















## Article

# Nowcasting India Economic Growth Using a Mixed-Data Sampling (MIDAS) Model (Empirical Study with Economic Policy Uncertainty–Consumer Prices Index)

Pradeep Mishra <sup>1,\*</sup>, Khder Alakkari <sup>2</sup>, Mostafa Abotaleb <sup>3</sup>, Pankaj Kumar Singh <sup>4</sup>, Shilpi Singh <sup>4</sup>,  
Monika Ray <sup>5</sup>, Soumitra Sankar Das <sup>6</sup>, Umme Habibah Rahman <sup>7</sup>, Ali J. Othman <sup>8</sup>,  
Nazirya Alexandrovna Ibragimova <sup>8</sup>, Gulfishan Firdose Ahmed <sup>9</sup>, Fozia Homa <sup>10</sup>, Pushpika Tiwari <sup>11</sup>  
and Ritisha Balloo <sup>12</sup>

- <sup>1</sup> College of Agriculture, Powarkheda, Jawaharlal Nehru Krishi Vishwavidyalaya, Jabalpur 461110, India
  - <sup>2</sup> Department of Statistics and Programming, Faculty of Economics, University of Tishreen, Lattakia P.O. Box 2230, Syria; khderalakkari1990@gmail.com
  - <sup>3</sup> Department of System Programming, South Ural State University, 454080 Chelyabinsk, Russia; abotalebmostafa@bk.ru
  - <sup>4</sup> R.D. Engineering College, Ghazibad 01001, India; p.mnavy@gmail.com (P.K.S.); shilpias20@rediffmail.com (S.S.)
  - <sup>5</sup> Regional Research and Technology Transfer Station (OUAT), Keonjhar 758002, India; monikarayouat@gmail.com
  - <sup>6</sup> Department of Statistics, Birsa Agricultural University, Kanke, Ranchi 834006, India; soumitradas30@gmail.com
  - <sup>7</sup> Department of Statistics, Assam University, Silchar 788011, India; umme.habibah.rahman17@gmail.com
  - <sup>8</sup> Department of Commodity Research and Commodity Expertise, Plekhanov Russian University of Economics, 117997 Moscow, Russia; Ali.mcisa@gmail.com (A.J.O.); naziryaibragimova@yandex.ru (N.A.I.)
  - <sup>9</sup> Department of Computer Science, College of Agriculture—JNKVV, Hoshangabad 461110, India; dr.gfahmed17@jnkvv.org
  - <sup>10</sup> Department of Statistics, Mathematics, and Computer Application, Bihar Agricultural University, Sabour, Bhagalpur 813210, India; fozia.homa.bhu@gmail.com
  - <sup>11</sup> M. Phil (NRM), Indian Institute of Forest Management (IIFM), Bhopal 462001, India; mishrapushpika013@gmail.com
  - <sup>12</sup> Department of Law and Management, University of Mauritius, Reduit 80837, Mauritius; ritishatballoo@gmail.com
- \* Correspondence: pradeepjnkvv@gmail.com



**Citation:** Mishra, P.; Alakkari, K.; Abotaleb, M.; Singh, P.K.; Singh, S.; Ray, M.; Das, S.S.; Rahman, U.H.; Othman, A.J.; Ibragimova, N.A.; et al. Nowcasting India Economic Growth Using a Mixed-Data Sampling (MIDAS) Model (Empirical Study with Economic Policy Uncertainty–Consumer Prices Index). *Data* **2021**, *6*, 113. <https://doi.org/10.3390/data6110113>

Academic Editor: Francisco Guijarro

Received: 14 October 2021

Accepted: 27 October 2021

Published: 2 November 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** Economics suffers from a blurred view of the economy due to the delay in the official publication of macroeconomic variables and, essentially, of the most important variable of real GDP. Therefore, this paper aimed at nowcasting GDP in India based on high-frequency data released early. Instead of using a large set of data thus increasing statistical complexity, two main indicators of the Indian economy (economic policy uncertainty and consumer price index) were relied on. The paper followed the MIDAS–Almon (PDL) weighting approach, which allowed us to successfully capture structural breaks and predict Indian GDP for the second quarter of 2021, after evaluating the accuracy of the nowcasting and out-of-sample prediction. Our results indicated low values of the RMSE in the sample and when predicting the out-of-sample 1- and 4-quarter horizon, but RMSE increased when predicting the 10-quarter horizon. Due to the effect of the short-term structural break, we found that RMSE values decreased for the last prediction point.

**Keywords:** nowcasting real GDP; economic policy uncertainty; consumer prices index; mixed-data sampling; almon weighting; structural break

## 1. Introduction

The official publication of many macroeconomic indicators (such as GDP) has been delayed, especially in developing countries, by Central Statistics Offices. This is a prob-

lem for economists and policy-makers who follow the state of the economy. GDP is the most important indicator for policy-makers, and is similar in describing the state of a country's economy to satellites that describe weather on earth. While meteorologists have today's weather information and only need to predict tomorrow's weather, economists need to know today's information. In the recent past, this has been described as nowcasting. Nowcasting technology allows us to exploit the information in high-frequency data (daily, weekly, monthly) to extract signals about the direction of change in low-frequency data (seasonal, yearly). This technology is only useful in an environment where high-frequency data are issued in real time. Nowcasting is used to generate real-time current GDP estimates as a low-frequency variable (annual, quarterly) based on information from high-frequency variables (monthly, daily) that are released early. Studies [1,2] gave the basis of nowcasting by studying daily GDP developments based on high-frequency data in the US economy. Giannone et al. [3] also developed a method for assessing the marginal impact of monthly data releases that have a jagged edge on current quarter real GDP forecasts. The usual methods used for forecasting rely on dynamic factor models that treat the underlying low-frequency variable of interest as a latent process with high-frequency data observations. These models are inferred using likelihood-based methods and Kalman filtering techniques [4]. Correspondingly, in order to nowcasting GDP, many studies have relied on a large set of high-frequency data using the MIDAS model according to a weighting function that fits the data [5–7]

For nowcasting low-frequency variables, researchers rely on a wide range of high-frequency early-release financial economic variables [8]. Among the most frequently used variables are exchange rates, price indices, retail trade, average income [9,10]. Macroeconomics relies on increasingly non-standard data extracted using machine learning (textual analysis) methods, with the analysis involving hundreds of time series. Some studies [11,12] investigated GDP growth forecasting in the U.S.A. using standard high-frequency time series and non-standard data generated by textual analysis of financial press articles and proposed a systematic approach to high-dimensional time regression problems.

In India, there are many studies of nowcasting GDP. Lyer and Gupta [13] used a dynamic factor model to forecast GDP growth in India on a quarterly basis from January 2000 to December 2018. The analysis included 6 quarterly indicators and 12 monthly high-frequency indicators derived from the monetary, financial, and real sectors of India. A work [14] built single-index dynamic factors (DFs) using a sequentially expanding list of 6, 9, and 12 high-frequency indicators. Another study [15] used, for nowcasting India's GDP growth, a dynamic factor model that incorporates a series of USA and Euro zone outputs in order to improve forecasts. A new framework was suggested [14] to forecast India's Gross Value Added that incorporates information of mixed-data frequencies and other data characteristics. Evening-hour luminosity was added as a crucial high-frequency indicator.

