial order 1 with a GCV value of 320.99. This model shows that the predictor variables were able to explain the response variable by 90.26%. Then the regression model equation can be estimated in **Table 5** as follows.

Table 5. Model Estimation Value

$\hat{\beta}$	Value
0	-259.4545
1	4.654996
2	18.26302
3	-2428.393
4	2773.808

Based on **Table 5**, the estimated value of Crude Oil Production (COP) model against the Indonesian Solar Wholesale Price Index (SWPI) with least square spline approach using order one and three knot points can be written through the following equation.

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 (x - \tau_1)_+ + \hat{\beta}_3 (x - \tau_2)_+ + \hat{\beta}_4 (x - \tau_3)_+$$
$$\hat{y} = -259.4545 + 4.654996x + 18.26302(x - 77.83)_+ - 2428.393(x - 82.53)_+ + 2773.808(x - 82.59)_+$$

Based on the definition of Equation ($\hat{y}$), then the equation can be written as follows:

$$\hat{y} = \begin{cases} -259.4545 + 4.654996x. \ x < 77.83 \\ -1680.865 + 22.918x. \ 77.83 \le x < 82.53 \\ 198734.409 - 2405.475x. \ 82.53 \le x < 82.59 \\ -30354.395 + 368.333x. \ x \ge 82.59 \end{cases}$$

3.3.1 Significance Testing of Model Parameters

Testing the model parameters to determine whether the parameters of the predictor variables have a significant effect on the Solar Wholesale Trade Production Index. The results are as follows.

Table 6. ANOVA Significance Testing of Model

Source	Degrees Of Freedom	Sum Of Square	Mean Square	F
Regression	5	99383.12	16563.85	36.38
Error	39	17755.17	455.26	
Total	44	117138.3		

Based on Table 6, it is known that the resulting F is 36.38. By comparing the F to $F_{0.05;5;39}$ which is 2.46, the decision to reject H_0 is made. This means that crude oil production has a significant effect against Indonesian Solar Wholesale Price Index.

3.3.2 Residual Assumption Test

1. Identical Residual Assumption Test

The identical assumption test is used to determine whether the residual variance is homogeneous or there is no case of heteroscedasticity. To determine whether there are symptoms of heteroscedasticity, the Glejser test is used. The Glejser test is performed by regressing the predictor variables on the absolute value of the residuals.

Table 7. Identical Residual Test

Term	F	p-value
Crude Oil Trade	2.137	0.093

Based on Table 7, the p-value = 0.093 is greater than the significance level (0.05), so the decision fails to reject H_0 so it can be concluded that the heteroscedasticity test assumption or the identical residual assumption is met.

2. Independent Residual Assumption Test

Independent assumption testing is used to determine whether there is no autocorrelation between residuals. Independent assumption testing is done using the Durbin-Watson test as follows.

Table 8. Independent Residual Test

d	$d_{L(0.05)}$	$d_{U(0.05)}$	$4-d_{U(0.05)}$	$4-d_{L(0.05)}$
2.0805	1.4754	1.5660	2.4224	2.4340

Based on Table 8, it is obtained that d_{value} generated in the Durbin-Watson test is 2,0805. If d compared with $d_{U(0,05)}$ dan $4-d_{U(0,05)}$, it can be seen that $d_{U(0,05)} < d < 4-d_{U(0,05)}$. This gives a decision to fail reject H_0 so it can be concluded that there are no symptoms of autocorrelation in the residual.

3. Normal Distribution Residual Assumption Test

After the independent and identical assumptions are met, there is still one more assumption that needs to be met by the residuals, namely the assumption of normal distribution. One test that can be done to determine the normality of the data is the Kolmogorov-Smirnov, obtained the following result.

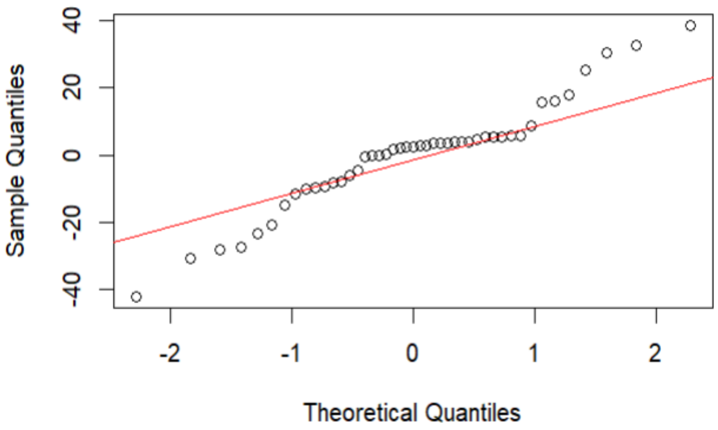

Figure 2. Normality Residual Plot

Based **Figure 2**, the p-value = 0.0819 is obtained which is greater than the significance level (0.05), so the decision to fail to reject H_0 is obtained so it can be concluded that the residuals are normally distributed so that the assumption of residual normality is met.

