INTERNATIONAL MONETARY FUND

Identifying Optimal Indicators and Lag Terms for Nowcasting Models

Jing Xie

WP/23/45

2023
MAR

WORKING PAPER

 WP/23/45

IMF Working Paper
Institute for Capacity Development

Identifying Optimal Indicators and Lag Terms for Nowcasting Models

Prepared by Jing Xie

Authorized for distribution by Paul Cashin
March 2023

ABSTRACT: Many central banks and government agencies use nowcasting techniques to obtain policy relevant information about the business cycle. Existing nowcasting methods, however, have two critical shortcomings for this purpose. First, in contrast to machine-learning models, they do not provide much if any guidance on selecting the best explantory variables (both high- and low-frequency indicators) from the (typically) larger set of variables available to the nowcaster. Second, in addition to the selection of explanatory variables, the order of the autoregression and moving average terms to use in the baseline nowcasting regression is often set arbitrarily. This paper proposes a simple procedure that simultaneously selects the optimal indicators and ARIMA(p,q) terms for the baseline nowcasting regression. The proposed AS-ARIMAX (Adjusted Stepwise Autoregressive Moving Average methods with exogenous variables) approach significantly reduces out-of-sample root mean square error for nowcasts of real GDP of six countries, including India, Argentina, Australia, South Africa, the United Kingdom, and the United States.

JEL Classification Numbers:	C32, C53, E37, E52, 011
Keywords:	Nowcasting; Mixed Frequency; Forecasting; Business Cycles
Author’s E-Mail Address:	jxie2@imf.org

*** Acknowledgements:**
I would like to thank Mr. Sam Ouliaris for providing instrumental guidance throughout this research, and am grateful to Mr. Paul Cashin, Mr. Fei Han, Ms. Ivy Sabuga, Mr. Alexander Borodin, and Institute for Capacity Development colleagues for their helpful comments. I am also grateful to Ms. Elisa Manarinjara for her support along the publication procedure.

WORKING PAPERS

Identifying Optimal Indicators and Lag Terms for Nowcasting Models

Prepared by Jing Xie

Contents

Section 1. Introduction

When major economic shocks such as the COVID-19 pandemic or the global financial crisis occur, governments typically use counter-cyclical policies to soften the severity of the negative shock to real gross domestic product (GDP). Such evidence-driven counter-cycle policy requires timely information on the state of the economy relative to trend. Unfortunately, the required data is often unavailable because of: (a) the so-called "ragged-edge" problem arising from publication lags or, more generally, missing data points, especially in the case of real GDP (Wallis, 1986) and (b) the mixed/incompatible frequencies with which key economic indicators are available (Armesto, Engemann, & Owyan, 2010).

Many central banks and government agencies use nowcasting techniques (e.g., Bridge, Mixed-Data Sampling and Dynamic Factor Model) to address these issues. Examples include the European Central Bank (Bańbura, et al., 2013), the Central Bank of Malta (Ellul and Ruisi, 2022), and the Federal Reserve Bank of Atlanta (Higgins, 2014). Nowcasting—the art of forecasting "the here and now"—enables real-time forecasting of lower frequency variables (such as real GDP and inflation) using more timely indicators that have similar or higher frequencies. Standard nowcasting models typically involve two steps: (a) forecasting the high frequency indicators in the preferred baseline nowcasting regression to eliminate the ragged edge problem; and (b) converting the high frequency indicators to the target frequency of the baseline regression. The way this conversion proceeds identifies the specific nowcasting procedure used (see Section 5).

A critical shortcoming of existing nowcasting methods, however, is that they do not provide adequate guidance on the selection of the right-hand side variables (typically exogenous) to include in the baseline regression. Moreover, the appropriate order of the autoregression (AR) and moving average terms (MA) to use in the baseline regression is rarely discussed and is often set arbitrarily. Indeed, to the best of our knowledge, not many nowcasting exercises use an ARIMA model with exogenous variables (i.e., ARIMAX model). Interestingly, medical researchers have for some time been successfully using ARIMAX models to nowcast influenza outbreaks with *Google Flu Trends* as exogenous variable, reporting significant reductions in mean absolute error (MAE) compared to using a more standard baseline model with previous flu levels only as explanatory variables (Preis & Moat, 2014).

This paper investigates the effectiveness of ARIMAX models for nowcasting key economic variables such as real GDP. We propose a simple procedure for selecting ─ from a larger set of economic variables ─ indicators that are economically meaningful (in the sense that their estimated coefficient is consistent with economic priors), statistically significant, and effective in terms of improving the accuracy of the nowcast.

For the example of India's real GDP, we show that applying a simple variable selection procedure that allows for ARIMA(p,q) terms in addition to optimally selected explanatory variables significantly enhances the nowcasting performance of the Bridge and U-MIDAS estimators relative to benchmark models formulated without using the proposed variable selection procedure.

The remainder of the paper is organized as follows: In Section 2 we review the automatic ARIMA estimation procedure available in EViews. Section 3 proposes an "Adjusted Stepwise ARIMAX Variable Selection Procedure (henceforth AS-ARIMAX)" to identify "optimal" ARIMA orders and exogenous variables for nowcasting[1]. The approach is implemented using EViews' automatic ARIMA selection procedure and customized codes. Sections 4 and 5 define the three benchmarks and two nowcasting models that we use for the empirical study. Sections 6 and 7 apply the AS-ARIMAX method to India's real GDP, yielding significant forecasting gains relative to the benchmark nowcasting models. Section 8 applies the AS-ARIMAX method to nowcast five additional countries' real GDP, to further prove the efficiency and applicability of the approach. Section 9 concludes.

Section 2. Automatic ARIMA Selection Procedure

Although EViews provides comprehensive tools for users to determine the orders of the ARIMA model using traditional (non-automated) Box-Jenkins methods, the procedure can be time-consuming and comes with significant risk of misidentification because of the difficulty of matching the data's correlogram with a specific ARIMA model. To improve efficiency and model identification, EViews also offers an automatic ARIMA model selection procedure to help users automatically determine the appropriate ARIMA specification. This procedure involves the following steps (EViews User's Guide I, pp538 - 540):

Step 1. Selecting appropriate transformations of the dependent variable

EViews runs the following two regressions to determine the appropriate transformation method:

$$D(y_t)^2 = \alpha_1 + \beta_1 y_t \qquad (1)$$

$$Dlog(y_t)^2 = \alpha_2 + \beta_2 log(y_t) \qquad (2)$$

Each of these regressions is a simple test for heteroskedasticity, with lower absolute *t-statistic* on β suggesting more homoskedasticity than heteroskedasticity. EViews uses a log transformation if the absolute *t*-statistic on β_2 is smaller than that on β_1. The natural log transformation is suitable for series with exponential growth rates that typically suffer from heteroskedasticity (since the change is non-constant). Given that the log transform linearizes the relationship, a β_2 that is lower than β_1 suggests regression (2) exhibits relatively more homoskedasticity. Thus, the log transformation is more appropriate.

Step 2. Selecting the level of differencing of the dependent variable

After deciding on the appropriate transformation method, one must decide the appropriate level of differencing to use on the dependent variable. EViews uses successive KPSS unit roots tests, with null hypothesis of stationarity, to determine the correct level of differencing. Based on the work by Hyndman and Khandakar (2008), EViews runs the successive unit roots tests as follows: the KPSS test is first run on the non-transformed data. If the test rejects the stationarity, the KPSS test is then rerun with differenced data. Such procedure continues until EViews can no longer reject the null hypothesis of stationarity.

Step 3. Selecting the exogenous regressors

[1] The author has developed the EViews code to implement the AS-ARIMAX procedure and would be happy to make it available to interested parties. Please contact Jing Xie (jxie2@imf.org) for such requests.

EViews allows users to specify exogenous regressors to include in the ARIMA selection process. By default, a constant term is included. We will define our proposed way of inputting the exogenous regressors in Section 3.

Step 4. Selecting the order of the ARIMA terms

Conditional on the user specified exogenous variables, EViews uses standard model selection criteria to determine the ARIMAX model that best fits a set of data. EViews offers standard information criteria (Akaike Information (AIC), Schwarz (SIC or BIC), and Hannan – Quinn (HQ)), along with the Mean Square Error (MSE), as model selection criteria. See below for the basic formula for these two types of model selection criterion.

