Nowcasting Philippine Economic Growth Using MIDAS Regression Modeling

By: Dr. Cesar C. Rufino
Distinguished Professor, SOE

ABSTRACT

Among the most anticipated data releases of the Philippine statistical system is the quarterly real gross domestic product. This all-important variable provides the basis for deriving the economic growth performance of the country on a year-on-year basis. Official publication of this statistics, however, comes at a significant delay of up to two months, upsetting the planning function of various economic stakeholders. Under this backdrop, 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, opens up the "mixed frequency" problem in forecasting, which is the data frequency asymmetry between the dependent and independent variables of regression models that will be 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 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

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 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 the variables are sampled in the same frequency. When any or all of the regressors is/are in higher frequency, the usual recourse, called temporal aggregation approach, is to temporally aggregate these variables, usually in terms of their sums or averages to conform with the sampling frequency of the regressand, thus, synchronizing the data sampling of the left hand and right hand side variables 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 to extract predictable information from the more frequently sampled regressors because of information loss and possible misspecification errors induced by the process of aggregation might compromise the forecast quality.

An alternative option called the individual coefficient approach, which is 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 effectively 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 (<u>Mi</u>xed <u>Da</u>ta <u>S</u>ampling) approach, introduced by Ghysell, 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 matched against models traditionally used in dealing with the "mixed frequency" problem.

Technical Specifications of the Models

Suppose the mixed frequency model under consideration is given as follows:

$$Y_t^L = \sum_{i=1}^{q} \beta_i W_{t-i}^L + \lambda f\left(\gamma, X_{j,t}^H\right) + \varepsilon_t \tag{1}$$

where,

Y_t^L- is the dependent variable sampled at low frequency
W_t^L- is the set of regressors sampled at the same (low) frequency as the regressand (possibly including lags of *Y* for autoregressive or ARDL form)
$X_{j,t}^H$- is the set of regressors sampled at a higher frequency
β_i, λ, and γ – are the parameters to be estimated
f (.) – is a function translating the higher frequency data into the low frequency

The following are the models that will be used in the study, each of which has a different way of translating the higher frequency data into their low frequency form, through their specific choice of *f (.)*.

Model 1: Temporal Aggregation

A conventional 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). For example, we can take a simple average:

$$\boldsymbol{X_t^L} = \frac{1}{m} \sum_{j=0}^{m-1} X_{t-j}^H \tag{2}$$

and carry out the following regression:

$$Y_t^L = \sum_{i=1}^{q} \beta_i W_{t-i}^L + \boldsymbol{\lambda X_t^L} + u_t \tag{3}$$

Where

m – is the number of periods in the higher frequency corresponding to a single period in the lower frequency. X_t^H is the high-frequency observation corresponding to the last observation in period *t*. As mentioned previously, the problem with this is that it assumes that the slope coefficients on each individual high-frequency observation of X are equal.

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.

Model 3: Bridge Equation

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." This method use popular forecasting models (such as VARs, ARIMA, Exponential Smoothing, etc.) for each of the high frequency indicators. These models are then used to provide forecasts for the missing higher-frequency (monthly) values. 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. A bridge equation can be written as:

$$Y_t^L = \beta_0 + \sum_{i=1}^{j} \beta_i(L) X_t^L + u_t \tag{4}$$

where X_t^L are the selected high frequency indicators (forecasted from the high frequency VAR, for example) aggregated at low frequency. The lag polynomial $\beta_i(L)$ embeds the parameters of the model for each relevant lags of each regressor. Many Central Banks use the Bridge Equation Model in coming up with advance releases of important statistics (see e.g., Runstler and Sedillot 2003; Zheng and Rossiter 2006). Ingenito and Trehahn (1996) used bridge equations to "nowcast" US real GDP based on nonfarm payrolls, industrial production and real retail sales.

MIDAS (Mixed Data Sampling) Regression

The key feature of MIDAS regression models is the use of a parsimonious and data-driven weighting scheme:

$$Y_t^L = \sum_{i=1}^{q} \beta_i W_{t-i}^L + \lambda \sum_{j=0}^{m-1} w_{t-j}(\gamma) X_{t-j}^H + u_t \tag{5}$$

where *w(.) is a weighting function* that transforms high-frequency data into low frequency data.