3.4 Interpretation the Best Model of Crude Oil Production (COP) Against Indonesian Solar Wholesale Price Index (SWPI)

The plot of estimated value and observations from the influence Crude Oil Production (COP) against the Indonesian Solar Wholesale Price Index (SWPI) with least square spline approach presented in **Figure 3** bellow:

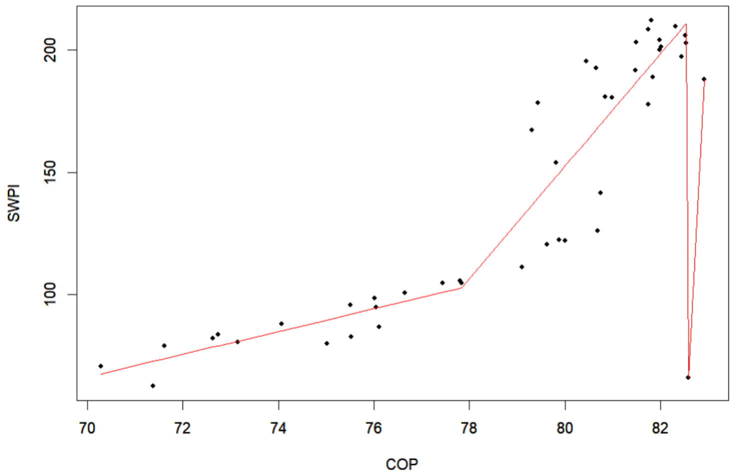

Figure 3. Value Plot Estimates and Observations

The best nonparametric regression model with the least square spline estimator approach to crude oil production against Indonesian solar wholesale price index is a model with polynomial order 1 and has 3 knot points. The following are the results of the interpretation of the best model that has been determined previously.

1. **Domain $x < 77.83$**

   When crude oil production is less than 77.83 million barrels per day, then every increase in crude oil production of 1 million barrels per day will result in an increase in SWPI diesel by 4.654996 units. The current increase in diesel prices is a result of the Russia-Ukraine war that began in March 2020. In addition, the COVID-19 pandemic has caused diesel prices to increase significantly despite the increase in crude oil production. Due to the Supply Chain Disruption phenomenon, the pandemic caused substantial disruption to the global supply chain affecting the transportation and availability of refined petroleum products such as diesel [20]. These lockdowns and restrictions cause logistical challenges that increase the costs associated with the distribution of diesel, which in turn increases the price of diesel.

2. **Domain $77.83 \le x < 82.53$**

When crude oil production is 77.83 million barrels per day to less than 82.53 million barrels per day, then every increase in crude oil production of 1 million barrels per day will result in an increase in diesel SWPI of 22.918 units. This increase in crude oil production occurs because in the time period of 77.83 to 82.53, or to be precise in July 2021 to February 2022, the increase in production in major countries such as countries such as the United States, Canada, and Brazil is projected to reach the highest annual production, increasing non-OPEC+ production by 1.8 million barrels per day [4]. However, there is a problem that diesel refining capacity is limited in oil-producing countries, even though crude oil production increases, there is not enough refining capacity to process all the crude oil production into diesel, which results in the price of diesel also increasing.

3. **Domain $82.53 \le x < 82.59$**

When crude oil production is 82.53 million barrels per day to less than 82.59 million barrels per day, then every increase in crude oil production of 1 million barrels per day will result in a decrease in diesel SWPI by 2405.475 units. This increase in oil production occurred because in the time period of 82.53 to 82.59, or to be precise in February 2022 to December 2023, increased production from countries such as the US, Brazil, Guyana, and Iran contributed to an increase in global supply. Although OPEC+ voluntarily reduced production, increased production from non-OPEC+ countries was able to make up for this reduction, resulting in an oversupply. With more oil available, the prices of crude oil and its derivatives including diesel prices are likely to decline [4].

4. **Domain $x \ge 82.59$**

When crude oil production is 82.59 million barrels per day or more, every additional 1 million barrels per day of crude oil production will result in an increase in diesel SWPI of 368.333 units. This is a result of the post-COVID-19 pandemic, after lockdowns and pandemic restrictions were relaxed, there was a surge in demand for energy including diesel [21]. So that the increase in diesel prices was inevitable even though major oil-producing countries such as Saudi Arabia and Russia had increased crude oil production [22].