Information Criteria*:* each of these three criteria are based upon the estimated log-likelihood of the fitted model, the number of parameters, and observations in the model. The model with the smallest information criterion is preferred.

$$Akaike\ Info\ Criterion(AIC): -2\left(\frac{l}{T}\right) + 2k\left(\frac{1}{T}\right)$$
$$Schwarz\ Criterion\ (SC): -2\left(\frac{l}{T}\right) + k\frac{log(T)}{T}$$
$$Hannan\ -\ Quinn\ Criterion\ (HQ): -2\left(\frac{l}{T}\right) + 2k\ \frac{log(T)}{T}$$

where l is the value of the log of the likelihood function, k is the number of parameters estimated using T observations.

Mean Square Error (MSE) Evaluation**:** this is also called in-sample forecast evaluation, in which each model is estimated using a sub-sample (i.e., first 80~90% of the data) and forecasted over the remaining data (i.e., 10~20%). Then the MSE is calculated according to

$$Mean\ Square\ Error\ (MSE) = \frac{1}{h}\sum_{t\,=\,T-h}^{T}(y_t\ -\ \hat{y}_t)^2$$

where h is the number of periods in the forecast sub-sample, y_t is the actual data, $\hat{y}_t$ is the forecast at time t, and T is the number observations in the sample. The model with the smallest MSE is selected.

The EViews automatic ARIMA selection procedure is conditional on the exogenous variables being pre-specified by the user. That is, the procedure determines only the autoregressive and moving average orders without any allowance for automatic exogenous variable selection. In the next section, we introduce an Adjusted Stepwise-ARIMAX (AS-ARIMAX) procedure that offers customized stepwise selection procedures for an arbitrary set of exogenous variables.

Section 3. Adjusted Stepwise ARIMAX Variable Selection Procedure and a Simple Example

Stepwise model selection procedures, which add or remove variables from a regression based on the statistical significance of the candidate variable, have been widely used to find the preferred baseline forecasting/nowcasting model. The process starts with either backward elimination from the most general model or forward inclusion from the smallest possible model. With forward selection, candidate variables are added to the model sequentially based on the significance level. The procedure checks whether all the variables are statistically significant and removes those that are not. With backward selection, all candidate variables are added to the model initially and then individual variables are deleted if they are insignificant. Note that the procedure will re-introduce a "dropped" variable if it subsequently determined to be statistically significant (Chowdhury & Turin, 2020).

Despite the popularity of stepwise model selection procedures in recent decades, criticisms have continued to arise. Smith (2018) argues that the fundamental problem with stepwise regression is that it may bypass explanatory variables that have causal effects on the dependent variables yet include nuisance (spurious) variables that are coincidentally statistically significant. Such an outcome typically results in good in-sample forecasting fit but poor out-of-sample forecasting.

To tackle such issues, we proposed a modified stepwise procedure that shifts the focus from statistical significance to the overall forecasting improvement that can be attributed to a specific exogenous variable (indicator). Beginning with no exogenous variables in the model except for the constant term, we test each variable separately and add it to the baseline model if it has an estimated coefficient that is consistent with economic priors and yields superior model forecasting performance.

Specifically, we decide whether a variable (X_t) is a suitable candidate based on **three criteria**:

- **Condition 1**: The X_t decreases the Akaike Information Criteria (AIC) value, compared to the model without X_t.
- **Condition 2**: The coefficient sign of X_t matches economic priors.
- **Condition 3**: X_t is statistically significant at the 5% confidence level.

The adjusted stepwise ARIMAX (AS-ARIMAX) variable selection procedure involves four **steps** (see *Appendix 1* for detailed procedure charts):

Step One: we add the first candidate indicator X_1 as an exogenous regressor to the automatic ARIMA procedure for the target variable (*Model 1-A)*. Then, we repeat the procedure without X_1(*Model 1-B*). If *Model 1-A* satisfies ***conditions 1-2***, we keep X_1, otherwise it is discarded.

Step Two: if X_1 is retained in step one, we add the second indicator, X_2, to the baseline model as an exogenous regressor, repeating the automatic ARIMA procedure for the target variable (*Model 2-A)* and repeat the procedure without X_2 (*Model 2-B*). If *Model 2-A* meets the ***condition 1-2***, we retain both X_1 and X_2.

If X_1 is removed from step one, we then repeat step two with X_2 as an exogenous regressor only. We keep X_2 if it meets ***conditions 1-2*** and discard it otherwise. We repeat Step 1 and 2 with the remaining candidate variables.

<u>Step Three:</u> we add the selected variables using steps 1 and 2 to the automatic ARIMA model selection procedure. We then evaluate the validity of ***condition 3*** for the selected variables and retain the variable if it is statistically significant at the 5% confidence level. Meanwhile, we also check ***condition 2*** to ensure that the coefficient sign of each variable consistently matches with economic priors.

<u>Step Four:</u> after ensuring all independent variables meet the three previous conditions, we need to manually check the significance level of the selected ARIMA orders to ensure those orders are meaningful in the model. We may start by removing the ARIMA term with the highest non-significant *t*-statistics, until all ARIMA are statistically significant at 15% level and regressors are statistically significant at 5% level while having intuitive coefficient signs.

Lastly, standard regression diagnostics are performed on the residuals of the preferred model, to guard against variable omission and non-normal error distributions.

To illustrate the proposed procedure, assume that we need to select variables to nowcast India's real GDP (Y_t) from the ten pre-selected, exogenous indicators shown in Table 1. The pre-selected data include commonly used macroeconomic variables in nowcasting models, covering the external, real, and monetary sides of the economy. They were obtained from official Indian government agencies (e.g., Reserve Bank of India, Ministry of Commerce and Industry, and Ministry of Statistics and Program Planning). All ten indicators are published monthly and are updated after the latest official data release for real GDP.

Table 1: Pre-Selected data to Nowcast India's Real GDP

Series Name	Description	Start	End	Source
credit_card	India: Credit Cards	Apr-2004	May-2022	RBI
ip	India: IP: General Index	Jan-1971	May-2022	MOSPI
pmi_manu	India PMI: Manufacturing	Mar-2005	Jul-2022	SPG
exports	India: Merchandise Exports, f.o.b.	Jan-1968	Jul-2022	MoCI
elec_genr	India: Electricity Generation	Jan-2005	Jun-2022	MoP
stock	India: Stock Prices: BSE Sensex/BSE 30 Index	Apr-1979	Jul-2022	BSE
for_inv	India: Foreign Investment Inflows	Sep-1997	May-2022	RBI
t_bill	India: 91-Day Treasury Bill Implicit Cut-Off Yield	Jan-1993	Jun-2022	RBI
ecma	India: Rupee/US$ Exchange Rate	Jan-1980	Jul-2022	RBI
reserve_assets	India: Official Reserve Assets	Feb-2005	Jun-2022	RBI

We start the exogenous selection procedure with the first indicator DLOG(CREDIT_CARD) and repeat the procedure for the remaining variables. Table 2 presents the detailed report for each indicator and how each variable meets the first two acceptance conditions. After Step 2 is completed, five indicators meet both conditions, namely: DLOG(CREDIT_CARD), DLOG(IP), DLOG(ELEC_GENR), D(T_BILL), and DLOG(ECMA), allowing us to move forward to Step 3.

Table 2 shows the detailed statistics for each variable selection criterion. A gray shaded area means that the indicator passed the indicated criterion. For example, DLOG(CREDIT_CARD) decreases the AIC value and is statistically significant. The last column tallies the number of conditions met by each variable. We retain only those variables that satisfy both conditions.