In this paper, we present a framework for the nowcasting of GDP growth in India using a Mixed-Data Frequency Model. Instead of relying on a large number of indicators in forecasting, which often leads to statistical complexity, we relied on two indicators as high-frequency data, in line with the economic theory: first, the economic policy uncertainty index for India [16] as a non-standard indicator based on textual analysis, and second, the consumer price index for India as a standard indicator. The main aim was to assess the extent to which the model is able to anticipate the significant negative effects of the Covid19 pandemic on economic growth in India, as most statistical models collapse during crises (global financial crisis) [17]. Studying the effects of uncertainty on India's economic growth is difficult because, as far as we know, these effects have not been studied before. At the same time, the proposed model provides real-time GDP growth nowcasting and continuous updates to the GDP growth in India, given that the data of Economic Policy Uncertainty index and Consumer Price Index are issued in real time, which gives a great benefit to trackers of the state of the economy.

## 2. Materials and Methods

The term MIDAS refers to regression analysis by sampling data measured at different frequencies. As a remedy to the problem of variation in statistical versions in economics, the methodology [18–20] addresses the situation where the dependent variable in the regression is sampled at a lower frequency than that of one or more of the regressors. The aim of the MIDAS approach is to incorporate the information present in the higher frequency data into the lower frequency regression in a parsimonious, yet flexible, fashion. An additional advantage of these models is that some additional value for the high-frequency variables may be available after observing the latest value of the low-frequency dependent variable sample. In this case, these additional observations can be used to update the predictions, which is called nowcasting. The following is the general form of the MIDAS regression equation:

$$Y_t = \beta X_t + \sum_{p=0}^{S-1} X_{(t-p)/S}^H \theta_p + \varepsilon_t \quad (1)$$

where  $Y_t$  is the variable low frequency during the period ( $t$ ),  $X_t$  is the set of regressors sampled at the same low frequency  $Y_t$ ,  $X_{(t-p)/S}^H$  is the set of regressors sampled at a higher frequency with  $S$  values for each low-frequency value,  $p$  is the high-frequency lag at  $t$ ,  $\theta$  indicates the partial effect parameters for each frequency interval  $S$ , and  $\varepsilon_t$  is the error term.

This approach estimates a distinct  $\theta$  for each  $S$  high-frequency lag regression. The MIDAS model estimation provides various weighting functions that reduce the number of parameters in the model by placing constraints on the effects of high-frequency variables to suit their properties and according to some assumptions.

### 2.1. Step Weighting

According to this function, the notation of Equation (1) becomes as follow [19]:

$$Y_t = \beta X_t + \sum_{p=0}^{k-1} X_{(t-p)/S}^H \theta_p + \varepsilon_t \quad (2)$$

where  $k$ : is the number of lagged high frequency, and  $p$  is the step length; the number of high-frequency coefficient increases with the number of lags. A linear trend between the variables is preferred when using this function.

### 2.2. Almon (PDL) Weighting

For each high-frequency lag up to  $k$ , the regression coefficient is modeled as a  $q$  dimensional lag polynomial in the MIDAS parameters  $\theta$ . Equation (1) follows [19]:

$$Y_t = \beta X_t + \sum_{p=0}^{k-1} X_{(t-p)/S}^H \sum_{j=1}^q p^j \theta_j + \varepsilon_t \quad (3)$$

where  $q$  is the almon polynomial order and the chosen number of lag  $k$ . This function is used if the data have multiple trends (quadratic, cubic).

### 2.3. Beta Weighting

Ghysels et al. [19] suggest a normalized beta-weighting function, according to the equation:

$$Y_t = \beta X_t + \sum_{p=0}^{k-1} X_{(t-p)/S}^H \left( \frac{i_L^{\lambda_1-1} (1-i_L)^{\lambda_2-1}}{\sum_{j=0}^k i_j^{\lambda_1-1} (1-i_j)^{\lambda_2-1}} + \lambda_3 \right) \gamma + \varepsilon_t \quad (4)$$

where  $\lambda_1$ ,  $\lambda_2$ ,  $\lambda_3$  are hyperparameters governing the shape of the weighting function,  $\gamma$  is a slope coefficient that is common across lags,  $k$  is the number of lags. This function can be used when the shape of the distribution of the variables is unknown or distorted.

### 3. Data and Theory

Our efforts for nowcasting real GDP growth in India as a low-frequency variable were based on the quarterly data available in the Federal Reserve Economic Data database from 2012 through the first quarter (Q1) of 2021. For the high-frequency (monthly) variables as independent variables, we used the India economic policy uncertainty index as a proxy for the non-standard indicators calculated according to a previously reported methodology [16]. The data are available from 2012 to July 2021 on the website [21]. The index is based on newspaper articles. Seven Indian newspapers are included: The Economic Times, The Times of India, the Hindustan Times, the Hindu, The Statesman, The Indian Express, and The Financial Express. The news articles containing at least one of the terms indicative of economic policy uncertainty were counted and classified into three groups: the first group included uncertain, uncertainties, or uncertainty. The second group included economic or economy. The third category included policy terms such as regulation, central bank, monetary policy, policymakers, deficit, legislation, and fiscal policy. The monthly Economic Policy Uncertainty article count was reduced by subtracting the number of all articles in the same newspaper and month. Each paper-specific series was normalized to standard deviation 1 prior to 2011. Once normalized, the seven newspaper-specific indices were then summed. The resulting series was normalized to mean 100 prior to 2011. Empirical studies [16,22–26] have confirmed that high levels of uncertainty lead to a slowdown in economic growth. Whereas uncertainty can lead companies to postpone hiring decisions, it affects workers' decisions, making them less willing to search for job opportunities, and it accompanies high volatility in macroeconomic variables, which makes the economic policy uncertainty index capture the slowdown in India's economic growth. The second high-frequency indicator as a standard index that we considered was the consumer price growth (inflation rate, total of all items) of India during the monthly period from 2012 to June 2021, drawn from the Federal Reserve Economic Data database. The rise in consumer prices can be a driver of economic growth and can have a negative impact on economic growth. Large and continuous rises in prices could lead to a decrease in the return on capital and weak real spending, which leads to a decrease in the confidence of investors and consumers in the economy and thus a decrease in growth. On the other hand, if price growth is controlled at acceptable levels, real consumer spending will increase, and thus the economy will achieve expansion growth. Accordingly, the importance of these indicators is reflected in the current and future forecast of the trends of the Indian economy.

### 4. Methodology

With our methodology, for nowcasting the Q2 of 2021 of India's economic growth, we divided the GDP growth data in India into two sets: the data from Q1 of 2012 to Q4 of 2018 for training, and the data from Q1 of 2019 to Q1 of 2021 to test the validity of the forecasts and the model ability for the nowcasting of developments in economic growth in India during 2019 and 2020.

