Nowcasting Philippine Economic Growth Using MIDAS Regression Modeling

Cesar C. Rufino
School of Economics, DLSU-Manila
cesar.rufino@dlsu.edu.ph

Abstract: Among the most anticipated data releases of the Philippine statistical system is the quarterly real gross domestic product. This all important statistic provides the basis of establishing the economic growth performance of the country on a year-on-year basis. Official publication of this statistic however comes at a significant delay of up to two months, upsetting the planning function of various economic stake holders. Under this back drop, data scientists coined the term "nowcasting" which refers to the prediction of the present, the very near future and the very recent past, based on information provided by available data that are sampled at higher frequencies (monthly, weekly, daily, etc.). Nowcasting, however unlocks the "mixed frequency" problem in forecasting, which is the data frequency asymmetry between the dependent and independent variables of regression models that are used in forecasting.

The central objective of this study is to demonstrate the viability of using a state-of-the art technique called MIDAS (Mixed Data Sampling) Regression to solve the mixed frequency problem in implementing the "nowcasting" of the country's economic growth. Different variants of the MIDAS model are estimated using quarterly Real GDP data and monthly data on Inflation, Industrial Production and Philippine Stock Exchange Index. These models are empirically compared against each other and against the models traditionally used by forecasters in the context of mixed frequency. The results indicate the feasibility of adopting the MIDAS framework in accurately predicting future growth of the economy using information from high frequency economic indicators. Certain MIDAS models considered in the study performed better than traditional forecasting models in both in-sample and out-of-sample forecasting performance.

Keywords: Nowcasting; MIDAS Regression; Mixed Frequency Problem; Temporal Aggregation; Ragged Edge Problem; Bridge Equations

1. INTRODUCTION

"Nowcasting" has been a "buzzword" in the current Economic forecasting literature. It refers to the prediction of the present, the very near future and the very recent past (Giannone, Reichlin and Small, 2008), which has a lot of decision making and planning implications. Its relevance to economic planning lies in the fact that the most important indicators of economic health (gross domestic product and its components – personal consumption expenditure, gross domestic capital formation, government expenditure, etc.) are sampled and published quarterly with substantial publication delays of even up to two months, thus upsetting the planning activities of various stakeholders of the economy (the central bank, legislators, fiscal planners, financial and business firms, and others who are immensely affected by the business cycle). On the other hand, many variables sampled at higher frequencies (monthly, weekly, daily, etc.) like industrial production, inflation, monetary aggregates, interest rates, stock market index, etc., that are known to carry predictive information on future economic growth are already available and the useful information they carry can be extracted to the fullest, even before the final quarterly indicators are released. The central objective of the "Nowcasting" research is in developing models and procedures that will make this information extraction process as effective and as reliable as possible.

Relationships of variables in Economics, Finance and other fields are traditionally modeled as a form of regression equations or systems of equations, wherein all variables are sampled in the same frequency. When any or all of the regressors is/are in higher frequency than the regressand, the usual recourse, called temporal aggregation approach, is to time aggregate these variables, usually in terms of their sums or averages to conform with the sampling frequency of the dependent variable thus synchronizing the data sampling of the left hand and right hand side variables of the equation to that of the lower frequency regressand, making the analysis viable. Although computationally convenient, this recourse of solving the mixed frequency problem does not conform to our desire of extracting predictable information from the more frequently sampled regressors because of information loss and possible misspecification errors induced by the process of aggregation and might compromise the forecast quality.

An alternative option, called the individual coefficient approach, the extraction of hidden information in the higher frequency regressors may be possible if the model is augmented by the individual components of the regressors, each with its own coefficient to be estimated. For example, if the regressand is quarterly and the regressor has m components (that is, m periods in a quarter, m = 3 if the regressor is monthly, m = 66 if the regressor is daily, etc.) of this variable. This will effectively introduce a multiplier for each component, which may be interpreted as the component's marginal contribution to the regressand during the specific quarter. This option is obviously unappealing because of parameter proliferation (with consequent loss in degrees of freedom), especially if m becomes large. In the temporal aggregation option, the multipliers are all equal to 1/m, when the aggregation scheme is averaging.