Table 2 Automatic ARIMA Stepwise Variable Selection - Steps 1 and 2 Result

			Condition 1	*Condition 2*	
Selected	**Iterations**	**Variable (X) Added**	**AIC value**	**Coefficient**	**# of Conditions Met**
√	1	dlog(CREDIT_CARD)	-4.977	0.312	2
√	2	dlog(IP)	-5.723	0.430	2
--	3	PMI_MANU	-5.707	0.000	1
--	4	dlog(EXPORTS)	-5.780	-0.042	1
√	5	dlog(ELEC_GENR)	-5.776	0.114	2
--	6	dlog(STOCK)	-5.762	-0.007	0
--	7	FOR_INV	-5.747	0.000	0
√	8	d(T_BILL)	-5.794	-0.004	2
√	9	dlog(ECMA)	-5.843	0.069	2
--	10	dlog(RESERVE_ASSETS)	-5.804	0.044	1

With the five variables selected from Steps 1 and 2, we can proceed with Step 3, which brings the selected variables to the automatic ARIMA procedure to assess **Condition 3** (statistical significance) and reassess **Condition 2** (coefficient sign consistent with economic priors).

Table 3 presents the result of Step 3, which shows only DLOG(ELEC_GENR) need to be remove due to statistical insignificancy. Other four variables remain to be valid with both condition 2 and 3.

Table 3 Automatic ARIMA Stepwise Variable Selection - Step 3 Result

First Iteration: *Remove DLOG(ELEC_GENR)*

	Condition 2	*Condition 3*
	Coefficient	**P-Value**
DLOG(CREDIT_CARD)	0.083	0.000
DLOG(IP)	0.382	0.000
DLOG(ELEC_GENR)	0.059	0.417
D(T_BILL)	-0.005	0.027
DLOG(ECMA)	0.069	0.027

Second Iteration: *No need to remove any regressors*

	Condition 2	*Condition 3*
	Coefficient	**P-Value**
DLOG(CREDIT_CARD)	0.072	0.000
DLOG(IP)	0.420	0.000
D(T_BILL)	-0.005	0.024
DLOG(ECMA)	0.074	0.022

This variable, along with the automatically selected ARIMA terms, are then used to formulate the baseline nowcasting model (Table 4).

As shown in Table 4, the ARIMA (4,1) model has been selected for the real GDP of India based on the Akaike Information Criterion (AIC). Note that the exogenous regressor in the baseline model fulfills the coefficient and significance level requirements.

Table 4 Automatic ARIMA Stepwise Variable Selection – Selected Baseline Model

Dependent Variable: DLOG_RGDP
Method: ARMA Maximum Likelihood (BFGS)
Sample: 2004Q3 2022Q1
Included observations: 71
Convergence achieved after 27 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.006153	0.002259	2.723768	0.0084
DLOG(CREDIT_CARD)	0.071922	0.017122	4.200483	0.0001
DLOG(IP)	0.420281	0.024645	17.05362	0.0000
D(T_BILL)	-0.004542	0.001954	-2.323883	0.0235
DLOG(ECMA)	0.074084	0.031424	2.357597	0.0217
AR(1)	-0.004592	0.146991	-0.031240	0.9752
AR(2)	-0.121867	0.119977	-1.015754	0.3138
AR(3)	0.001750	0.136442	0.012827	0.9898
AR(4)	0.703361	0.111025	6.335172	0.0000
MA(1)	-0.394784	0.163224	-2.418671	0.0186
SIGMASQ	0.000120	2.27E-05	5.272565	0.0000

R-squared	0.912335	Mean dependent var	0.016324
Adjusted R-squared	0.897725	S.D. dependent var	0.037240
S.E. of regression	0.011909	Akaike info criterion	-5.835296
Sum squared resid	0.008510	Schwarz criterion	-5.484740
Log likelihood	218.1530	Hannan-Quinn criter.	-5.695891
F-statistic	62.44261	Durbin-Watson stat	1.933290
Prob(F-statistic)	0.000000		

Notice that some of the selected AR terms are not statistically significant in the baseline model. Step 4 is now used to remove the insignificant AR terms one-by-one, starting from the AR term with the highest *p*-value. After removing the AR(3), AR(1), and AR(2) terms, we obtain an adjusted baseline model with RHS variables that satisfy all the selection criteria.

Table 5 Automatic ARIMA Stepwise Variable Selection – Adjusted Baseline Model

Dependent Variable: DLOG_RGDP
Method: ARMA Maximum Likelihood (BFGS)
Sample: 2004Q3 2022Q1
Included observations: 71
Convergence achieved after 8 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.005763	0.002811	2.050101	0.0445
DLOG(CREDIT_CARD)	0.074599	0.015350	4.859851	0.0000
DLOG(IP)	0.417048	0.020870	19.98309	0.0000
D(T_BILL)	-0.004621	0.001905	-2.426252	0.0181
DLOG(ECMA)	0.074911	0.030221	2.478769	0.0159
AR(4)	0.748953	0.101580	7.373048	0.0000
MA(1)	-0.451573	0.096209	-4.693643	0.0000
SIGMASQ	0.000123	2.24E-05	5.504864	0.0000

R-squared	0.909958	Mean dependent var	0.016324
Adjusted R-squared	0.899953	S.D. dependent var	0.037240
S.E. of regression	0.011779	Akaike info criterion	-5.890504
Sum squared resid	0.008741	Schwarz criterion	-5.635554
Log likelihood	217.1129	Hannan-Quinn criter.	-5.789118
F-statistic	90.95306	Durbin-Watson stat	1.838124
Prob(F-statistic)	0.000000		

The final step is to conduct residual-based diagnostic tests to guard against omitted variable bias, which presents no evidence of serial correlation nor heteroskedastic. We will demonstrate how to improve this model using a more comprehensive list of candidate variables in the section 6.

Section 4. Benchmark Models

We use three alternate benchmark models to assess the effectiveness of the proposed AS-ARIMAX approach to the country case of India:

1) The random walk model (with Autoregression of order 1)
2) The professional forecasters survey from the Reserve Bank of India
3) The combinatorial variable selection

We show below that our Bridge and Unrestricted Mixed-frequency Data Sampling (U-MIDAS) estimations for real GDP – both formulated using the AS-ARIMAX approach – outperform the three benchmark models, delivering much lower root mean square error (RMSE).

We now explain the three benchmark models in detail.

Benchmark 1: Univariate autoregression model

The first-order autoregressive AR (1) model has been used very frequently as a benchmark to compare the relative performance of nowcasting models. For example, Giannone, *et al.* (2013) used AR (1) as the benchmark in their study on nowcasting China's real GDP; Bok, *et al.* (2017) used naïve as the benchmark in their report on nowcasting using big data for the United States.

The AR (1) model is (Bragoli & Fosten, 2017):

$$y_t^Q = \rho y_{t-1}^Q + \epsilon_t^Q$$

where y_t^Q is the quarter-on-quarter growth rate of quarterly real GDP, y_{t-1}^Q is the previous period value of the y_t^Q, ϵ_t^Q is a zero mean idiosyncratic term, and ρ is the autoregressive parameter satisfies $|\rho| < 1$.

Benchmark 2. Reserve Bank of India Professional Forecasters

The Reserve Bank of India (RBI) conducts and publishes a survey of 30 professional forecasters on the annual growth rate of Indian real GDP by industry. We use the mean of these professional forecasts. Since the forecast survey focuses on the annual growth rate, we convert the forecast output to the quarter-on-quarter rate to be consistent with the target variable. The RBI forecast series has been seasonally adjusted using X-13 procedure to ensure the consistency with target variable.

Benchmark 3. EViews Regression Variable Selection: Combinatorial

EViews has five variable selection methods (Uni-directional, Stepwise, Swap-wise, Combinatorial, Auto-Search/GETS and Lasso Selection). Among the five methods provided, the combinatorial method provides the most thorough evaluation as it evaluates all the possible combinations of added variables, selecting the combination with the largest R-squared (EViews User Guide II, pp89).

Section 5. Nowcasting Methodology

The main idea we are proposing here is to first apply the AS-ARIMAX selection procedure to select the indicators that meet the three conditions mentioned previously (i.e., AIC value, intuitive coefficient sign, and statistically significant). Then, we use the selected indicators in standard nowcasting models to assess their predictive performance. We use two nowcasting models in what follows: Bridge and the Unrestricted Mixed-Frequency Data Sampling (U-MIDAS).