4. CONCLUSIONS

Based on nonparametric regression modeling with the Least Square Spline estimator approach, using polynomial orders of 1 and 3 knots, this model explains 90.26% of the variability in crude oil production relative to the Solar Wholesale Price Index (SWPI) in Indonesia. The selection of optimal knots was carried out using the GCV method, resulting in a minimum GCV value of 320.9889. The test results show that crude oil production significantly affects the SWPI and meets the classical assumption tests. However, this model still has limitations in capturing complex patterns and the influence of other variables outside the model, such as energy policies or geopolitical influences. For future research, it is recommended to expand the model by considering more knot points and various polynomial orders to capture more complex relationship patterns between crude oil production and SWPI.

REFERENCES

[1] A. Widianto, “UJI KEMAMPUAN CAMPURAN BAHAN BAKAR SOLAR-BIODIESEL DARI MINYAK BIJI JARAK TERHADAP UNJUK KERJA DAN OPOSITAS MESIN DIESEL 4 LANGKAH,” *Jurnal Teknik. Mesin*, vol. 03, no. May, pp. 38–46, 2014.
[2] V. B. Kusnandar, “PEMERINTAH NAIKKAN HARGA SOLAR, PERTALITE, DAN PERTAMAX (RON 92) MULAI 3 SEPTEMBER 2022,” katadata.co.id. Accessed: May 25, 2024. [Online]. Available: https://databoks.katadata.co.id/energi/statistik/27a4de9db0d4024/pemerintah-naikkan-harga-solar-pertalite-dan-pertamax-ron-92-mulai-3-september-2022
[3] A. B. Astuti, A. N. Guci, V. Iqbal, A. Alim, and L. N. Azizah, “NON HIERARCHICAL K-MEANS ANALYSIS TO

CLUSTERING PRIORITY DISTRIBUTION OF FUE; SUBSIDIES IN INDONESIA," *BAREKENG Journal of Mathematics and Its Applications.*, vol. 17, no. 3, pp. 1663–1672, 2023, doi: https://doi.org/10.30598/barekengvol17iss3pp1663-1672 NON.

[4] International Energy Agency, "OIL MARKET REPORT," iea.org. Accessed: May 25, 2024. [Online]. Available: https://www.iea.org/analysis?type=report.

[5] A. K. Rahma, Q. Z. Abidin, J. K. Prasetyo, D. Amelia, and N. Chamidah, "PREDICTION OF CRUDE OIL PRICES IN INDONESIA USING FOURIER SERIES ESTIMATOR AND ARIMA METHOD," *BAREKENG Journal of Mathematics and Its Applications.*, vol. 18, no. 3, pp. 1673–1682, 2024, doi: https://doi.org/10.30598/barekengvol18iss3pp1673-1682.

[6] Q. Zhang, Y. Hu, J. Jiao, and S. Wang, "THE IMPACT OF RUSSIA–UKRAINE WAR ON CRUDE OIL PRICES: AN EMC FRAMEWORK," *Humanities and Social Sciences Communication*, vol. 11, no. 1, 2024, doi: 10.1057/s41599-023-02526-9.

[7] P. P. Kartia, N. P. Hapsari, A. G. Nuswantoro, and H. B. Pamungkas, "ANALISIS DAMPAK KENAIKAN BBM TERHADAP BIAYA TRANSPORTASI PERDAGANGAN INTERNASIONAL," *Digital Bisnis: Jurnal Publikasi Ilmu Manajemen dan E-Commerce*, vol. 2, no. 4, pp. 439–453, 2023.

[8] M. Faturahman, "ESTIMASI PARAMETER MODEL REGRESI SPLINE," *Jurnal Eksponensial*, vol. 2, no. 1, pp. 55–58, 2011.

[9] Y. Rahmah, "IMPLIKASI PEMBATASAN KUOTA PRODUKSI MINYAK OLEH ORGANIZATION OF THE PETROLEUM EXPORTING COUNTRIES (OPEC) JOINT TECHNICAL COMITTEE HINGGA APRIL 2020 TERHADAP INDUSTRI MINYAK DI INDONESIA ".," Universitas Satya Negara Indonesia, 2020.