The MIDAS Regression approach represents an intuitively appealing middle ground between the two options discussed above. The MIDAS (Mixed Data Sampling) approach, introduced by Ghysels, Sta Clara and Valkanov (2004) allows for non-equal weights (multipliers) for the components that are parsimoniously reparametrized through a weighing scheme anchored on the use of lag polynomials. The way lag polynomials are employed in defining the weighing scheme for the multiplier represents a specific MIDAS regression model.

In this study, MIDAS regression models are estimated and empirically matched against models traditionally used in dealing with the "mixed frequency" problem. The main conjecture is that the Midas Approach is better than the traditional regression models used in solving the "mixed frequency" and the "ragged edge problem" and can be relied upon in undertaking the "nowcasting" of the country's economic growth.

2. METHODOLOGY

Seven forecasting models are to be compared empirically based on their out-of-sample forecasting performance. Three are the traditional models used by practitioners and government economic planners, and the other four are variants of the MIDAS model.

2.1 Model 1: Temporal Aggregation

One way to address mixed frequency samples is to use some type of aggregation, perhaps summing or taking average of high-frequency data that occur between samples of the lower-frequency variable (Clements and Galvão.

2008). This is done to come up with a regression model with conformable sampling frequencies of both the dependent and the independent variables.

2.2 Model 2: VAR Forecasting Model

The forecasting capability of Vector Auto Regressive (VAR) models offers another way of using available high frequency predictors in forecasting low frequency target variables. This is done by first converting the high frequency variables into the sampling frequency of the target variable, after which, an unrestricted VAR model (Sims 1982) is constructed featuring the target variable and the time aggregated predictors, forming the vector. Forecasts are then made out-of-sample for all of the variables in the vector which are all considered endogenous. The focus of interest in this exercise is the forecast for the target variable.

2.3 Model 3: Bridge Equation Model

Another intuitive alternative in using higher-frequency data (e.g., monthly) to forecast lower frequency series (e.g., quarterly) would be to estimate a "bridge equation" (see Baffigi, et.al. (2004) and Diron (2008)). This method use popular forecasting models (such as VARs, ARIMA, ARDL, Exponential Smoothing, etc.) for each of the high frequency indicators. These models are then used in generating forecasts for the missing out-of-sample higher-frequency (monthly) observations. The forecasts are then aggregated to provide estimates of the quarterly values of the regressors of the bridge equation. A bridge equation is nothing but a low frequency (quarterly) regression with the aggregated (quarterly) forecasts of high frequency (monthly) regressors. Many Central Banks use the Bridge Equation Model in coming up with advance releases of important statistics (see e.g., Runstler and Sedillot 2003). Ingenito and Trehahn (1996) used bridge equations to "nowcast" US real GDP based on nonfarm payrolls, industrial production and real retail sales.

2.4 MIDAS (Mixed Data Sampling) Regressions

A key feature of MIDAS regression models is the use of a parsimonious and data-driven weighting scheme using lag polynomials. MIDAS estimation offers several different weighting functions/schemes which define a specific MIDAS regression model (Ghysels, et. al., 2004)

2.5. Model 4: Almon or PDL MIDAS

MIDAS regression shares some features with distributed lag models (Ghysels, et. al., 2004). In particular, one parametrization used is the Almon lag model_(also known as Polynomial Distributed Lag), which is widely used in classical distributed lag modeling. The weighting scheme for the contribution of each higher frequency variable to the low frequency regressand is determined by a suitable polynomial of a certain order (Almon 1964).

2.5 Model 5: Beta Weighting MIDAS

An alternative method is based on following a higher transcendental function called the Beta function. This function involves estimation of three parameters, but can be restricted by imposing constrains on the parameters of the function to come up with a more parsimonious parametrization (Andreou, et. al. 2010) The number of parameters estimated can therefore be 1, 2 or 3 (depending on the types of restrictions imposed). Notice also that with this weighting scheme, the number of parameters also does not increase with the number of lags, but the estimation involves a highly non-linear estimation procedure (Foroni, et.al, 2012).

