S³H Working Paper Series

Number 02: 2021

Forecasting Pakistan's GDP Growth With Leading Indicators: A MIDAS Approach

Sidra Tahir

Tanweer Ul Islam

June 2021

School of Social Sciences and Humanities (S³H)
National University of Sciences and Technology (NUST)
Sector H-12, Islamabad, Pakistan

S3H Working Paper Series

S³H Working Paper Series

Number 02: 2021

Forecasting Pakistan's GDP Growth With Leading Indicators: A MIDAS Approach

Sidra Tahir

Graduate, School of Social Sciences and Humanities, NUST
E-mail: sidratahirahmed95@gmail.com

Tanweer Ul Islam

Assistant Professor, School of Social Sciences and Humanities, NUST
E-mail: tanweer@s3h.nust.edu.pk

June 2021

School of Social Sciences and Humanities (S³H)
National University of Sciences and Technology (NUST)
Sector H-12, Islamabad, Pakistan

Table of Contents

List of Tables

Abstract

This study forecasts the quarterly GDP growth rates for Pakistan using leading macro-economic indicators covering a period from January 2002 to December 2020. We use Mixed Data Sampling (MIDAS) regressions, which directly links the quarterly observations with the high-frequency monthly indicators without any aggregation techniques and compare the forecasting performance of MIDAS with the conventional Autoregressive Distributed Lag (ARDL) model. The main purpose of this study is to test the power of leading indicators in providing early estimates of quarterly GDP growth rates. The forecast horizon covers all the 4 quarters of the annual year 2020. The forecast evaluation criteria to compare the forecast of these models are RMSE, MAE, MAPE and Theil Inequality Coefficient. Diebold Mariano Test is also conducted to statistically check the forecasting accuracy of two models. Our results show that, MIDAS performs better than the benchmark model, the ARDL. Among the MIDAS variants considered in this study, U-MIDAS turns out to be the best option for forecasting.

Keywords: GDP**,** MIDAS, forecasting, leading macro-economic indicators, accuracy.

1. Introduction

Economic forecasts are an important element of rational economic policy as stable budgetary plans for government expenditures and revenues rely on cost-effective macroeconomic projections. Central Banks, policy-making institutions and financial markets analysts are in adequate need to have reliable early information on the current state of the economy as soon as possible. The purpose is to keep track of the macroeconomic variables and to make valuable judgements on the future state of the economy. However, official quarterly data on GDP in Pakistan are not available, and furthermore, annual GDP is published with a considerable delay.

The complicated economic situations prevailing nationally and internationally outspreads the uncertainties in hand, consequently, early short-term forecasting of the key macro-economic variables is essential in order to understand the economic situations for constructing rational policy decisions. But, sadly, having a reliable up-to-date forecast is not an easy task and it is even harder under crisis situations. Currently, when financial markets have become more volatile than ever, witnessing a steeper decline. Moreover, intense fluctuations in business cycles caused by the external shocks adds a little more instability to the economy. The early information thus available in the form of nowcasts and forecasts results in necessary repercussions on the policy actions been taken.

Business cycles identify the periodic irregular up & downs in economic activities. They are well explained as the difference between the GDP and the underlying trend. The economy is in expansion when the underlying trend lies above GDP and in recession when the trend lies below GDP. Business cycles are recognized under four phases namely contraction, trough, expansion and peak. Peaks and troughs are the turning points in the economy indicating the highest and the lowest level an economy can reach under different economic circumstances. Other than real GDP, there are variables that influence the business cycle. These variables are classified under three heads:

I. Leading Indicators
II. Lagging Indicators
III. Coincident Indicators

Leading indicators are the ones giving early signals to the economy. They have the power to predict when an economy is entering an expansion or a contraction phase. While lagging indicators provide information about the economic fluctuations after their existence. Coincident indicators provide information about the present state of the economy. Information obtained through these indicators helps in understanding the business cycles and forecasting the GDP growth. Samuelson

(1939) carried out lots of research in describing market fluctuations under business cycles by combining several hypotheses into a rational framework.

The use of leading indicators has gained prominence over time as policymakers, investors and the business community are interested in knowing the early signals of recession or recovery. The timely availability of financial indicators, prices, survey-based indicators, etc. helps in giving early signals of the real economic situations and is considered as their strength (Drechsel and Scheufele, 2010). The approach of using leading indicators for economic forecasting is based on the view that economies experience the occurrence of repetitive sequences within business cycles and the sequences highlight the generation of the business cycle itself (Lahiri and Moore, 1991).

Gross Domestic Product (GDP) is an indicator which measures the size of an economy by observing its overall performance. Therefore, forecasting GDP growth is of paramount importance for the decision-makers. For this purpose, economists, econometricians and statisticians have developed different forecasting models. But a specific forecasting model should be one, which can address two types of issues: i) reasonable predictor's selection and ii) data handling at different frequencies (Koenig et al., 2003, Armesto et al., 2010 and Andreou et al., 2013). Generally, data for the macroeconomic variables have been observed at different frequencies. For instance, data for the GDP growth is available quarterly/annually with a considerable delay while data for the leading indicators like industrial production index, consumer confidence index, broad money, etc. are monthly available and data for the financial variables are available at a weekly or daily frequency. This mixed frequency data sampling seems to be a problem as filtration of data may result in loss of useful information (Silvestrini and Veredas, 2008). To handle these data irregularities, to predict GDP growth rate, one single equation approach has been analysed. The approach is named Mixed Data Sampling (MIDAS) regression.

A mixed data sampling (MIDAS) model proposed by Ghysels, Santa-Clara, and Valkanov (2004) is best at handling different frequencies by incorporating high-frequency variables into a model for forecasting low-frequency variable. MIDAS follows a parsimonious weighing scheme system in order to reduce the number of parameters to be estimated by employing the functional lag polynomials. As there is no time aggregation, the MIDAS model directly uses real-time data without any artificial handling. The pros of using MIDAS regression are: i) it saves from data interpolation that helps in preventing data rejection and thus enhancing the forecast accuracy ii) it is less prone to specification errors and iii) it takes into account the real-time data i.e. the latest released high-frequency data thus performing real-time forecasts (Ghysels and Wright, 2009, Andreou et al., 2011, Frale and Monteforte, 2011, Liu et al., 2012 and Dias et al., 2015).

Covid-19 appeared to be a global problem with highlighting the facts like uncertainty in the duration of this pandemic, the discovery of vaccine and the economic revival of major economies being exogenous in nature, mandatory for the developing countries like Pakistan's economic recovery. None of these factors is under the direct or endogenous control of Pakistan's economy. Coronavirus crises are seemed to be more underlined in the "real" sector of an economy leaving a massive effect as opposed to financial markets (Chohan, 2020). Considering the shutdown of the economy for the past few months and irregular lockdown restrictions affects the production activities and this calls for a need to forecast quarterly GDP growth rate under these uncertain times. Pakistan's official data for GDP is annually published by the Pakistan Bureau of Statistics. [1]

This study undertakes a comprehensive analysis of predicting quarterly GDP growth rates for Pakistan to provide early estimates to the planning commission for making sound economic policies and plans. It aims to forecast quarterly GDP for Pakistan using MIDAS Models by allowing many potential high-frequency macroeconomic leading indicators common in the economic literature. It chooses the set of leading indicators based on their importance according to the previous studies and the availability of data. It classified the list of indicators into different categories as prices, financial indicators, real economic indicators, etc. following the study of Drechsel and Scheufele (2010). To forecast GDP for the annual year 2020, monthly data on several leading indicators including prices, financial, fiscal, credit and real economic variables are taken from January 2002 to December 2020 for a total of 228 monthly observations.

The rest of the paper is arranged as follows. Section 2 gives an overview on forecasting with MIDAS literature. Section 3 explains data and methodology. Section 4 discusses the results and section 6 concludes the results.

