JURNAL ILMIAH MANAJEMEN BISNIS DAN INOVASI UNIVERSITAS SAM RATULANGI (JMBI UNSRAT)

SPURIOUS REGRESSION ANALYSIS ON TIME SERIES DATA FROM FACTORS AFFECTING INDONESIAN HUMAN DEVELOPMENT INDEXS IN 1990 – 2017

Rinto Rain Barry, Innocentius Bernarto
Universitas Pelita Harapan

A R T I C L E I N F O

Keywords*: spurious regression, stationary, cointegration, error correction model, equilibrium*

Kata kunci: *Regresi Spurious, stasioner, kointegrasi, error correction model, equilibrium*

Corresponding author:
Innocentius Bernarto
bernarto227@gmail.com

***Abstract*:** In a spurious regression conditions occur linear regression equations that are not stationary on the mean and variance. If the variables are not stationary, there will be cointegration, so it can be concluded that there is a long-term equilibrium relationship between the two research variables and in the short term there is a possibility of an imbalance, so to overcome it in this study using the Error Correction Model. The purpose of this study is to apply a cointegration test to see whether there is a long-term non-equilibrium relationship between the time series between the Human Development Index and life expectancy at birth, average school year for adults aged 25 years and over and gross national income per capita. The data used in this study are time series data between 1990-2017. The statistical management is carried out using Eviews 10. Based on the results obtained, it was concluded that 81.7% and it can be said that the types of independent variables included in the model are already good, because only 18.3% of the diversity of the dependent variable is influenced by the independent variables outside this research model.

Abstrak: Pada regresi spurious terjadi kondisi persamaan regresi linier yang tidak stasioner pada mean dan variansinya. Apabila variabel tidak stasioner, maka akan terjadi kointegrasi, maka dapat disimpulkan bahwa ada hubungan kesetimbangan (*equilibrium*) jangka panjang antara kedua variabel penelitian dan dalam jangka pendek ada kemungkinan terjadi ketidakseimbangan, sehingga untuk mengatasinya dalam penelitian ini menggunakan Model Koreksi Kesalahan/*error correction model* (ECM). Tujuan penelitian ini adalah menerapkan uji kointegrasi untuk melihat apakah terdapat hubungan ketidakkeseimbangan (*non-equilibrium*) jangka panjang data deret waktu (*times series*) antara Indeks Pembangunan Manusia (*human development index*) dengan harapan hidup saat lahir (*life expectancy at birth*), rata-rata tahun sekolah untuk orang dewasa berusia 25 tahun dan lebih (*mean of years of schooling for adults aged 25 years and more*) dan pendapatan nasional kotor per-kapita. Data yang digunakan dalam penelitian ini adalah data deret waktu (*time series*) antara tahun 1990-2017. Pengelolaan statistiknya dilakukan menggunakan program Eviews 10.Berdasarkan hasil yang didapat, disimpulkan bahwa 81.7% dan dapat dikatakan bahwa jenis variabel bebas yang dimasukkan dalam model sudah baik, sebab hanya 18.3% keragaman variabel terikat yang dipengaruhi oleh variabel bebas di luar model penelitian ini.

INTRODUCTION

Time series data are generally stochastic in nature, that is, they have an element of opportunity or probability, which means they have non-stationary trends, or the data has unit roots. If the data has a unit root, then the value will tend to fluctuate not around the average value with constant variation making it difficult to estimate a model and it is feared to produce spurious regression. Unit Root Test is one of the concepts that is increasingly popular lately used to test the time series data stationary. This test was developed by Dickey and Fuller, using the Augmented Dickey Fuller test (Dickey and Fuller, 1979). Regression models that involve time series data sometimes produce false or dubious results, in the sense that the results look superficially good, but further investigation can look suspicious (Gujarati and Porter, 2010, Aljandali and Tatahi, 2018).

Many observers or researchers often fall asleep and are lulled by what is called R square syndrome. Researchers are often fooled by the R square value that is so convincing and less responsive to diagnostic tests or tests of classical assumptions, especially auto correlation, heteroscedasticity, and linearity of the analytical tools they are using. Whereas high R square is only one of the criteria for choosing a regression equation. But that is not a prerequisite for observing whether, or not the formulation of a model, because in fact the high R square value of the regression results or the estimation of a model is a warning that the estimation results are subject to spurious regression.

Examples of the most severe cases of this type that we might face as stated by Andersen and Carlson (1974) are the "St. Louis Model" cases, where this model contains five behavioral equations, and found three of them turned out to be R square greater than d (Durbin-Watson test). And it was noted that one of the problems found was a series correlation error and, also the significance test was invalid. In this section, it is possible to obtain a regression equation related to economic time series that shows typical behavior, with a seemingly high level of compatibility, when in fact the independent variable does not have any explanatory power (Granger and Newbold, 1986, p. 205). As a case study, this study analyzes the spurious regressions that occur in time series data of the factors that influenced the Indonesian human development index in 1990 - 2017.

Formulation of the problem

As a case study related to overcoming spurious regression in the equation model of this study, where this research wants to test whether there is a significant influence between life expectancy at birth as variable X1, the average school year for adults aged 25 years and over as variable X2, gross national income per-capita as variable X3 on the Human Development Index as variable Y.

Research Purposes

The purpose of this study was arranged to find out:

1. Whether or not the influence of independent variables in this study on the dependent variable and how much influence.
2. The presence or absence of spurious regressions that occur in the research model conducted, and how to resolve or overcome them using the form of a Dynamic Linear Model, namely the Error Correction Model, starting from the unit root test, continued co-integration test.
3. Find the right research model and not produce a spurious regression.