2.6 Model 6: Step Weighting MIDAS

Perhaps the simplest weighting scheme is a step function, where the distributed lag pattern is approximated by several discrete steps. Step-weighing lowers the number of estimated coefficients since it restricts consecutive lags to have the same coefficient (Armesto, et. al., 2010). For example, if the distributed lag order is 12 and the number of steps is 4, the first 4 lags have the same coefficient; the next four lags have the same coefficient and so on, all the way up to 12

2.7 Model 7: Unrestricted MIDAS (U-MIDAS)

This MIDAS variant is appropriate if the differences in sampling frequencies are small (say, monthly and quarterly data). When the difference in sampling frequencies between the regressand and the regressors is large, distributed lag functions are typically employed to model dynamics avoiding parameter proliferation (Foroni, et.al, 2012). In macroeconomic applications, however, differences in sampling frequencies are often small. In such a case, it might not be necessary to employ distributed lag functions and parameters can be estimated by OLS (Ghysels, Sinko and Valkanov, 2007).

2.8 Estimating the Models

The empirical counterparts of Models 1 to 7 are constructed as part of the tasks completed in this study. All of the operational models are estimated using Eviews 9.5 software released just recently, which is the only commercial software available that supports estimation of MIDAS regression. All data to be used, quarterly, monthly, and daily statistics are accessed through PSA, BSP, and PSE websites. The following variables over the period 2002-2016 comprise the database of the study:

Quarterly (2002q1-2016q4) – Economic Growth (year-on-year continuously compounded growth of Seasonally Adjusted Gross Domestic Product, in real terms (the regressand) computed for as:

$$ecogrowth_t = 400 * \log(rgdp_t / rgdp_{t-1})\%$$

Where: $rgdp_t =$ Seasonally adjusted real Gross Domestic Product for quarter t.

Monthly (2002m1 to 2016m12):

- **Inflation**: $\text{infl}_t = 100 * \log(cpi_t / cpi_{t-1})\%$
- **Growth of Industrial Production**: $ipg_t = 100 * \log(ip_t / ip_{t-1})\%$
- **PSEI Return**: $pseig_t = 100 * \log(\text{PSEI}_t / PSEI_{t-1})\%$
- **Interest Rate**: $IR_t =$ 91 days T-Bills Rates during month t
- **Exchange Rate** (Peso to US Dollar) **Return**: $erg_t = 100 * \log(er_t / er_{t-1})\%$

3.0 RESULTS

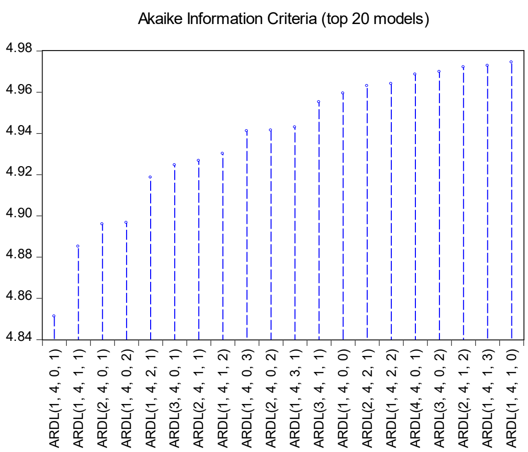
Preliminary analyses are done to establish the statistical properties of the quarterly and monthly data. The most important concern is to determine the order of integration of each of the time series used in the study, their cointegration and their unidirectional and bi-directional causality. Correlations and cross-correlations are analyzed to establish the strength of statistical association among the variables, particularly the explanatory impact of the high frequency variables to quarterly economic growth. It is determined through these analyses that a dynamic modeling format featuring *Inflation, Industrial Production Growth and PSEI Returns* are the key high frequency predictors of Economic Growth. The choice for these predictors is supported by most growth theories and studies in the literature.

3.1 ARDL Form of the Models

It is expected that the effects of its predictors to Economic Growth are not instantaneous. The explanatory contributions of the regressors are manifested in the target variable with a lag; hence the ARDL (AutoRegressive Distributed Lag) is an appropriate specification of the relationship. However, the central problem is in the determination of the optimal lag of all variables in the model. Different lag configurations for the variables constitute different ARDL models from which we are going to select the optimal specification. We adopt the procedure of model selection based on the AIC (Akaike Information Criterion).