Accordingly, our methodology included several stages. To visualize the data features (patterns, unusual observations, changes over time), we first plotted the data, then translated them with descriptive statistics and the normal distribution of the data using the following statistic:

$$Jarque - Bera = \frac{n}{6} \left( S^2 + \frac{1}{4}(K - 3)^2 \right) \quad (5)$$

where  $n$  is the number of observations,  $S$  is skewness,  $K$  is kurtosis. Because most economic variables show a stochastic trend and may be exposed to shocks that do not allow them to return to their pre-shock average, economic time series may undergo discrete changes at a particular date or gradual changes over time called "breaks", often associated with economic policy or structural changes in the economy. We need to study the extent to which these variables are affected by using the Breakpoint Unit Root test previously proposed [27–30]. Assuming that breaks follow the dynamic course of innovations, we tested the stationarity of our variables according to the following equation:

$$y_t = c + \alpha t + \theta DU_t(T_b) + \gamma DT_t(T_b) + \omega D_t(T_b) + \delta y_{t-1} + \sum_{i=1}^p \varphi_i \Delta y_{t-i} + \varepsilon_t \quad (6)$$

where  $\alpha$ ,  $\theta$ ,  $\gamma$ ,  $\omega$  are trend and break parameters,  $c$  is a constant,  $p$  is the lag order of the autoregressive process. The test was carried out under the null hypothesis  $\delta = 0$  (not stationary with or without break) against the alternative of  $\delta < 0$  (stationary with or without break). In the next step, in order to avoid the problem of multicollinearity, we tested the degree of a linear relationship between the high-frequency independent variables, which was obtained according to:

$$r = \frac{\sum X_1 X_2 - n \bar{X}_1 \bar{X}_2}{(n-1) * \sigma_{X_1} \sigma_{X_2}} \quad (7)$$

where  $X_1$ ,  $X_2$  are independent variables,  $\sigma_{X_1}$ ,  $\sigma_{X_2}$  indicate the standard deviation. With no multicollinearity problem, we estimated the MIDAS model for nowcasting the Q2 of 2021 and used the fourth-degree Almon PDL weighting for capturing volatility in the data and giving an appropriate slowdown period for each high-frequency variable according to the Akaike Information Criterion (AIC), which is given as follow [31]

$$-2 \log L(\hat{\theta}) + 2k \quad (8)$$

where  $\hat{\theta}$  is the maximum value of the likelihood function,  $k$  is the number of variables. We also included in the model the AR( $p$ ) term for the low-frequency variable in order to benefit from the historical information of the variable in addition to the information of the high-frequency variables in forecasting, which is an advantage for these models (AR-MIDAS) over the dynamic factor model. After estimating the model, we evaluated the model forecast ability by testing the random error term  $\varepsilon_t$ : the error term values must be distributed in a normal distribution, according to the testing Equation (5). In addition, the error term values should not be self-correlated; we tested this using the auto correlation function:

$$\rho(p) = \frac{Cov(\varepsilon_t, \varepsilon_{t+p})}{\sigma^2} \quad (9)$$

where  $p$  is the number of lags. The random error term series must be stationary, and we tested this with the augmented dickey fuller (ADF) equation [32]:

$$\Delta \varepsilon_t = c + \alpha.t + \delta \varepsilon_{t-1} + \beta_{p-1} \Delta \varepsilon_{t-p+1} \quad (10)$$

where  $c$  is a constant,  $\alpha$  is a coefficient on a time trend,  $p$  is the lag order of the autoregressive process. The ADF test was carried out under the null hypothesis  $\delta = 0$  (not stationary) against the alternative hypothesis of  $\delta < 0$  (stationary). After obtaining the predictions, the accuracy of the predictions was confirmed in- and out-of-sample and on the short and long term, depending on Root Mean-Square Error (RMSE):

$$\sqrt{\frac{\sum_{t=1}^n (\hat{y}_t - y_t)^2}{n}} \quad (11)$$

where  $\hat{y}_t$  is the forecast value,  $y_t$  is the actual value,  $n$  is the number of fits observed. The closer the value of this indicator is to zero, the more expected values are identical to the actual one [33].

## 5. Empirical Results

We created a monthly (high-frequency) visualization of India's quarterly real GDP growth, economic policy uncertainty (EPU), and consumer price rate (CPI). As an advantage over the literature reviews that used nowcasting models, the use of indicators of uncertainty and consumer prices may provide important and quick results in picking up signals of change in GDP growth, but because of the overlapping patterns and factors affecting GDP growth, the use of only two variables is a limit that affects the nowcasting results. Figure 1 shows that the margin of GDP volatility before 2020 (Stationary) was less than afterwards; it was the largest, at 2.9%, in 2016 Q1 and the lowest, at 0.8%, in 2017Q1. Then, there was a sharp decline in 2021 Q2 at a rate of  $-26\%$ . This was due to the effects of the COVID-19

pandemic and the great leap achieved by the economic growth during the third quarter of 2021, driven by the decline in the quarter before it [22–32]. This made forecasting of the growth rate more difficult, especially after the second wave of the COVID-19 pandemic. Figure 1 also shows that the largest increase in consumer prices, which reached 2013 M2, was due to the increase in the prices of vegetables and fruits. We also note that the growth rate of consumer prices took an upward trend from the beginning of 2018 until the end of 2019 and remained at high levels during the subsequent period due to the rise in prices in various economic and health sectors. We note from Figure 1 that the economic policy uncertainty index reached high levels in 2012 and 2020.



**Figure 1.** Evolution of the quarterly real GDP growth (2012 Q1–2021 Q1), the monthly economic policy uncertainty index (2012 M1–2021 M7) (EPU), and the monthly consumer price rate (2012 M1–2021 M6) (CPI) in India.

The data visualization shows that when (EPU) (CPI) rises, GDP falls, and thus our current 2021 Q2 forecast suggests an improvement. The following table reports the most important descriptive statistics and normal distribution of the variables:

The Table 1 shows that both GDP growth and economic policy uncertainty are not distributed according to the normal distribution. Considering GDP, the large difference between the maximum and the minimum values that occurred during 2020 caused the distribution to become significantly kurtosis and skewed to the left with strongly negative growth. We also note the significant difference between the maximum (283.68) and the minimum (32.88) values of the EPU index, which occurred in 2012M6 and 2016M6, respectively. This also led to a high kurtosis greater than 3 (normal). We notice that the outlier value in 2012M6 skewed the distribution to the right. As for the CPI rate, the data present a normal distribution, and the average during the studied period was 6.73% with a standard deviation of 2.37, which indicates an increase in price in general for the study period [19–31]. To find the extent of the impact of these values on the variable, we used the Breakpoint unit root test and obtained the following results.

**Table 1.** Descriptive statistics and normal distribution of the variables.

Variables	Normality	Mean	Standard Deviation	Maximum	Minimum	Skewness	Kurtosis
	J-B						
GDP%	343.6 ***	1.64	5.98	23.12	−25.29	−1.39	17.66
EPU	61.6 ***	94.09	46.82	283.68	32.88	1.37	5.31
CPI%	2.89	6.37	2.57	12.06	1.08	0.32	2.54

\*\*\* Indicates significance at 1%. \*\* significance at 5%. \* significance at 1%.