LITERATURE REVIEW

Human Development Index

The main priority for the development of a country, is not economic growth, but human development and the ability of its people. To support and know this the Human Development Index was created, which can also be used to determine national policy options. For example, two countries where the level of Gross National Income per capita is not necessarily the same as human development, could be different results (Engineer, King, and Roy, 2008). With these contrasting results, it can help the policy priorities that will be taken by the government. The average size of human development, such as: a long and healthy life, extensive knowledge and decent living standards is known as the Human Development Index (Hou, Walsh and Zhang, 2015).

The Human Development Index is the normalized geometry average for each of the three dimensions, as meant by the United Nations Development Program published in the Human Development Reports in 2019, namely:

1. The health dimension is measured by life expectancy at birth.
2. Education dimension is measured by the average school year for adults aged 25 years and over and the expected school year for school children entering age.
3. Gross national income per capita is used to measure the standard of living dimension. Whereas Human Development Index uses the income logarithm to find out the amount of Gross National Income, which reflects a country's income.

Gatt (2005) states that the Human Development Index is useful for:

1. measuring success in building the quality of life for residents of a country.
2. determine the rank or level of development of a region or country.
3. measuring the Government's performance and, also as a determination of the General Allocation Fund.

Spurious Regression

Spurious regression is a regression that does not reveal the truth. This is because time series data have their own behavior, which is not infrequently influenced by trends or trends. If at least one variable is non stationary, then regressed, then it can be as if the independent/dependent variable significantly influences the independent variable and has a high R square, even though it turns out that the relationship is only because the two variables have the same trend. Sure, the resulting regression is meaningless. Non-stationary data we often encounter in economic data, such as: Gross Domestic Product, prices, population, and so on. Engle and Granger (1987), and Granger and Newbold (1988) argue that many levels of time series in economics are not stationary; the sample is well represented by an integrated ARIMA type process as stated by Box and Jenkins (1970), and Stock (2015) and often seems to run random. Furthermore, it is said that in a regression equation related to time series often has a high R square, but also displays highly correlated residues, which is very low for Durbin-Watson statistical results. Under these conditions, they argue that the significance test of the regression coefficient is very misleading. Experiments in the sampling conducted provide strong evidence that the significance test for the null hypothesis has no relationship and hence a false relationship occurs, including when the serial data generated is statistically random.

It can be concluded that summary the characteristics of Spurious Regression include:

1. High R square value.
2. Durbin-watson value is smaller/lower than R square.

3. The value of t-statistics (partial) results is high, thus rejecting the H_0 hypothesis, even though the trend between independent and dependent variables is not related at all.
4. The F-test value is large and significant.
5. Not stationary at the stage of Level.
6. Standard error and low p-value.
7. The mean is constant, but the variance is not constant.

Related to the problems in the background of this study and in line with the development of econometric methods, according to Insukindro (1991, p. 75-87) there are two methods that can be used to avoid spurious regression, namely:

1. The first method, without data stationarity test, is by forming a Dynamic Linear Model (MLD), such as: Partial Adjustment Model (PAM), Error Correction Model (ECM), Buffer Stock Model / buffer stock model (BSM) or shock absorber model (SAM). The use of Dynamic Linear Model (MLD), in addition to avoiding spurious regression (spurious regression), can also be used to observe or see the long-term relationship between variables, as expected by the related theory.
2. The second method, by using data stationerity test or using a cointegration approach. This approach is basically a test of theory and an important part in the formulation and estimation of MLD. We can process time series data using ordinary least square (OLS) if and only if the data is stationary / unchanged. Data is said to be stationary, if it meets the following things:
   a. The average value of E (X_t) is constant throughout the period t.
   b. The variance value Var (X_t) is constant throughout the entire period t.
   c. Covariance value Cov (X_t, $X_t + k$) is constant throughout the whole period and all $k \neq 0$.

Research Hypothesis

The following are hypotheses in this study:

1. H_0: There is no significant effect between life expectancy at birth on the Human Development Index.
   H_a: There is a significant influence between life expectancy at birth on the Human Development Index.
2. H_0: There is no significant effect between the average school year for adults aged 25 years and over on the Human Development Index.
   H_a: There is a significant influence between the average school year for adults aged 25 years and over on the Human Development Index.
3. H_0: There is no significant effect between gross national income per capita on the Human Development Index.
   H_a: There is a significant influence between gross national income per capita on the Human Development Index.
4. H_0: There is no significant effect between life expectancy at birth, average school years for adults aged 25 years and over, together with gross national income per capita on the Human Development Index.
   H_a: There was no significant effect between life expectancy at birth, average school years for adults aged 25 years and over, together with gross national income per capita on the Human Development Index.

Research Model

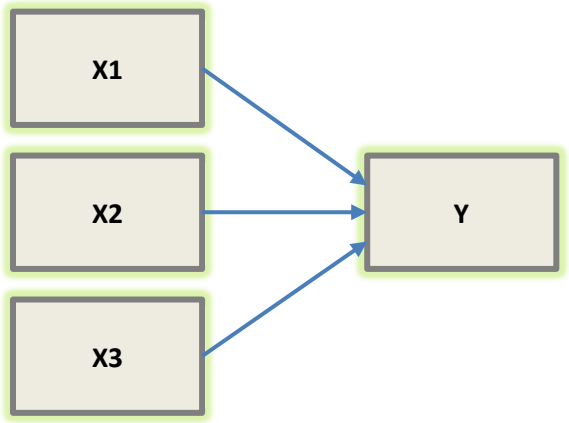


Figure 1. Research Model

Figure explanation:
X1 = Life expectancy at birth (years)
X2 = Average school year for adults aged 25 years and over (years)
X3 = Gross national income per capita (dollar)
Y = Human Development Index (index)

RESEARCH METHODOLOGY

Research Types and Data

This study uses the econometric software EViews 10 with time series data consisting of one dependent variable, namely: Human Development Index (Y), and three independent variables, namely: Life expectancy at birth (X1), Average year school for adults aged 25 years and over (X2), Gross national income per capita (X3). These variables use annual real data in the period 1990 to 2017, obtained from the United Nations Development Program at http://www.hdr.undp.org/en/data.