Out of a total of 500 ARDL models evaluated, the top 20 of these models with the smallest AIC scores are shown in the table below. The best among them is the ARDL (1, 4, 0, 1) – autoregressive order is 1 and the distributed lag orders for Inflation, Industrial Production Growth and PSE Returns are respectively, 4, 0 and 1.

Table 1. Top 20 ARDL models Using the Akaike Information Criterion



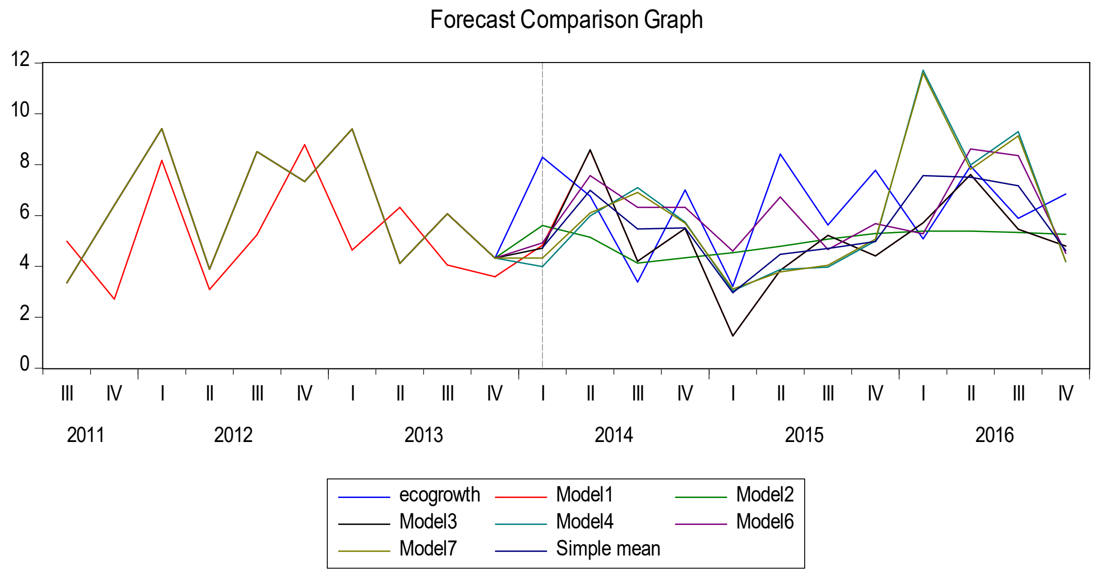
3.2 Empirical Comparison of the Out-of-sample Forecasting Performance of the Models

The different models considered in this study are estimated as ARDL (1,4,0,1), tested and empirically compared as to their capability to effectively track, out of sample the actual growth data. Presented in Table 2 are the results of this comparison based on each model's ability to encompass the forecasting ability of the other models, as well their scores on different forecast evaluation criteria.

Table 2. Forecast Evaluation
Evaluation sample: 2014Q1 2016Q4

Evaluation statistics

Forecast	RMSE	MAE	MAPE	SMAPE	Theil U1	Theil U2
Model1	2.224218	1.777170	27.77141	33.75438	0.185607	0.747321
Model2	2.007592	1.725676	26.10817	29.20138	0.172795	0.699839
Model3	2.239918	1.787617	27.89750	33.94568	0.187012	0.747411
Model4	3.270208	2.662702	45.27516	41.92192	0.250342	0.866209
Model6	1.898047	1.633489	28.71923	27.35359	0.147565	0.437027
Model7	3.183358	2.584310	43.73610	40.59814	0.244464	0.866315
Simple mean	2.171492	1.808920	28.73343	30.26478	0.176487	0.674793



The obvious winner in this forecast comparison is a MIDAS model – the Step-weighing MIDAS (Model 6), as it obliterated all other competing models in all evaluation criteria, except the MAPE (mean absolute percentage error). Model 2 or the VAR model consistently placed 2nd in all criteria, except MAPE where it ranked 1st. The outstanding performance of the MIDAS model with respect to the RMSE considered as the benchmark criterion in forecasting, accentuate its superiority as it is the only model with RMSE of less than 2.0. Incidentally, estimation for Model 5, the Beta Function weighing MIDAS failed to converge.

