Parameter: Journal of Statistics
Available at https://bestjournal.untad.ac.id/index.php/parameter
Vol. 2 No. 2 2022, 30 - 36
DOI: *https://doi.org/10.22487/27765660.2022.v2.i2.15913*
Journal of Statistics
eISSN: 2776-5660

COMPARISON OF COCHRANE-ORCUTT AND HILDRETH-LU METHODS TO OVERCOME AUTOCORRELATION IN TIME SERIES REGRESSION (CASE STUDY OF GORONTALO PROVINCE HDI 2010-2021)

Khusnudin Tri Subhi[1*], Azka Al Azkiya[2]
[1]Badan Pusat Statistik Kabupaten Sigi
[2]Department of Statistics, Faculty of Mathematics and Natural Sciences, IPB University

***e-mail**: *tri.subhi@bps.go.id*

ABSTRACT

Time series data is data that is prone to autocorrelation. Autocorrelation is a violation of assumptions in Ordinary Least Square regression. The presence of autocorrelation can make parameter estimates, not BLUE (Best, Linear, Unbiased Estimator). Several methods to overcome autocorrelation include Cochrane-Orcutt and Hildreth-Lu methods. Therefore, this study aimed to compare the Cochrane-Orcutt and Hildreth-Lu methods to deal with autocorrelation in the time series regression of the Gorontalo Human Development Index case in 2010 2021. We used HDI data for Gorontalo Province from 2010-to 2021, taken from the BPS-Statistics Indonesia Gorontalo Province. The method we used was Cochrane-Orcutt and Hildreth-Lu in the case of regression using Ordinary Least Squares (OLS) parameter estimation. The results obtained are that the Cochrane-Orcutt and Hildreth-Lu could overcome autocorrelation. The results of the Durbin Watson test after using both methods show no autocorrelation. However, the Hildreth-Lu method resulted in a lower Root Mean Square Error (RMSE) of 0.147 compared to the RMSE of the OLS model of 0.165 and the RMSE of the Cochrane-Orcutt model of 0.196. Therefore, the Hildreth-Lu method was the best method to overcame autocorrelation in this case.

Keywords: *autocorrelation, Cochrane-Orcutt, Hildreth-Lu, regression, time series*.

Cite: Subhi, K. T., & Azkiya, A. A., (2022). Comparison of Cochrane-Orcutt and Hildreth-Lu Methods to Overcome Autocorrelation in Time Series Regression (Case Study of Gorontalo Province HDI 2010-2021). *Parameter: Journal of Statistics*, *2*(2), 30-36. https://doi.org/10.22487/27765660.2022.v2.i2.15913

*Received: May 2022; Reviewed: May 2022; Published: June 2022

INTRODUCTION

Autocorrelation is a violation of assumptions in regression. In time series regression, autocorrelation events are common. However, autocorrelation can make parameter estimates, not BLUE (Best, Linear, Unbiased Estimator). Therefore, handling autocorrelation becomes very important. A previous study discussed handling autocorrelation was done by (Aprianto, 2020). This research tells us about the Cochrane-Orcutt Method to Overcome Autocorrelation in Estimating Ordinary Least Squares Parameters. In this study, a comparison of two methods of handling autocorrelation was carried out, namely Cochrane-Orcutt and Hildreth-Lu.

The case study used in this study is the HDI of Gorontalo Province because Gorontalo Province is a province with HDI that continues to increase even in the Covid-19 pandemic; namely, in 2020, it increased by 0.19 points from 2019, and in 2021 there was an increase of 0.32 points compared to 2020 (BPS, 2021). According to UNDP, the Human Development Index (HDI) measures human development achievements based on several fundamental components of quality of life. HDI is built through a basic three-dimensional approach to measuring the quality of life. These dimensions include longevity and health, knowledge, and a decent life. According to (Asnidar, 2018), advanced human development determines the ability of the population to absorb and manage sources of economic growth, both in terms of technology and institutions, as an essential means to achieve economic growth.

Therefore, the purpose of this study was to compare the Cochrane-Orcutt and Hildreth-Lu methods to deal with autocorrelation in the time series regression of the Gorontalo HDI case in 2010-2021.

MATERIALS AND METHODS

The data used in this study comes from the Statistics Gorontalo Province, namely Gorontalo Province HDI data from 2010 to 2021 (BPS, 2021). The method used in this study is linear regression or ordinary least squares and the Cochrane-Orcutt method, and the Hildreth-Lu method to overcome the autocorrelation problem.