Table 2 shows that the date of 2020 Q2 was set as a point of structural break. We determined the significance of all parameters of the model. The break occurred at the intercept and trend levels. We also found that t-statistics stationary is significant at 1%, and therefore, we concluded that the point structural break caused a short-term change in GDP and was stationary at that level. To complete our steps before building the model, we estimated the correlation matrix for high-frequency variables to make sure that there was no multicollinearity.

**Table 2.** Breakpoint Unit Root Test for GDP.

GDP	Stationary	GDP (−1)	C	TREND	INCPT BREAK	TREND BREAK	BREAKDUM	Integrated
t-statistics	−54.25 ***	−4.9 ***	10.76 ***	−1.97 *	23.23 ***	−16.95 ***	−46.80 ***	I(0)
Break Date	2020 Q2							

\*\*\* Indicates significance at 1%. \*\* significance at 5%. \* significance at 1%.

The results of the correlation matrix estimations showed that the degree of linear correlation (0.544) was low among the variables, and thus we ruled out a problem of multicollinearity. We then built the model and obtained the following results, presented in Table 3, divided into three main sections. The first section represents the AR (1) equation for the low-frequency variable. We note the significance of the parameters and the negative impact of GDP on their previous values. The second and third sections show the polynomial coefficient for high-frequency variables. We note that two lag periods were subtracted from high-frequency variables, since we needed two months out of each quarter to predict the low-frequency variable. Table 3 shows that 5 lag periods were selected for EPU, and 15 lag periods were chosen for CPI according to the AIC that achieved the least values for model sum of squared residuals (Figure 2). We note the significant effect of the partial coefficient of the monthly variables in each quarter of GDP. Table 4 shows that the effects of EPU varied but had a negative impact on the growth of GDP in each quarter, while the impact was positive for the growth of CPI (see Appendix A). The Durbin Watson stat indicates no first-order autocorrelation between the residuals [19,29]. Here, we needed to test autocorrelation between residuals of more than one degree, due to the use of lags.

**Table 3.** Residuals test results.

Test	Normality	ACF2	ADF
	J-B		Non Constant–Non Trend
Residuals (PROB)	0.824	0.645	000 ***

\*\*\* Indicates significance at 1%. See Appendix A for statistical details.

## Sum of Squared Residuals (top 20 models)

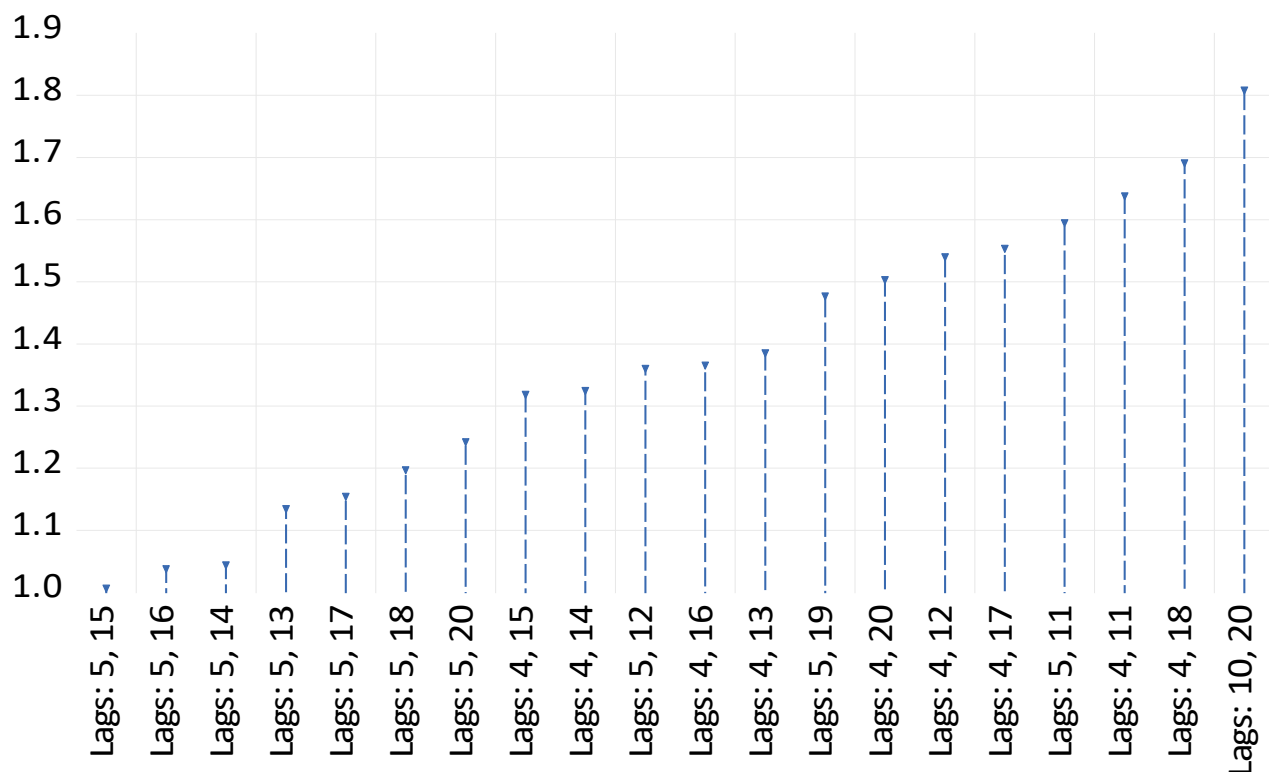


Figure 2. Choosing the model that achieves the lowest values sum of squared residuals.

Table 4. Estimation using Mixed-Data Sampling.

Dependent Variable: GDP (2013Q4 2018Q4)			
Variable	Coefficient	Std. Error	t-Statistics
C	2.80203	0.41732	6.714348 ***
GDP (−1)	−0.583927	0.186453	−3.131770 ***
EPU(−2) Lags: 5			
PDL1	−0.092474	0.021263	−4.348970 ***
PDL2	0.110246	0.026672	4.133369 ***
PDL3	−0.039758	0.009813	−4.051492 ***
PDL4	0.004328	0.001083	3.996321 ***
CPI(−2) Lags: 15			
PDL1	0.47354	0.076658	6.177316 ***
PDL2	−0.232206	0.040478	−5.736635 ***
PDL3	0.028698	0.005617	5.109536 ***
PDL4	−0.001012	0.000226	−4.475439 ***
<b>Adg. R-Squared</b>	0.580478	<b>Log Likelihood</b>	2.022063

\*\*\* Indicates significance at 1%. \*\* Significance at 5%. \* Significance at 1%.

Before issuing nowcasting, we checked that the standard assumptions for the residuals were correct:

The results showed that the residuals were distributed according to a normal distribution, and we noted that there was no autocorrelation between different orders. We found that



the residual series was stationary. Now, according to the previous results, we performed multi-step ahead nowcasting to forecast 2021 Q2 and obtained the following results.

The visual representation showed (Figure 3) that the expected values using the model were close to the actual data points. The most prominent observation is that the model succeeded in capturing the point of structural change in 2020 Q2, but not the actual value that it reached, as indicated in Figure 4.