4.0 CONCLUSION

The mixed frequency problem in economic forecasting and structural analysis has recently attracted considerable following in the literature. This is true among policy makers and planners who are hard pressed in making updated assessment of the performance of the economy, under limited and at times missing information. Most important data releases related to economic growth are normally done quarterly (e.g., gross domestic product and its components in the national accounts). Moreover, these releases often come with substantial publication delays (which cause the so-called "ragged-edge problem" – missing values for some of the variables, especially at the end of the sample), whereas other equally important statistics, which are reported more frequently are already available, even before the publication gaps are filled. These problems of "mixed frequency", "ragged edge" and asynchronous data availability motivate this study, whose objective is to demonstrate the viability of the MIDAS Regression modeling - a "state-of-the-art" approach capable of generating "nowcasts" of the country's economic growth. In this study seven forecasting models, including four variants of the MIDAS model are estimated, tested and empirically evaluated for their out-of-sample forecasting performance over a forecast horizon of 12 quarters (2014q1 to 2016q4). Model estimation is over the period 2002q1 to 2013q3 setting aside the remaining available data for forecast assessment. The results indicated the outstanding performance of a variant of the MIDAS model which is the Step-weighing MIDAS in practically all evaluation criteria, except one. This demonstration led us to conclude the superiority of the MIDAS approach in "nowcasting" the year-on-year quarterly economic growth of the Philippine economy.

5.0 ACKNOWLEDGEMENTS

The author would like to acknowledge the funding support of the Angelo King Institute under the AKI Grants for the Research Thrust of the

DLSU School of Economics for the Academic Year 2016-2017.

6.0 REFERENCES

Andreou, E.; Ghysels, E. and Kourtellos, A.(2010). Regression Models with Mixed Sampling Frequencies. Journal of Econometrics, October 2010b, *158*(2), pp. 246-61.

Baffigi, A., R. Golinelli, and G. Parigi (2004): Bridge models to forecast the Euro area GDP, International Journal of Forecasting, 20(3), 447-460.

Armesto, M. T., K. M. Engemann, and M.T. Owyang (2010). Forecasting with Mixed Frequencies. Federal Reserve Bank of St Louis Review 92(6): 521-536.

Clements, M. and A. Galvão. (2008). Macroeconomic Forecasting with Mixed-Frequency Data: Forecasting Output Growth in the United States. Journal of Business & Economic Statistics 26: 546–554.

Diron, M., (2008). Short-term forecasts of euro area real GDP growth: an assessment of real time performance based on vintage data. J. Forecast. 27 (5), 371–390.

Giannone, D., Reichlin, L. and Small, D. Nowcasting: The Real-Time Informational Content of Macroeconomic Data. Journal of Monetary Economics, May 2008, *55*(4), pp. 665-76.

Foroni, C., M. Marcellino, and C. Schumacher (2012): U-MIDAS: MIDAS regressions with unrestricted lag polynomials, CEPR Discussion Papers, 8828.

Forsberg, L., and E. Ghysels (2007): "Why do absolute returns predict volatility so well?", Journal of Financial Econometrics, 5(1), 31-67.

Ghysels, E., Santa-Clara, P., and Valkanov, R. (2004), The MIDAS Touch: Mixed Data Sampling Regression Models, CIRANO Working Papers 2004s-20, CIRANO, Montreal, Canada.

Ghysels, E., A. Sinko and R. Valkanov (2007). MIDAS Regressions: Further Results and New Directions. Econometric Reviews 26 (1): 53–90.

Ghysels, E., and Wright, J. (2006), Forecasting Professional Forecasters, Finance and Economics Discussion Series 2006-10, Board of Governors of the Federal Reserve System (U.S.), Washington, DC.

Ingenito, R. and B. Trehan (1996). Using monthly data to predict quarterly output, Economic Review, Federal Reserve Bank of San Francisco, pages 3-11.

Rünstler, G., and Sédillot, F. (2003), Short-Term Estimates of Euro Area Real GDP by Means of Monthly Data, Working Paper 276, European Central Bank.

Sims, C. (1982). “Macroeconomics and Reality”. Econometrica 48 (1): 1–48