Analysis of Residential Property Price Index (RPPI) Using Multi Input Transfer Function

Y Chiristy*, M Novita, and S M Soemartojo
Department of Mathematics, Universitas Indonesia, Depok, Jawa Barat, Indonesia, 16424

*E-mail : yesicachristy71@gmail.com

Abstract. Residential property or home is one of the basic human needs as a shelter, to be able to continue to live. As the population in Indonesia increases, express the need for residential property will also increase. The movement of residential property prices in Indonesia can be observed from the Residential Property Price Index (RPPI) issued by Bank Indonesia. Residential property prices can be influenced by several factors. For example in economic, the movement of RPPI can be influenced by several factors such as inflation, Composite Stock Price Index (RPPI) and loan interest rates. Forecasting the RPPI can be done with times series modeling. Modeling RPPI using three influencing variables requires a multivariate time series model. The multi-input transfer function model is a multivariate model that can be used in modeling RPPI. In the multi-input transfer function model there is an output series ($y_t$) which is the RPPI which is estimated to be influenced by several input series ($x_{jt}$), which are inflation, IHSG and loan interest rates. Based on the model obtained in this study, it can be seen that the prediction of RPPI at the $t$-time is influenced by the amount of inflation in the previous two months until previous five months, influence by loan interest rates without delay so that it is influenced by the $t$-time until previous three months, influenced by the IHSG at $t$-time until previous four months, and also influenced by itself at one month before to four months before.

1. Introduction
In today’s competitive era, the companies are looking for solutions to predict house prices in the future because it is considered as an important thing for the success of a residential property project. One of the benefits is to provide consultation on the process or plans of a client who wants to buy a house in the future. The information in regard to the development of residential properties from time to time is included in Residential Property Price Index (RPPI). The Residential Property Price Index (RPPI) is an economic indicator that provides information on residential property developments both in the current and future months. One of the information that can be obtained from Residential Property Price Index (RPPI) is how the development of sales in residential properties. Residential Property Price Index (RPPI) data is obtained from Residential Property Price Survey (RPPS) which is conducted regularly by Bank Indonesia to obtain the information on residential property developments, both in the current month and the prediction for the following month.

The movement in Residential Property Price Index (RPPI) can be caused by several factors. For example, when the growth rate is hampered, it happens due to a high loan interest rates. Stohldreier [8] conducted a study on the factors that influence the house prices or Residential Property Price Index (RPPI) in China from the supply and demand side using the unbalanced dynamic panel method.
According to Lee [17] population, stock prices and construction costs have a positive effect only on house prices in Switzerland. Moreover, it is also found that loan interest rates have a negative effect on house prices. A research conducted by Anari and Kolari [2] also shows that inflation affects the house prices. The same thing was also found by Dougherty-Van Order [9] regarding inflation or the increase in price of goods and services continuously affects the house prices.

Residential Property Price Index (RPPI) modeling can be done using time series modeling. Time series is a range of observations which is collected sequentially in a time interval or periodically [3]. Time series data can be recorded based on daily, weekly, monthly, annual, or other specific time periods with the same period of time [6]. The examples are stock price data, interest rates, or Indonesian rupiah exchange rates. The time series model was first introduced by Box and Jenkins in 1970, which became known as Box-Jenkins ARIMA. Time series, which is a measurement of one variable that varies from time to time and being collected periodically is called a univariate time series. Meanwhile, the multivariate time series is a measurement of several variables which vary from time to time and being collected periodically. Time series modeling not only can be done for univariate data but also for multivariate data. One of the goals of modeling using time series is to forecasting or predicting.

Transfer function models is a multivariate time series forecasting models that combines several characteristics of one-variable ARIMA models with several characteristics of regression analysis. If the time series $y_t$ is related to one or more other time series $\{x_t\}$, then a time series model can be created to estimate the value of $y_t$ based on the information of $x_t$. The difference between transfer function model and linear regression models is in the type of data used. In the transfer function model, time series data used are not independent of each other between its periods, while in the linear regression model, the opposite data is used. Transfer function analysis is an alternative to solve the problem if there is more than one time series or multivariate time series.

Transfer function model is a model that describes the value of future prediction of a time series (named as the output series or $y_t$) based on the past values from the series itself ($y_t$) and also based on one or more associated time series (named as the input series or $x_t$) with the output series. The input series is mapped against the output series using a function, which is called a transfer function.