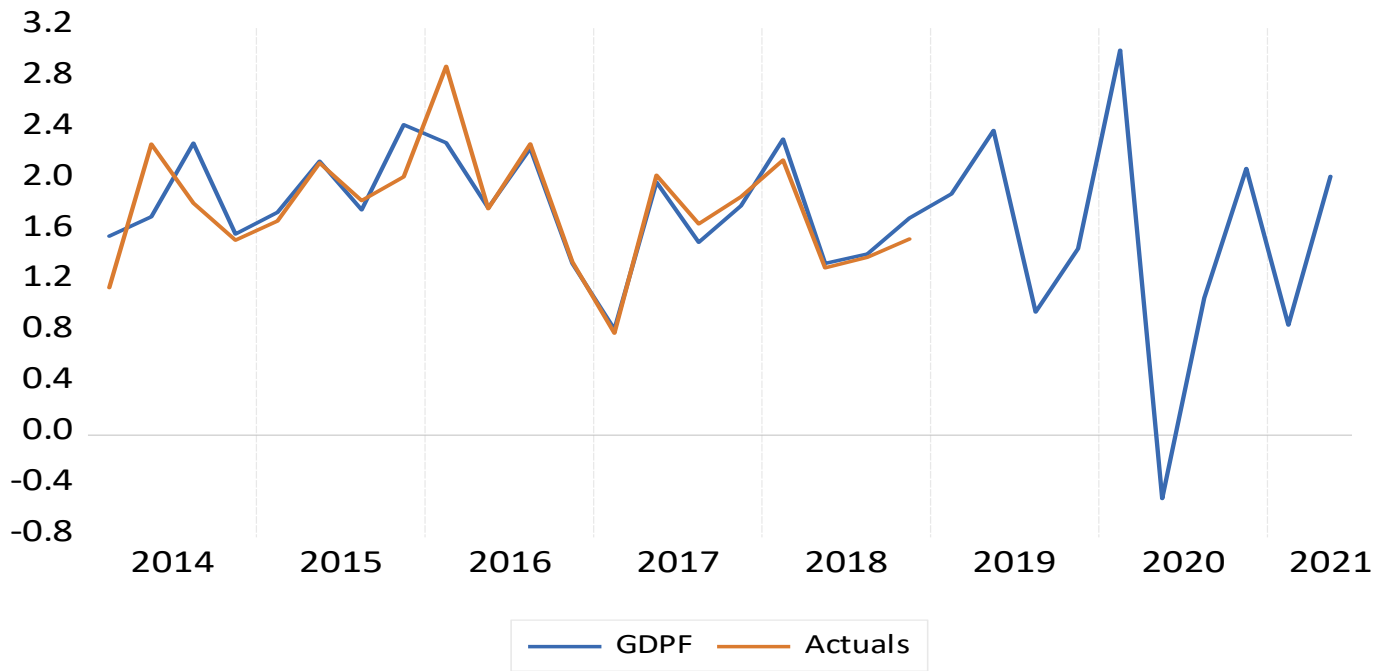


Figure 3. Multi-step ahead nowcasting GDP in and out of sample.

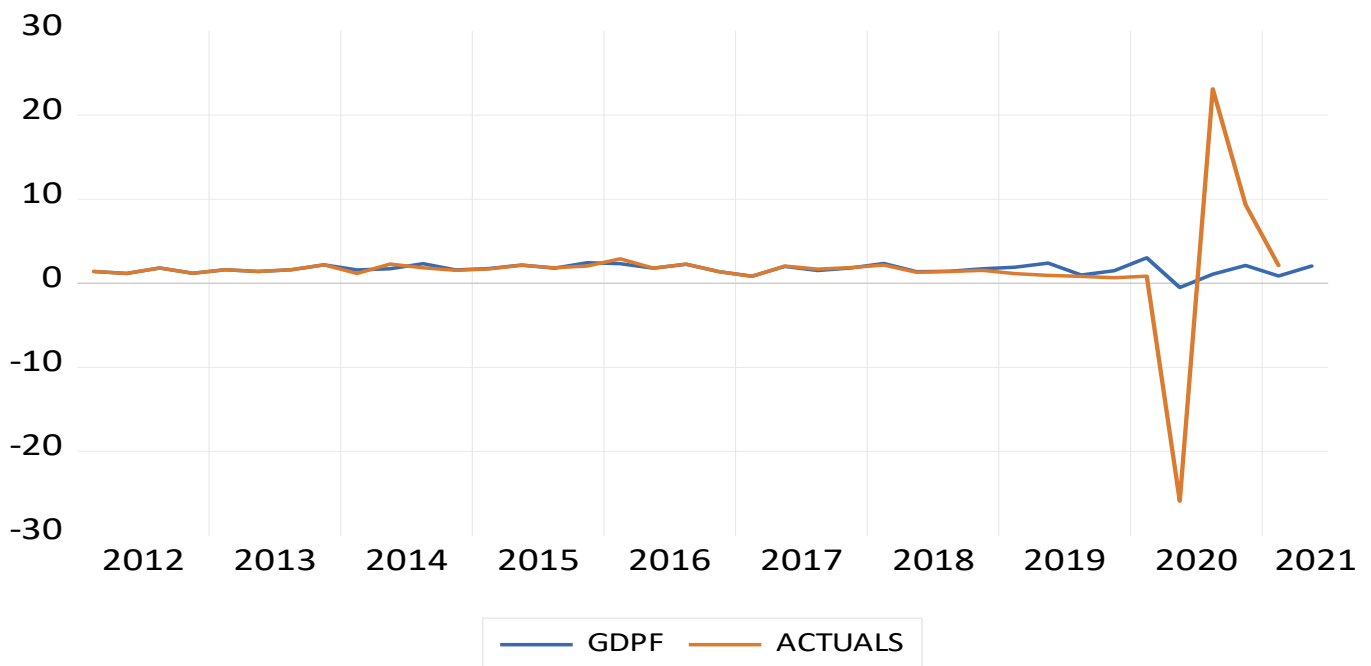


Figure 4. Multi-step ahead nowcasting and actual of GDP out of sample.

The presence of such large changes makes any statistical model invalid, but our model accurately predicted when the structural Break would occur. We calculated the

RMSE (Table 5) to evaluate the predictions in and out of sample on the horizon (short, medium, long):

**Table 5.** Evaluating the performance of the model in and out of sample.

Horizon (Quarter)	In of Sample (Training Data)	Out of Sample (Testing Data)			
		H = 1	H = 4	H = 10	H = Last Point
RMSE	0.258	0.778	0.937	11.51	1.24

We see from the table that RMSE values [32] were less than 1 for in-sample forecasting, the model achieved satisfactory results for out-of-sample forecasting at the horizons of 1 and 4 quarters, and the RMSE value became large for forecasting during 2020, as the value included the structural break that GDP growth was exposed to during 2020, but the results of Table 2 show that the structural break was short-term, as the model achieved a lower value of RMSE at the last forecast point. According to our expectations that GDP growth will achieve 2.1% in 2021 Q2, affected by an increase in CPI and a decrease in EPU, we were able to include in the forecast the month of July, in which the level of the index increased, which might herald a new decline in 2021Q3. It is possible to include any new data for indicators and update the forecasting based on that. This model can also be used to understand the effect of targeting a variable on GDP growth. For example, suppose that the economic policy in India aims to stabilize the CPI (i.e., zero growth) for the next three months. This information can be included in the model and obtain GDP growth forecasts for the third quarter; this indicated a drop in growth of  $-0.64\%$ .