Research Equation Model

The specification of the relationship of the variables stated in this research model is as follows:

$$Y_t = \alpha_0 + \alpha_1 X1_t + \alpha_2 X2_t + \alpha_3 X3_t$$

Where,

Y_t = Human Development Index in period t.
$X1_t$ = life expectancy at birth in period t.
$X2_t$ = The average school year for adults aged 25 years and over in period t.
$X3_t$ = Gross national income per capita in period t.
$\alpha_0, \alpha_1, \alpha_2, \alpha_3$ = Short-term coefficient.

Data Analysis

The data analysis used in this study using the Error Correction Model (ECM) Method as an econometric calculation tool and using descriptive analysis method aims to identify long-term and short-term relationships that occur because of the cointegration between research variables. Before carrying out ECM estimation and descriptive analysis, several steps must be carried out, such as: data stationarity test, determining lag length and degree of cointegration test (Cheung and Lai, 1995).

Before conducting a regression with the ECM test, what needs to be done first is knowing whether the variables used are stationary or not. If the data is not stationary, a spurious regression will be obtained, an autocorrelation phenomenon arises, also cannot generalize the results of the regression for a different time. In addition, if the data to be used is stationary, OLS regression can be used, but if it is not stationary, the data needs to be seen for stationarity through the degree of integration test. And furthermore, data that are not stationary at the stage of Level are likely to be cointegrated so that a cointegration test is needed. Then if the data has been cointegrated, then ECM testing can be done.

To find out whether the time series data used is stationary or not stationary, a unit root test is used. Unit root tests were performed using the Dicky Fuller (DF) method, with the following hypotheses:

H_0: there is a unit root or non-stationary data, if ADF value > Critical Value ($\alpha = 5\%$).
H_1: there is no unit root or stationary data, if the ADF value < Critical Value ($\alpha = 5\%$).

The t-statistic results from the estimation results on the method will be compared with the McKinnon critical value at the critical point of 1%, 5%, and 10%. If the t-statistic value is smaller than the McKinnon critical value, then H_0 is accepted, meaning that the data has a root unit, or the data is not stationary. If the t-statistic value is greater than the McKinnon critical value, then H_0 is rejected, meaning that there is no unit data or stationary data.

Data testing was performed using a unit root test developed by Dickey-Fuller, or better known as the Augmented Dickey-Fuller test (Cheung and Lai, 1995).
There are three ADF test models that can be used to test stationarity, namely:

1. Model without intercept and without trend.
2. Models that use intercept only.
3. Models that use intercept and trend.

This test is carried out to determine the degree or order of differential how the data under study will be stationary. This test is carried out on the unit root test, if it turns out that the data is not stationary in the first degree (Insukrindo, 1992), the test is carried out in the form of first differentiation. The following test is the stationarity test with the Dickey-Fuller (DF) test at the first stage of differentiation, using a 5% test of critical values.

Error Correction Model Regression Procedure
The method procedure of the Error Correction Model / error correction model (ECM) is as follows:

1. Unit Root Test (unit root test)
   The concept used to discuss a time series data is the unit root test. If a time series data (time series) does not work, then it can be requested that the data is discussing the root problem (root unit problem).
   The root unit can be seen by comparing the t-statistic results of the regression with the value of the Augmented Dickey Fuller test.
   The Equation Model for the Y variable as an example is as follows:

   $$\Delta Y = a_1 + a_2 T + \Delta Y_{t-1} + \alpha_i \sum_{i=1}^{m} \Delta Y_{t-1} + e_t$$

   Where:
   $\Delta Yt\text{-}1 = (\Delta Yt\text{-}1 - \Delta Yt\text{-}2)$ and so on

m = length of time lag based on i = 1, 2, ..., m.
The null hypothesis is still $\delta = 0$ or $\rho = 1$.
The ADF t-statistic value is the same as the Dickey-Fuller t-statistical value (DF).

2. Integration Degree Test
If the unit root test above, the observed time series data is not stationary, then the next step is to test the degree of integration to find out at what degree of integration the data will be stationary. For example, the test of the degree of integration of the Y variable in the model is as follows:

$\Delta Y_t = \beta_1 + \delta \Delta Y_{t-1} + \alpha_i \sum_{i=1}^{m} \Delta Y_{t-1} + e_t$
$\Delta Y_t = \beta_1 + \beta_2 T + \delta \Delta Y_{t-1} + \alpha_i \sum_{i=1}^{m} \Delta Y_{t-1} + e_t$

The t-statistic value of the regression results from the two equations above is compared with the t-statistic value in the Dickey-Fuller (DF) table. If the value of δ in both equations is equal to one ($\delta = 1$), then the variable ΔYt is said to be stationary at the first degree (1st difference). But if the value of δ does not differ from zero, then the variable ΔYt is not stationary at the first degree of integration. Therefore, the test is continued to the second, third degree integration test and so on until we get the stationary variable ΔYt data.

3. Cointegration Test
The most commons used cointegration tests are the Engle-Granger (EG) test, the Augmented Engle-Granger (AEG) test and the Cointegrating Regression Durbin-Watson (CRDW) test. To get the calculated values for EG, AEG and CRDW, the data to be used must be integrated to the same degree. OLS testing of the equation of the research model below:

$Y_t = a_0 + a_1 \Delta X1_t + a_2 \Delta X2_t + a_3 \Delta X3_t + e_t$
From the equation above, save the residual (error terms). The next step is to estimate the autoregressive equation model of the residual based on the following equations:
$\Delta \mu_t = \lambda \mu_{t-1}$
$\Delta \mu_t = \lambda \mu_{t-1} + \alpha_i \sum_{i=1}^{m} \Delta \mu_{t-1}$

With the hypothesis test:
H_0: $\mu = I(1)$, meaning that there is no cointegration.
H_1: $\mu \neq I(1)$, meaning that there is cointegration.