2. Literature Review

Statisticians and econometricians are always interested in improving the estimations carried out by conventional models that deal with the time series data. And, over the years handling information in unbalanced data sets has gained potential attention. As most of the appropriate information is saved in high-frequency data sets. Time aggregation and averaging may lead to the loss of useful

[1] The periodicity of the estimates of GDP is based on new base year 2005-2006. The development of experimental series of estimates of quarterly GDP is in progress.

data. Ghysels, Santa-Clara and Valkanov (2004) proposed (Mi)xed (Da)ta (S)ampling MIDAS regression models that directly put in data sets sampled at different frequencies. For instance, models combining annually/quarterly, quarterly/monthly or monthly/daily data together. MIDAS models are a simple, parsimonious yet flexible class of time series models based on distributed lag polynomials. From a decade, they have now been employed to nowcast and forecast macroeconomic variables (output, inflation, unemployment rate etc.), providing favourable results for short-term forecasting. Presently, MIDAS is included in the toolkit of short-term forecasts of some Central banks. This section reviews some important empirical studies on MIDAS models as well as forecasting studies conducted in Pakistan.

2.1. Empirical Review on Forecasting using MIDAS Regressions

Ghysels, Santa-Clara and Valkanov (2004) addressed different functional forms to estimate the MIDAS models with a few numbers of parameters. They further examined the asymptotic properties of MIDAS regression estimations, and their comparison with the basic distributed lag regressions, by reviewing the challenging econometrics issues. Clements and Galvao (2009) extended the MIDAS regression by including an autoregressive term, termed as MIDAS-AR. They examined the short-term forecasting performance of MIDAS models by Ghysels, for the US output growth by considering the monthly predictors. Significant improvement was observed in forecasting current and next quarter output growth by using monthly data on the current quarter. Sinko (2008) introduced some new MIDAS specifications by extending the univariate MIDAS model that include unequally spaced observations under mixed data sampling structures. He discussed two lag parameterizations of MIDAS regression i.e. the Exponential Almon and Beta lag polynomials by investigating the response of U.S stock returns to macroeconomic shocks.

Armesto, Engemann and Owyang (2010) highlighted variation in the results of forecasting by comparing different Time Aggregating high-frequency data. Time averaging, step weighting and MIDAS models represented a trade-off between parsimony and flexibility for different sets of data. They explicitly modelled Mixed Data Sampling (MIDAS) as more beneficial for Intra-Period forecasting. Foroni, Marcellino and Schumacher (2011) discussed the pros and cons of unrestricted lag polynomials in unrestricted MIDAS regressions termed as U-MIDAS from linear high-frequency models. They showed that U-MIDAS outperforms the MIDAS regression models, by empirically nowcasting Euro area and US GDP using monthly indicators, highlighting if the difference between the sampling frequencies is small like quarterly/monthly. Foroni and Marcellino (2012) discussed all the main approaches to deal with mixed frequency data. With empirical applications that is forecasting quarterly euro area GDP components using monthly indicators, they compared U-MIDAS TO MIDAS and showed that U-MIDAS performed better. They highlighted the importance of monthly information during a crisis and focused on the issues in mixed data sampling related to structural models. Foroni and Marcellino (2013) clarified the robustness of MIDAS regression models in miss-specification by contrasting MIDAS, bridge equation models and the state-space approaches. The other two approaches may lead to the problem of parameter proliferation. The complexities associated with the estimation of models; MIDAS seemed to be computationally simpler.

Franco and Mapa (2014) employed mixed frequency modelling by MIDAS. To analyse the dynamics of inflation and GDP growth in the Philippines, monthly macro indicators as time-varying parameters were estimated. A one-step ahead forecast had been generated by incorporating more indicators. They found that the inclusion of more variables results in desirable forecasts that are lower RMSE and MAE and the most important variables are consumer and wholesale price indices, stock prices, exchange rate, real money supply and merchandise exports in anticipating the growth rates. Jiang, Guo and Zhang (2017) predicted China's GDP growth rates using monthly and daily dynamic predictors from a total set of mixed frequency data of 44 macroeconomic and 54 financial variables. MIDAS presented better forecast accuracy comparing to traditional predicting models.

Chikamatsu et al., (2018) performed a comparison of different mixed frequency approaches i.e. bridge equation approach, MIDAS and factor-augmented to forecast Japanese quarterly GDP using monthly indicators ranging from hard data to soft data as predictors. The results showed that out-of-sample forecast performance is superior to an in-sample mean benchmark. They also highlighted the gains from forecast model combinations and considered supply-side data as a useful tool for nowcasting revised annual GDP.

Rufino (2019) employed the use of MIDAS regressions to nowcast the Philippines economic growth by high-frequency datasets as explanatory variables, solving the problem of mixed frequencies. Different variants of the MIDAS model were used to estimate quarterly growth rates based on monthly data of inflation, industrial production and stock exchange index. These models were compared against each other and against conventional models indicating the relative superiority of the MIDAS framework in anticipating the growth rates. To nowcast, the world's annual GDP growth rates. Mikosch and Solanko (2019) evaluated the predictor power of a large set of high-frequency indicators for nowcasting quarterly Russian GDP. They selected different forecast evaluation periods and listed the top-performing indicators under each period. Their forecasted results showed a very small difference between the performance of bridge models, MIDAS models and U-MIDAS models. However, during the recession period from 2008 to 2011, U-MIDAS and MIDAS models outperform bridge models. Ghunay (2020) discussed the significance of different functional forms of the lag polynomials and lag length structure in the MIDAS approach. He analysed the short-term forecasts of Turkish GDP by a set of monthly indicators and noted that the performance of real domestic turnover and industrial production indicators stands out. U-MIDAS successfully tracks the GDP growth with around five lags when all the month's information becomes available for a quarter.

A set of 44 microeconomic indices are used by Bhaghoe et al., (2021) over a time period 2012Q1 2020Q2 for the nowcasting quarterly GDP growth of Suriname by employing FA-MIDAS and MF-VAR. Initially, the first set of monthly regressors explained the variation in GDP growth with no lags while the second set incorporated lagged observation for up to two months. Three sample periods have been selected for model estimations 2012Q1–2019Q3, 2012Q1–2020Q1 and 2012Q1–2020Q2, performing the comparison between models based on RMSE. They concluded that FA-MIDAS and MF-VAR models delivered suitable results.

2.2. Empirical Review on Forecasting: Evidence from PAKISTAN

Hussain, Hyder and Rehman (2018) nowcasted Large Scale Manufacturing (LSM) growth as a proxy for GDP using different explanatory variables. By employing DFM and Penalize regression models to extract a piece of unique information, they endeavoured to nowcast trends and cycles separately. The results show the significance of Penalize Regression models in tracing cycles while DFM in tracing trends in LSM growth.

When the world is seeking to cope with the pandemic, Chohan (2020) studied the impact of Covid-19 on the real economy for developing countries. He presented forecasting through scenario analysis using the aggregate demand approach for Pakistan and concluded that massive declines to be observed in the economy during 2020 while reformation of economic activities will be expected in the fiscal year 2021.

Another study conducted by Syed and Lee (2021) forecast CPI inflation, GDP growth proxied by Large Scale Manufacturing (LSM) or Industrial Production Index (IPI) growth and overnight repurchase rate for Pakistan using sophisticated machine learning models such as Lasso and Ridge regression. They incorporated 161 predictors and evaluated the forecasting performance using RMSE and MASE by comparing them with benchmark ARDL. Machine learning models outperform the benchmark.

2.3. Literature Summary

The research studies conducted above suggest different results. It is worthy to note that MIDAS models outperform classical traditional forecasting techniques. MIDAS approach is more acceptable and reliable for short-term forecasts. Presently, different MIDAS variants are being used to perform the forecasting exercise. U-MIDAS outpace all other models where differences in the sampling frequencies are minor like quarterly/monthly. There exists massive international literature in forecasting GDP with leading indicators. For different countries, GDP forecasting has been done using mixed data sampling regression models. The significance of leading indicators is well established in International literature. But, there is no empirical study so far in Pakistan for forecasting the quarterly GDP growth rates by incorporating a set of leading indicators using Mixed Data Sampling Regression (MIDAS) analysis. Also, a comparison with the conventional ARDL model has been performed to evaluate the performance of these forecasting models. Therefore, this research study extends the analysis by predicting the quarterly growth rates for the first time in Pakistan that will be a contribution to the literature.