## 6. Conclusions

This paper presents a framework for nowcasting India's GDP using the MIDAS–Almon (PDL) weighting model, relying on the information in high-frequency data, instead of including a large number of variables that lead to statistical complexity. Two indicators were relied on, namely, Economic Policy Uncertainty (EPU) for India as a proxy for non-standard indicators and the Consumer Price Index (CPI) as a proxy for standard indicators. The model showed a negative effect of EPU and a positive effect of CPI on GDP growth, as high periods of EPU correspond to a decrease in GDP growth, and high periods of CPI correspond to a rise in GDP growth in accordance with the economic theory. In addition to the information in the high-frequency variables, we made use of the historical information in the low-frequency variable and performed the estimation on the training data. The prediction results showed low values of the RMSE in the sample and when predicting the 1- and 4-quarter horizon for out of sample, but the RMSE increased when predicting the 10-quarter horizon. Due to the effect of a short-term structural break, we found that the RMSE values decreased for the last prediction point. The nowcasting results indicated that GDP growth will achieve 2.1% in 2021 Q2, affected by an increase in CPI and a decrease in EPU. Compared with the results of the literature reviews for nowcasting India GDP growth, the models previously used did not allow capturing the structure break in GDP growth during the COVID-19 pandemic, despite the large number of high-frequency variables used, which suggests that future studies should include EPU when nowcasting changes in India GDP growth. This exercise is expected to be useful to India's economic planners and policy-makers in institutions and sites that are concerned with economic forecasts for the present and the near future of economic growth, which can be obtained with specific constraints.

**Author Contributions:** Conceptualization, P.M., K.A., M.A. and P.K.S.; methodology, S.S., M.R., S.S.D. and U.H.R.; software, A.J.O., N.A.I., G.F.A., F.H., P.T. and R.B.; validation, all the authors; formal analysis, P.M., K.A., M.A. and P.K.S.; investigation, S.S., M.R., S.S.D. and U.H.R.; resources, all the authors; data curation, P.M. and K.A.; writing—original draft preparation, P.M. and K.A.; writing—review and editing, all the authors; visualization, P.M.; supervision, P.M., K.A., M.A. and P.K.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** No funding received for this research.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are openly available in [https://www.policyuncertainty.com/india\\_monthly.html](https://www.policyuncertainty.com/india_monthly.html) and <https://fred.stlouisfed.org/series/NAEXKP01INQ657S> (accessed on 13 October 2021).

**Acknowledgments:** The work was supported by Act 211 Government of the Russian Federation, contract No. 02.A03.21.0011. The work was supported by the Ministry of Science and Higher Education of the Russian Federation (government order FENU-2020-0022).

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

Null Hypothesis: GDP has a unit root  
Trend Specification: Trend and intercept  
Break Specification: Trend and intercept  
Break Type: Innovational outlier

Break Date: 2020Q2  
Break Selection: Minimize Dickey-Fuller t-statistic  
Lag Length: 0 (Automatic - based on Schwarz information criterion, maxlag=9)

	t-Statistic	Prob.*
<b>Augmented Dickey-Fuller test statistic</b>	<b>-54.25358</b>	<b>&lt; 0.01</b>
Test critical values:		
1% level	-5.719131	
5% level	-5.175710	
10% level	-4.893950	

\*Vogelsang (1993) asymptotic one-sided p-values.

Augmented Dickey-Fuller Test Equation  
Dependent Variable: GDP  
Method: Least Squares  
Date: 08/21/21 Time: 12:43  
Sample (adjusted): 2012Q2 2021Q1  
Included observations: 36 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
GDP(-1)	-0.099953	0.020274	-4.930055	0.0000
C	2.077006	0.192898	10.76737	0.0000
TREND	-0.019784	0.010018	-1.974873	0.0575
INCPTBREAK	36.53110	1.571926	23.23972	0.0000
TRENDBREAK	-8.724858	0.514548	-16.95637	0.0000
BREAKDUM	-55.06525	1.176393	-46.80857	0.0000
R-squared	0.993633	Mean dependent var		1.653602
Adjusted R-squared	0.992571	S.D. dependent var		6.068686
S.E. of regression	0.523060	Akaike info criterion		1.692772
Sum squared resid	8.207762	Schwarz criterion		1.956692
Log likelihood	-24.46989	Hannan-Quinn criter.		1.784887
F-statistic	936.2884	Durbin-Watson stat		1.217931
Prob(F-statistic)	0.000000			

**Figure A1.** Breakpoint unit root test results.

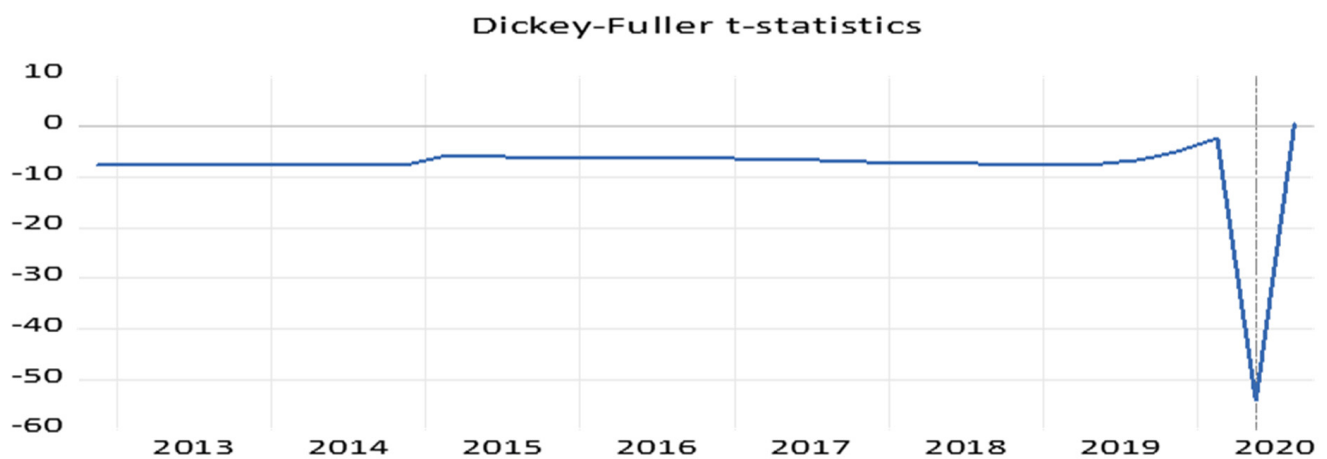


Figure A2. Structural Break Point according to t-statistics by Dickey-Fuller for GDP.

