RESEARCH ARTICLE | OCTOBER 17 2018

Comparison of methods for mixed data sampling (MIDAS) regression models to forecast Indonesian GDP using agricultural exports FREE

Dina Tri Utari ; Hafizah Ilma

AIP Conf. Proc. 2021, 060016 (2018)
https://doi.org/10.1063/1.5062780

Comparison of Methods for Mixed Data Sampling (MIDAS) Regression Models to Forecast Indonesian GDP Using Agricultural Exports

Dina Tri Utari[1,a)] and Hafizah Ilma[1,b)]

[1]*Departement of Statistics, Islamic University of Indonesia, Sleman, Yogyakarta, Indonesia*

[a)]Corresponding author: dina.t.utari@uii.ac.id
[b)]15611161@students.uii.ac.id

Abstract. Indonesia is an agrarian country and has agricultural land and abundant resources. Therefore, it is expected that Indonesia can utilize existing natural resources to increase the export value of agricultural products, which will impact the value of the GDP. This can improve the welfare of Indonesian society in general, and farmers in particular. However, the data to forecast the growth of the Indonesian GDP using the export value of the agricultural sector has unequal frequency data. Therefore, a special regression model, namely, the Mixed Data Sampling (MIDAS) regression model introduced by Ghysels, Santa-Clara, and Valkanov (2004) is applied. The advantages of MIDAS, in addition to overcoming the problem of data with mixed frequency, is to minimize the number of estimated parameters and make the regression model simpler. A weighting function is used to reduce the number of parameters in the MIDAS regression. The weighting function can have a number of functional forms. Ghysels, Santa-Clara, and Valkanov suggest the Exponential Almon function and the Beta function, then compare their performance with the distributed lag model. This research proves that, based on the Root Mean Square Error, the MIDAS Beta regression model yields a better model estimation than either the MIDAS Exponential Almon or the distributed lag model in the case of forecasting the growth of the Indonesian GDP using the export value of the agricultural sector.

Keywords: Agriculture, exponential, regression.

INTRODUCTION

Some econometric models have been developed to forecast time series data using explanatory variables. One application of this is to forecast macroeconomic variables, something that has become an important task for governments, central banks, financial companies, and other entities. The result of the macroeconomic forecasting depends on the current business or economic conditions. Unfortunately, many important macroeconomic indicators are not sampled at the same frequency. For example, Gross Domestic Product (GDP) data are sampled every three months (quarterly), export and import value data are sampled every month, and most stock data are sampled daily.

The Mixed Data Sampling (MIDAS) regression model was introduced by Almon for dealing with data with different frequencies or mixed frequencies without losing information from the data.[1] It is designed to find a balance between retaining the individual timing information of the high-frequency data and reducing the number of parameters that need to be estimated.

Research using the MIDAS regression model mostly focuses on developed country economies to forecast macroeconomic data, e.g., Anthony and Clements and Galvao, but only a few have focused on the economies of developing countries.[2,3] A country's economy is influenced by exports. The revenue from the exports of a country will directly increase the income of the country. An increase in the income of a country will also result in an increase in GDP.

The 8th Annual Basic Science International Conference
AIP Conf. Proc. 2021, 060016-1–060016-6; https://doi.org/10.1063/1.5062780
Published by AIP Publishing. 978-0-7354-1739-7/$30.00

Therefore, it is interesting to examine the MIDAS regression model to forecast the Indonesian GDP. It is based on macroeconomic variables as important indicators of economic development, but usually macroeconomic data is available at low frequency.

In this paper, we apply the MIDAS regression model to forecast the growth of the Indonesian GDP using the value of Indonesian agricultural exports. This study focuses on the export value of the agricultural sector, considering that Indonesia is an agricultural country and has abundant agricultural land and resources. Therefore, it is expected that Indonesia can utilize the available natural resources to increase the values of exports, especially for agriculture, and this will increase the GDP.

The rest of this study is organized as follows. First, it will describe the background of the research. Second, it discusses the models and the methodology used in this study. Third, it analyzes the data. The empirical results are provided in after.

METHODOLOGY

In this study, we use Indonesian quarterly GDP data and monthly agricultural export data. The relationship between these two variables is that exports will directly increase a country's income. It is expected that an increase in a country's income will also result in an increase of its GDP. We use the Mixed Data Sampling (MIDAS) regression model to deal with a period or frequency difference issues of GDP and export variables.

The exponential Almon weight function and the Beta weight function are used to estimate the parameters in the MIDAS regression model.[1] Then the results of the model with these two functions are compared with the lag distributed model. The model that gives the smallest error is then used to estimate the GDP of Indonesia.