Nowcasting Model 1: Bridge Model

The first model we use is the Bridge model, which relies on linear regressions that link ("bridge") high-frequency explanatory variables with the low-frequency target variables. To nowcast quarterly GDP, the high-frequency indicators (e.g., monthly) are converted to the lower, target frequency (e.g., quarterly) using the sum or average of the observations in the quarter. The Bridge model is then estimated using ordinary least squares (OLS). If the high-frequency indicators have publication lags, an auxiliary regression is used to forecast the high frequency indicators so that each low frequency period has a complete set of high frequency values. Note that the inclusion of the right-side variables or indicators in the Bridge model is not based on casual relations (as compared to a more structural model), but on a pre-assessment or prior that they contain timely updated information on the future direction of the dependent variable (e.g., real GDP). Because of its simplicity and transparency, numerous policy institutions have used bridge equations to guide policy decisions (e.g., Federal Reserve Bank of San Francisco (Ingenito & Trehan, 1996), Euro Area (Baffigia, Golinellib, & Parigia, 2004), and Norges Ban (Foroni & Marcellino, 2013)).

The ARIMAX Bridge model can be represented as

$$y_{t_q} = \alpha + \sum_{i=1}^{p} \varphi_i \, Y_{t-i} + \sum_{i=1}^{q} \theta_i \, \varepsilon_{t-i} + \sum_{i\,=\,1}^{j} \beta_i \, x_{it_q} + u_{t_q}$$

where β_i is the coefficient of the exogenous regressor, t_q= 1,…,T indicates time in quarters, x_i is a high-frequency indicator, and u_{t_q} is an i.i.d. error term. Moreover, $\sum_{i=1}^{p} \varphi_i \, Y_{t-i}$ is the autoregressive (AR) term of order *p* (i.e., AR (*p*)) and $\sum_{i=1}^{q} \theta_i \, \varepsilon_{t-i}$ is the moving average (MA) of q (i.e., MA (q)).

Nowcasting Model 2: U-MIDAS

The second nowcasting model we use is the unconstrained mixed-frequency model (U-MIDAS). The Mixed-Frequency Data Sampling (MIDAS) model is a tightly parameterized reduced form regression in which variables are sampled at a different frequency (Ghysels, Sinko, & Valkanov, 2007). To guard against parameter proliferation issues, the MIDAS model uses distributed lag polynomials that depend on a smaller number of parameters. The MIDAS approach is suitable if the frequency mismatch is large (e.g., when using daily indicators to nowcast a quarterly variable). By contrast, the unrestricted MIDAS model (U-MIDAS) is used when the frequency mismatch is not large. Unlike standard MIDAS, it does not use functional distributed lags. Foroni, Marcellino and Schumacher (2012) studied the performance of U-MIDAS and found that U-MIDAS generally performs better than MIDAS when mixing quarterly and monthly data (i.e., small frequency mismatch).

In this paper, we apply the AS-ARIMAX procedure to the U-MIDAS model. We convert the higher frequency indicators to quarterly frequency using split-sampling. The model can be expressed in its simplest form as follows:

$$y_{t_m} = \alpha + \sum_{i=1}^{p} \varphi_i \, Y_{t-i} + \sum_{i=1}^{q} \theta_i \, \varepsilon_{t-i} + \sum_{i=1}^{j} \beta_i \, x_{it_m} + \sum_{i=1}^{j} \beta_i \, x_{it_{m-1}} + \sum_{i=1}^{j} \beta_i \, x_{it_{m-2}} + u_{t_m}$$

where β_i is the coefficient of the exogenous regressor, t_q= 1,…,T indicates time in quarters, x_i is a high-frequency indicator, x_{it_m} is the first skip-sampled quarterly high-frequency variable, and $x_{it_{m-1}}$ and $x_{it_{m-2}}$ are the second and third skip-sampled variables, j is the number of high-frequency indicators.

Section 6. Nowcasting Indian Real GDP

We now demonstrate the effectiveness of the AS-ARIMAX approach for nowcasting the real GDP of India. We follow Bragoli and Fosten (2017) and show that the AS-ARIMAX approach outperforms the three benchmark models proposed in Section 4.

Target variable: Real GDP

Indian real Gross Domestic Product (GDP) data is published by Central Statistics Organization (CSO), India. We also use Real Gross Domestic Product at Basic Prices as our target variable, to make it consistent with the target variable of the professional forecasts published by the Reserve Bank of India.

Input variables

We started with the indicators list proposed by Bragoli and Foster (B&F, 2017), applying two selection criteria to pre-filter the list of indicators:

1) **Frequency**: the frequency of the selected indicators must be the same or higher than the target frequency. In this study, our target variable (i.e., real GDP) is available quarterly. Therefore, we choose indicators with at least quarterly frequency.

2) **Availability**: the selected indicators must have published data after the latest publication of the target variable. In this study, the target variable is available until 2022Q1. Therefore, we use indicators that have published data after 2022Q1 (i.e., 2022M03).

We removed two indicators upfront: India Crude Oil Production and Steel Production as these series failed to meet the availability requirement. In addition, India's "Industrial Performance Assessment indicator" is unavailable from our data sources. We used an industry index as a replacement, representing the industry performance of eight core industries in India: Coal, Crude Oil, Natural Gas, Petroleum Refinery products, fertilizers, steel, cement, and electricity. In addition to the headline indicators used by B&F, we wish to gauge the predictive ability of each sub-component of the industrial, stock indexes, and foreign trade. Therefore, we added sub-components of industrial production, the eight core industries index, export, foreign investment flows, and the NSE stock index. After these adjustments, the exogenous variable candidate list contains eighty variables (Appendix 2).

Data Transformation

To ensure the stationarity of the variables in the OLS regression, we apply the Augmented Dicky-Fuller unit root test to each variable. If the test rejects the null hypothesis of a unit root, we treat the series as stationary. Otherwise, we apply an appropriate transformation to ensure stationarity (i.e., first difference or log difference).

Seasonal Adjustment

We gather the seasonal adjustment status of each variable from the data source and use the X-13 seasonal adjustment procedure (United States Census Bureau, 2022) as required to seasonally adjust the series. Detailed seasonal adjustment status for each series can be found in Appendix 2.

Frequency Conversion

All the selected indicators have monthly frequency. Given the target variable is quarterly, we need to convert the high-frequency indicators from monthly to the quarterly frequency. For the Bridge model, we use the aggregation approach by summing or averaging the monthly data. We decide the conversion method based on the nature of the indicators. Specifically, we characterize each variable as “flow” or “stock/index”. We use “sum observation” for “flow” variables and “average observation” for “stock/index” variables. Detailed descriptions of each indicator’s stock or flow classification can be found in Appendix 2.

Model evaluation

To evaluate the model’s performance, we use a realistic forecast evaluation methodology with a “pseudo real-time” historical series construction that reflects the operational procedures typically used in a forecasting unit of a central bank (Bok, Caratelli, Giannone, Sbordone, & Tambalotti, 2017). We emulate a nowcasting protocol in which the baseline model is re-estimated regularly based on all the information available at that specific time.

To keep the procedure as straightforward as possible, we assume that the monthly indicators have a one-month lag, meaning we need to forecast each regressor by one period to ensure sufficient data for the nowcasting exercise. Suppose we are currently at the end of 2018Q2, and we wish to nowcast the real GDP (**rgdp**), which is only available until 2018Q1, using monthly indicators available till 2018M05. We use all the available quarterly data to construct a baseline model (estimated using data out to 2018Q1).

Then, we forecast all the monthly indicators for one month to 2018M06 using an auxiliary model to ensure that we have sufficient quarterly data for a nowcast in 2018Q2. After forecasting the monthly indicators and converting the forecasted series to quarterly frequency, we nowcast real GDP to 2018Q2. We record the nowcast value for 2018Q2 and repeat the same steps for 2018Q3. The procedure is repeated until we have the nowcasting result for our target evaluation end date. We then evaluate the forecast accuracy of different models using RMSE and the Theil U2 statistics, as these contain information most applicable to model selection procedures.