Dependent Variable: GDP  
 Method: MIDAS  
 Date: 08/25/21 Time: 15:14  
 Sample (adjusted): 2013Q4 2018Q4  
 Included observations: 21 after adjustments  
 Method: PDL/Almon (polynomial degree: 4)  
 Automatic lag selection, max lags: 20  
 Chosen selection: 5 15

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.802030	0.417320	6.714348	0.0000
GDP(-1)	-0.583927	0.186453	-3.131770	0.0095

Page: I Series: EPU(-2) Lags: 5

PDL01	-0.092474	0.021263	-4.348970	0.0012
PDL02	0.110246	0.026672	4.133369	0.0017
PDL03	-0.039758	0.009813	-4.051492	0.0019
PDL04	0.004328	0.001083	3.996321	0.0021

Page: I Series: CPI(-2) Lags: 15

PDL01	0.473540	0.076658	6.177316	0.0001
PDL02	-0.232206	0.040478	-5.736635	0.0001
PDL03	0.028698	0.005617	5.109536	0.0003
PDL04	-0.001012	0.000226	-4.475439	0.0009

R-squared	0.769263	Mean dependent var	1.802676
Adjusted R-squared	0.580478	S.D. dependent var	0.468793
S.E. of regression	0.303640	Akaike info criterion	0.759803
Sum squared resid	1.014167	Schwarz criterion	1.257195
Log likelihood	2.022063	Hannan-Quinn criter.	0.867750
Durbin-Watson stat	2.275827		

IEPU(-2)	Lag	Coefficient	Distribution
	0	-0.017658	
	1	0.003608	
	2	-0.002711	
	3	-0.010650	
	4	0.005756	

ICPI(-2)	Lag	Coefficient	Distribution
	0	0.269020	
	1	0.115827	
	2	0.007891	
	3	-0.060856	
	4	-0.096483	
	5	-0.105061	
	6	-0.092656	
	7	-0.065339	
	8	-0.029179	
	9	0.009755	
	10	0.045395	
	11	0.071672	
	12	0.082516	
	13	0.071858	
	14	0.033629	

Figure A3. MIDAS model results.

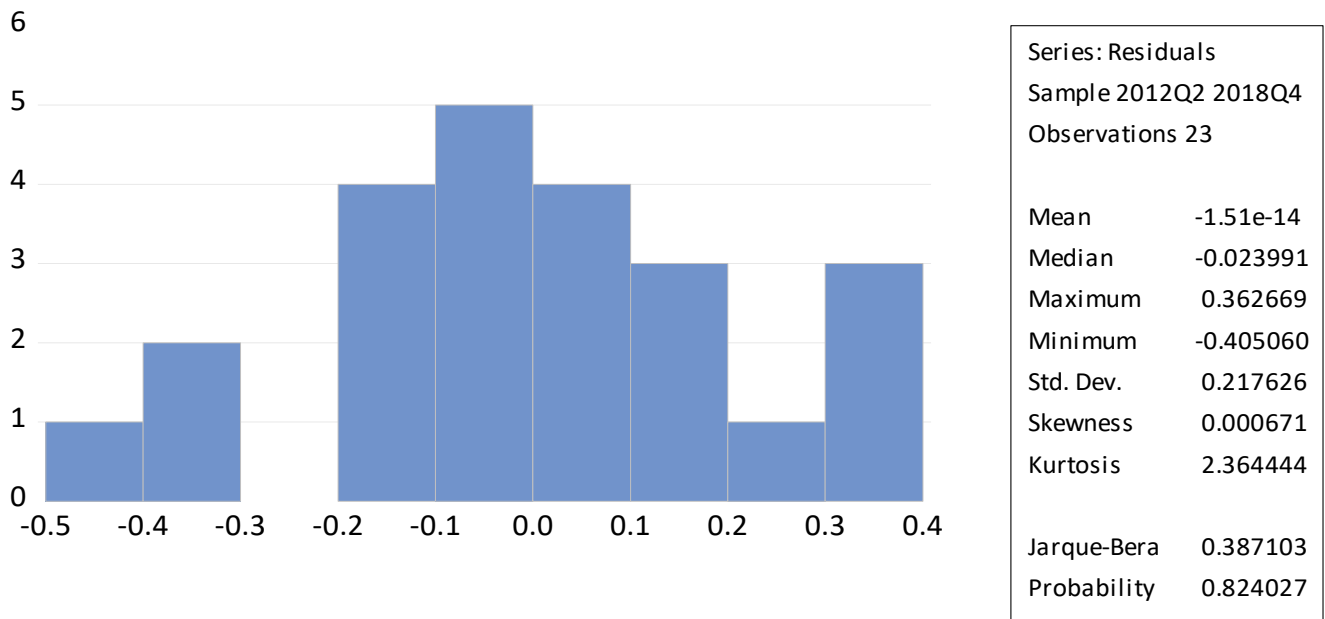


Figure A4. MIDAS Model Residuals Distribution graph.

Date: 08/25/21 Time: 15:16

Sample (adjusted): 2013Q2 2018Q4

Q-statistic probabilities adjusted for 1 dynamic regressor

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob*	
		1	-0.164	-0.164	0.7022	0.402
		2	-0.080	-0.110	0.8778	0.645
		3	0.142	0.114	1.4589	0.692
		4	-0.096	-0.064	1.7389	0.784
		5	-0.061	-0.069	1.8566	0.869
		6	-0.219	-0.285	3.4781	0.747
		7	0.173	0.111	4.5529	0.714
		8	-0.272	-0.302	7.3933	0.495
		9	-0.099	-0.124	7.7928	0.555
		10	0.107	-0.109	8.3023	0.599
		11	-0.118	-0.109	8.9697	0.625
		12	0.012	-0.152	8.9767	0.705

\*Probabilities may not be valid for this equation specification.

Figure A5. Auto and Partial correlation function for MIDAS Model residuals.

Null Hypothesis: RESID01 has a unit root  
 Exogenous: None  
 Lag Length: 0 (Automatic - based on SIC, maxlag=4)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-5.371431	0.0000
Test critical values:		
1% level	-2.674290	
5% level	-1.957204	
10% level	-1.608175	

\*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation  
 Dependent Variable: D(RESID01)  
 Method: Least Squares  
 Date: 08/25/21 Time: 15:20  
 Sample (adjusted): 2013Q3 2018Q4  
 Included observations: 22 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RESID01(-1)	-1.169064	0.217645	-5.371431	0.0000
R-squared	0.578693	Mean dependent var		-0.003993
Adjusted R-squared	0.578693	S.D. dependent var		0.337021
S.E. of regression	0.218754	Akaike info criterion		-0.157349
Sum squared resid	1.004919	Schwarz criterion		-0.107756
Log likelihood	2.730842	Hannan-Quinn criter.		-0.145667
Durbin-Watson stat	1.965918			

Figure A6. Augmented Dickey–Fuller test results for MIDAS Model residuals.

Forecast: GDPF  
 Actual: GDP  
 Forecast sample: 2014Q1 2021Q2  
 Included observations: 30

Root Mean Squared Error	0.258888
Mean Absolute Error	0.170259
Mean Absolute Percentage Error	9.142257
Theil Inequality Coef.	0.070305
Bias Proportion	0.002899
Variance Proportion	0.048888
Covariance Proportion	0.948213
Theil U2 Coefficient	0.318201
Symmetric MAPE	8.915890

Figure A7. Accuracy forecasting using MIDAS in sample.