When viewed from the number of input variables, transfer function model is divided into two, namely the single input transfer function model and the multi input transfer function model. The single input transfer function model is a bivariate time series model, because there is one output variable and one input variable. Meanwhile, the multi-input transfer function model is a multivariate time series model because there is one output variable which is influenced by more than one input variable.

Therefore, in this study, the transfer function model will be used with the output result is the Residential Property Price Index (RPPI) and the factors suspected of influencing it as an input, so the model will be used in this study is a multi-input transfer function model.

2. Materials and Method
   - Output Series
     RPPI (Residential Property Price Index)
     The Residential Property Price Index (RPPI) contains information of the movement and growth of residential property prices in certain cities in Indonesia. The Residential Property Price Index (RPPI) is an economic indicator that provides information on residential property developments both in the current and future months. This data which is routinely issued by Bank Indonesia includes information about the movement of residential property prices from time to time.
   - Input Series:
     a. IHSG (Composite Stock Price Index)
     Stock price is the price that occurs on the stock exchange at a certain time determined by market performers, namely market supply and demand. The stock price determines the
Maximizing shareholder wealth translates to maximizing the stock price of the company, the stock price at any given time will depend on the cash flow that investors are expected to receive in the future if investors buy shares [5]. The Composite Stock Price Index (IHSG) is one of the stock market indices used by the Indonesia Stock Exchange (IDX).

b. Inflation
Inflation is a tendency to increase the prices of goods and services in general which takes place continuously [20]. If the price of goods and services in the country increases, then inflation will increase. The continuous increase in the prices of goods and services causes a decrease in the value of money because people's purchasing power will decrease [20].

c. Loan Interest Rate
The loan interest rate is the amount charged on the principal of the loan by the lender to the borrower for use of the asset. According to Brigham [5], the price that must be paid for loan capital, and dividends and capital gains resulting from equity capital, is called the loan interest rate.

The steps that will be taken in this study are:

1. The Prewhitening of Input and Output Series
The prewhitening process in the input series is carried out to make it easier for being managed, or in other words, it eliminates all existing patterns in the series so that it will become a white noise.

\[ \alpha_t = \frac{\theta_x(B)}{\delta_x(B)} x_t \]  

(1)

The form of the equation above is the result of prewhitening for the input \( x_t \) series. The series of \( \alpha_t \) is obtained from the prewhitening process of input \( x_t \) series which is named as prewhitened input. Prewhitening on the output \( y_t \) series does not necessarily change into a white noise series. The prewhitening process on the output \( y_t \) series will produce a new series, namely the prewhitened output:

\[ \beta_t = \frac{\theta_y(B)}{\delta_x(B)} y_t \]  

(2)

2. Cross-correlation of Input and Output Series
After obtaining the input and output series which have been through the prewhitening process, then the next step is to calculate the cross correlation between the two prewhitened series. Cross correlation function (CCF) is used to measure the strength and direction of the correlation between two random variables at a \( k \) time difference. The cross correlation between \( x_t \) and \( y_t \) determines the level of correlation between the value \( x_t \) at \( t \) time and the value \( y_t \) at \( t + k \) time.

3. The Identification of Impulse Response Weight
The next step after prewhitening process in the input and output series, then getting a cross correlation between the two series by estimating the impulse response weight \( v(B) \). The calculation of the impulse response weight which is carried out in this stage aims to calculate the noise series from each input series model, the weight of impulse response is obtained through the equation:

\[ v_k = r_{\alpha\beta}(k) \frac{s_y}{s_x} \]  

(3)

4. The Determination \((b, r, s)\) of Transfer Function Models of Input and Output Series.
After estimating the impulse response weight, it can be determined \((b, r, s)\) for each input series in the transfer function model by looking at the cross-correlation plot between the prewhitened
input and output. The ‘b’ value indicates that \( y_t \) begins to be influenced by \( x_t \) in the period of \( t + b \). The value of \( s \) states how long the \( y_t \) series still affected by the values of \( x_t \) or \( y_t \) input series which affected by \( x_{t-b-1}, x_{t-b-2}, \ldots, x_{t-b-s} \). The \( r \) value states that \( y_t \) is affected by its past value, that is \( y_{t-1}, \ldots, y_{t-r} \).