1. The Method of Least Squares

Ordinary least squares is a method that minimizes the sum of squares of the observed deviations. The sum of the squares of the deviations is commonly called the sum of squares of error (SSE). It is defined as (Mendenhal et al., 2012):

$$SSE = \sum (y_i - \hat{y}_i)^2 = \sum (y_i - a - bx_i)^2 \tag{1}$$

Least-squares estimators of α and β

$$b = \frac{S_{xy}}{S_{xx}} \text{ and } \alpha = \bar{y} - b\bar{x} \tag{2}$$

Where the quantities S_{xy} and S_{xx} are defined as

$$S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n} \tag{3}$$

$$S_{xx} = \sum (x_i - \bar{x})^2 = \sum x_i{}^2 - \frac{(\sum x_i)^2}{n} \tag{4}$$

For the method to see the model's accuracy, we use Root Mean Square Error (RMSE). The RMSE define as follows (Wiranda & Sadikin, 2019):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{5}$$

2. Durbin Watson Test

Most time series regression problems contain data with positive autocorrelation. The Durbin-Watson test is a statistical test of existence. Positive autocorrelation of regression model errors. Especially hypotheses considered in the Durbin-Watson test are:

$$H_0: \phi = 0 H_1: \phi = 0$$

The Durbin-Watson **test statistic** is as follow (Montgomery et al., 2015)

$$d = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2} = \frac{\sum_{t=2}^{T} e_t^2 + \sum_{t=2}^{T} e_{t-1}^2 - 2\sum_{t=2}^{T} e_t e_{t-1}}{\sum_{t=1}^{T} e_t^2} \approx 2(1 - r_2) \quad (6)$$

Where e_t, t = 1, 2, ..., T are the residuals from the subsequent OLS regression. From y_t on x_t. In the formula. Above r_1 is the lag 1 autocorrelation between the residuals. In the case of uncorrelated error, the value of the Durbin-Watson statistic should be about 2.

3. Double Exponential Smoothing

In the early stages, smoothing and forecasting are carried out using the Double Exponential Smoothing method. This method is because the data used has a trend pattern. According to Montgomery et al. (2015), the formula of Double Exponential Smoothing is as follows:

$$\begin{gathered} \bar{y}_T = \lambda y_T + (1-\lambda)\bar{y}_{T-1} \\ \bar{y}_T^{(2)} = \lambda y_T^{(1)} + (1-\lambda)\bar{y}_{T-1} \end{gathered} \quad (7)$$

with an initial value of:

$$\begin{gathered} \bar{y}_0^{(1)} = \hat{\beta}_{0,0} - \frac{1-\lambda}{\lambda}\hat{\beta}_{0,1} \\ \bar{y}_0^{(2)} = \hat{\beta}_{0,0} - 2\left(\frac{1-\lambda}{\lambda}\right)\hat{\beta}_{0,1} \end{gathered} \quad (8)$$

4. Cochrane Orcutt

Meanwhile, the Cochrane-Orcutt and Hildreth-Lu methods are used to deal with autocorrelation. According to Montgomery et al. (2015), the procedure for the Cochrane-Orcutt method is to transform the response variable so that $y_t' = y_t - \phi y_{t-1}$ is obtained so that the model is obtained:

$$\begin{gathered} y_t' = y_t - \phi y_{t-1} \\ y_t' = \beta_0 + \beta_1 x_t + \varepsilon_t - \phi(\beta_0 + \beta_1 x_{t-1} + \varepsilon_{t-1}) \\ y_t' = \beta_0(1-\phi) + \beta_1(x_t - \phi x_{t-1}) + \varepsilon_t - \phi\varepsilon_{t-1} \\ y_t' = \beta_0' + \beta_1 x_t' + a_t \end{gathered} \quad (9)$$

Where $\beta_0' = \beta_0(1-\phi)$ and $x_t' = x_t - \phi x_{t-1}$. Note that the error expression in α_t transformed or reparameterized model is an independent random variable.

5. Hildreth Lu

Another method that can handle autocorrelation is Hildreth-Lu. The Hildreth-Lu procedure of simultaneously estimating the autocorrelation of the disturbances and the coefficients provides more efficient parameter estimates and consistent estimates of the standard errors (Feldstein, 1996). In contrast to the Cochrane-Orcutt method, the Hildreth-Lu method models the data according to the following equation (Hildreth & Lu, 1960):

$$y_t - \rho y_{t-1} = \alpha(1-\rho) + \beta(X_t - \rho X_{t-1}) + \varepsilon_{t-1} \quad (10)$$

The procedure is repeated so that the value of ρ is obtained which causes the smallest sum of squares for error (SSE). SSE measures the "residual" variety within the data that is not clarified by the independent variable x.