3. Data and Methodology

3.1. Data

For forecasting GDP growth rate, a macroeconomic dataset of vital monthly macroeconomic leading indicators is used. This dataset is classified into five groups namely Prices, Real economic indicators, Financial Indicators, Money and Credit and Fiscal variables. The first group consists of prices such as consumer price index (CPI), wholesale price index (WPI) and core inflation non-food non-energy (NFNE), the second group includes real economic indicators such as large-scale manufacturing index (LSM), broad money (M2), exports, imports and petrol prices, the third group contains the financial indicators including exchange rate (ER) and KSE 100 index, the fourth group entails money and credit variables such as foreign direct investment (FDI), foreign portfolio investment (FPI), money call rate (MR), workers remittances and credit to the private sector while the fifth group covers fiscal variable namely total taxes. By total taxes, it means Direct plus Indirect taxes. While indirect taxes include sales, excise and customs taxes.[2]

In this research study, the quarterly data of Real Gross Domestic Product (GDP) in Million Rupees from 2002 Q1 to 2020 Q4 for a total of 76 observations and monthly data from January 2002 to December 2020 for a total of 228 monthly observations are taken. This is the period for which monthly time series are consistently available for all the variables. The complete dataset as an information set is used to predict the quarterly GDP growth rate for Pakistan. The data for all the variables have been provided by the State Bank of Pakistan (SBP). Quarterisation of annual GDP is performed by the SBP following the study of Arby (2008) (Hanif, Iqbal and Malik, 2018). For performing forecast with the leading indicators, no theoretical background exits in literature. This study is carried out by identifying the indicators on the basis of principal components and correlation analysis.

Due to the noisy nature of macroeconomic variables in Pakistan, all individual data series have been seasonally adjusted by using Census X-13 (successor to X-12 ARIMA and X-11) before performing any further analysis.

3.2. Methodology

This study chooses to build a quarterly real GDP growth rates model based on a set of macroeconomic indicators. The two types of exercise have been performed by using this set of data. Firstly, Principal Component Analysis (PCA) has been extracted to overcome the problem of degrees of freedom that arises due to many variables and their lags. Secondly, variables having a strong association with GDP, available at a high frequency and considered to be the determinants of economic growth are undertaken to perform the forecasting exercise. Consumer price index (CPI), broad money (M2), exports, workers' remittances, private sector credit, exchange rate (ER) and foreign direct investment (FDI) etc. are examined to be the major determinants of GDP growth as supported by the literature (Ajmair et al., 2018, Ali and Saif, 2017, Rehman, 2016, Shahbaz, Ahmad and Chaudhary, 2008 and Iqbal and Zahid, 1998).

The forecasting performance being studied is based on the regression analysis expressed by different types of regression models. MIDAS regression models along with its variants such that MIDAS with principal component analysis (MIDAS-PCA) and MIDAS with leading indicators

[2] Source: State Bank of Pakistan.

(MIDAS-LI) regressions where MIDAS are further classified as Unrestricted (U-MIDAS) and Restricted (R-MIDAS) are compared to the benchmark ARDL model. The first exercise to forecast GDP growth is carried out by employing the Principal Component Analysis (PCA).[3]

3.2.2. MIDAS

This section provides a detailed analysis of the estimation of the equation used for forecasting GDP growth rate with MIDAS. MIDAS is further classified as Unrestricted MIDAS (U-MIDAS) and Restricted MIDAS (R-MIDAS). Previously, Bridge Models are the first models employed in forecasting the current state of an economy by undertaking the monthly information available during the quarter. They transformed the high-frequency monthly indicator into a low-frequency quarterly indicator to match the common frequency. Likewise, Bridge equations, MIDAS regressions are also an extension of the distributed lag models (Schumacher, 2016).But MIDAS models come with an innovation to use all the indicators at their own frequency. That is, it allows regressand and regressor to be sampled at different frequencies. MIDAS approach relates low-frequency regressand to the lagged observations of the high-frequency regressors in a parsimonious way. The parameter proliferation problem is solved by these models as MIDAS reduces the coefficients by placing higher weights on closer lags as compared to the distant ones. Ghysels et al. (2004) suggested different functional forms of the lag polynomials to estimate the coefficients by imposing weights. MIDAS model basic equation is as follow:

$$y_t = \alpha + \sum_{i=1}^{p} \beta_i L^i y_t + \gamma \sum_{k=0}^{K} B\,(k;\theta) L^{\frac{k}{m}} x_t^{(m)} + \mu_t \qquad \dots (1)$$

Here, $x_t^{(m)}$ and y_t represents high and low-frequency indicators respectively. This is an autoregressive distributed lag model (ADL-MIDAS) as given by (Clements and Galvao, 2009) where y_t is regressed on its own lags along with x_t. The superscript *m* thus refers to daily, weekly or monthly observations being observed in a single week, month or quarter. β_i, the regression coefficients represent the lags of the dependent variable (GDP growth) to be included in the model. $\gamma B(k,\theta) L^{\frac{k}{m}} x_t^{(m)}$, this term shows the relation of a high-frequency indicator with the low-frequency indicator, where γ is the regression coefficient and $B(k,\theta)$ denotes the functional form of the lags

[3] Real GDP is converted into growth rate by standard growth rate formula: $\frac{y_2 - y_1}{y_1}$ ***100.**

and the number of high frequency lags to be incorporated in the model. $L^{\frac{k}{m}}$ is the lag operator, such that: $L^{\frac{k}{m}} x_t^{(m)} = x_{t-\frac{k}{m}}^{(m)}$. $B(k, \theta)$ is a polynomial function that determines the weights for temporal aggregation, vector θ consists of a specific number of parameter and μ_t is the white noise process (Andreou et al., and Armesto et al., 2010). The principle of parsimony is satisfied under MIDAS regression models because an important feature of weighting polynomial function $B(k, \theta)$ is to estimate the limited number of parameters irrespective of the number of high-frequency lags or the value of *m*.

For further clarification, moving ahead towards the Restricted Mixed Data Sampling Regression (R-MIDAS) model, which is the same as the above described in eq. (1). Suppose that the variable y_t is quarterly available while another variable $x_t^{(m)}$ is available monthly, say 3 times in a single quarter that is m=3. The low-frequency dependent variable y_t of the model depends on the history of lagged observations of $x_{t-\frac{k}{m}}^{(m)}$. A simple R-MIDAS model is:

$$y_t = \beta_0 + \beta_1 \sum_{k=0}^{K} B\,(k; \theta) x_{t-\frac{k}{3}}^{(3)} + \mu_t \qquad \dots (2)$$

As the summation begins from k=0, it represents the inclusion $x_t^{(m)}$ in the model, this is the third month of each quarter. K represents the number of lags so it should be properly specified, AIC criterion has been used for optimal lag selection. For instance, in eq. 2, it is assumed that K=m-1 is 3-1=2. The expanded equation is as follows:

$$y_t^{(q)} = \beta_0 + \beta_1 \left\{ B(0; \theta) x_t^{(3)} + B(1; \theta) x_{t-\frac{1}{3}}^{(3)} + B(2; \theta) x_{t-\frac{2}{3}}^{(3)} \right\} + \mu_t \qquad \dots (3)$$

Here, the dependent variable y_t is regressed on the high-frequency variable $x_t^{(3)}$, which represents the third month of each quarter, the first lagged observation of high-frequency variable $x_t^{(3)}$ that is $x_{t-\frac{1}{3}}^{(3)}$representing the second month of each quarter and the second lagged observation of high-frequency variable $x_t^{(3)}$ that is $x_{t-\frac{2}{3}}^{(3)}$representing the third month of each quarter (Sinko, 2008 and Andreou et al., 2010). The model equation (3) can be extended by the inclusion of an autoregressive term (Lagged observations of y_t) and multiple high-frequency variables. The purpose of this research is to forecast quarter-on-quarter GDP growth, using monthly indicators. The parsimony of the coefficients is one of the features of MIDAS regression analysis and it can be achieved by imposing restrictions on the lagged coefficients $B(k;\theta)$ (Sinko, 2008). There are different specifications of lagged coefficients that exist in the literature. But this research has particularly focused on two popular types of restrictions namely PDL/Almon and Step function methods.