Based on the OLS regression results on the research model equation, a CRDW value will be calculated (the calculated DW value in the equation) to then be compared with the CRDW table. While from the equations above, we will get the calculated EG and AEG values which will also be compared with the DF and ADF table values.

Based on the OLS regression results on the research model equation, a CRDW value will be calculated (the calculated DW value in the equation) to then be compared with the CRDW table. While from the equations above, we will get the calculated EG and AEG values which will also be compared with the DF and ADF table values.

4. Test the error correction model (ECM)

If it passes the cointegration test, it will then be tested using the Dynamic Linear Model (MLD) to determine the possibility of structural changes, because the long-term balance relationship between the independent variable and the dependent variable from the results of the cointegration test will not apply at any time.

In brief, the process of working the ECM on the modified research equation model is as follows:

$Y_t = a_0 + a_1\Delta X1_t + a_2\Delta X2_t + a_3\Delta X3_t + a_4e_{t-1} + e_t$

ANALYSIS AND DISCUSSION

Regression Early Equations of Research

The linear regression results obtained from the initial equation model from this study are as follows:

Dependent Variable: Y
Method: Least Squares
Date: 11/01/19 Time: 23:51
Sample: 1990 2017
Included observations: 28

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X1	0.010026	0.002172	4.615823	0.0001
X2	0.014587	0.003142	4.642889	0.0001
X3	6.18E-06	1.41E-06	4.372945	0.0002
C	-0.460577	0.186919	-2.464052	0.0213

R-squared	0.997464	Mean dependent var	0.620179
Adjusted R-squared	0.997147	S.D. dependent var	0.051920
S.E. of regression	0.002773	Akaike info criterion	-8.806075
Sum squared resid	0.000185	Schwarz criterion	-8.615761
Log likelihood	127.2851	Hannan-Quinn criter.	-8.747894
F-statistic	3146.692	Durbin-Watson stat	0.652771
Prob(F-statistic)	0.000000		

Based on the results of the initial regression equation above, it is known that:
- High R-squared value of 0.997464 = 99.7%.
- Durbin-watson value (0.652771) is smaller than R-squared (0.997464).
- All t-statistic (count) results are greater than t-table (2,069), so the H_0 hypothesis is rejected, even though the trend between independent and dependent variables is not related at all.
- The F-test results are large and significant (0.000000).

So, from the regression equation results above it can be concluded the occurrence of spurious regression, and therefore we need a way to solve and overcome it in terms of using a Dynamic Linear Model (MLD), namely the Error Correction Model (ECM) regression test. And the steps that must be done before carrying out a regression test are the Correction Error Model Stationary test, then the Cointegration test.

Stationary Data Test Results

1. Stationary Test variable Y (Human Development Index)

a. Unit root test (unit root test) variable Y at the stage of Level
The results are as follows:

Null Hypothesis: Y has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=6)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.194929	0.2125
Test critical values:	1% level	-3.699871	
	5% level	-2.976263	
	10% level	-2.627420	

*MacKinnon (1996) one-sided p-values.

From the stationary test results at the stage of Level, ADF t-statistics (-2.194929) > Test critical values at α = 5% (-2.976263) and, also Prob. (0.2125) > 0.05, then H_0 is accepted, meaning there is a unit root or data not stationary.
Likewise, non-stationary data can be seen from the graph of Y variables at the stage of Level, as shown in Figure 2 below:

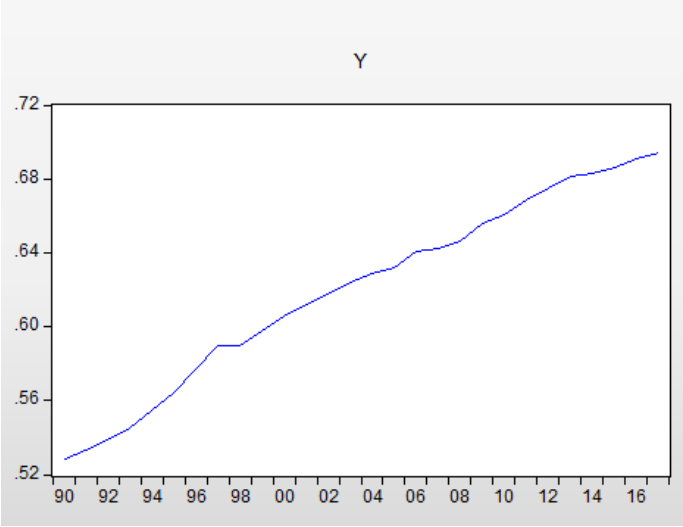


Figure 2 **Graph for the Y variable at the stage of Level**

Seeing the results of the unit root test and graph of variable Y it turns out that it is not stationary at the stage of Level, then the 1st Difference stage integration test is performed.

b. Unit root test (unit root test) variable Y at the 1st Difference stage
Obtained the following results:

Null Hypothesis: D(Y) has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=6)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-4.398451	0.0019
Test critical values:	1% level	-3.711457	
	5% level	-2.981038	
	10% level	-2.629906	

*MacKinnon (1996) one-sided p-values.