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{11}$$

Where y_i is observed value and $\hat{y}_i$ is the value estimated by the regression line.

RESULTS AND DISCUSSION

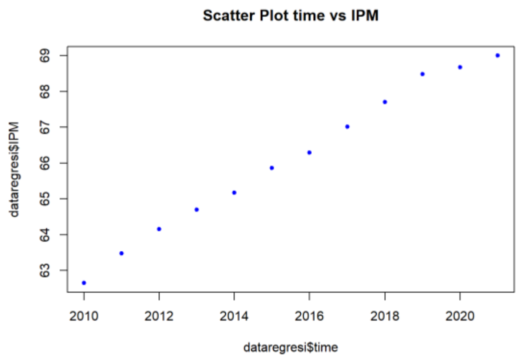


Figure 1. Scatter Plot between year and HDI

Based on Figure 1, it can be seen that there is a positive relationship between year and HDI. Based on the scatter plot above, it shows that the HDI value of Gorontalo Province has improved from year to year. In addition, the correlation between the two variables is very high, with a value of 0.996. Followed by forecasting for the next 10 years using Double Exponential Smoothing, the results are as follows:

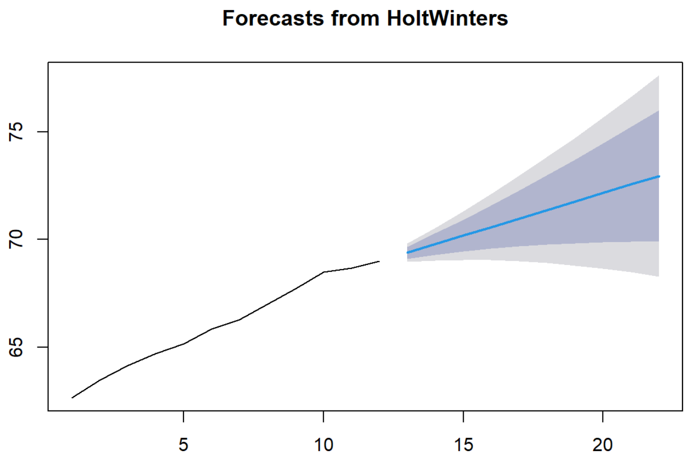


Figure 2. Forecasts from HoltWinters Double Exponential Smoothing

Based on the results of smoothing and forecasting using Double Exponential Smoothing in Figure 2, the data tends to have an uptrend, including forecasting for the next 10 years. Starting in 2024, it is predicted that the HDI value of Gorontalo province will reach 70,18546 or reach the high HDI classification according to UNDP (UNDP, 2018). When the year 2029-2031 is the target year for the SDGs, the HDI value of the province of Gorontalo has reached 72. If the data is carried out by simple linear regression, the regression model formed is as follows. The regression equation formed is as follows:

$$y = 1117.5481818 + 0.5872727x \tag{12}$$

Based on the results of the F-test, we obtained a p-value of $3.136 \times 10^{-12} < 0.05$ so reject H0. So, there is sufficient evidence to say that the year has an effect on HDI in Gorontalo Province. Meanwhile, the intercept and slope t-test also show that the two regression parameters have a p-value <0.05.

Table 1. Partial T-Test

	Estimate	**Std. Error**	**t value**	**Pr (>|t|)**
(Intercept)	-1,12E+06	3,06E+04	-36.58	5.55e-12 ***
time	5,87E+02	1,52E+01	38.74	3.14e-12 ***
Multiple R-squared	0.9934			
Adjusted R-squared	0.9927			
F-statistic:	1501 on 1			
P-value	3,14E-09			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

The R^2 value is 0.9934, or the year variable can explain the diversity of HDI variables by 99.34%. The results of the regression analysis showed promising results. However, it is necessary to check assumptions. For example, we check the regression assumptions using the plot approach:

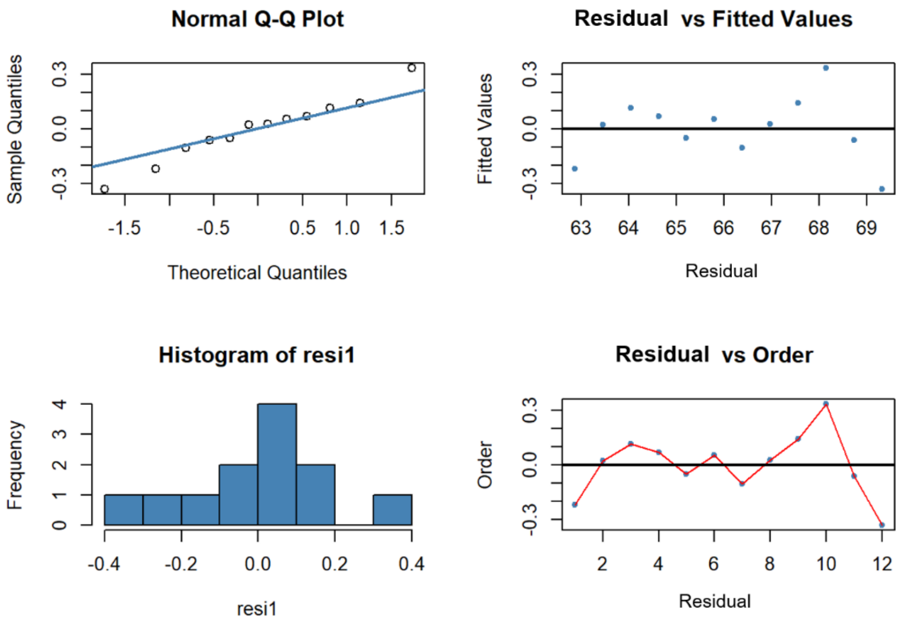
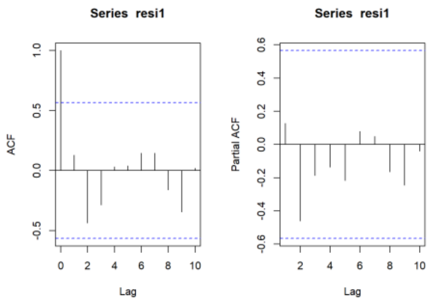


Figure 3. Residual diagnostic plot

From Figure 3, the Normal Q-Q Plot shows that the rest of the data tends to spread normally. The histogram plot of the residuals also shows that the residuals tend to be normally distributed. However, in the residual vs fitted values and residual vs order plots, there is a pattern formed. The existence of a formed pattern can be an early indication of autocorrelation. To ensure this, the autocorrelation detection performed by looking at the Autocorrelation Function (ACF), and Partial Autocorrelation Function (PACF) plots as follows:



(a) (b)

Figure 4. (a) ACF plot and (b) PACF plot

Figure 4 tells us that based on the ACF and PACF plots, the data are still within the safe limits of autocorrelation, so based on the graph, there is no autocorrelation. However, the Durbin-Watson test (DW-test) can be performed to be more accurate. The results of the Durbin-Watson test show a p-value of 0.03741 <0.05 or reject H0. This result indicates sufficient evidence that there is autocorrelation in the data at the 5% significance level. Handling autocorrelation is needed so that all regression analysis assumptions are met so that the model formed becomes accurate. Cochrane-Orcutt and Hildreth-Lu methods can handle autocorrelation.

1. Handling of Autocorrelation

1. Cochrane-Orcutt

Based on the output of handling autocorrelation using Cochrane Orcutt, the corrected regression equation is:

$$y = -1062.042646 + 0.559755x \tag{13}$$

In addition, the results of the Durbin-Watson test after the data were transformed using the Cochrane-Orcutt method showed a p-value of 9.403e-02 or 0.094. This value exceeds 5%, so it does not reject H0, or there is no autocorrelation in the corrected or transformed model using the Cochrane-Orcutt method. This value is obtained when the optimum rho is 0.340921.