MIDAS estimation offers a number of different weighting functions/schemes which define a specific MIDAS regression model

Model 4: Almon or PDL MIDAS

MIDAS regression shares some features with distributed lag models. In particular, one parametrization used is the **Almon lag weighting** (also known as Polynomial Distributed Lag weighting), which is widely used in classical distributed lag modeling. The weighting scheme can be written as follows:

$$Y_t^L = \sum_{i=1}^{q} \beta_i W_{t-i}^L + \sum_{i=1}^{p} \gamma_i \sum_{j=0}^{k} j^{i-1} X_{t-j}^H + u_t \tag{6}$$

where,
k - is the chosen number of lags (which may be longer or shorter than *m*)
p - is the order of the polynomial

Notice that the number of coefficients to be estimated depends on the polynomial order ***(p)*** and not on the number of lags *(k)* chosen.

Model 5: Beta Weighting MIDAS

An alternative method is based on the following Beta function:

$$Y_t^L = \sum_{i=1}^{q} \beta_i W_{t-i}^L + \lambda \sum_{j=0}^{k} Z_{j,t} + u_t \tag{7}$$

where,

$$Z_{j,t} = \left(\frac{{\psi_j}^{\gamma 1-1} (1-\psi_j)^{\gamma 2-1}}{\sum_{i=0}^{k} {\psi_i}^{\gamma 1-1} (1-\psi_i)^{\gamma 2-1}} + \gamma_3 \right) X_{t-j}^H \tag{8}$$

where $\psi_j = \frac{j-1}{k-1}$

This function involves estimation of three parameters, but we can restrict them by imposing either: $\gamma_1 = 1$ or $\gamma_3 = 0$ or $\gamma_1 = 1$ and $\gamma_3 = 0$. The number of parameters estimated can, therefore, be 1, 2, or 3 (depending on the types of restrictions we impose). 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 (Ghysels, Rubia, and Valkanov, 2009).

Model 6: Step Weighting MIDAS

Perhaps the simplest weighting scheme is a step function, where the distributed lag pattern is approximated by a number of discrete steps. The Step weighting can be written as:

$$Y_t^L = \sum_{i=1}^{q} \beta_i W_{t-i}^L + \sum_{j=0}^{k} \phi_{t-j} X_{t-j}^H + u_t \quad (9)$$

where, $\phi_j = \gamma_k$

k – is a number of lags (k may be longer or shorter than m)
$k = \frac{j}{\eta}$ and η is the number of steps.

Step-weighing lowers the number of estimated coefficients since it restricts consecutive lags to have the same coefficient (Forsberg and Ghysels, 2007). For example, if k=12 and η=4, the first four lags have the same coefficient, the next four lags have the same coefficient and so on, all the way up to k=12.

Model 7: U-MIDAS

U-MIDAS or ***Unrestricted Midas*** 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. Introduced by Foroni, Marcellino, and Schumacher (2012), U-MIDAS does not depend on any specific functional lag polynomial since in most macroeconomic applications differences in sampling frequencies are often small, usually quarterly-monthly. In such a case, it might not be necessary to employ distributed lag functions and parameters can be estimated by OLS. In essence, the **U-MIDAS** approach can be written as:

$$Y_t^L = \sum_{i=1}^{q} \beta_i W_{t-i}^L + \sum_{j=0}^{m-1} \gamma_{t-j} X_{t-j}^H + u_t \quad (10)$$

where we estimate a different slope coefficients for each high-frequency lag.

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 (2002m1–2016m12):

- **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 Return during month *t*
- **Exchange Rate** (Peso to US Dollar) **Return**: $erg_t = 100 * \log(er_t / er_{t-1})\%$

RESULTS

Preliminary analysis of the quarterly correlation matrix reveals the potential of some of the temporally aggregated monthly variables significant growth drivers. As shown in Table 1, inflation, growth of industrial production, and possibly stock index returns (pseig) produce relatively high contemporaneous correlation with economic growth.

Correlation p-value	Econ Growth	FX Returns	Inflation	IP Growth	Interest Rate
FX Returns	0.008518	1.000000			
	0.9485	-----			
Inflation	-0.286533	0.077610	1.000000		
	0.0264	0.5556	-----		
IP Growth	0.283804	0.074094	-0.148502	1.000000	
	0.0280	0.5737	0.2575	-----	
Interest Rate	-0.174246	-0.066392	0.404472	0.022381	1.000000
	0.1830	0.6143	0.0013	0.8652	-----
PSE Returns	0.202960	-0.468602	-0.150008	0.005459	0.071218
	0.0497	0.0002	0.2526	0.9670	0.5887

Table 1. Correlation Matrix of the Potential Explanatory Variables for Mixed Frequency Regressions of Economic Growth

In the above matrix, each cell exhibits the sample correlation coefficient between the row variable and the column variable, together with the p-value of the test for zero correlation. The first column is quite revealing as it indicates the variables with significant correlation with economic growth—Inflation, Industrial Production growth, and PSE Returns. Economic intuition may lead us to believe that these variables carry predictive contents despite their asymmetric sampling frequency with economic growth, may be considered as the key explanatory variables for growth.