The MIDAS Regression Model

To determine the MIDAS regression model, the dependent variable Y_t is assumed to have been sampled with a fixed sampling frequency, called the interval of reference. The independent variable is then denoted by $X_t^{(m)}$, with m being the frequency of the independent variable observed during the interval of reference.

With the help of this notation, the MIDAS regression model can be expressed as

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

where $t = 1, 2, \ldots, T$, with the parameter β_1 capturing the overall impact of the lagged variable $X_t^{(m)}$ on Y_t since the dependent variable is only sampled once between t and $t+1$.

Model (1) can be rewritten with the help of a standard lag operator:

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

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 the lag operator.

One of the most important parts of the MIDAS regression model is the determination of the parameters of the weighting function $b(k;\theta)$. In addition, parameter determination is also important to ensure that the weighting function always sums to one. Another purpose of the weighting function in MIDAS regression is to achieve flexibility and maintain model simplicity.

The method of determining the parameters used is called 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 for theta $\theta = (\theta_1, \theta_2)$. The Exponential Almon weighting function is expressed as follows:

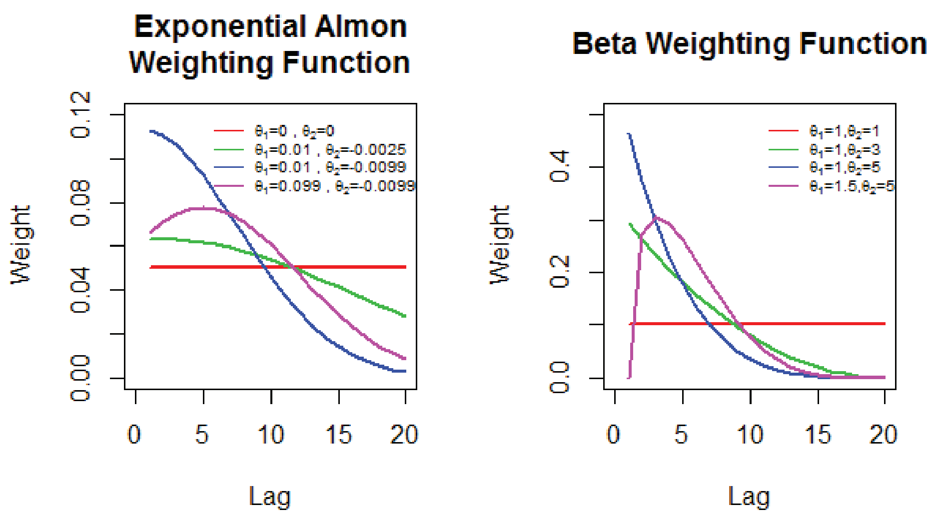
$$b(k;\theta) = \frac{exp\,(\theta_1 k + \theta_2 k^2)}{\sum_{l=0}^{K-1} exp(\theta_1 l + \theta_2 l^2)} \qquad (3)$$

In contrast to the Exponential Almon function that is formed based on the Almon polynomial function in the distributed lag model, the Beta function used is independently formed. The Beta function has two parameters theta $\theta = (\theta_1, \theta_2)$ that is

27 February 2024 22:55:13

$$b(k;\theta_1,\theta_2) = \frac{x_i^{\theta_1-1}(1-x_i)^{\theta_2-1}}{\sum_{l=0}^{K-1} x_l^{\theta_1-1}(1-x_l)^{\theta_2-1}} \tag{4}$$

where $x_i = \frac{(k-1)}{K}$.



(a) (b)

FIGURE 1. Weighting Function for (a) exponential almon weighting function, and (b) beta weighting function.

Fig. 1(a) illustrates the flexibility of the Exponential Almon weighting function in the case of two parameters. First, it appears that for $\theta = (\theta_1, \theta_2)$, we get the same weight (red panel). Second, the weight can be decreased slowly (green panel) or quickly (blue panel). Finally, the Exponential Almon weighting function can produce a single hump shape, as shown in the magenta panel.

On the other hand, Fig. 1(b) displays various shapes for various values of θ_1 and θ_2. The Beta weighting function can also produce many shapes, as shown in Fig. 1(b). For example, it appears that for $\theta_1 = \theta_2 = 1$ we get the same weight (red panel). The green panel shows a case of slow weighting decrease corresponding to $\theta_1 = 1$ and $\theta_2 > 1$. As the value of θ_2 increases, weighing decreases faster, as shown in the blue panel. Finally, the magenta panel illustrates a single hump shape pattern formed for $\theta_1 = 1.5$ and $\theta_2 = 5$. The flexibility of Beta functions is well known. It is often used in Bayesian econometrics to impose flexible prior distributions.