Section 7. Empirical Results

Starting with the eighty indicators described in Appendix 2, the AS-ARIMAX indicator selection procedure shortlisted five variables for the baseline model with AR(1), MA(1) and MA(3) terms (see Table 6). The selected indicators are:

- **HVI**, defined as India's Eight Core Industry Infrastructure Index (SA, Apr.11-Mar.12=100)
- **IP_LEATHER**, defined as India's Industrial Production in Leather and Related Products (SA, Apr.11-Mar.12=100)
- **IP_CAPITAL**, defined as India's Industrial Production in Capital Goods (SA, Apr.11-Mar.12=100)
- **IP_TEXTILES**, defined as India's Industrial Production in Textiles (SA, Apr.11-Mar.12=100)
- **NFGEB**, defined as India's Central Government: Expenditure (SA, 10 Mil. Rupees)

Given manufacturing's importance to the Indian economy (i.e., representing more than 23% of India's 2021 real GDP), industrial production (IP) sub-components account for three of five indicators selected. The selected IP indicators cover capital goods and the wearing-apparel industry. Additionally, the selection of the "Eight Core Industry Infrastructure Index" presents the importance of core industries (such as refinery products, electricity, and steel) to Indian's economy development. The last indicator reflects central government expenditure, which is a sound indicator for the fiscal policy and will likely to impact economic activity more generally.

Table 6 Selected Baseline Model

Dependent Variable: DLOG_RGDP
Method: ARMA Maximum Likelihood (BFGS)
Sample: 2005Q3 2022Q1
Included observations: 67
Convergence achieved after 21 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.018805	0.004931	-3.813584	0.0003
DLOG(HVI)	0.470237	0.058540	8.032805	0.0000
DLOG(IP_LEATHER)	0.053820	0.013773	3.907548	0.0002
DLOG(IP_CAPITAL)	0.047824	0.009175	5.212350	0.0000
IP_TEXTILES	0.000269	5.03E-05	5.346868	0.0000
DLOG(NFGEB)	0.033673	0.008986	3.747099	0.0004
AR(1)	0.718138	0.118836	6.043101	0.0000
MA(1)	-1.085777	0.266486	-4.074428	0.0001
MA(3)	0.581108	0.386306	1.504270	0.1380
SIGMASQ	5.72E-05	2.61E-05	2.191669	0.0325

R-squared	0.960367	Mean dependent var	0.015998
Adjusted R-squared	0.954109	S.D. dependent var	0.038280
S.E. of regression	0.008200	Akaike info criterion	-6.553206
Sum squared resid	0.003833	Schwarz criterion	-6.224148
Log likelihood	229.5324	Hannan-Quinn criter.	-6.422997
F-statistic	153.4661	Durbin-Watson stat	2.236925
Prob(F-statistic)	0.000000		

The procedure ensures that the selected indicators are statistically significant at a 5% confidence interval. Moreover, the estimated coefficients on the selected indicators have the correct sign.

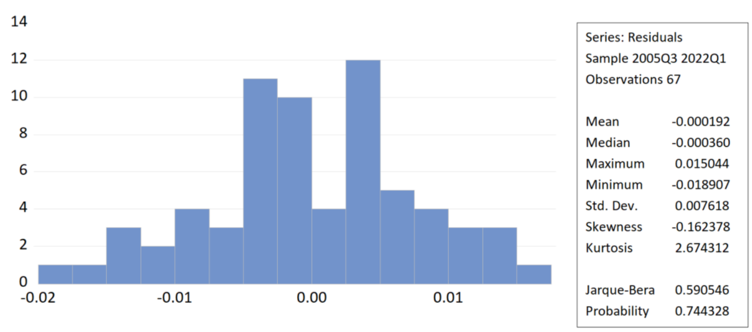
Table 7 presents the correlogram and Q statistics of the LM test with 12 lags. Since all the p-values of Q-stat is above the 5 percent significance level, we conclude that no serial correlation is present in the baseline model.

Table 7 Testing for Serial Correlation: Q Statistics

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob*
		1	-0.124	-0.124	1.0839	
		2	0.074	0.059	1.4717	
		3	-0.017	-0.001	1.4926	
		4	0.007	0.000	1.4959	0.221
		5	0.147	0.152	3.1078	0.211
		6	0.073	0.112	3.5126	0.319
		7	0.111	0.120	4.4534	0.348
		8	-0.068	-0.048	4.8194	0.438
		9	-0.025	-0.058	4.8683	0.561
		10	0.096	0.074	5.6139	0.585
		11	0.164	0.172	7.8395	0.449
		12	-0.078	-0.093	8.3459	0.500

The *p*-value for the Jarque-Bera test exceeds the significance level, indicating accepting the null hypothesis of normal distribution. We can also see a clear bell-shaped distribution from the histogram of the fitted residuals.

Table 8 Testing for Normality: Jarque-Bera Test



The Breusch-Pagan-Godfrey heteroskedasticity test shown in Table 9 indicates no evidence of heteroskedasticity given that the *p*-value of the *F*-statistics are all above 0.05.

Table 9 Testing for Heteroskedasticity: Breusch-Pagan-Godfrey test

Heteroskedasticity Test: Breusch-Pagan-Godfrey
Null hypothesis: Homoskedasticity

F-statistic	1.984319	Prob. F(5,61)	0.0937
Obs*R-squared	9.372984	Prob. Chi-Square(5)	0.0951
Scaled explained SS	5.736579	Prob. Chi-Square(5)	0.3327

Given that we have confirmed the validity of the selected baseline model, the next step is to use the baseline model to generate nowcasting results for Indian real GDP. To compare the Bridge and U-MIDAS estimations with our three benchmark models, we used a one-period ahead out-of-sample evaluation methodology starting from 2018Q1 and ending in 2022Q1. We also created a simpler forecast evaluation approach, in which we estimated the model up to 2022Q1, and then created in-sample forecast from 2018Q1 to 2022Q1.

Table 10 shows the RMSE values for the benchmark and the nowcasting models during the same evaluation period. The results indicate that the two nowcasting models, created using the AS-ARIMAX variable selection procedure, outperform all three-benchmark models. Among the two nowcasting models, U-MIDAS performs better with both evaluations, while Bridge and U-MIDAS performs almost the same with out-of-sample evaluation (difference equals to 0.001).

Comparing the best-performed benchmark model (i.e., combinatorial), by using the AS-ARIMAX procedure, U-MIDAS decreases the out-of-sample RMSE by more than **twenty-six percent**. Comparing the worst performing benchmark model (i.e., RBI forecast), the AS-ARIMAX nowcasting model improves the out-of-sample RMSE by more than **seventy percent**.

Table 10 Forecast Evaluation Comparison: RMSE

	In-Sample	Out-of-Sample
AR(1)	0.074	0.097
RBI Forecast	0.102	0.102
Combinatorial	0.016	0.033
Bridge	0.012	0.025
U-MIDAS	0.010	0.025
Accuracy Gains*	36.7%	26.0%

Note: Accuracy Gains in RMSE = - 100% * (U-MIDAS – Combinatorial)/Combinatorial

The result of the Theil coefficient is largely consistent with that of RMSE: nowcasting models using AS - ARIMAX procedure outperformed all three benchmarks. The combinatorial is still the best-performing benchmark model using the Theil U2 statistic. Meanwhile, U-MIDAS persist as the best performing model among the five models in both in-sample and out-of-sample evaluation. Compared with the combinatorial model, the U-MIDAS model reduces the out-of-sample Theil U2 by more than **sixty percent.**

Table 11 Forecast Evaluation Comparison: Theil U2

	In-Sample	Out-of-Sample
AR(1)	1.073	1.071
RBI Forecast	1.136	1.136
Combinatorial	0.042	0.548
Bridge	0.035	0.390
U-MIDAS	0.013	0.195
Accuracy Gains*	69.9%	64.4%

The results from both statistics provide strong suggestive evidence of the significant efficiency gains from using the AS-ARIMAX variable selection procedure in determining the high-frequency (HF) regressors in the nowcasting model.

We also demonstrate the efficiency gains by calculating the forecast error (i.e., 100*(Forecasted value – Actual Value)) and visualizing the differences between forecasted and actual real GDP.