2. Hildreth-Lu

The Hildreth-Lu method can be defined based on the following plot looking for optimum rho by Plot rho optimum:

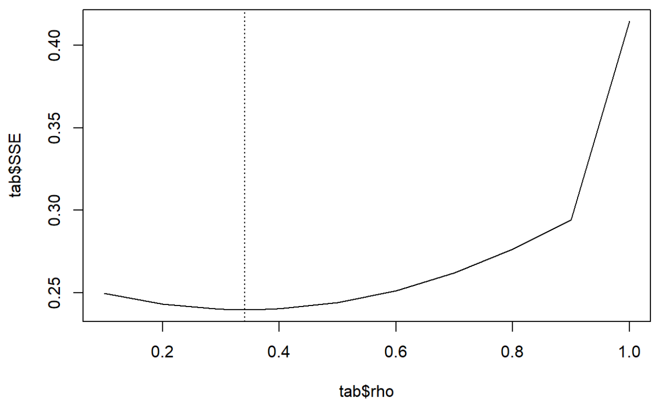


Figure 5. ρ Optimum

Based on the rho and plots above, the optimum rho is when = 0.341. Furthermore, the model was re-analyzed with = 0.341. The best Hildreth-Lu model produces the following regression equation:

$$y = -701.03275 + 0.55982x \tag{14}$$

Furthermore, the Durbin-Watson test was carried out again, whether there was still autocorrelation in the model after correction using the Hildreth-Lu method. Based on the output of the Durbin-Watson test, a p-value of 0.09386 > 0.05 was obtained so that it did not reject H0 or there is insufficient evidence to say autocorrelation in the data at the 5% significance level.

2. Comparisons between OLS, Cochrane-Orcutt, and Hildreth-Lu

To see the accuracy of the model, the Mean Square Error (MSE) and Root Mean Square Error (RMSE) are calculated as Table 2.

Table 2. Value of RMSE and MSE of Models

Model	RMSE	MSE
Original Model (OLS)	0.1654883	0.02738636
Cochrane Orcutt Model	0.1955602	0.03824381
Hildreth-Lu Model	0.1476328	0.02179545

Based on the results of the MSE and RMSE above, the smallest MSE value is 0.021, and the smallest RMSE is 0.147. The smallest MSE and RMSE values occur in the Hildreth-Lu model. So, in the case of the 2010-2021 Gorontalo Province HDI, the best model is to use the Hildreth-Lu model.

CONCLUSION

Time series data is data that is prone to autocorrelation. The autocorrelation identified through several approaches, namely the graphical method and the formal test, namely Durbin-Watson. There was autocorrelation In Gorontalo Province HDI data. Autocorrelation could be handled in two ways, namely the Cochrane-Orcutt method and the Hildreth-Lu method. The Hildreth-Lu method is a method that produces a model with the smallest MSE and RMSE so that Hildreth-Lu is the most effective.

REFERENCES

Aprianto, A. (2020). Citation: Metode Cochrane-Orcutt Untuk Mengatasi Autokorelasi Pada estimasi Parameter Ordinary Least Squares. *Buletin Ilmiah Mat, Stat, dan Terapannya (Bimaster) 9*(1), 95-102.

Asnidar. (2018). Citation: Pengaruh indeks pembangunan manusia (IPM) dan inflasi terhadap pertumbuhan ekonomi di Kabupaten Aceh Timur. *Jurnal Samudera Ekonomika*, 2(1): 1-12.

Badan Pusat Statistik. (2022, Feb 22). *Indeks Pembangunan Manusia menurut Provinsi 2019-2021*. Retrieved from https://www.bps.go.id/indicator/26/494/1/-metode-baru-indeks-pembangunan-manusia-menurut-provinsi.html.

Feldstein, M. (1996). Citation: Social security and saving: new time-series evidence. *National Tax Journal*, 9(2).

Hildreth, C., Lu, J.Y. (1960). *Demand Relations with Autocorrelated Disturbances*. Michigan (US): Michigan State University, Agricultural Experiment Station, Department of Agricultural economics.

Mendenhall, W., Beaver, R.J., Beaver, B.M. (2012). *Introduction to Probability & Statistics: 14th Edition.* Boston (US): Brooks/Cole, Cengage Learning.

Montgomery, D.C., Jennings, C.L., & Kulahci, M. (2015). *Introduction to Time Series Analysis and Forecasting: 2nd Edition*. Hoboken (US): John Wiley and Sons.

United Nations Development Programme. (2022). *New threats to human security in the Anthropocene Demanding greater solidarity*. New York(US): UNDP.

United Nations Development Programme. (2018). *Human Development Indices and Indicators: 2018 Statistical Update*. New York (US): UNDP.

Wiranda, L., & Sadikin, M. (2019). Citation: Penerapan Long Short-Term Memory Pada Data Time Series Untuk Memprediksi Penjualan Produk PT. Metiska Farma. *JANAPATI (Jurnal Nasional Pendidikan Teknik Informatika)*, 10, 1–13.