All of the variables in all regressions are stationary as evidence by the results of individual unit root tests shown in Table 2 below. These results potentially prevent the occurrence of spurious regressions. This conjecture will be empirically validated through the conduct of cointegration assessment using appropriate procedure.

Table 2. Stationarity and Unit Root Tests of the Key Variables

Test Statistic	Eco. Growth	Inflation	IP Growth	PSE Returns
KPSS *Statistic*[1]	0.208503	0.296980	0.180490	0.058695
	(p>0.10)	(p>0.10)	(p>0.10)	(p>0.10)
ADF *Statistic*[2]	-7.09498	-3.94489	-8.91786	-6.01192
	(p<0.000)	(p<0.000)	(p<0.000)	(p<0.000)
PP *Statistic*[2]	-7.09148	-2.40887	-12.01738	-6.025076
	(p<0.000)	(p<0.000)	(p<0.000)	(p<0.000)
Order of Integration	*I(0)*	*I(0)*	*I(0)*	*I(0)*

[1]H_o : Variable is stationary [2]H_o : Variable has a Unit root

In order to empirically demonstrate the presence of long run relationship(s) among the three key variables and economic growth, testing for cointegration is necessary. Since all variables are integrated of order 1 (i.e., *I (1)*), the Johansen cointegration tests cannot be used, instead we use the Pesaran Bound test for cointegration under the ARDL approach. Table 3 confirms the presence of long run equilibrium relationship between economic growth and its postulated determinants.

Table 3. ARDL Bound Test for the Presence of Cointegration

Bounds Test for Cointegration **Null Hypothesis: No cointegrating relationships exist**

Test Statistic	Value	Significance Level	I(0)	I(1)
F-statistic	21.89180***	10%	2.72	3.77
		5%	3.23	4.35
		2.5%	3.69	4.89
		1%	4.29	5.61

EC = ecogrowth - (-0.0551*infl + 0.2776*ipg + 0.5095*pseig)

Long Run Coefficients

Variable	Coefficient	Std. Error	t-Statistic	p-value
Inflation	-0.055089	0.272754	-0.201973	0.1411
IP Growth	0.277642	0.157034	1.768031	0.0860
PSE Returns	0.509460	0.162470	3.135730	0.0035