3.2.2.1. PDL/ALMON

Almon Lag weighting also called Polynomial Distributed Lag (PDL) weighting is broadly used to impose restrictions on the lag coefficients of autoregressive models. It is used to specify $\beta_i B(k;\theta)$ together in MIDAS regression (Foroni et al., 2011), defined as:

$$\beta_1 B(k;\theta_0,\theta_1,\dots,\theta_p) = \theta_0 + \sum_{p=0}^{P} \theta_p k^p \qquad \dots (4)$$

Where p is the Almon order polynomial, k is the number of lags. This polynomial can also be written in matrix form as:

$$\beta_1 \begin{bmatrix} B(0;\theta) \\ B(1;\theta) \\ B(2;\theta) \\ \vdots \\ B(k;\theta) \\ \vdots \\ B(K;\theta) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0^2 & \cdots & 0^p \\ 1 & 1 & 1^2 & \cdots & 1^p \\ 1 & 2 & 2^2 & \cdots & 2^p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & k & k^2 & \cdots & k^p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & K & K^2 & \cdots & K^P \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_P \end{bmatrix}$$

Using matrix representation, represented above, high-frequency data regressors are completely transformed and then the use of Almon lags in MIDAS models can be achieved via OLS estimation. Once the weights are estimated via OLS, they can be rescaled to obtain a slope coefficient. PDL is non-normalized, that is, the sum of estimated weights is not equal to 1. Importantly, the number of coefficients to be estimated depends on the polynomial order of degree *p* and not on the number of high-frequency lags.

3.2.2.2. STEP Function Weighting

MIDAS with Step functions, introduced by Forsberg and Ghysels (2004). The simplest method employs the step functions:

$$\beta\eta_i(k;\theta_1,\dots,\theta_p) = \theta_1 I_{i\in[a_0,a_1]} + \sum_{p=2}^{P} \theta_p I_{i\in[a_{p-1},a_p]} \qquad \dots (5)$$

$$a_0 = 1 < a_1 < \cdots < a_p = N$$

Where, k is the number of high-frequency lags; n is the step length and $\emptyset_m = \theta_i$ for $k = int(\frac{m}{\eta})$By using the step functions, this approach restricts the coefficients of high-frequency data. Within each given step high frequency lagged observations are sharing values of $\emptyset$. For instance, with η=3 (step length) the first three lagged high-frequency lags employ the same coefficient θ_0, the next three lags employ θ_1, further next three employ θ_2 and so on till maximum lags k employed reaches. The inclusion of more steps in the regression makes the model parsimonious yet defining the purpose of MIDAS regressions (Ghysels, Sinko and Valkanov, 2005). Particularly, in this step weighting model, the number of high-frequency coefficients increase with the number of high-frequency lags, but in comparison to the individual coefficient approach (U-MIDAS) the number of coefficients is reduced by a factor $\frac{1}{\eta}$.

3.2.2.3. U-MIDAS

Another alternative of the MIDAS model is the Unrestricted Mixed Data Sampling (U-MIDAS) model based on unrestricted lag polynomials suggested by Foroni, Marcellino, and Schumacher (2011). The autoregressive distributed lag U-MIDAS model can be written as:

$$y_t^{(q)} = \alpha + \sum_{i=1}^{p} \beta_i L^i y_t + \sum_{k=0}^{K} \gamma_k L^{\frac{k}{m}} x_t^{(m)} + \mu_t \quad \dots (6)$$

Eq. 6 is the same as Eq. 1 (the basic MIDAS model), the only difference comes with the weighting schemes of the high-frequency variable that is γ in U-MIDAS while $B(k;\theta)$ in R-MIDAS. Imposing restrictions by using different lag specifications results in estimating few numbers of parameters regardless of the number of lags of high-frequency variables is an essential feature of R-MIDAS models. Even though the U-MIDAS model preserves the timing information by directly incorporating high-frequency variables, but it may lead towards the problem of parameter proliferation as extending the model (4) to multiple lags increases the number of parameters to be estimated (Foroni et al., 2011). U-MIDAS are useful in macroeconomic applications when the differences in the sampling frequencies are small, for instance, quarterly and monthly. Foroni et al., (2015) noted that difference between restricted and unrestricted MIDAS and suggested to directly estimate the coefficients without confining them via any polynomial form in nowcasting the GDP applications. This is possible when frequency differences are nominal. Unlike, the financial applications where these differences are quite large, in a monthly-daily example, 22 coefficients for variable x, need to be estimated alone.

3.2.4. Empirical Models

The main MIDAS models used in this study to forecast GDP growth rate are as follows:

- GDP Forecasting Model with PCA's: (R-MIDAS)

$$\mathrm{GGDP}_t^{(q)} = \beta_0 + \beta_1 \mathrm{GGDP}_{t-1}^{(q)} + \beta_2 \sum_{k=0}^{K} B_1(k,\theta)\mathrm{PC1}_{t-\frac{k}{m}}^{(m)} + \beta_3 \sum_{i=0}^{K} B_2(i,\theta)\mathrm{PC2}_{t-\frac{i}{m}}^{(m)} + \beta_4 \sum_{j=0}^{K} B_3(j,\theta)\mathrm{PC3}_{t-\frac{j}{m}}^{(m)} + \mu_t \qquad \ldots (10)$$

- GDP Forecasting Model with Selected Indicators: (R-MIDAS)

$$\begin{aligned}\mathrm{GGDP}_t^{(q)} = {} & \beta_0 + \beta_1 \mathrm{GGDP}_{t-1}^{(q)} + \beta_2 \sum_{k=0}^{K} B_1(k,\theta)\mathrm{LCPI}_{t-\frac{k}{m}}^{(m)} + \beta_3 \sum_{i=0}^{K} B_2(i,\theta)\mathrm{LX}_{t-\frac{i}{m}}^{(m)} \\ & + \beta_4 \sum_{j=0}^{K} B_3(j,\theta)\mathrm{LWR}_{t-\frac{j}{m}}^{(m)} + \beta_5 \sum_{l=0}^{K} B_4(l,\theta)\mathrm{ER}_{t-\frac{l}{m}}^{(m)} \\ & + \beta_6 \sum_{n=0}^{K} B_5(n,\theta)\mathrm{LFDI}_{t-\frac{n}{m}}^{(m)} + \beta_7 \sum_{p=0}^{K} B_6(p,\theta)\mathrm{LPSC}_{t-\frac{p}{m}}^{(m)} \\ & + \beta_8 \sum_{q=0}^{K} B_7(q,\theta)\mathrm{LM2}_{t-\frac{q}{m}}^{(m)} + \mu_t \qquad \ldots (11)\end{aligned}$$

Where,

$$\theta = (\theta_1, \theta_2)$$

And,

$\mathrm{GGDP}^{(q)}$: the quarterly Gross Domestic Product growth rate as a low-frequency variable

t: time as low-frequency variable

$\mathrm{LCPI}^{(m)}$: the log of the monthly Consumer Price Index as a high-frequency variable

$\mathrm{LX}^{(m)}$: the log of the monthly Exports as a high-frequency variable

$\mathrm{LWR}^{(m)}$: the log of the monthly Worker Remittances as a high-frequency variable

$ER^{(m)}$: the monthly Exchange Rate as a high-frequency variable

$LFDI^{(m)}$: the log of the monthly Foreign Direct Investment as a high-frequency variable

$LPSC^{(m)}$: the log of the monthly Private Sector Credit as a high-frequency variable

$LM2^{(m)}$: the log of the monthly Broad Money as a high-frequency variable

- GDP Forecasting Model with PCA's: (U-MIDAS)

$$\mathrm{GGDP}_t^{(q)} = \beta_0 + \beta_1 \mathrm{GGDP}_{t-1}^{(q)} + \sum_{k=0}^{K} \beta_{k,2} \mathrm{PC1}_{t-\frac{k}{m}}^{(m)} + \sum_{i=0}^{K} \beta_{I,3} \mathrm{PC2}_{t-\frac{i}{m}}^{(m)} + \sum_{j=0}^{K} \beta_{j,4} \mathrm{PC3}_{t-\frac{j}{m}}^{(m)} + \mu_t \quad \ldots (12)$$