From the stationary test results above, that the ADF t-statistic (-4.398451) < Test critical values at α = 5% (-3.212696) and also Prob. (0.0019) < 0.05, then H_0 is rejected or H_1 is accepted, meaning there is no unit root or the data is stationary.
Likewise, the data is stationary as seen from the graph of Y variable at the 1st Difference stage, as shown in Figure 3 below:

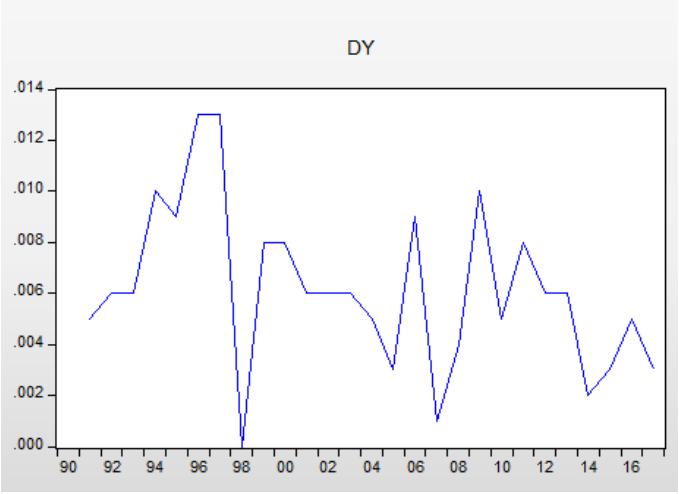


Figure 3 **Graph for the Y variable at the 1st Difference stage**

2. Stationary Test variable X1 (Life expectancy at birth)

a. Unit root test (unit root test) variable X1 at the stage of Level
The results are as follows:

Null Hypothesis: X1 has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=6)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.213716	0.2062
Test critical values:	1% level	-3.699871	
	5% level	-2.976263	
	10% level	-2.627420	

*MacKinnon (1996) one-sided p-values.

From the stationary test results at the stage of Level, ADF t-statistics (-2.213716) > Test critical values at α = 5% (-2.976263) and, also Prob. (0.2062) > 0.05, then H_0 is accepted, meaning there is a unit root or non-stationary data.

Likewise, the data is not stationary, which can be seen from the graph of variable X1 at the stage of Level, as shown in Figure 4 below:

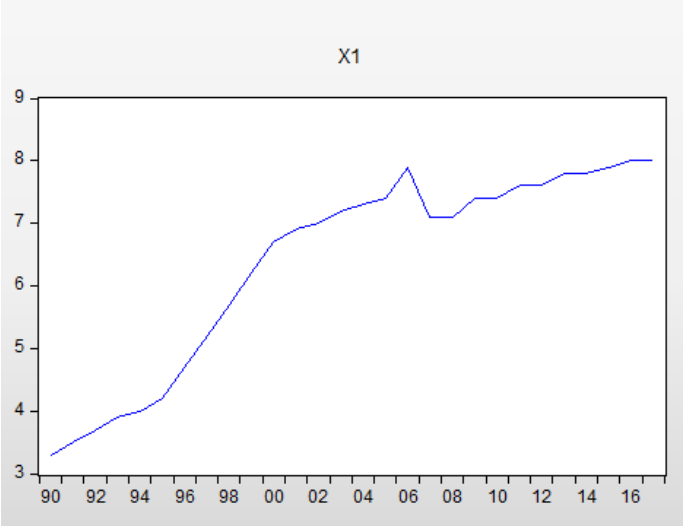


Figure 4 **Graph of variable X1 at the stage of Level**

Seeing the results of the unit root test and graph X1 variables were apparently not stationary at the stage of Level, then the 1st Difference stage integration test was conducted.

b. Unit root test (unit root test) variable X1 at the 1st Difference stage
Obtained the following results:

Null Hypothesis: D(X1) has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=6)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-4.241938	0.0028
Test critical values:	1% level	-3.711457	
	5% level	-2.981038	
	10% level	-2.629906	

*MacKinnon (1996) one-sided p-values.

From the results of the stationary test at the 1st Difference stage, we can see the ADF t-statistic (-4.241938) < Test critical values at α = 5% (-2.981038) and also Prob. (0.0028) < 0.05, then H_0 is rejected or H_1 is accepted, meaning there is no unit root or the data is stationary.

Likewise, the data is stationary, which can be seen from the graph of the DX1 variable at the stage of the 1st Difference, as shown in Figure 5 below:

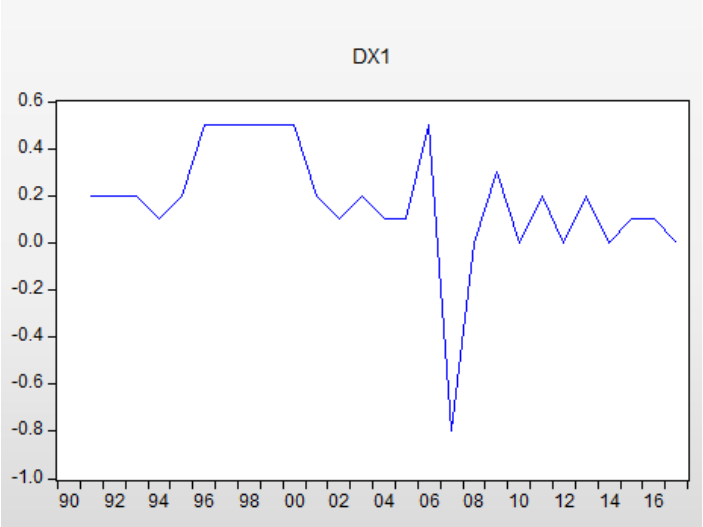


Figure 5 **Graph for the DX1 variable at the 1st Difference stage**

3. Stationary Test variable X2 (Average school year for adults aged 25 years and over)

a. Unit root test variable X2 at the stage of Level
The results are as follows:

Null Hypothesis: X2 has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=6)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-6.175083	0.0000
Test critical values:	1% level	-3.699871	
	5% level	-2.976263	
	10% level	-2.627420	

*MacKinnon (1996) one-sided p-values.

From the results of the stationary tests at the stage of Level, the ADF t-statistic (-6.175083) < Test critical values at α = 5% (-2.976263) and also Prob. (0.0000) < 0.05, then H_0 is rejected or H_1 is accepted, meaning there is no unit root or the data is stationary at the stage of Level, so there is no need to test the degree of integration of the 1st Difference stage.