5. The Estimation and Determination of The Noise Series Model \((n_t)\)

Impulse response weight estimation is used to obtain the estimated value of the noise series. To estimate the noise series of the multi-input transfer function model by as many as ‘p’ input, modifications can be done [10], so that it becomes:

\[
\hat{n}_t = y_t - \hat{v}_j(B)x_{jt} \quad (4)
\]

\[
\hat{n}_t = y_t - \hat{v}_1(B)x_{1t} - \hat{v}_2(B)x_{2t} - \cdots - \hat{v}_p(B)x_{pt} \quad (5)
\]

After obtaining the values of the noise series based on equation (5), so the noise series \((n_t)\) then will be modeled with ARIMA so that the order of \( p \) and \( q \) will be obtained. The corresponding noise \((n_t)\) series model can be stated by:

6. The Assessment of Transfer Function Model Parameters

By maximizing the log likelihood function in equation (6), an estimate will be obtained for each parameter contained in the transfer function model.

\[
L(\delta, \omega, \phi, \theta, \sigma_a^2 | b, x, y, x_0, y_0, e_0) = \left( 2 \pi \sigma_a^2 \right)^{- \frac{n}{2}} \exp \left[ - \frac{1}{2\sigma_a^2} \sum_{t=1}^{n} e_t^2 \right] \quad (6)
\]

7. Diagnostic Test for Transfer Function Model

a. Based on ACF and PACF from the residual value of \( e_t \)

   For the corresponding model, each of ACF dan PACF from \( e_t \) will show the random pattern or it has been a white noise.

b. Based on the cross correlation of \( \alpha_{jt} \) and \( e_t \)

   At this stage, there will be an examination in the cross correlation between the prewhitened input series, namely the \( \alpha_{jt} \) series and the \( e_t \) residual series to determine whether the two series are independent of each other.

3. Result and Discussion

In the analysis of time series, if a series is non stationary, the differentiation will be made to convert it to a stationary series. This stage also applies in the formation of transfer function model. The input series \((X_t)\) and the output series \((Y_t)\) that will be used in the formation of transfer function model also must be a stationary series. If the initial variables of the input and output series are non stationary, then differentiation is done to each of the series so that they reach the stationary.

| Variabel   | t-statistic | \(p\)-value |
|------------|-------------|-------------|
| IHSG       | -1.287511   | 0.6328      |
| IHSG diff-1| -8.576618   | 0.0000      |

Based on Table 1 which is the result of the Augmented-Dickey Fuller test for the Composite Stock Price Index (IHSG) input series in differencing-1, it is found that the customer price index (CPI) series has reached stationary with \( p\)-value = 0.0000 < \( \alpha = 5% \). So it can be concluded that the input series of the Composite Stock Price Index (IHSG) has been stationary. In the same way, a stationary test was also carried out on each variable, with the following results presented in Table 2.
Table 2. Input and Output Series of Stationary Test Results.

| Variable          | t-statistic | p-value | Result       |
|-------------------|-------------|---------|--------------|
| Inflation         | -8.724379   | 0.0000  | Stationary   |
| Loan Interest Rate| -0.051757   | 0.9506  | Non Stationary |
| RPPI              | -2.788500   | 0.0638  | Non Stationary |

According to the Table 2, it can be seen that in addition to the inflation input series that has been stationary in the actual data, the other four variables have not yet stationary. For this reason, the first differencing process is carried out for each variable that is have not stationary yet. The results of the stationarity test for the first differencing can be seen in the Table. 3 and the result is that all variables have reached stationary.

Table 3. The Result of Differentiation -1 for nonstationary series.

| Variable          | t-statistic | p-value | Result       |
|-------------------|-------------|---------|--------------|
| IHSG Diff-1       | -8.576618   | 0.0000  | Stationary   |
| Loan Interest Rate Diff-1 | -2.897081   | 0.0495  | Stationary   |
| RPPI Diff-1       | -3.9657     | 0.01417 | Stationary   |

3.1 Prewhitening of Input and Output series

After ensuring the input and output series are stationary already, the next step to do is to determine some temporary ARIMA models for the input series based on significant lags in the ACF and PACF plots.

Table 4. Some ARIMA Models for IHSG.

| Model             | AIC    | L-Jung Box | P-value | Result             |
|-------------------|--------|------------|---------|--------------------|
| ARIMA (1,1,1)     | 13.19488 | 9.0776     | 0.9102  | White Noise        |
| ARIMA (1,1,0)     | 13.17442 | 9.2975     | 0.9305  | White Noise        |
| ARIMA (0,1,0)     | 13.14177 | 10.895     | 0.8965  | White Noise        |

According to the Table 4, it can be seen that all of the ARIMA model candidates for the Composite Stock Price Index (IHSG) input series have reached the white noise process based on the results of the L-Jung Box test on each residual of the proposed model. The best model will be selected based on the model that has the smallest AIC value, then the selected model is ARIMA (0,1,0). So the ARIMA model for the Composite Stock Price Index (IHSG) input series is:

\[ x_{1t} = 24.81947 + \alpha_{1t} \]