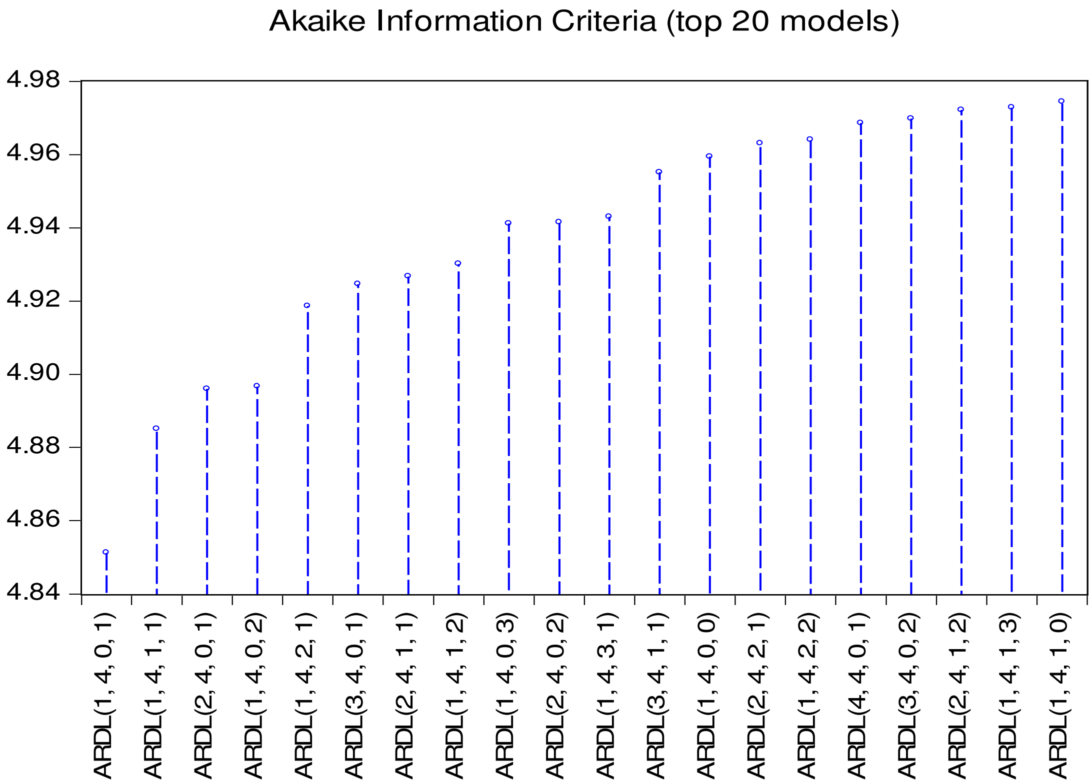
***significant at 0.01 level

ARDL Forms of the Models

It is expected that the effect of its predictors to Economic Growth is 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 Table 5. 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 4, 0, and 1, respectively.

Table 4. Top 20 ARDL Models Using the Akaike Information Criterion



To implement **Model 1** as an ARDL (1, 4, 0, 1) with time aggregated monthly regressors into quarterly frequency, we use a recent inclusion in the Eviews 9.0 suites of commands—the ARDL estimation and the results are presented in Table 6.

Performing the different diagnostic procedures on this model, the following are noted: No residual autocorrelation up to the 4^{th} order, no heteroscedasticity, Ramsey-RESET confirms correct specification, and no structural change. It is important to check for structural change within the sample horizon as its presence will affect the quality of the forecasts. Presented in Table 7 is the result of the Quandt-Andrews Unknown Breakpoint Test for structural change.

Table 6. Estimated ARDL (1, 4, 0, 1)

Dependent Variable: ECOGROWTH
Method: ARDL
Included observations: 44 after adjustments
Maximum dependent lags: 4 (Automatic selection)
Model selection method: Akaike info criterion (AIC)
Dynamic regressors (4 lags, automatic): INFL IPG PSEIG
Number of models evaluated: 500
Selected Model: ARDL(1, 4, 0, 1)

Variable	Coefficient	Std. Error	t-Statistic	Prob.*
ECOGROWTH(-1)	-0.150920	0.141117	-1.069474	0.2924
INFL	-0.306245	0.493605	-0.620424	0.5391
INFL(-1)	1.824895	0.783209	2.330024	0.0259
INFL(-2)	-1.881618	0.888671	-2.117339	0.0416
INFL(-3)	-1.125012	0.916205	-1.227904	0.2279
INFL(-4)	1.424577	0.502055	2.837489	0.0076
IPG	0.319544	0.168774	1.893325	0.0669
PSEIG	0.269582	0.127310	2.117521	0.0416
PSEIG(-1)	0.316767	0.133321	2.375974	0.0233
C	5.380362	1.857334	2.896821	0.0065

R-squared	0.620667	Mean dependent var	5.189676
Adjusted R-squared	0.520256	S.D. dependent var	3.580807
S.E. of regression	2.480194	Akaike info criterion	4.851267
Sum squared resid	209.1463	Schwarz criterion	5.256765
Log likelihood	-96.72788	Hannan-Quinn criter.	5.001645
F-statistic	6.181235	Durbin-Watson stat	2.135071
Prob(F-statistic)	0.000040		

*Note: p-values and any subsequent tests do not account for model selection.

Table 7. Quandt-Andrews Unknown Breakpoint Test For Structural Change

Null Hypothesis: No breakpoints within 15% trimmed data
Varying regressors: All equation variables
Equation Sample: 2002Q4 2013Q4
Test Sample: 2004Q3 2012Q2
Number of breaks compared: 32

Statistic	Value	p-value.
Maximum LR F-statistic (2009Q2)	3.042404	0.1331
Maximum Wald F-statistic (2009Q2)	15.21202	0.1331
Exp LR F-statistic	0.594499	0.4925
Exp Wald F-statistic	5.127762	0.1064
Ave LR F-statistic	1.001204	0.4255

Ave Wald F-statistic	5.006022	0.4255

Note: probabilities calculated using Hansen's (1997) method

Empirical Comparison of the Models Out-of-Sample Forecasting Performance

The different models considered in this study are estimated, tested, and empirically compared as to their capability to effectively track, out of sample, the actual growth data. Presented in Table 8 are the results of this comparison, first, based on each model's ability to encompass the forecasting ability of the other models in the upper panel, and second, their scores on the different evaluation statistics.