- GDP Forecasting Model with Selected Indicators: (U-MIDAS)

$$\begin{aligned} \mathrm{GGDP}_t^{(q)} = {} & \beta_0 + \beta_1 \mathrm{GGDP}_{t-1}^{(q)} + \sum_{k=0}^{K} \beta_{k,2} \mathrm{LCPI}_{t-\frac{k}{m}}^{(m)} + \sum_{i=0}^{K} \beta_{i,3} \mathrm{LX}_{t-\frac{i}{m}}^{(m)} + \sum_{j=0}^{K} \beta_{j,4} \mathrm{LWR}_{t-\frac{j}{m}}^{(m)} \\ & + \sum_{l=0}^{K} \beta_{l,5} \mathrm{ER}_{t-\frac{l}{m}}^{(m)} + \sum_{n=0}^{K} \beta_{n,6} \mathrm{LFDI}_{t-\frac{n}{m}}^{(m)} + \sum_{p=0}^{K} \beta_{p,7} \mathrm{LPSC}_{t-\frac{p}{m}}^{(m)} \quad \ldots (13) \\ & + \sum_{q=0}^{K} \beta_{q,8} \mathrm{LM2}_{t-\frac{q}{m}}^{(m)} + \mu_t \end{aligned}$$

3.2.5. Forecast Evaluation

The prime objective of using these time series models is to forecast the values for GDP at future times. In univariate time series models, a purpose to forecast is based on the available history of the series up to time t. But, in MIDAS models, a forecast can be performed multiple-step ahead (y_{t+h}) for a low-frequency variable (y_t) based on its past values and the lagged high- frequency values $x_{t-\frac{k}{m}}$ of other time series. To assess the precision of those forecasts, this study observed few forecast evaluation statistics including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Theil Inequality Coefficient. "Forecast Error" is defined as the difference between the actual value and the predicted value i.e. $e_{t+1} = y_{t+1} - \hat{y}_{t+1}$. Where, y_{t+1} is the observed value while $\hat{y}_{t+1}$ is the predicted forecast point. The formulas for the forecast evaluation statistics are given as follow:

1. Root Mean Square Error (RMSE):

$$\sqrt{\sum_{t=T+1}^{T+h} \frac{(e_t)^2}{h}}$$

2. Mean Absolute Error (MAE):

$$\sum_{t=T+1}^{T+h} \frac{|e_t|}{h}$$

3. Mean Absolute Percentage Error (MAPE):

$$100 \sum_{t=T+1}^{T+h} \frac{|\, e_t/y_t \,|}{h}$$

4. Theil U- Statistic:

$$\frac{\sqrt{\sum_{t=T+1}^{T+h} (e_t)^2/h}}{\frac{\sqrt{\sum_{t=T+1}^{T+h} \hat{y}_t^{\,2}}}{h} + \frac{\sqrt{\sum_{t=T+1}^{T+h} y_t^{\,2}}}{h}}$$

Where; $e_t = y_t - \hat{y}_t$

3.2.6. Diebold Mariano (DM) Test:

To check, either the two forecasts available for a series of interests are equally good or not, Diebold Mariano (DM) proposed a test statistic (Mariano, 1995). The loss associated with the forecast error, is given by $\mathrm{g}(e_{it})$,defining loss as a function of error. Generally, $\mathrm{g}(e_{it})$,is the squared error loss or absolute error loss of e_t. The loss difference between the two forecasts is given by:

$$\mathrm{d} = \mathrm{g}(\mathrm{e}_1) - \mathrm{g}(\mathrm{e}_2)$$

Two forecasts have equal accuracy if and only if the loss differential has zero expectation for all t. Testing the Null hypothesis as:

$$\mathrm{H}_0\text{: } \mathrm{d(mean)} = 0$$

Versus, Alternative hypothesis:

$$\mathrm{H}_0\text{: } \mathrm{d(mean)} \neq 0$$

It means, Null Hypothesis is that the two models have similar accuracy. While, the Alternative Hypothesis is that the two models have different accuracies

4. Results and Discussion

The results of this study present the out-of-forecast accuracy one-step-ahead (h=1) and in-sample forecasting from different forecast approaches namely U-MIDAS, MIDAS along with two functional forms (PDL/ALMON and Step weighting functions) compared to the ARDL benchmark. Given a number of leading indicators and different model groups, the table contains the summary statistics for the distribution of MAE, MAPE, RMSE and Theil Inequality Coefficient[4] from all the models with a different set of predictors as in Stock and Watson (2012) and Foroni (2014). There are several dimensions of short-term forecasting performances in this analysis. But, two dimensions affecting the performance are the lag length selection and lag polynomial functional form of MIDAS equations (Mahmut, 2020). All these forecasted values have been generated by single equations regression models (Eq. 10-15) mentioned in the previous chapter.

Table 4.1 Out-of-Sample Forecasting Results (Static Forecasting)

- PCAs

Year:		U-MIDAS		PDL/ALMON		STEP		ARDL	
Quarter	Observed	Forecast	Error	Forecast	Error	Forecast	Error	Forecast	Error
2020 : Q1	0.4329	0.3880	0.0449	0.0300	0.4029	0.5881	-0.1552	0.4774	-0.0445
2020 : Q2	-0.9286	-0.5056	-0.4342	0.8365	-1.7651	0.4602	-1.3888	0.0350	-0.9636
2020 : Q3	-2.1019	-0.2472	-1.8547	-0.7802	-1.3217	0.5068	-2.6087	0.5418	-2.6437
2020 : Q4	-7.2439	-1.9737	-5.2702	1.5591	-8.803	0.5657	-7.8096	1.9244	-9.1683
RMSE		2.8042		4.5219		4.1758		4.6852	
MAE		2.1510		3.0732		2.9905		3.2050	
MAPE		81.4%		116.8%		104.3%		91.59%	
Theil Inequality Coef.		0.5946		0.9516		0.9624		0.9520	

- Indicators

Year:		U-MIDAS		PDL/ALMON		STEP		ARDL	
Quarter	Observed	Forecast	Error	Forecast	Error	Forecast	Error	Forecast	Error
2020 : Q1	0.4329	-0.4078	0.8407	0.2506	0.1823	0.0161	0.4168	-0.9000	1.3329
2020 : Q2	-0.9286	-0.3985	-0.5301	0.4123	-1.3409	0.4848	-1.4134	0.7101	-1.6387
2020 : Q3	-2.1019	-0.4798	-1.6221	0.3729	-2.4748	0.1595	-2.2614	0.6136	-2.7155
2020 : Q4	-7.2439	1.1904	-8.4343	1.3736	-8.6175	1.2940	-8.5379	1.4810	-8.7249
RMSE		4.3231		4.5333		4.4772		4.7993	
MAE		2.8568		3.1539		3.1574		3.6030	
MAPE		111.2%		105.8%		118.4%		183.4%	
Theil Inequality Coef.		0.9589		0.9647		0.9645		0.9987	

[4] The value of Theil equality coefficient lies between 0 and 1.

Table 4.1 shows the out-of-sample forecasting results of all the models compared against each other and conventional model as followed by (Rufino, 2019) based on two different sets of regressors. First, the common components extracted using PCAs and below the results based on selected indicators. Overall, as expected, the forecast errors are small under PCAs due to the employment of more and more information. Comparing PCAs forecast results with the variables, the RMSE, MAE, MAPE and Theil Inequality Coefficient are small from all the models with PCAs while only MAPE is small for variables under PDL. There exists a little difference between the values, comparing to PCAs where indicators are showing small errors.

Generally, PCAs perform significantly well and offer consistent improvement comparative to variables. This exists a relatively large literature on forecasting/nowcasting with PCA confirming the fact that forecast errors are minimum by incorporating a large set of information (Angelini et al., 2011, Kuzi et al., 2013 and Heinisch, K., and Scheufele, R., 2018).