So that the prewhitened series will be obtained from prewhitening process for the IHSG input series are:

\[ \alpha_{1t} = x_{1t} - 24.81947 \]

output series of prewhitening, as follows:

\[ \beta_{1t} = y_{1t} - 0.686895 \]

Using the same method as the previous IHSG input series, a prewhitenined series model for other inputs will be obtained in Tabel 5. as follows :
Table 5. Prewhitened Input and Output for Other Input Series.

| Input Series | Prewhitened Input | Prewhitened Output |
|--------------|------------------|--------------------|
| Inflation    | $a_{2t} = (1 - 0.462762B + 0.421816B^2)x_{2t}$ $\beta_{2t} = (1 - 0.462762B + 0.421816B^2)y_{2t}$ | $-0.378479$ $-0.694385$ |
| Loan Interest Rate | $a_{3t} = (1 - 0.834332B)x_{3t} + 0.031504$ | $\beta_{3t} = (1 - 0.834332B)y_{3t}$ |

3.2 Cross Correlation of Prewhitened Input and Output Series

In Figure 1.(a) we can see a plot of the cross-correlation between the prewhitened inflation input and the prewhitened output. From this plot it can be said that the 2 month delay before inflation affects the Residential Property Price Index (RPPI), so it can be concluded that the value of $b = 2$ is because the plot has a significant cross correlation at lag = 2. Furthermore, to determine the $r$ value seen from the correlation plot pattern cross it is concluded that $r = 2$ because it looks like a sine wave pattern. Then the value of $s$ is obtained, namely $s = 1$, because after lag 0 there is 1 significant lag. In the same way $b$, $r$, $s$ will be determined for the other inputs from the cross-correlation plot.

3.3 Identification of Impulse Response Weight

Impulse response weight can be obtained using the formula in the Equation (3). This step useful for estimating the noise series on the next step.
3.4 Determination \((b, r, s)\) of the input and output series transfer function models.

So it can be concluded that each \((b, r, s)\) for each existing input series is as follows, which is presented in Table 4.6. From the cross-correlation plot, four models with different orders \((b, r, s)\) can be proposed, then each model will be compared through a diagnostic test at the final stage.

Table 6. Proposed Models with the order \((b, r, s)\) of each variable based on the cross correlation plot.

| Variable       | Model-1 order \((b,r,s)\) | Model-2 order \((b,r,s)\) | Model-3 order \((b,r,s)\) | Model-4 order \((b,r,s)\) |
|----------------|--------------------------|--------------------------|--------------------------|--------------------------|
| Loan Interest  | (0,1,0)                  | (0,0,0)                  | (0,0,0)                  | (0,1,0)                  |
| Rate Inflation | (2,2,1)                  | (2,2,1)                  | (2,2,1)                  | (2,2,1)                  |
| IHSG           | (0,0,0)                  | (0,0,0)                  | (0,2,0)                  | (0,2,0)                  |

3.5 The Estimation and Determination of the Noise Series Model

The estimated noise series is obtained by equation (5). The AR model (1) for the noise series has achieved white noise based on the L-Jung Box test, so the AR (1) model for the noise series will be used in the multi input transfer function model.

![Figure 2. Plot (a) ACF and (b) PACF for the noise series.](image)

In Figure 2 can be seen the ACF and PACF plot of the noise series which can be used to determine the ARIMA model of the noise series.
3.6 Diagnostic Test for Transfer Function Model

In Table 7, it can be seen that the ACF and PACF coefficients for the four models along with the Portmanteau test results on the Q-statistic value at each lag have met the diagnostic test. Also the cross correlation between noise model and prewhitened each input series independent. So it can be concluded that the residuals of the four proposed models are not autocorrelated at each lag, which can be seen from the p-value for each lag, so it have been a white noise.

Table 8. Mean Absolute Percentage Error for the Proposed Models

| Model  | MAPE  |
|--------|-------|
| Model-1 | 0.571 |
| Model-2 | 0.596 |
| Model-3 | 0.612 |
| Model-4 | 0.607 |

In Table 8. it can be seen that the model with the smallest Mean Absolute Percentage Error (MAPE) is model-1 with a MAPE value = 0.571, which means that model-1 on average makes forecasting errors on the data used by 0.571% so that the model is good enough to do forecasting.

