Regular Proceeding International Seminar on Innovation in Mathematics and Mathematics Education
2nd ISIM-MED joint Conference on Innovation and Technology for Mathematics and Mathematics Education 23nd ATCM : "I*nnovative Technology in Mathematics : New Way for Learning,Teaching, Researching Mathematics*" : Yogyakarta, 20-24 November 2018
ISBN 978-602-73403-4-3

Forecasting Indonesian GDP growth using mixed data sampling (MIDAS) regression

D T Utari[1] and D Rosadi[2]

[1]Department of Statistics, Faculty of Mathematics and Natural Science, Universitas Islam Indonesia, Yogyakarta, Indonesia
[2]Department of Statistics, Faculty of Mathematics and Natural Science, Universitas Gadjah Mada, Yogyakarta, Indonesia

E-mail: dina.t.utari@uii.ac.id

Abstract. MIDAS regression model is a time series model that can be used to model the data with different frequencies or mixed frequencies without losing information from the data. In this paper, we discuss the MIDAS regression model using Exponential Almon function and compare its performance with the distributed lag model. We also apply the model for forecasting purpose. For empirical study, we apply the model to forecast Indonesian Gross Domestic Product (GDP) data. Compared to the distributed lag model, it turns out that the MIDAS regression model gives a smaller error value in forecasting GPD.

1. Introduction

The frequency difference in time series data has been a challenge for researchers who work on time series analysis, especially on time series regression. This is because time series regression involves independent and dependent variables with the same frequencies.

A common solution to solving the issue of data with different frequencies is to transform the data so that the dependent and independent variables have the same frequencies. However, this transformation process may cause some lost useful information and difficulties on detecting variable.

In 2004, Ghysels, Santa-Clara, and Valkanov found a model that can overcome the issue of time series data with different frequencies or mixed frequencies without losing information from the data [2]. This model is called Mixed Data Sampling (MIDAS) regression model. The advantages of MIDAS regression model in addition to overcoming the problem of data with mixed frequency is to minimize the parameters being estimated and make the regression model become simpler. The MIDAS regression model can retain information in different data frequencies between dependent and independent variables and reduce the number of parameters being estimated. Therefore, this model is more appropriate for forecasting in comparison with other classical models.

In this paper, we apply the MIDAS regression model to forecast the Indonesian GDP growth using the return of the Jakarta Composite Index (JCI). This is based on the fact that the macroeconomic variables are important indicators of economic growth, but usually, the variable macroeconomic data are in low frequencies, such as the quarterly period for GDP, monthly inflation, and monthly employment growth. On the other hand, the financial variables such as the return of JCI which contains useful information for future economic growth [5].

ISBN 978-602-73403-4-3

The first section of this paper describe the background of the research, while the second section discuss about method used in this research. The third section explains the source of data and tools for analyzing the data. The empirical results and discussions are then provided in the fourth section. Finally, the conclusions are presented in fifth section respectively.

2. Method

One of the main objectives of the MIDAS regression model is to deal with the problem of frequency differences between dependent and independent variables without having to eliminate important information from both variables. The implementation of this MIDAS regression model is to forecast quarterly Indonesian GDP using its own lags and monthly return of JCI.

After fitting the MIDAS regression model and distributed lag model, the selected model is the model with smaller mean square error. The MIDAS regression model has a particular shape associated with the polynomial weight function specification. The polynomial weighting function is used to help incorporate independent variables that have a high frequency in the MIDAS regression model.

In the MIDAS regression model, the dependent variable Y_t is assumed to have a fixed frequency or period, called as the interval of reference. Meanwhile, the independent variable $X_t^{(m)}$ has a higher frequency. Let Y_t have a quarterly frequency, then at least $X_t^{(m)}$ has a monthly frequency where $m = 3$.