Figure 1 Realistic Forecast Evaluation: Three Benchmark Models

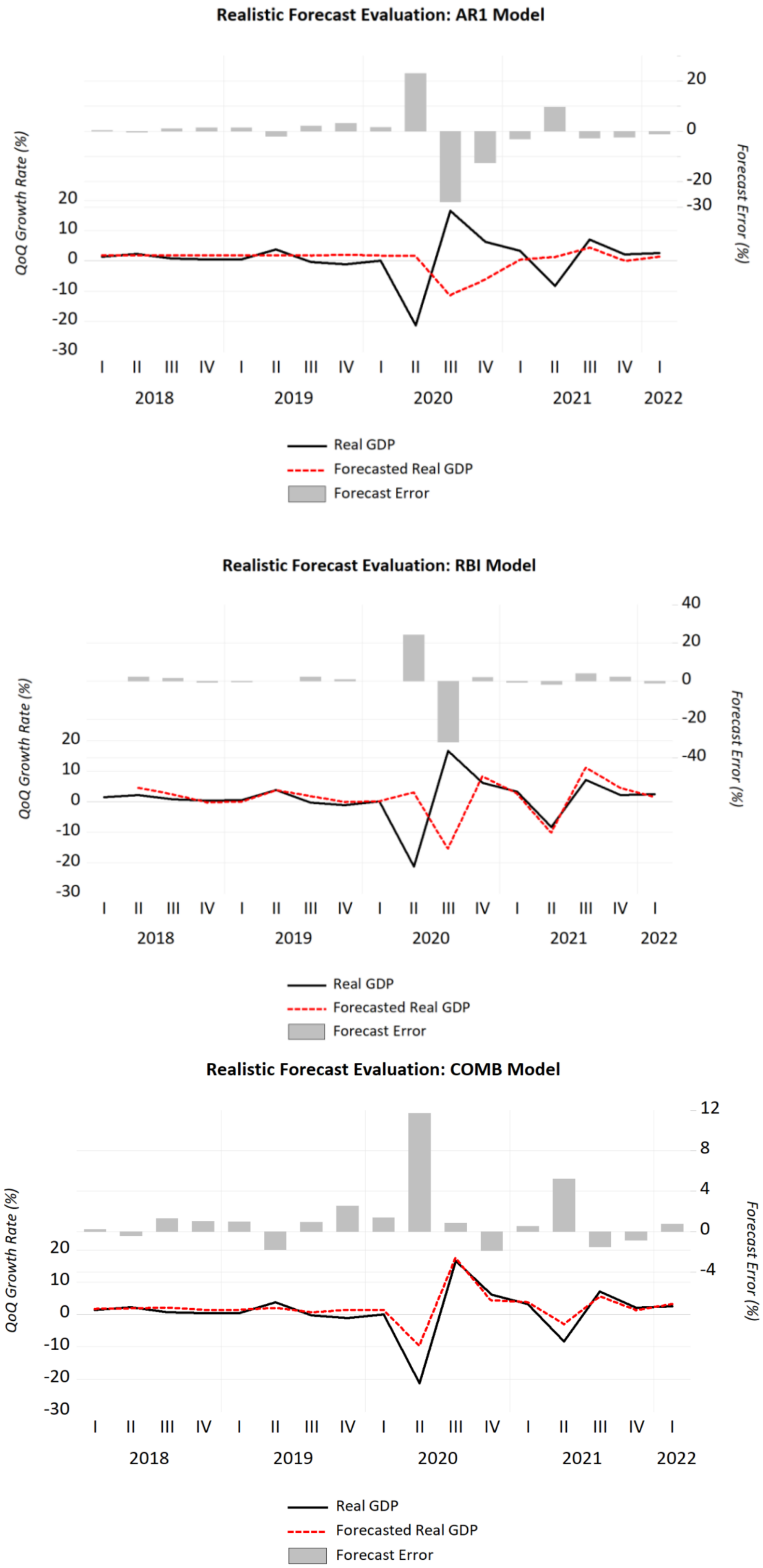


Figure 1 shows the realistic evaluation (i.e., out-of-sample forecast) graph of the three benchmark Models. The forecast error of all three benchmark models increased significantly as the COVID-19 pandemic hit India, resulting in a strict nationwide lockdown during 2020Q2. Note that the AR1 and Reserve Bank of India (RBI) forecasts only capture the impact of COVID-19 with a lag: both models predict a negative growth rate in 2020Q3 when the actual economy recovered with the ease of national lockdown. The AR(1) model did not capture the impact of the second COVID-19 wave in 2021Q2. The RBI forecast captured such an impact quite precisely. The combinatorial approach (COMB) generates forecasts that are most aligned with the actual compared to other benchmarks. However, it fails to reflect the intensity of the negative impact. For example, the actual quarter-on-quarter growth rate in 2020Q2 is -21.3%, while the COMB approach suggests it to be only -9.6%.

Figure 2 Realistic Forecast Evaluation: Bridge and U-MIDAS Models

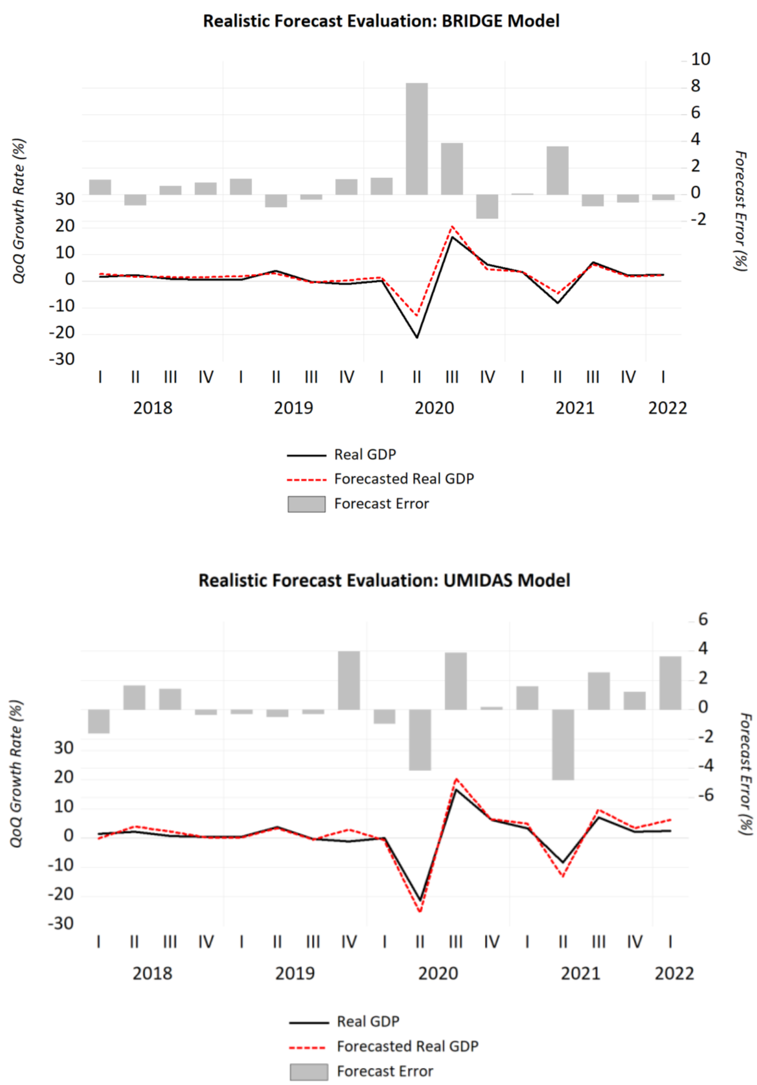


Figure 2 shows the forecasting performance of the two nowcasting models (Bridge and U-MIDAS). Unlike all three benchmark models, the Bridge and U-MIDAS models capture the negative shocks emanating from both

COVID-19 waves in direction and intensity. The forecast gap from Bridge and U-MIDAS is much smaller than the COMB model (the best performing benchmark).

Section 8. Other Country Examples

We also implemented the proposed approach in five additional countries to further demonstrate the efficacy of the AS-ARIMAX procedure. The countries selected (Argentina, Australia, South Africa, the United Kingdom, and the United States) are diverse in terms of geographic location and income level.