Parameter Estimation of MIDAS Regression Model

In the estimation of the MIDAS regression model, it is necessary to select the number of lags used. In the MIDAS regression model, the number of lags used relates to the Exponential Almon weighting function $b(k;\theta)$. Furthermore, the lags in the Exponential Almon weighting function are used to help incorporate independent variables with high frequencies in the MIDAS regression model. As a result, dependent and independent variables with the same frequencies will be obtained.

When using the Exponential Almon or Beta weighting function, the last independent variable observed is expressed as $L^{0/m}$, since it is expected that it will have the greatest influence on the variable being explained. Similarly, observation 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.,

$$f(X,\emptyset) = \beta_0 + \beta_1 \sum_{k=0}^{K-1} b(k;\theta)\, L^{k/m} X_t^{(m)}$$

Then (3) becomes $Y_t = f(X, \emptyset) + \varepsilon_t^{(m)}$ where $t = 1, 2, \ldots, T$. Estimating the parameters using Nonlinear Least Squares minimizes the sum of the squares of the errors, that is,

$$\emptyset = \text{argmin}_{\emptyset} \sum_{t=1}^{T} \left(Y_t - \left(\beta_0 + \beta_1 \sum_{k=0}^{K-1} b(k; \theta) \, L^{k/m} X_t^{(m)} \right) \right)^2$$

The above minimization of the sum of the squared errors cannot be done analytically. An iteration method is required. One of the ways to do this is an iterative technique which is a quasi-Newton method: the Broyden–Fletcher–Goldfarb–Shanno algorithm.

DATA

The dataset includes Indonesian GDP data based on 2000 Constant Prices, i.e., 2010Q1 to 2017Q2 and monthly agricultural exports from January 2010 to July 2017.

To estimate the parameters of the MIDAS regression model, Indonesian GDP data from 2010Q1 to 2014Q4 and agricultural export data from January 2010 to December 2014 are used. Furthermore, the forecasting of the Indonesian GDP for the next 10 quarters, from 2015Q1 until 2107Q2, is done with the known data of agricultural exports from January 2015 until July 2017.

RESULTS AND DISCUSSION

In the MIDAS regression model, the assumption of stationary data must be made. One way to cope with non-stationary data is to do a transformation, so it is expected to obtain stationary data so that there can be done the analysis of the next step.

From Fig. 2, it can be concluded that the data structure of GDP and exports over time has constant or constant fluctuations of data, and the fluctuations in the data are around the mean value. Therefore, the data is considered stationary.

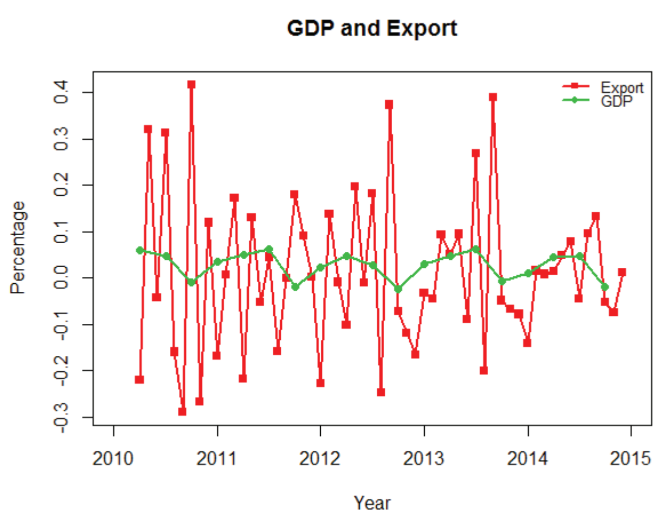


FIGURE 2. Indonesian GDP and Agricultural Export in 2010 – 2015.

This assumption is reinforced by the Augmented Dickey–Fuller (ADF) test with the null hypothesis that data is not stationary. The p-value of the test is 0.01, which is smaller than 0.05, indicating the possibility of rejecting the null hypothesis. In other words, this test concludes that the data of Indonesian GDP and exports is stationary.

In this empirical study, the parameter estimation for the MIDAS regression model with an exponential Almon function as well as with a Beta function is done by using the initial values $\beta_0 = 1$, $\beta_1 = 2$, $\theta_1 = 0.01$, and $\theta_2 = -0.01$, and five lags, i.e., $K = 5$, were selected. The optimal lag determination is based on the smallest values of the AIC and BIC. The values of the AIC and BIC for the MIDAS regression model are as follows.