Table 8. Forecast Evaluation

Evaluation sample: 2014Q1 2016Q4

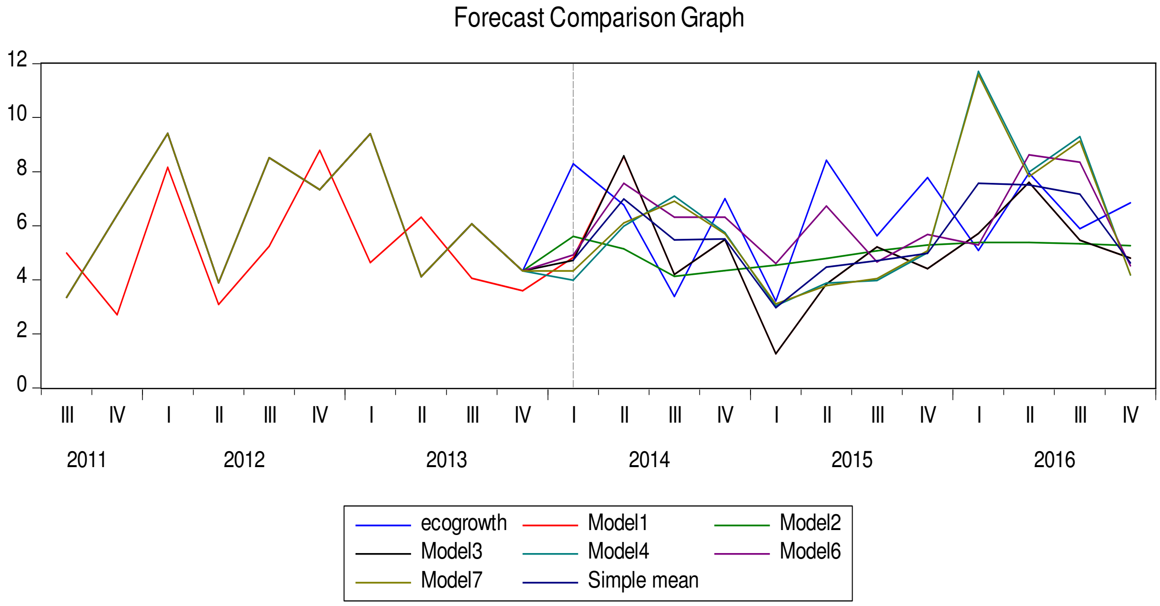
Encompassing Tests

Null hypothesis: Forecast of model i includes all information contained in others

Forecast	F-stat	p-value
Model1	2.380380	0.1605
Model2	1.280371	0.3810
Model3	2.333357	0.1659
Model4	10.79025	0.0059
Model6	1.797065	0.2477
Model7	9.044966	0.0092

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 out-of-sample forecast comparison is a MIDAS model – the Step-weighing MIDAS (Model 6), not only that it encompasses the other models, it also 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.

SUMMARY AND CONCLUSION

As of late, the mixed frequency and rugged-edge problems in economic forecasting and structural analysis have been attracting a considerable following in the literature. This is true most especially among policy makers and planners who are hard pressed in making an updated assessment of the performance of the economy, under limited and at times missing or incomplete 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 using 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 out-of-sample forecast evaluation.

The results indicate the outstanding performance of a variant of the MIDAS model which is the Step-weighing MIDAS in practically all of the 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. Hopefully, this study may be able to supply the missing element in the periodic Angelo King Institute (AKI) press releases about the condition of the economy—advanced estimates of the current and future economic performance statistics, which unlike other most think tank groups' forecasts are based on a rigorous cutting-edge macro econometric framework.

\REFERENCES

Andreou, E.; Ghysels, E. and Kourtellos, A. "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.

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

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.

Ghysels, E., A. Rubia, and R. Valkanov (2009), "Multi-period forecasts of volatility: direct, iterated and mixed-data approaches". University of North Carolina WP.

Ghysels, E., P. Santa-Clara, and R. Valkanov. (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.

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