For each country, we pre-selected 30 indicators covering their external environment, surveys, consumptions, financial, trade, labor markets, and productions, all readily available and updated after the latest actual GDP figure. We then apply the AS-ARIMAX procedure to determine the optimal baseline model. The computed nowcasting models are compared with a univariate autocorrelation AR(1) model and EViews's combinatorial variable selection approach.

Table 12 reports the realistic out-of-sample forecast evaluation results (RMSE) for the benchmark and nowcasting models. The gray shaded area indicates the best model based on RMSE. The result is largely consistent with the output from the India country example, in which the Bridge and U-MIDAS models outperformed both benchmark models. Compared with the combinatorial model, the U-MIDAS model reduces the out-of-sample RMSE on average of **sixty-seven percent** across all five countries (see Appendix III for a detailed report on each country).

Table 12 Forecast Evaluation Comparison for Other Countries: RMSE

RMSE	**Argentina**	**Australia**	**South Africa**	**United Kingdom**	**United States**
AR1	0.055	0.030	0.080	0.105	0.045
Combinatorial	0.013	0.017	0.032	0.032	0.016
BRIDGE	0.007	0.010	0.016	0.028	0.013
UMIDAS	0.007	0.005	0.003	0.012	0.005
Accuracy gains*	45%	70%	89%	62%	69%

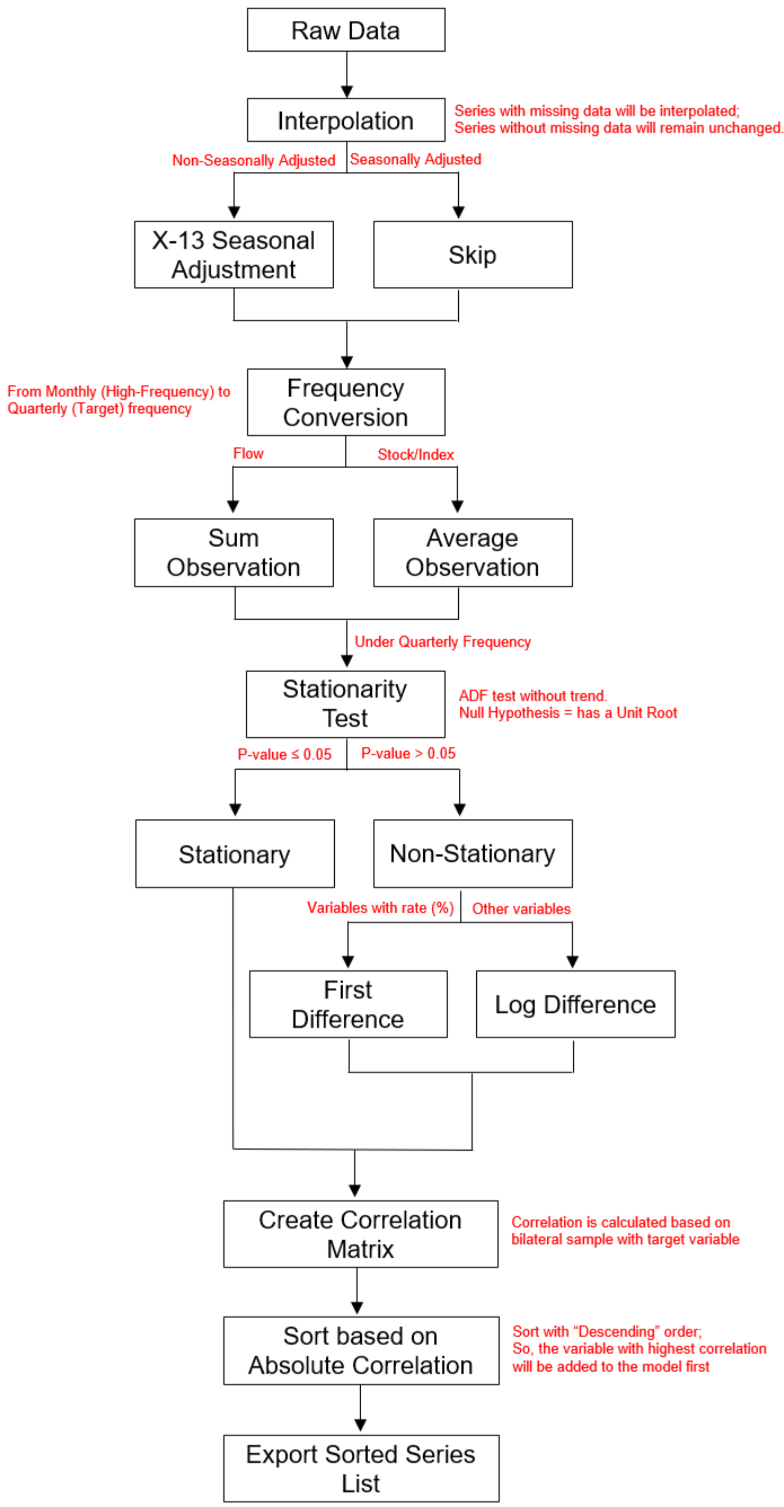
**Accuracy gains between best nowcasting model and best benchmark model*

Section 9. Conclusion

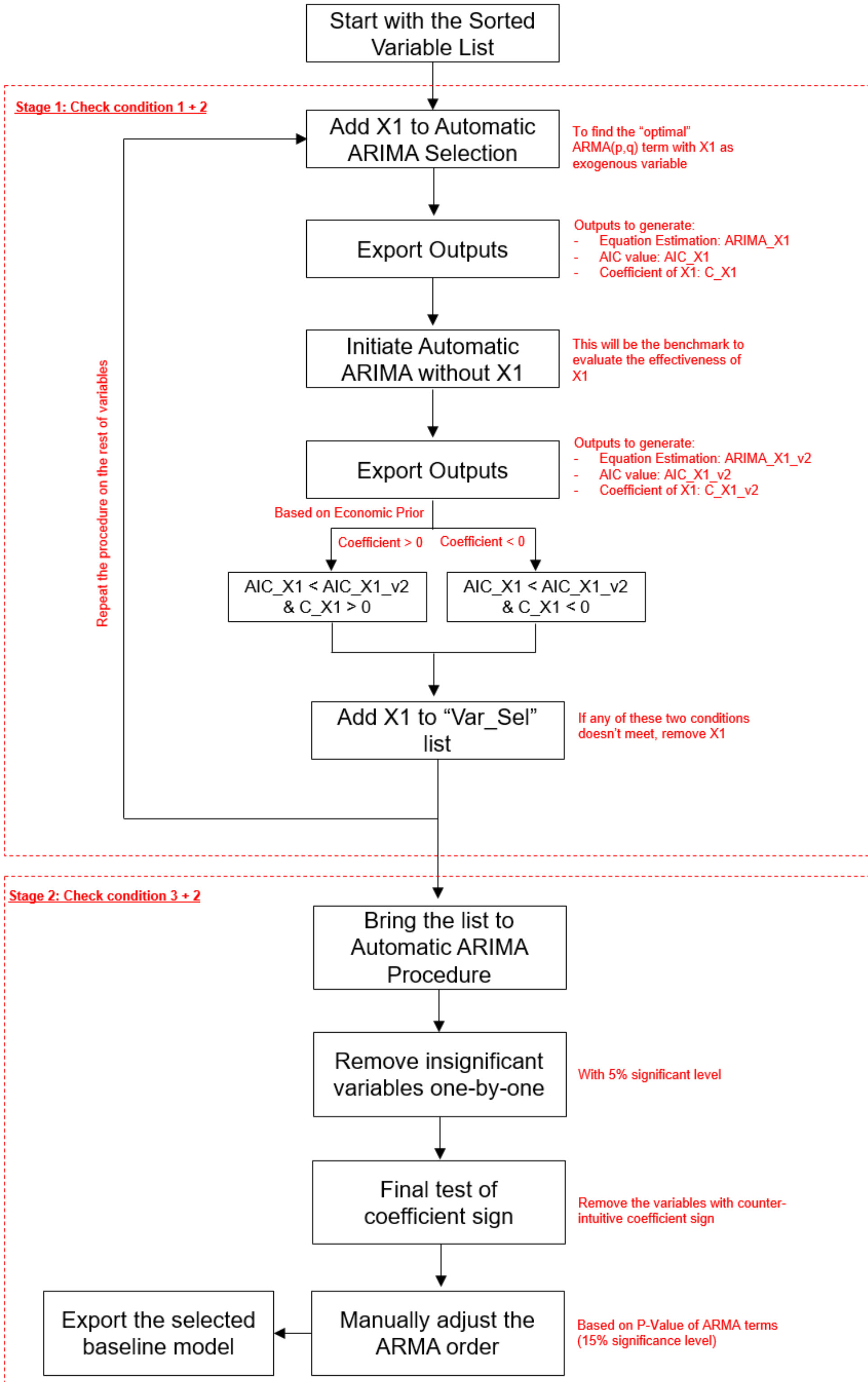
This paper focuses on how to choose the best nowcasting model given a set of exogenous indicators to select from. The AS-ARIMAX model selection procedure proposed in this paper ensures the inclusion of significant ARIMA terms in the model, while also assessing three critical conditions for the explanatory variables that are likely to improve forecasting/nowcasting performance. We show that the AS-ARIMAX approach yields reliable nowcasting in both direction and intensity during the COVID-19 crisis period compared to more traditional approaches. Using Indian real GDP data, we show that the AS-ARIMAX selection procedure reduces the RMSE by at least **twenty-six** percent compared to the three competing nowcasting methods. We also verify its effectiveness for five other countries by showing that the AS-ARIMAX selection procedures reduce the RMSE of the baseline model by an average of **sixty-seven** percent compared to two benchmark forecasting models. Given these impressive forecasting gains, the effectiveness of the AS-ARIMAX approach using other macroeconomic variables (e.g., the inflation rate) will be assessed in future work.