4. Stationary Test variable X3 (Gross National Income per Capita)

a. Unit Root Test variable X3 at the stage of Level
The results are as follows:

Null Hypothesis: X3 has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=6)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		1.201969	0.9973
Test critical values:	1% level	-3.699871	
	5% level	-2.976263	
	10% level	-2.627420	

*MacKinnon (1996) one-sided p-values.

From the stationary test results above, that the ADF t-statistic (1.201969) > Test critical values at α = 5% (-2.976263) and, also Prob. (0.9973) > 0.05, then H_0 is accepted, meaning there is a unit root or non-stationary data.

Likewise, the data is not stationary, which can be seen from the X3 variable graph at the stage of Level, as in Figure 6 below:

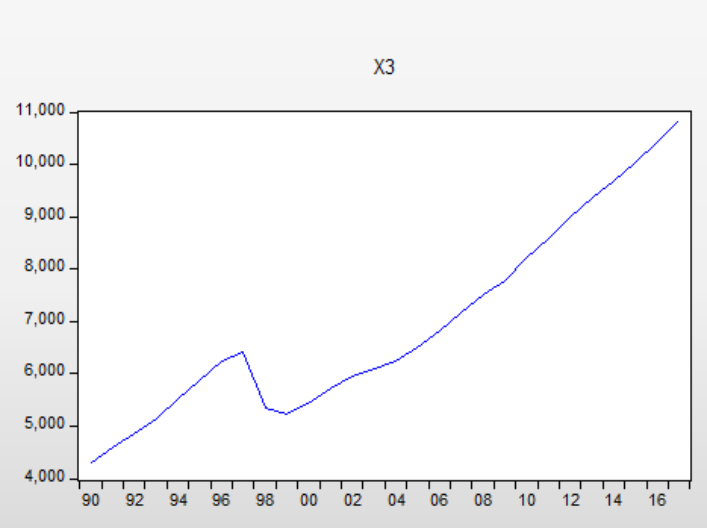


Figure 6 **Graph for the X3 variable at the stage of Level**

Seeing the unit root test results and graph X3 variables were apparently not stationary at the stage of Level, then the 1st Difference stage integration test was conducted.

b. Unit Root Test variable X3 at the 1st Difference stage
Obtained the following results:

Null Hypothesis: D(X3) has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=6)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-3.347667	0.0228
Test critical values:	1% level	-3.711457	
	5% level	-2.981038	
	10% level	-2.629906	

*MacKinnon (1996) one-sided p-values.

From the results of the stationary test at the 1st Difference stage, we can see the ADF t-statistic (-3.347667) < Test critical values at α = 5% (-2.981038) and also Prob. (0.0228) < 0.05, then H_0 is rejected or H_1 is accepted, meaning there is no unit root or the data is stationary.

Likewise, the data is stationary, which can be seen from the DX3 variable graph at the 1st Difference stage, as shown in Figure 7 below:

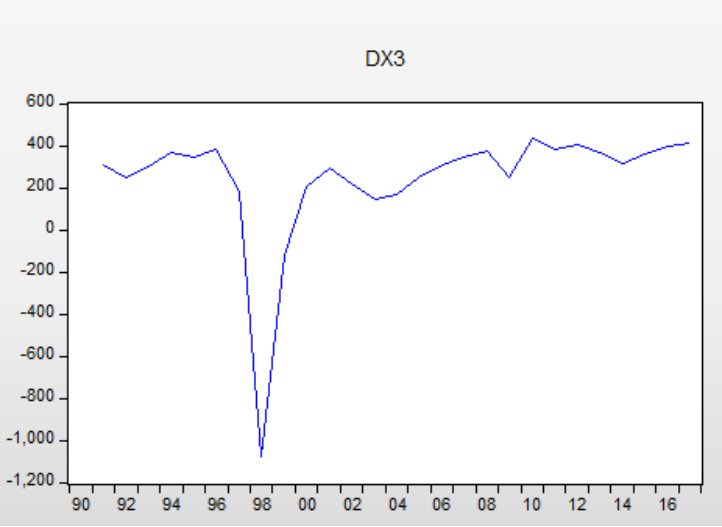


Figure 7 **Graphs for the DX3 variable at the 1st Difference stage**

After unit root tests for all variables, the table of the results of unit root test results for the research variables is obtained:

Table 1 Results of Root Test Unit Variable Research

Variable	Root Test Unit			
	Level		1st Difference	
	ADF	Prob.	ADF	Prob.
Y	-2.194929	0.2125	-4.398451	0.0019
X1	-2.213716	0.2062	-4.241938	0.0028
X2	-6.175083	0.0000		
X3	1.201969	0.9973	-3.347667	0.0228

Cointegration Test Results

After knowing that the data is not stationary, the next step is to identify whether the data is cointegrated. For this reason, a cointegration test is needed. Cointegration test is used to give an early indication that the model used has a long-term relationship (cointegration relation).

Cointegration test results are obtained by forming a residual obtained by regressing the independent variable on the dependent variable OLS. The residual must be stationary at the stage of Level to be said to have cointegration.