Table 4.2 Out-of-Sample Forecasting Results (Dynamic Forecasting)

- PCAs

Year:		U-MIDAS		PDL/ALMON		STEP		ARDL	
Quarter	Observed	Forecast	Error	Forecast	Error	Forecast	Error	Forecast	Error
2020 : Q1	0.4329	1.3086	-0.8757	0.1148	0.3181	-0.4697	0.9026	0.5929	-0.16
2020 : Q2	-0.9286	0.8780	-1.8066	1.0346	-1.9632	1.0704	-1.999	1.1645	-2.0931
2020 : Q3	-2.1019	-2.3246	0.2227	-0.5495	-1.5524	-0.7854	-1.3165	0.0803	-2.1822
2020 : Q4	-7.2439	1.4400	-8.684	1.1561	-8.4	0.4794	-7.7233	2.0618	-9.3057
RMSE		4.4579		4.3853		4.0679		4.8930	
MAE		2.8972		3.0584		2.9853		3.4352	
MAPE		131.8%		118.6%		148.2%		123.6%	
Theil Inequality Coef.		0.8279		0.9469		0.8940		0.9733	

- Indicators

Year:		U-MIDAS		PDL/ALMON		STEP		ARDL	
Quarter	Observed	Forecast	Error	Forecast	Error	Forecast	Error	Forecast	Error
2020 : Q1	0.4329	-1.1789	1.6118	-0.7532	1.1861	-1.8655	2.2984	-0.3383	0.7712
2020 : Q2	-0.9286	-1.8415	0.9129	0.8250	-1.7536	0.7515	-1.6801	0.7284	-1.657
2020 : Q3	-2.1019	-3.3575	1.2556	0.1997	-2.3016	-1.1736	-0.9283	0.4440	-2.5459
2020 : Q4	-7.2439	2.2560	-9.4999	-0.1943	-7.0496	1.8733	-9.1172	2.1175	-9.3614
RMSE		4.8799		3.8560		4.7982		4.9360	
MAE		3.3200		3.3727		3.5060		3.5839	
MAPE		165.3%		167.4%		220.4%		151.7%	
Theil Inequality Coef.		0.7993		0.8800		0.9052		0.9951	

Table 4.3 In-Sample Forecasting Results

	U-MIDAS	PDL/ALMON	STEP	ARDL
▪ PCAs				
RMSE	0.8739	0.9251	0.8859	1.1287
MAE	0.6229	0.6936	0.6181	0.9351
MAPE	90.15	98.78	93.86	96.03
Theil Inequality Coef.	0.2291	0.2378	0.2361	0.2959
▪ Indicators				
RMSE	0.8396	0.9061	0.8600	0.9434
MAE	0.5767	0.6568	0.6018	0.7007
MAPE	72.64	87.57	84.02	85.67
Theil Inequality Coef.	0.2126	0.2209	0.2144	0.2399

Comparing the model performances within PCAs, indicators and both, U-MIDAS outperforms R-MIDAS and ARDL in both PCAs and variables and U-MIDAS-PCA outpaces each other model. This is in line with the study of Foroni and Marcellino, 2012 which showed that U-MIDAS performed better than MIDAS. The RMSE from U-MIDAS-PCA is the smallest recorded as 2.80 relative to 4.52, 4.17 and 4.68. U-MIDAS models are relatively better in cases where frequency differences are small (Mikosch and Solanko, 2019, Mustafa M. Al-Qawasmi, 2014 and Schumacher, 2011).

The forecasted growth rates for the four quarters of 2020 are 0.39%, -0.51%, -0.24% and -1.97% respectively. Adding these figures, the forecasted annual growth rate for the year 2020 turns out to be -2.3%. While the forecasted growth rate for the FY-20 (July-June)[5] is -0.1%. SBP reports this real growth rate of about -0.4% for the FY-20. The reasons for these negative and zero growth rates are very much obvious, Pakistan's economy devastated abruptly due to lockdowns, unemployment, cancellation of orders, deferment of investment payments and plans and a halt in production and real activities.[6] Other models forecasting performances are good too for the majority of quarters. But, most of the models failed in forecasting the last quarter of the year 2020, that is, the second quarter of FY-21. The reason to this is that the high frequency data obtained is firsthand information, the models may improve their forecasting performances once information data sets are revised.

Common components extracted restores large datasets to few numbers of high-frequency explanatory variables rather than any non-linear polynomial specifications, directly estimating the

[5] The Pakistan's Fiscal year begins from 1 July of the previous calendar and concludes on 30 June.
[6] Source: SBP's third quarterly report 2019-2020.

coefficients outcomes in better forecast performances is in line with Foroni (2015). In addition to small frequency differences, the use of lots of information results in lower RMSE, MAE, MAPE and Theil Inequality Coefficient. After U-MIDAS, R-MIDAS performs better than the benchmark, both in PCAs and variables. U-MIDAS, R-MIDAS and ARDL models represents a tradeoff between parsimony and flexibility and this is in line with Armesto, Engemann and Owyang (2010). The polynomial lag specification step function is considered better than the PDL under restricted MIDAS. PDL shows a low Theil inequality coefficient in PCAs and low MAPE in variables as compared to STEP. MAE and Theil inequality coefficient in variables are almost the same under both R-MIDAS specifications. Static forecasting is used for the one-step-ahead forecast that is h=1 and forecast has been performed by Static forecasting. Other than this, Dynamic forecasting has also been employed to forecast multiple periods that are h=4 (covering 4 quarters for 2020). Table 4.2 presents the out-of-sample dynamic forecasting. Comparing the Static and Dynamic forecasting for out-of-sample performance, one step ahead forecasting reports better results. MIDAS models are particularly more reliable for short-term and intra-period forecasting/nowcasting (Schumacher, 2016).

Table 4.3 shows the in-sample forecasting performance of all the models for the forecast sample 2015Q1-2020Q4 with PCAs and selected indicators. As the estimation sample covers the year 2020, structural break due to COVID-19 crisis has been captured by including a dummy variable. U-MIDAS, R-MIDAS and ARDL models with leading indicators have shown low RMSE, MAE, MAPE and Theil inequality coefficient rather than PCAs. U-MIDAS-LI outperforms all other models, RMSE, MAE, MAPE and Theil Inequality Coefficient are 0.83, 0.57, 72% and 0.21 respectively. Once the estimation sample is complete, the forecasted sample improves the predictions even with the selected indicators. The reason for this is that the indicators (CPI, FDI, Exports, ER, etc.) are strong determinants of GDP growth. And the inclusion of these essential variables results in desirable forecasts as found by Franco and Mapa (2014). U-MIDAS-PCA, U-MIDAS-LI,R-MIDAS-PCA and R-MIDAS-LI show better forecast results by reporting lower RMSE, MAE, MAPE and Theil inequality coefficient as compared to the benchmark. Under R-MIDAS, the Step function outpaces PDL in both PCAs and indicators just like in out-of-sample forecast performance. The in-sample and out-of-sample results are alike in this study, MIDAS regressions outperform benchmark, as noted by Shi (2010). This in-sample forecasting has been performed by Dynamic forecasting. The graphs of in-sample forecasting have been shown in Appendix B.

This would be interesting to analyze that which of this approach performs better for short-term and long-term forecasting. But, as this study is undertaken for short-term forecasting, the results that MIDAS outperforms the benchmark time averaging model are in line with the study of Jiang, Guo and Zhang (2017). And, in MIDAS, U-MIDAS-PCA performs relatively well in forecasting quarter-on-quarter GDP growth for Pakistan as compared to other models.

4.1. Diebold Mariano Test:

Table 4.4 presents the Diebold Mariano (DM) test results, to check the statistical significance of the difference in the forecasting performance of the two models. When the resulted t-value is negative, it means that MIDAS has lower RMSE on average. And if the negative number reported is also statistically significant, the difference in the RMSE of the two models is also significant (Bhaghoe, Ooft and Franses, 2019 and Asimakopoulos, Paredes and Warmedinger, 2013). Since in this research study primarily we are comparing MIDAS (U-MIDAS and R-MIDAS) with time averaging model ARDL. The results in the table confirm the results mentioned earlier. MIDAS outperforms the benchmark.