The multi-input transfer function model for predicting the Residential Property Price Index is:

\[ y_t = 1,046 y_{t-1} - 0,09 y_{t-2} + 1,036 y_{t-3} - 0,992 y_{t-4} + (-0,00049)(1 - 0,999B)^2(1 + 0,952B + 0,994B^2)x_{1t} + (-0,025 + 0,07B)(1 - 0,999B)^2x_{2t-2} + (2,911)(1 - 0,999B)(1 + 0,952B + 0,994B^2)x_{3t} + (1 + 0,952B + 0,994B^2)(1 - 0,999B)e_t \]

3.7 Forecasting Results

Based on Table 9, which contains the results of forecasting of the RPPI using the model-1, it is known that the RPPI in January was 226.25. Next in February it is 226.12. Then in March it was 225.79. Based on the model that has been

4. Conclusion

Forecasting results from the RPPI using the model-1 above, it is known that the RPPI in January was 226.25. In the February it is 226.12. Then, it was 225.79 in March. Based on the model that has been
obtained, it can be seen that the RPPI forecast at \( t \)-time is influenced by the amount of inflation in the previous two months \((t-2)\) time to the previous five months \((t-5)\) time. RPPI also influenced by loan interest rates without delay so that it is influenced by the \( t \)-time to the previous three months \((t-3)\) time.

Then it was also influenced by the IHSG without delay, which was influenced by the \( t \)-time until the previous four months \((t-4)\) time. The RPPI index at \( t \)-time is also affected by itself at \((t-1)\) time or one month earlier to the previous four months.

As for things that might be done to continue the following study is to observe other influencing variables regarding to RPPI and adjust to environmental conditions

References

[1] Atmaja L S 1997 Memahami Statistika Bisnis Edisi Pertama Cetakan Pertama Penerbit Andi, Yogyakarta

[2] Anari A and Kolari J 2002 House prices and inflation Real Estate Economics 30(1) 67-84

[3] Box G E, Jenkins G M, Reinsel G C, and Ljung G M 2015 Time series analysis: forecasting and control John Wiley & Sons

[4] Box G E P and G H Jenkins 1976 Time-Series Analysis: Forecasting and Control Revised Edition San Francisco: Holden Day

[5] Brigham E F and Joel F Houston 2001 Fundamentals of Financial Management Alih Bahasa Dodo Suharto dan Herman Wirawan Edisi Kedelapan Buku Satu Erlangga Jakarta

[6] Cryer J D 1986 Time series analysis (Vol. 286) Boston: Duxbury Press

[7] Cryer J D and Kellet N 1991 Time series analysis Royal Victorian Institute for the Blind Tertiary Resource Service.

[8] Stohldreier M T 2012 The determinants of house prices in Chinese cities Mater Thesis of University of Zurich.

[9] Dougherty A and Van Order R 1982 Inflation, housing costs, and the consumer price index The American Economic Review 72(1) 154-164

[10] Edlund P O 1984 Identification of the multi-input box-Jenkins transfer function model Journal of Forecasting 3(3) 297-308

[11] Elton E J, Gruber M J, and Blake C R 1995 Fundamental economic variables, expected returns, and bond fund performance The Journal of Finance 50(4) 1229-1256

[12] Granger C W J and P Newbold 1977 Identification of Two-way Causal Systems Frontiers in Quantitative Economics Vol. IIIA (Intriligator, H.D., ed.) Amsterdam: North- Holland

[13] Haugh L D and G E P Box 1977 Identification of Dynamic Regression (Distributed Lag) Models Connecting Two Time Series Journal of the American Statistical Association 72 121 130

[14] Hogg R V, McKean J, and Craig A T 2005 Introduction to mathematical statistics Pearson Education

[15] Huffle S V and Vandewalle J 1989 Comparison of total least squares and instrumental variable methods for parameter estimation of transfer function models International journal of control 50(4) 1039-1056

[16] Jonathan D C and Kung-Sik C 2008 Time series analysis with applications in R SpringerLink Springer eBooks

[17] Lee C L 2009 Housing price volatility and its determinants International Journal of Housing Markets and Analysis

[18] Lii K S 1985 Transfer function model order and parameter estimation Journal of Time Series Analysis 6(3) 153-169

[19] Liu L M and Hanssens D M 1982 Identification of multivariate-input transfer function models Communications in statistics-theory and methods 11(3) 297-314

[20] Lipsy R G And Steiner 1981 Economics Sixth edition Harper and Row publisher New York

[21] Ljung G H and G E P Box 1979 The Likelihood Function of Stationary Autoregressive

[22] Moving Average Models Biometrika 66, 265-270