The following equation is the form of the MIDAS regression model:

$$Y_t = \beta_0 + \beta_1 \left(b(0;\theta)X_{t-0/m}^{(m)} + b(1;\theta)X_{t-1/m}^{(m)} + \cdots + b(K-1;\theta)X_{t-(K-1)/m}^{(m)} \right) + \varepsilon_t^{(m)} \quad (1)$$

where $t = 1, 2, \ldots, T$, with the parameter β_1 contains the overall impact of the lag $X_t^{(m)}$ from Y_t, whereas ε is a model error that is normally distribution. Standardizing lag operator, the model (1) can be written as follows:

$$Y_t = \beta_0 + \beta_1 B\left(L^{1/m};\theta\right)X_t^{(m)} + \varepsilon_t^{(m)} \quad (2)$$

where $B\left(L^{1/m};\theta\right) = \sum_{k=0}^{K-1} b(k;\theta)\, L^{k/m}$ must sum to one and $L^{1/m}$ is lag operator.

The determination of the parameters of the weighting function $b(k;\theta)$ becomes important in the MIDAS regression model. The use of weight function in MIDAS regression aims to maintain the simplicity of the model.

The weight function used is called an Exponential Almon since it is closely related to the Almon polynomial function in the distributed lag model [3]. Ghysels, Santa-Clara, and Valkanov suggested using two parameters of theta $\theta = (\theta_1, \theta_2)$. Various shapes can be made for weighting functions with two parameters, including descending, rising, or single hump shape. The Exponential Almon weighting function is expressed as follows:

$$b(k;\theta) = \frac{\exp(\theta_1 k + \theta_2 k^2)}{\sum_{l=0}^{K-1} \exp(\theta_1 l + \theta_2 l^2)}.$$

Therefore, equation (2) is written as

$$Y_t = \beta_0 + \beta_1 \left[\sum_{k=0}^{K-1} \frac{\exp(\theta_1 k + \theta_2 k^2)}{\sum_{l=0}^{K-1} \exp(\theta_1 l + \theta_2 l^2)} \right] L^{k/m} X_t^{(m)} + \varepsilon_t^{(m)}. \quad (3)$$

From the Exponential Almon weight function, then the lag selection will be used in the MIDAS regression model. The selection of appropriate lags aims to help incorporate X_t variables that have higher frequencies in the MIDAS regression model. As a result, we will get the variables Y_t and X_t with the same frequency.

In the Exponential Almon weight function, the last independent variable is expressed as $L^{0/m}$, which has the greatest influence in the model. Similarly, the observations of $L^{(K-1)/m}$ will correspond to the first observation in the quarterly period.

Suppose $\emptyset$ is a set of unknown parameters in the model i.e. $\emptyset = \{\beta_0, \beta_1, \theta_1, \theta_2\}$. The estimated parameters will be solved using the Nonlinear Least Square (NLS) method. The form of the function is not linear so the iterative method Quasi Newton with the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm is used for estimating the parameters.

The Quasi-Newton method replaces derivative computing with direct computing functions. The quasi-newton method differs in how the approximate Hessian matrix is formed and updated. The simplest Quasi-Newton method sets the approximate Hessian matrix as the identity matrix which is a symmetric and definite matrix. The most famous way of updating the Hessian matrix is with the BFGS algorithm. This algorithm is known for robustness and convergence. The general equation of the Quasi-Newton method is [6]:

$$B^{r+1} s^r = y^r$$

where $s^r = \emptyset^{r+1} - \emptyset^r$ and $y^r = \nabla S(\emptyset^{r+1}) - \nabla S(\emptyset^r)$, with ϕ is the set of unknown parameters in the model and $S(\phi)$ is the sum square error, so the update matrix of the Hessian matrix B is

.

$$B^{r+1} = B^r - \frac{B^r s^r (s^r)^T B^r}{(s^r)^T B^r s^r} + \frac{y^r (y^r)^T}{(s^r)^T y^r}$$

3. Data

The data used in this study are the Indonesian GDP data based on 2000 Constant Price i.e. 2000Q1 to 2014Q4 and JCI closing price data as the benchmark index for Indonesian share price of monthly period available from January 2000 to December 2014. The Indonesian GDP data is sourced from Badan Pusat Statistik, while JCI data is obtained from Yahoo Finance.