Table 4.4 Diebold Mariano (DM) Test Results

Out-of-sample forecasting performance	Static Forecasting		Dynamic Forecasting	
	PCAs t-values (p-values)	Indicators t-values (p-values)	PCAs t-values (p-values)	Indicators t-values (p-values)
U-MIDAS versus PDL	0.7585	-0.5294*	1.5457	2.1877
	(0.6207)	(0.3166)	(0.8901)	(0.9418)
U-MIDAS versus STEP	0.8729	0.0857	1.8146	1.2171
	(0.7765)	(0.5315)	(0.9164)	(0.8447)
U-MIDAS verses ARDL	-0.3001*	-3.3512*	-0.6110*	-2.6337*
	(0.0322)**	(0.0220)**	(0.0402)**	(0.0396)**
PDL versus STEP	1.2538	1.4009	0.3854	-2.3472*
	(0.8506)	(0.8721)	(0.6372)	(0.0503)**
PDL versus ARDL	-0.2836*	-3.5213*	-1.1326*	-0.8655*
	(0.3975)	(0.0194)**	(0.1698)	(0.2252)
STEP versus ARDL	-0.8866*	-4.0808*	-0.8982*	-0.9672*
	(0.2203)	(0.0266)**	(0.2176)	(0.7976)

* reports negative t values and the significance level at 5% is represented by **

In static forecasting, U-MIDAS and R-MIDAS (PDL and STEP) models are showing negative numbers in comparison to ARDL, indicating on average the lower RMSE. The negative number reported by the U-MIDAS-PCA is also statistically significant, rejecting the Null hypothesis as mentioned in the previous chapter, of having the same predicting accuracy between the models. While with indicators, the negative numbers reported by U-MIDAS, PDL and STEP function are also statistically significant, which means the difference between the RMSE's of models is also significant. This indicates that by taking high-frequency data into account when forecasting the quarter-on-quarter GDP growth, the resulted forecast from MIDAS statistically differs from the forecast reported from the ARDL. Rejecting the null hypothesis, $H_{0,}$ that is the two forecasts have equal accuracy. No statistical significance has been observed between PDL and STEP functions. This means that both polynomial specifications are performing the same. U-MIDAS-LI versus MIDAS-LI with PDL also reports a negative value, demonstrating lower RMSE for U-MIDAS-LI. Dynamic Forecasting with PCAs results is the same as the Static forecasting, U-MIDAS-PCA and MIDAS-LI with PDL and STEP functions report lower RMSE with negative numbers. While, U-MIDAS-LI reports a statistically significant negative number in comparison with ARDL rejecting the Null hypothesis H_0, PDL and STEP sign tend to be negative but not significant. PDL versus STEP shows a negative statistically significant number indicating that the predicting accuracy of both the models is not the same.

5. Conclusion

In this research study, we empirically investigate the power of leading indicators in forecasting the quarterly GDP growth rate for Pakistan. The goal is to analyse the forecasting performance of Mixed Data Sampling (MIDAS) regressions namely U-MIDAS and R-MIDAS and to make a comparison with the traditional benchmark model ARDL. This study incorporates leading macro-economic variables available at a monthly frequency from January 2002 to December 2020, for anticipating the quarterly GDP growth rates for the year 2020. Using high-frequency data in its original frequency without imposing any time aggregation technique, which may result in the discard of relevant information, adds a significant contribution to the existing nowcasting/forecasting literature on econometrics time series forecasting models in Pakistan.

The results show that, according to all the forecast accuracy measures that are RMSE, MAE, MAPE and Theil Inequality Coefficient, MIDAS models outperform the benchmark in both in-sample and out-of-sample forecasting, recording the lowest values of all these statistics. The first set of results obtained by PCA outperform the selected indicators approach in the out-of-sample forecast. Inclusion of lots and lots of potential timely information results in better forecasting accuracy of the models. While, in MIDAS, U-MIDAS-PCA outpaces each other model displaying the lowest of RMSE as 2.80. These results are in line with (Mikosch and Solanko 2019, Al-Qawasmi, 2014 and Schumacher, 2011) as when frequency differences are small, U-MIDAS is the most appropriate forecasting tool. The forecasted value for the annual year 2020 stands out to be -2.3%, for FY-20 it is -0.1% while SBP reports -0.4%. The declined growth rate is the result of devastations caused by Covid-19. The forecasted growth rates from U-MIDAS for the first two-quarters of FY-21 are -0.24% and -1.97%. Comparing the MIDAS lag polynomial specifications results in an improved forecast by Step weighting function other than the PDL in the case of Pakistan. Step shows lower forecast errors for most of the models. Lag length and polynomial specification selection are crucial in MIDAS regression analysis.

The second set of results attained by the selective indicators approach are in line with the results from PCA in out-of-sample forecasting. MIDAS models outperform ARDL, U-MIDAS performs remarkably well, and STEP forecast errors are lower than PDL. However, comparing the two sets of results, PCA forecast accuracy measures are small than selective indicators. In-sample forecasting shows minimum forecast evaluation statistics by selective indicators. U-MIDAS-LI outpaces all other models. As the variables are the strong determinants of GDP, completion of estimation sample improves the in-sample forecasting.

This study opens avenues for future research. Anticipating the GDP growth rates by the information contained in macroeconomic fundamentals can be extended by incorporating more high-frequency indicators and by disaggregating the GDP components into Expenditure and Production sides. It is possible that forecasting the disaggregated GDP components could be more beneficial in tracking the current situation of the economy in more depth under these unforeseen circumstances.

References

Ajmair, M., Gilal, M. A., Farooq, S., and Hussain, K. (2018). Factors determining Economic Growth in Pakistan: An ARDL Bound Testing with General to Specific Approach. Pakistan Journal of Applied Economics, 28(1).

Ali, A., and Saif, S. (2017). Determinants of Economic Growth in Pakistan: A Time Series Analysis. European Online Journal of Natural and Social Sciences, 6(4), 686.

Al-Qawasmi, M. M. (2014). Forecasting Palestinian gross domestic product using mixed data sampling regression techniques (Doctoral dissertation, Birzeit University).

Andreou, E., Ghysels, E., and Kourtellos, A. (2010). Regression models with mixed sampling frequencies. Journal of Econometrics, 158 (2), 246-261.

Andreou, E., Ghysels, E., and Kourtellos, A. (2011). Forecasting with mixed-frequency data. In The Oxford handbook of economic forecasting.

Andreou, E., Ghysels, E., and Kourtellos, A. (2013). Should macroeconomic forecasters use daily financial data and how? Journal of Business and Economic Statistics, 31(2), 240-251.

Angelini, E., Camba-Mendez, G., Giannone, D., Reichlin, L., and Rünstler, G. (2011). Short-term forecasts of euro area GDP growth.

Armesto, M. T., Engemann, K. M., and Owyang, M. T. (2010). Forecasting with mixed frequencies. Federal Reserve Bank of St. Louis Review, 92(6), 521-36.

Asimakopoulos, S., Paredes, J., and Warmedinger, T. (2013). Forecasting fiscal time series using mixed frequency data (No. 1550). ECB Working Paper.

Bhaghoe, S., and Ooft, G. (2021). Nowcasting Quarterly GDP Growth in Suriname with Factor-MIDAS and Mixed-Frequency VAR Models (No. 176). The Johns Hopkins Institute for Applied Economics, Global Health, and the Study of Business Enterprise.

Bhaghoe, S., Ooft, G., and Franses, P. H. (2019). Estimates of quarterly GDP growth using MIDAS regressions (No. EI2019-29).

Chikamatsu, K., Hirakata, N., Kido, Y., and Otaka, K. (2018). Nowcasting Japanese GDPs. Bank of Japan.

Chohan, U. W. (2020). Forecasting the economic impact of coronavirus on developing countries: case of Pakistan.

Clements, M. P., and Galvão, A. B. (2009). Forecasting US output growth using leading indicators: An appraisal using MIDAS models. Journal of Applied Econometrics, 24(7), 1187-1206.

Dias, F., Pinheiro, M., and Rua, A. (2015). Forecasting Portuguese GDP with factor models: Pre- and post-crisis evidence. Economic Modelling, 44, 266-272.

Drechsel, K., & Scheufele, R. (2010). Should we trust in leading indicators? Evidence from the recent recession (No. 10/2010). IWH Discussion Papers.

Foroni, C. (2012). Econometric models for mixed-frequency data (Doctoral dissertation).

Foroni, C., and Marcellino, M. (2014). A comparison of mixed frequency approaches for nowcasting Euro area macroeconomic aggregates. International Journal of Forecasting, 30(3), 554-568.

Foroni, C., and Marcellino, M. G. (2013). A survey of econometric methods for mixed-frequency data. Available at SSRN 2268912.