## References

1. Ghysels, E.; Horan, C.; Moench, E. Forecasting through the rearview mirror: Data revisions and bond return predictability. *Rev. Financ. Stud.* **2018**, *31*, 678–714. [[CrossRef](#)]
2. Evans, M. Where Are We Now? Real-Time Estimates of the Macroeconomic. *Int. J. Cent. Bank* **2005**, *1*, 127–175.
3. Giannone, D.; Reichlin, V.; Small, D. Nowcasting: The real-time informational content of macroeconomic data. *J. Monet. Econ.* **2008**, *55*, 665–676. [[CrossRef](#)]
4. Banbura, M.; Giannone, D.; Modugno, M.; Reichlin, L. Now-casting and the real-time data flow. In *Handbook of Economic Forecasting*; Elsevier: Amsterdam, The Netherlands, 2013; Volume 2, pp. 195–237.
5. Olli-Matti, L.; Annika, L. *Nowcasting Finnish GDP Growth Using Financial Variables: A MIDAS Approach*; BoF Economics Review; Bank of Finland: Helsinki, Finland, 2020.
6. Gunay, M. *Nowcasting Turkish GDP with MIDAS: Role of Functional Form of the Lag Polynomial*; Working Paper; Central Bank of the Republic of Turkey: Ankara, Turkey, 2020.
7. Barsoun, F.; Stankiewicz, S. *Forecasting GDP Growth Using Mixed-Frequency Models with Switching Regimes*; Working Paper Series; University of Konstanz: Konstanz, Germany, 2013.
8. Jarret, C.; Meunier, B. *Nowcasting World GDP Growth with High-Frequency Data*; Working paper; Banque de France: Paris, France, 2020.
9. Rufino, C.C. Nowcasting Philippine Economic Growth Using MIDAS Regression Modeling. *DLSU Bus. Econ. Rev.* **2019**, *29*, 14–23.
10. Bok, B.; Caratelli, D.; Giannone, D.; Sbordon, A.M.; Tambalotti, A. Macroeconomic Nowcasting and Forecasting with Big Data. *Annu. Rev. Econ.* **2018**, *10*, 615–643. [[CrossRef](#)]
11. Thorsrud, L.A. Words are the New Numbers: A Newsy Coincident Index of the Business Cycle. *J. Bus. Econ. Stat.* **2018**, *38*, 393–409. [[CrossRef](#)]
12. Babii, A.; Ghysels, E.; Striaukus, J. Machine Learning Time Series Regressions with an Application to Nowcasting. *J. Bus. Econ. Stat.* **2021**, 1–23. [[CrossRef](#)]
13. Lyer, T.; Gupta, S. *Nowcasting Economic Growth in India: The Role of Rainfall*; ADB Economics Working Paper Series; Asian Development Bank: Mandaluyong, Philippines, 2019.
14. Bhadury, S.; Ghosh, S.; Kumar, P. *Nowcasting Indian GDP Growth Using a Dynamic Factor Model*; Reserve Bank of India: Mumbai, India, 2020.
15. Bragoli, D.; Fosten, J. Nowcasting Indian GDP. *Oxf. Bull. Econ. Stat.* **2018**, *80*, 259–282. [[CrossRef](#)]
16. Baker, S.R.; Bloom, N.; Davis, S.J. Measuring Economic Policy Uncertainty. *Q. J. Econ.* **2016**, *131*, 1593–1636. [[CrossRef](#)]
17. Ghysels, E.; Pedro, S.; Rossen, V. *The MIDAS Touch: Mixed Data Sampling Regression Models*; University of North Carolina and UCLA Discussion Paper; 2004. Available online: <https://rady.ucsd.edu/faculty/directory/valkanov/pub/docs/midas-touch.pdf> (accessed on 13 October 2021).
18. Ghysels, E.; Pedro, S.; Rossen, V. Predicting volatility: Getting the most out of return data sampled at different frequencies. *J. Econom.* **2006**, *131*, 59–95. [[CrossRef](#)]
19. Andreou, E.; Ghysels, E.; Andros, K. Should macroeconomic forecasters use daily financial data and how? *J. Bus. Econ. Stat.* **2013**, *3*, 240–251. [[CrossRef](#)]
20. Economic Policy Uncertainty. Available online: [www.policyuncertainty.com](http://www.policyuncertainty.com) (accessed on 10 June 2021).
21. Bloom, N. The Impact of Uncertainty Shocks. *Econometrica* **2009**, *77*, 623–685. [[CrossRef](#)]
22. Basu, S.; Bundick, B. *Uncertainty Shocks in a Model of Effective Demand*; American Economic Association: Nashville, TN, USA, 2012.
23. Stock, J.H.; Watson, M.W. Disentangling the Channels of the 2007–09 Recession. *Brook. Pap. Econ. Act.* **2012**, *2012*, 81–135. [[CrossRef](#)]
24. Jurado, K.; Ludvigson, S.; Ng, S. Measuring Uncertainty. *Am. Econ. Rev.* **2015**, *105*, 1177–1216. [[CrossRef](#)]
25. Ahir, H.; Bloom, N.; Furceri, D. *The World Uncertainty Index*; International Monetary Fund: Washington, DC, USA, 2018; pp. 1–33.
26. Perron, P. The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis. *Econometrica* **1989**, *57*, 1361. [[CrossRef](#)]
27. Perron, P.; Timothy, J.; Vogelsang, T. Testing for a Unit Root in a Time Series with a Changing Mean: Corrections and Extensions. *J. Bus. Econ. Stat.* **1992**, *10*, 467–470.
28. Vogelsang, T.J.; Perron, P. Additional Tests for a Unit Root Allowing for a Break in the Trend Function at an Unknown Time. *Int. Econ. Rev.* **1998**, *39*, 1073. [[CrossRef](#)]
29. Perron, P. Dealing with Structural Breaks, in *Palgrave Handbook of Econometrics. Econom. Theory* **2006**, *1*, 278–352.
30. Dickey, D.A.; Fuller, W.A. Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica* **1981**, *49*, 1057–1072. [[CrossRef](#)]
31. Mishra, P.; Al Khatib, A.M.G.; Sardar, I.; Mohammed, J.; Karakaya, K.; Dash, A.; Ray, M.; Narsimhaiah, L.; Dubey, A. Modeling and forecasting of sugarcane production in India. *Sugar Tech* **2021**. [[CrossRef](#)]
32. Bhadury, S.; Beyer, R.; Pohit, S. *A New Approach to Nowcasting Indian Gross Value Added*; Working Paper NO. 115; National Council of Applied Economic Research: New Delhi, India, 2018.
33. Mishra, P.; Yonar, A.; Yonar, H.; Kumari, B.; Abotaleb, M.; Das, S.S.; Patil, S.G. State of the art in total pulse production in major states of India using ARIMA techniques. *Curr. Res. Food Sci.* **2021**. [[CrossRef](#)]