Regress the following equation:

$$Y = \beta_0 + \beta_1 X1 + \beta_2 X2 + \beta_3 X3 + e_t$$

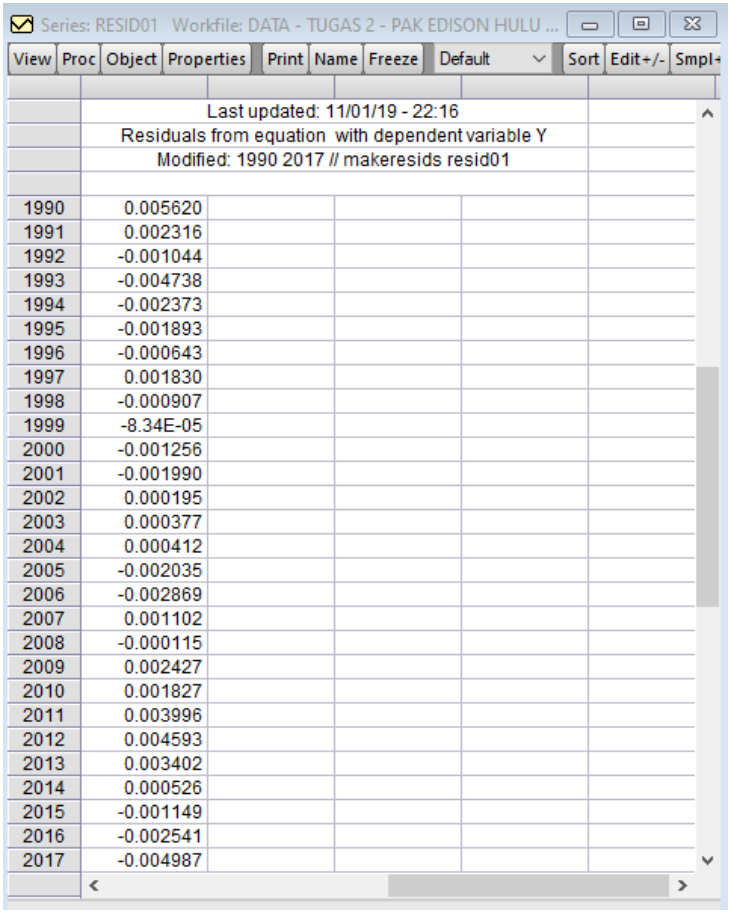
Obtained the following results:

Dependent Variable: Y
Method: Least Squares
Date: 11/01/19 Time: 22:11
Sample: 1990 2017
Included observations: 28

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X1	0.010026	0.002172	4.615823	0.0001
X2	0.014587	0.003142	4.642889	0.0001
X3	6.18E-06	1.41E-06	4.372945	0.0002
C	-0.460577	0.186919	-2.464052	0.0213

R-squared	0.997464	Mean dependent var	0.620179
Adjusted R-squared	0.997147	S.D. dependent var	0.051920
S.E. of regression	0.002773	Akaike info criterion	-8.806075
Sum squared resid	0.000185	Schwarz criterion	-8.615761
Log likelihood	127.2851	Hannan-Quinn criter.	-8.747894
F-statistic	3146.692	Durbin-Watson stat	0.652771
Prob(F-statistic)	0.000000		

Then get residuals, the results are as follows:



1990	0.005620
1991	0.002316
1992	-0.001044
1993	-0.004738
1994	-0.002373
1995	-0.001893
1996	-0.000643
1997	0.001830
1998	-0.000907
1999	-8.34E-05
2000	-0.001256
2001	-0.001990
2002	0.000195
2003	0.000377
2004	0.000412
2005	-0.002035
2006	-0.002869
2007	0.001102
2008	-0.000115
2009	0.002427
2010	0.001827
2011	0.003996
2012	0.004593
2013	0.003402
2014	0.000526
2015	-0.001149
2016	-0.002541
2017	-0.004987

Figure 8 **Resid01 results data**

Then do a stationary test of resid01 variable, the result is as follows:

Null Hypothesis: RESID01 has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=6)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.544651	0.1167
Test critical values:	1% level	-3.699871	
	5% level	-2.976263	
	10% level	-2.627420	

*MacKinnon (1996) one-sided p-values.

The residuals should be stationary at the stage of Level to be said to have cointegration. After a Dickey-Fuller test was conducted to test the resulting residuals, it was found that the residuals were not stationary at the stage of the Level as seen from the t-statistic value that was not significant at the Test critical values

of 5% (-2.544651 > -2.976263) and Prob. (0.1167) > 0.05. Thus, it can be said that the data is not co-integrated.

So that the data can be co-integrated in the long run, then the regression equation model is made log against variables Y, X1, X2 and X3.

The results of the regressions are as follows:

Dependent Variable: LOG(Y)
Method: Least Squares
Date: 11/01/19 Time: 22:48
Sample: 1990 2017
Included observations: 28

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-8.921178	1.606970	-5.551551	0.0000
LOG(X1)	0.075898	0.022597	3.358773	0.0026
LOG(X2)	1.851362	0.440335	4.204439	0.0003
LOG(X3)	0.059570	0.023777	2.505314	0.0194

R-squared	0.996933	Mean dependent var	-0.481203
Adjusted R-squared	0.996549	S.D. dependent var	0.085141
S.E. of regression	0.005002	Akaike info criterion	-7.626567
Sum squared resid	0.000600	Schwarz criterion	-7.436252
Log likelihood	110.7719	Hannan-Quinn criter.	-7.568386
F-statistic	2599.997	Durbin-Watson stat	0.621076
Prob(F-statistic)	0.000000		

Then take the residuals back, then create a resid02 variable.
Then the results are as follows:

Null Hypothesis: RESID02 has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=6)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-3.777314	0.0083
Test critical values:	1% level	-3.699871	
	5% level	-2.976263	
	10% level	-2.627420	

*MacKinnon (1996) one-sided p-values.

From the results of the stationary resid02 test, the ADF t-statistic (-3.777314) < Test critical values at α = 5% (-2.976263) and also Prob. (0.0083) < 0.05, then H_0 is rejected or H_1 is accepted, so it can be said that the data has been cointegrated.