Foroni, C., Marcellino, M. G., and Schumacher, C. (2011). U-MIDAS: MIDAS regressions with unrestricted lag polynomials.

Foroni, C., Marcellino, M., and Schumacher, C. (2015). Unrestricted mixed data sampling (MIDAS): MIDAS regressions with unrestricted lag polynomials. Journal of the Royal Statistical Society. Series A (Statistics in Society), 57-82.

Frale, C., and Monteforte, L. (2011). FaMIDAS: A mixed frequency factor model with MIDAS structure. Bank of Italy Temi di Discussione (Working Paper) No, 788.

Franco, R. J. G., and Mapa, D. S. (2014). The Dynamics of Inflation and GDP Growth: A Mixed Frequency Model Approach.

Ghysels, E., and Wright, J. H. (2009). Forecasting professional forecasters. Journal of Business and Economic Statistics, 27(4), 504-516.

Ghysels, E., Santa-Clara, P., and Valkanov, R. (2004). The MIDAS touch: Mixed data sampling regression models.

Ghysels, E., Sinko, A., and Valkanov, R. (2007). MIDAS regressions: Further results and new directions. Econometric reviews, 26(1), 53-90.

Gunay, M. (2020). Nowcasting Turkish GDP with MIDAS: Role of Functional Form of the Lag Polynomial (No. 2002).

Hanif, N., Iqbal, J., and Malik, J. (2013). Quarterisation of national income accounts of Pakistan.

Heinisch, K., and Scheufele, R. (2018). Bottom-up or direct? Forecasting German GDP in a data-rich environment. Empirical Economics, 54(2), 705-745.

Hussain, F., Hyder, K., and Rehman, M. (2018). Nowcasting LSM Growth in Pakistan. STATE BANK OF PAKISTAN, 1.

Iqbal, Z., and Zahid, G. M. (1998). Macroeconomic determinants of economic growth in Pakistan. The Pakistan Development Review, 125-148.

Jiang, Y., Guo, Y., and Zhang, Y. (2017). Forecasting China's GDP growth using dynamic factors and mixed-frequency data. Economic Modelling, 66, 132-138.

Koenig, E. F., Dolmas, S., and Piger, J. (2003). The use and abuse of real-time data in economic forecasting. Review of Economics and Statistics, 85(3), 618-628.

Kuzin, V., Marcellino, M., and Schumacher, C. (2013). Pooling versus model selection for nowcasting GDP with many predictors: Empirical evidence for six industrialized countries. Journal of Applied Econometrics, 28(3), 392-411.

Lahiri, K., & Moore, G. H. (Eds.). (1992). Leading economic indicators: new approaches and forecasting records. Cambridge University Press.

Liu, P., Matheson, T., and Romeu, R. (2012). Real-time forecasts of economic activity for Latin American economies. Economic Modelling, 29(4), 1090-1098.

Mikosch, H., and Solanko, L. (2019). Forecasting quarterly Russian GDP growth with mixed-frequency data. Russian Journal of Money and Finance, 78(1), 19-35.

Rehman, N. U. (2016). FDI and economic growth: empirical evidence from Pakistan. Journal of Economic and Administrative Sciences.

Rufino, C. C. (2019). Nowcasting Philippine Economic Growth Using MIDAS Regression. DLSU Business and Economics Review, 29(1), 14-23.

Samuelson, P. A. (1939). A Synthesis of the Principle of Acceleration and the Multiplier. Journal of Political Economy, 47(6), 786-797.

Schumacher, C. (2016). A comparison of MIDAS and bridge equations. International Journal of Forecasting, 32(2), 257-270.

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

Shahbaz, M., Ahmad, K., and Chaudhary, A. R. (2008). Economic growth and its determinants in Pakistan. The Pakistan Development Review, 471-486.

Silvestrini, A., and Veredas, D. (2008). Temporal aggregation of univariate and multivariate time series models: a survey. Journal of Economic Surveys, 22(3), 458-497.

Sinko, A. (2008). Some applications of mixed data sampling regression models.

Stock, J. H., and Watson, M. W. (2012). Generalized shrinkage methods for forecasting using many predictors. Journal of Business and Economic Statistics, 30(4), 481-493.

Syed, A. A. S., and Lee, K. H. (2021). Macroeconomic forecasting for Pakistan in a data-rich environment. Applied Economics, 53(9), 1077-1091.

S^{3}H Working Paper

01: 2018 Impact of the 18th Amendment on Energy Security and Governance Parameters in Energy Sector of Pakistan by Hafiz Fawad Khan and Faisal Jamil (2018), 33 pp.

02: 2018 Role of Hi-Tech Trade, Foreign Investment and Intellectual Property in Stimulating Innovation and Economic Growth in the South Asian Economies by Nimra Shahid and Zafar Mahmood (2018), 37 pp.

03: 2018 Factor Affecting International Outsourcing Decision: Evidence from Plant-level Data of Pakistan's Surgical Instruments Industry by Sidra Nazir and Zafar Mahmood (2018), 29 pp.

01: 2019 Income Differentials between Police and Taxation Departments in Risk Prone Peshawar City by Sanam Khan and Faisal Jamil (2019), 23 pp.

02: 2019 Pakistan's Experience with the IMF by Ashfaque Hasan Khan (2019), 21 pp.

03: 2019 DEMYSTIFYING *RIBA*: *Riba* in the Method of the Muslim Jurists by Muhammad Zahid Siddique and Muhammad Mushtaq Ahmad (2019), 33 pp.

04: 2019 Should Microfinance Institutions Diversify or Focus? Evidence from Pakistan by Mohammad Hassam Afgun, Ammar Bin Zafar, Arish Batool and Ashfaque Hasan Khan (2019), 16 pp.

05: 2019 A Critical Evaluation of the IMF Program: A Case Study of Pakistan by Oroba Tasnim Siddiqui, Sehrish Shoaib, Maham Bilal and Ashfaque Hasan Khan (2019), 29 pp.

06: 2019 Non-linear Model of Aggregate Credit Risk for Banking Sector of Pakistan: A Threshold Vector Autoregressive Approach by Muhammad Anwaar Alam Khokhar and Ather Maqsood Ahmed (2019), 30 pp.

07: 2019 Modern Money and Islamic Banking in the Light of Islamic Law of *Riba* by Muhammad Zahid Siddique (2019), 23 pp.

08: 2019 Offsetting the Beggar-thy-neighbour Effect of Chinese Exchange Rate Policy on Pakistani Textile Exports by Saba Arif and Zafar Mahmood (2019), 41 pp.

01: 2020 Preferences for Truthfulness: An Experimental Analysis by Fizzah Najm and Muhammad Zahid Siddique (2020), 27 pp.

02: 2020 Investigating the Nexus between Fiscal Decentralization, Social Development and Economic Growth in Pakistan by Tuaha Adil and Faisal Jamil (2020), 28 pp.

03: 2020 Consumer's Perception towards Electricity Theft: A Path Analysis by Zainab Babar, Faisal Jamil and Wajiha Haq (2020), 24 pp.

04: 2020 An Investigation into the Trade Pattern of Goods Exported from Pakistan to China through FTA Analysis by Imran Ali Khan, Sameer Sajjad and Ayesha Nazuk (2020), 36 pp.

05: 2020 The Effects of Import and Export Determinants on the Balance of Trade: A Case of Pakistan by Laila Yamin, Ayesha Javaid, Bahlol Khan Orakzai and Zafar Mahmood (2020), 31 pp.

01: 2021 Hawks and Doves: Seizing Pecuniary Benefits in a Conflict Situation Maha Fayyaz and Zahid Siddique (2021), 29 pp.

Chinese Studies:

CS-01: 2016 China's Development Experience by Syed Hasan Javed (2016), 15 pp.

Development Studies:

DS-01: 2016 Rehabilitation of 2010 Flood Affected People in Pakistan: Role of Development Partners by Sheeba Farooq (2016), 39 pp.

DS-01: 2019 Spatial and Temporal Changes in the Tropospheric Ozone Concentration due to Developmental Projects under China-Pakistan Economic Corridor (CPEC) by Ramsha Munir and Umer Khayyam (2019), 29 pp.