TABLE 1. Value of AIC and BIC Model of MIDAS Regression Exponential Almon

Lag	AIC	BIC
Lag 0 – 2	-73.71840	-68.99620
Lag 0 – 3	-69.02082	-64.56896
Lag 0 – 4	-73.43332	-68.98146
Lag 0 – 5	-69.03390	-64.58204

TABLE 2. Value of AIC and BIC Model of MIDAS Regression Beta

Lag	AIC	BIC
Lag 0 – 2	-71.60078	-66.87859
Lag 0 – 3	-68.05646	-63.60461
Lag 0 – 4	-73.43371	-68.98185
Lag 0 – 5	-72.42327	-67.97141

Based on the smallest AIC and BIC values in Table 1, it is found that the MIDAS regression Exponential Almon with the optimal lag is the model with lags 0–2, with estimation results for the parameters of the model $\widehat{\beta_0} = 0.025981$, $\widehat{\beta_1} = 0.151466$, $\widehat{\theta_1} = 1.268359$, and $\widehat{\theta_2} = -0.428522$. The optimal lags in the MIDAS regression Beta model are, from Table 2, lags 0–4 with $\widehat{\beta_0} = 0.02620$, $\widehat{\beta_1} = 0.09705$, $\widehat{\theta_1} = 1.32180$, and $\widehat{\theta_2} = -1.32180$.

Comparison of MIDAS Regression Models and Distributed Lag Model

One of the transformations to make both variables have the same frequencies is to find the average value of the monthly export data, turning it into quarterly data. The results of this transformation can then be employed in the distributed lag model, using the same lags as in the MIDAS Exponential Almon regression model, lags 0 to 2. Below is the distributed lag model using the two-degree Almon approach.

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

The obtained parameter estimation for the distributed lag model is $\widehat{\theta_0} = 0{,}20032452$, $\widehat{\theta_1} = -0{,}25191952$, $\widehat{\theta_2} = 0{,}09313472$, $\widehat{\beta_0} = 0{,}20032452$, $\widehat{\beta_1} = 0{,}04153972$, and $\widehat{\beta_2} = 0{,}06902435$. The following table gives the RMSE values for the MIDAS regression model and the distributed lag model.

TABLE 3. Value of RMSE

Model	RMSE
MIDAS Eksponensial Almon	0.02722339
MIDAS Beta	0.02320064
Distributed Lag	0.03783329

From Table 3 it can be seen that the MIDAS regression model can reduce the number of parameters that are estimated. In addition, based on the RMSE values obtained, The MIDAS Beta regression model yielded a better estimation than either the MIDAS Exponential Almon or the distributed lag model. Furthermore, the MIDAS Beta regression model can be used to forecast the Indonesian GDP growth rate.

Forecasting Indonesian GDP

In the forecasting stage, the Indonesian GDP for the next ten steps from 2015Q1 to 2017Q2 will be forecast Table 4. In this case, the agricultural export data from January 2015 to July 2017 should be known.

TABLE 4. Forecasting Indonesian GDP

Period	Indonesian Actual GDP (Billion Rupiah)	Indonesian Forecasted GDP (Billion Rupiah)
2015 Q1	2,728,289.00	2,649,505.00
2015 Q2	2,868,797.00	2,700,321.00
2015 Q3	2,992,674.00	2,760,465.00
2015 Q4	2,941,958.00	2,814,156.00
2016 Q1	2,931,446.00	2,979,010.00
2016 Q2	3,075,135.00	3,015,920.00
2016 Q3	3,205,452.00	3,077,004.00
2016 Q4	3,194,776.00	3,170,352.00
2017 Q1	3,227,075.00	3,457,200.00
2017 Q2	3,366,764.00	3,546,984.00

Based on the obtained output values, it can be concluded that the estimated error of the Indonesian GDP forecast is 4.17% since the value of MAPE amounted to 4.17%.

CONCLUSIONS

1. In this empirical study, a parameter estimation for MIDAS regression models, either with an exponential Almon function or a Beta function, has been done by setting initial values $\beta_0 = 1$, $\beta_1 = 2$, $\theta_1 = 0.01$, and $\theta_2 = -0.01$, and using five lags ($K = 5$ was found to be optimal).

2. Based on the RMSE values obtained, the MIDAS Beta regression model yielded a better estimation than the MIDAS Exponential model with the Almon function or the distributed lag model. Furthermore, the MIDAS Beta regression model can be used to forecast the Indonesian GDP growth rate.

ACKNOWLEDGMENTS

The authors express their immense thanks to Vaidotas Zemlys, as a journal writer who has taken the time to discuss with the authors on the concept and program of MIDAS regression model. This research is a research funded by Directorate and Community Service Universitas Islam Indonesia.

REFERENCES

1. S. Almon, Econometrica **33**, 178 –196 (1965).
2. S. T. Anthony, "Financial Variables as Predictors of Real Output Growth", Working Paper, East Asian Bureau of Economics Research, 2007.
3. M. P. Clements and A. B. Galvao, J. Appl. Economet. **24**, 1187-1206 (2009).