To estimate parameters of the MIDAS regression model, the Indonesian GDP data from 2000Q1 to the 2012Q2 and JCI data from January 2000 to June 2012 are used. Furthermore, forecasting of the Indonesian GDP values for the next 10 quarters, from 2012Q3 until 2104Q4 is conducted using the JCI data from July 2012 until December 2014.

4. Empirical Result and Discussion

To calculate the growth rate of GDP and return of JCI, the following formula is used:

$$r_t = ln\left(\frac{P_t}{P_{t-1}}\right) * 100$$

where

r_t : GDP growth rate or return of JCI

P_t : value of GDP at time t or closing value of JCI at time t

P_{t-1} : value of GDP at time $t-1$ or closing value of JCI at time $t-1$

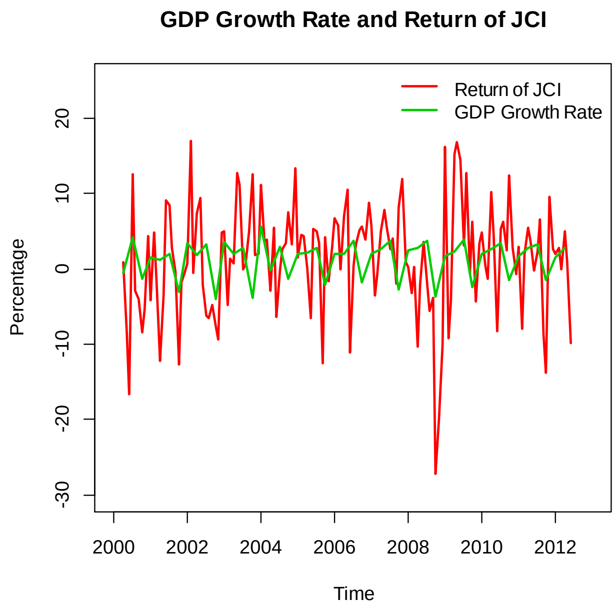

Figure 1. Indonesian GDP growth rate and return of JCI in 2000 – 2012

Figure 1. shows that the structure of the Indonesian GDP growth rate and the return of JCI had both constant and near zero fluctuations. Therefore, the data were considered stationary. The p-value of the Augmented Dickey-Fuller (ADF) test is 0.01, so it can be concluded that the data of Indonesian GDP growth rate and the return of JCI is stationary.

In this empirical study, 10 lags ($K = 10$) were selected in fitting the MIDAS regression. The AIC and BIC are used to select the optimal lag.

Table 1. Value of AIC and BIC model of MIDAS regression

Lag	AIC	BIC
Lag 0 – 5	224.6038	233.9598
Lag 0 – 6	219.2729	228.5236
Lag 0 – 7	221.4874	230.7381
Lag 0 – 8	219.2729	228.5236
Lag 0 – 9	219.8951	229.0383

Table 1. shows the AIC and BIC for several lag categories. The lag 0 – 6 has the smallest AIC and BIC values in comparison with other lags. Hence the estimated MIDAS regression function for the Indonesian GDP growth rate and the return of JCI is given below,

$$\widehat{Y}_t = 1.03 + 0.17 \left(\sum_{k=0}^{6} \frac{exp(3.83k - 0.59k^2)}{\sum_{l=0}^{6} exp(3.83l - 0.59l^2)} \right) L^{k/3} X_t^{(3)}.$$