Test Results Of The Error Correction Model (ECM)

Knowing cointegrated data means there is a long-term relationship (equilibrium) between variables. In the short term there is the possibility of an imbalance (non-equilibrium), and therefore we need the error correction model (ECM) test as follows:

D(LOG(Y)) = b0 + b1D(LOG(X1)) + b2LOG(X2) + b3D(LOG(X3)) + RESID02(-1) + e

The result of the regressions is as follows:

Dependent Variable: D(LOG(Y))
Method: Least Squares
Date: 11/22/19 Time: 16:37
Sample (adjusted): 1991 2017
Included observations: 27 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.074875	0.096810	0.773419	0.4475
D(LOG(X1))	0.090752	0.013982	6.490591	0.0000
LOG(X2)	-0.016653	0.022955	-0.725480	0.4758
D(LOG(X3))	0.068349	0.010826	6.313470	0.0000
RESID02(-1)	-0.451772	0.102408	-4.411478	0.0002

R-squared	0.845543	Mean dependent var	0.010125
Adjusted R-squared	0.817460	S.D. dependent var	0.005685
S.E. of regression	0.002429	Akaike info criterion	-9.037055
Sum squared resid	0.000130	Schwarz criterion	-8.797085
Log likelihood	127.0002	Hannan-Quinn criter.	-8.965699
F-statistic	30.10867	Durbin-Watson stat	1.775683
Prob(F-statistic)	0.000000		

The equation after the Error Correction Model (ECM) test is:

D (LOG (Y)) = 0.090752 * D (LOG (X1) - 0.016653 * (LOG (X2)) + 0.068349 * D (LOG (X3)) - 0.451772 * RESID02 (-1)

The ECM estimation results above, show in the short and long term that:

a. Life Expectancy Variables at Birth (X1) have a positive and significant effect on the Human Development Index (Y).
b. The average school year variable for adults aged 25 years and over (X2) has a negative and not significant effect on the Human Development Index (Y).
c. The variable gross national income per capita (X3) has a positive and significant effect on the Human Development Index (Y).

By looking at the results of the Adjusted R-squared value of 0.817460 or 81.7%, it can be said that the types of independent variables included in the model are good, because only 18.3% of the diversity of the dependent variable is influenced by the independent variables outside this research model.

CONCLUSIONS AND SUGGESTIONS

Conclusion

Spurious regression must be completed so that the results of the estimation are correct, where the model in this study that underwent spurious regression using the ECM model and by applying the cointegration test can be completed.

As a case study, this research can look at the relationship between long-term time series data on life expectancy at birth, average school year for adults aged 25 years and over and gross national income per capita on the Human Development Index.

The estimation results above illustrate that in the short term there is an influence between life expectancy at birth, the average school year for adults aged 25 years and over, and the gross national income per capita on the Human Development Index.

Finally based on the short-term equation using ECM method produces RESID02 coefficient. This coefficient measures the regression response for each period that deviates from equilibrium.

Suggestion

The suggestions that can be given for further research are:

1. Consider the other variables that can be examined for their effects, in addition to the variables used in this study.
2. Can use other methods, other than ECM that can be developed and researched to overcome spurious regressions, such as: a partial adjustment model (PAM), a buffer stock model (BSM), or a shock absorber model (SAM).

REFERENCES

Aljandali, A. and Tatahi, M. (2018). *Economic and Financial Modelling with Eviews. A Guide for Students and Professionals*. Switzerland: Springer International Publishing.

Andersen, L. C., and Carlson, K. M. (1974). St. Louis model revisited. *Int. Econ. Rev. 15*, 305-327.

Basuki, A. T. and Prawoto, N. (2016). *Analisis Regresi Dalam Penelitian Ekonomi dan Bisnis*. Jakarta: PT. Raja Grafindo Persada.

Box, G. and Jenkins, G. (1970). *Time Series Analysis: Forecasting and Control*. San Francisco, CA: Holden-Day.

Cheung, Y.W., and Lai, K. S. (1995). Lag Order and Critical Values of the Augmented Dickey–Fuller Test. *Journal of Business & Economic Statistics, 13*(3), 277–280.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association, 74*(366), 427. doi:10.2307/2286348

Engineer, M., King, I. and Roy, N. (2008). The human development index as a criterion for optimal planning. *Indian Growth and Development Review 1* (2), 172-192.

Engle, R.F. and C.W.J. Granger (1987). Co-integration and error correction: representation, estimation and testing. *Econometrica 55*, 251–76.

Gatt, E. (2005). The Human Development Index and Small States. *Bank of Valletta Review*, No. 32, Autumn 2005.

Granger, C. W. J., Newbold, P. (1986). *Forecasting Economic Time Series* (Second Edition). San Diego, California: Academic Pres, Inc.

Gujarati, D. N. and Porter, D. C. (2010). *Essentials of Econometrics* (Fourth Edition). New York: McGraw-Hill Companies, Inc.

Hou, J., Walsh, P. P., & Zhang, J. (2015). The Dynamics of Human Development Index. *The Social Science Journal, 52*(3), 331–347. Doi: 10.1016/j.soscij.2014.07.003.

Insukindro (1991). Regresi Linier Lancung dalam Analisis Ekonomi: Suatu Tinjauan dengan Satu Studi Kasus di Indonesia. *Jurnal Ekonomi dan Bisnis Indonesia, 6*: 75-88.

Lemy, D. M., Bernarto, I., & Tulung, J.E. (2013). Pengaruh Bauran Komunikasi Pemasaran Jasa Terhadap Proses Pengambilan Keputusan Mahasiswa Memilih Perguruan Tinggi Swasta di Sulawesi Utara. *Jurnal Manajemen: Derema*

Phillips, P.C.B. (1986). Understanding Spurious Regressions in Econometrics. *Journal of Econometrics 33* (1986) 311-340. North-Holland.

Startz, R. (2019). *EViews Illustrated*. Irvine, CA: IHS Global Inc. M

Stock, J. H. (2015). Time Series: Economic Forecasting. *International Encyclopedia of the Social & Behavioral Sciences, 2nd edition, Volume 24*, 337–340.