[10] F. K. S. Sitohang, S. Sifriyani, and S. Mahmuda, "APPLICATION OF NONPARAMETRIC REGRESSION SPLINE TRUNCATED FOR MODELING THE HEIGHT OF YEOP CHAGI KICKS OF TAEKWONDO ATHLETES IN SAMARINDA CITY," *BAREKENG: Journal of Mathematics and Its Applications*, vol. 18, no. 2, pp. 0657–0666, 2024, doi: 10.30598/barekengvol18iss2pp0657-0666.

[11] A. Islamiyati *et al.*, "THE USE OF PENALIZED WEIGHTED LEAST SQUARE TO OVERCOME CORRELATIONS BETWEEN TWO RESPONSES," *BAREKENG Journal of Mathematics and Its Applications*, vol. 16, no. 4, pp. 1497–1504, 2022, doi: 10.30598/barekengvol16iss4pp1497-1504.

[12] V. L. Trinh and C. K. Chung, "RENEWABLE ENERGY FOR SDG-7 AND SUSTAINABLE ELECTRICAL PRODUCTION, INTEGRATION, INDUSTRIAL APPLICATION, AND GLOBALIZATION," *Cleaner Engineering and Technology*, vol. 15, p. 100657, 2023.

[13] N. Pigitha, A. Djuraidah, and A. H. Wigena, "APPLICATION OF PENALIZED SPLINE-SPATIAL AUTOREGRESSIVE MODEL TO HIV CASE DATA IN INDONESIA," *BAREKENG: Journal of Mathematics and Its Applications.*, vol. 17, no. 1, pp. 527–534, 2023, doi: https://doi.org/10.30598/barekengvol17iss1pp0527-0534.

[14] R. Marchelina, N. N. Bakar, and E. Asfa'ani, "KARAKTERISTIK SOLUSI KUADRAT TERKECIL," *MAp (Mathematics and Applications) Journal*, vol. 5, no. 2, pp. 127–136, 2023, doi: https://doi.org/10.15548/map.v5i2.7045.

[15] S. Sifriyani, A. T. R. Dani, M. Fauziyah, and Z. Mar'ah, "MIXED ESTIMATORS OF TRUNCATED SPLINE-EPANECHNIKOV KERNEL ON NONPARAMETRIC REGRESSION AND ITS APPLICATIONS," *BAREKENG Journal of Mathematics and Its Applications.*, vol. 17, no. 4, pp. 2023–2032, 2023, doi: 10.30598/barekengvol17iss4pp2023-2032.

[16] M. Maharani and D. R. S. Saputro, "GENERALIZED CROSS VALIDATION (GCV) IN SMOOTHING SPLINE NONPARAMETRIC REGRESSION MODELS," in *Journal of Physics: Conference Series IOP Publishing*, IOP Publishing, 2021, p. 12053. doi: 10.1088/1742-6596/1808/1/012053.

[17] L. P. S. Pratiwi, "PEMILIHAN TITIK KNOT OPTIMAL MODEL SPLINE TRUNCATED DALAM REGRESI NONPARAMETRIK MULTIVARIABEL DENGAN GCV," *Jurnal Matematika.*, vol. 10, no. 2, pp. 78–90, 2020, doi: 10.24843/JMAT.2017.v07.i01.p77.

[18] E. Y. Sutopo and A. Slamet, STATISTIK INFERENSIAL. Yogyakarta, Indonesia: Penerbit Andi, 2017.

[19] A. Nasrum, UJI NORMALITAS UNTUK PENELITIAN. Denpasar, Indonesia: Jayapangus Press Books, 2018.

[20] F. Sahid and R. R. Fauzy, "DAMPAK KENAIKAN HARGA BBM TERHADAP RBP DI KONTRAKTOR X KARENA KELANGKAAN BBM DI PANDEMIK COVID-19," *J-MAS (Jurnal Manajemen dan Sains)*, vol. 8, no. 2, pp. 1298–1302, 2023.

[21] Kemeneterian ESDM, "COVID-19 MELUAS DAN LOCKDOWN NEGARA KONSUMEN MINYAK, ICP MARET 2020 TERTEKAN LAGI JADI US$ 34,23 PER BAREL," Ditjen Migas. Accessed: Jun. 01, 2024. [Online]. Available: https://migas.esdm.go.id/post/read/covid-19-meluas-dan-lockdown-negara-konsumen-minyak-icp-maret-2020-tertekan-lagi-jadi-us-34-23-per-barel

[22] R. Ruchuan, T. Xiong, and Y. Bao, "THE RUSSIA-SAUDI ARABIA OIL PRICE WAR DURING THE COVID-19 PANDEMIC RICHIE," *Energy Economic.*, vol. 102, no. January, 2021, doi: https://doi.org/10.1016/j.eneco.2021.105517.