4.1. Comparison of MIDAS Regression Model and Distributed Lag

In order that the variables Y_t and X_t have the same frequency, then one solution is finding the average of the variable X_t. Then the transformed data are applied to the distributed lag model, using the same lag as is in the regression model, lag 0 to lag 6. The distributed lag model using two-degree Almon approach is given below,

$$Y_t = \sum_{k=0}^{6} \beta_k X_{t-k} + \varepsilon_t = \sum_{k=0}^{6} \sum_{j=0}^{2} \theta_j k^j X_{t-k} + \varepsilon_t$$

The estimated distributed lag equation for Indonesian GDP growth rate and the return of JCI is given below,

$$\widehat{Y_t} = 0.05X_t + 0.02X_{t-1} + \cdots + 0.04X_{t-6}.$$

Table 2. Value of Root Mean Square of Error (RMSE)

Model	**RMSE**
MIDAS Regression Model	2.24
Distributed Lag Model	2.50

Table 2. shows the RMSE for both models. The MIDAS regression model has smaller RMSE than the distributed lag. This result of this case study confirms that the MIDAS regression is better for modelling data with variables with mixed frequency [2].

4.2. Forecasting Indonesian GDP

The Indonesian GDP from 2012Q3 to 2014Q4 are used to forecast the next 10 months.

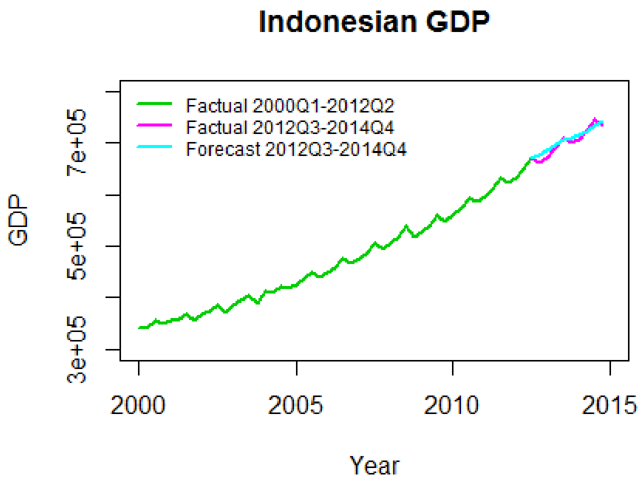

Figure 2. Indonesian GDP and forecast result

Figure 2. shows the results of forecasting for the third quarter of 2012 until the fourth quarter of 2014 increased in every quarter. An increasing of GDP in the next period is expected to be an indicator that state productivity is also high and there is an increase in the welfare of its society. The mean absolute percentage error (MAPE) of Indonesian GDP forecasting is 1.14%.

5. Conclusion

Based on the research that has been done, we can conclude that the best model that can predict the Indonesian GDP is the MIDAS regression model. This is based on the results that the MIDAS regression model produces a smaller error (RMSE) compared to the distributed lag model.

References

[1] Almon S 1965 *The Distributed Lag Between Capital Appropriations and Expenditures* (Econometrica) p 178 – 196

[2] Ghysels E Santa-Clara P and Valkanov R 2004 *The MIDAS Touch: Mixed Data Sampling Regression Models* (Anderson Graduate School of Management UCLA)

[3] Ghysels E Santa-Clara P and Valkanov R 2006 *Predicting Volatility: Getting The Most Out of Return Data Sampled at Different Frequencies* (Journal of Econometrics) p 59 – 95

ISBN 978-602-73403-4-3

[4] Ghysels E Kvedaras V and Zemlys V 2015 *Mixed-Frequency Data Sampling Regression Models: The R Package midasr* (Journal of Statistical Software)
[5] Stock S A and Watson M W 2003 *Forecasting Output and Inflation: The Role of Asset Prices* (Journal of Economic Literature) p 788 – 829
[6] Yuan G and Lu X 2011 *An active set limited memory BFGS algorithm for bound constrained optimization* (Applied Mathematical Modelling) p 3561 – 3573