Annex I. AS-ARIMAX Procedure Charts

Part 1. Data Processing



Part 2. Variable Selections



Annex II. Candidate Indicators and Three Main Attributes

No.	Series_name	Description	Frequency Conversion	Seasonal Adjustment	Coefficient Sign
1	exports	India: Merchandise Exports, f.o.b. (NSA, Mil.US$)	flow	nsa	1
2	imports	India: Merchandise Imports, c.i.f. (NSA, Mil.US$)	flow	nsa	-1
3	pmi_manu	India PMI: Manufacturing (SA, 50+=Expansion)	stock	sa	1
4	pmi_serv	India PMI: Services Business Activity (SA, 50+=Expansion)	stock	sa	1
5	elec_genr	India: Electricity Generation (NSA, Gwh)	flow	nsa	1
6	eight_core_infra	India: Eight Core Industry Infrastructure Index (NSA, Apr.11-Mar.12=100)	stock	nsa	1
7	m1	India: Money Supply: M1 (EOP, NSA, Bil.Rupees)	stock	nsa	1
8	ecma	India: Rupee/US$ Exchange Rate (EOP)	stock	nsa	1
9	t_bill	India: 91-Day Treasury Bill Implicit Cut-Off Yield (% per annum)	stock	sa	-1
10	stock	India: Stock Prices: BSE Sensex/BSE 30 Index (EOP, 1978-79=100)	stock	nsa	1
11	cpi	India: Consumer Price Index (NSA, 2012=100)	stock	nsa	1
12	wpi	India: Wholesale Price Index: All Items (NSA, Apr.11-Mar.12=100)	stock	nsa	1
13	ip_us	IP: Total Index (SA, 2017=100)	stock	sa	1
14	pmi_us	US Markit PMI: Manufacturing (SA, 50+= Expansion)	stock	sa	1
15	pmi_ea	Euro Area PMI: Manufacturing (SA, 50+=Expansion)	stock	sa	1
16	ip_ea	EA19: IP: Industry Including Construction (SWDA, 2015=100)	stock	sa	1
17	credit_card	India: Credit Cards (Bil.Rupees)	flow	nsa	1
18	private_credit	India: Private Sector Credit (EOP, NSA, 10 Mil.Rupees)	flow	nsa	1
19	reserve_assets	India: Official Reserve Assets (NSA, EOP, Mil.US$)	stock	nsa	1
20	for_inv	India: Foreign Investment Inflows (NSA, Mil US$)	flow	nsa	1
21	ip_basic	India: IP: Basic Goods (SA, Apr.11-Mar.12=100)	stock	sa	1
22	ip_dur	India: IP: Consumer Durable Goods (SA, Apr.11-Mar.12=100)	stock	sa	1
23	ip_non_dur	India: IP: Consumer Nondurable Goods (SA, Apr.11-Mar.12=100)	stock	sa	1
24	ip_inter	India: IP: Intermediate Goods (SA, Apr.11-Mar.12=100)	stock	sa	1
25	ip_capital	India: IP: Capital Goods (SA, Apr.11-Mar.12=100)	stock	sa	1
26	ip_manu	India: IP: Manfacturing (SA, Apr.11-Mar.12=100)	stock	sa	1
27	ip_metals	India: IP: Basic Metals (SA, Apr.11-Mar.12=100)	stock	sa	1
28	ip_paper	India: IP: Paper and Paper Products (SA, Apr.11-Mar.12=100)	stock	sa	1
29	ip_chemicals	India: IP: Chemicals and Chemical Products (SA, Apr.11-Mar.12=100)	stock	sa	1
30	ip_coke	India: IP: Coke and Refined Petroleum Products (SA, Apr.11-Mar.12=100)	stock	sa	1
31	ip_tobacco	India: IP: Tobacco Products (SA, Apr.11-Mar.12=100)	stock	sa	1
32	ip_fabricated	India: IP: Fabricated Metal Products, ex Machinery & Eqpt(SA, Apr.11-Mar.12=100)	stock	sa	1
33	ip_leather	India: IP: Leather and Related Products (SA, Apr.11-Mar.12=100)	stock	sa	1
34	ip_rubber	India: IP: Rubber and Plastics Products (SA, Apr.11-Mar.12=100)	stock	sa	1
35	ip_motor	India: IP: Motor Vehicles, Trailers & Semi-Trailers (SA, Apr.11-Mar.12=100)	stock	sa	1
36	ip_wearing	India: IP: Wearing Apparel (SA, Apr.11-Mar.12=100)	stock	sa	1
37	ip_wood	India: IP: Wood & Prods of Wood & Cork ex Furn; Etc (SA, Apr.11-Mar.12=100)	stock	sa	1
38	ip_machinery	India: IP: Machinery and Equipment N.E.C. (SA, Apr.11-Mar.12=100)	stock	sa	1
39	ip_mining	India: IP: Mining (SA, Apr.11-Mar.12=100)	stock	sa	1
40	ip_electricity	India: IP: Electricity (SA, Apr.11-Mar.12=100)	stock	sa	1
41	ip_textiles	India: IP: Textiles (SA, Apr.11-Mar.12=100)	stock	sa	1
42	eight_core_coal	India: Eight Core Industries Index: Coal (SA, Apr.11-Mar.12=100)	stock	sa	1
43	eight_core_cement	India: Eight Core Industries Index: Cement (SA, Apr.11-Mar.12=100)	stock	sa	1
44	eight_core_crude Oil	India: Eight Core Industries Index: Crude Oil (SA, Apr.11-Mar.12=100)	stock	sa	1
45	eight_core_elec	India: Eight Core Industries Index: Electricity (SA, Apr.11-Mar.12=100)	stock	sa	1
46	eight_core_fertilizer	India: Eight Core Industries Index: Fertilizers (SA, Apr.11-Mar.12=100)	stock	sa	1
47	eight_core_nat_gas	India: Eight Core Industries Index: Natural Gas (SA, Apr.11-Mar.12=100)	stock	sa	1
48	eight_core_petro	India: Eight Core Ind Index: Petroleum Refinery Products(SA, Apr.11-Mar.12=100)	stock	sa	1
49	eight_core_steel	India: Eight Core Industries Index: Steel (SA, Apr.11-Mar.12=100)	stock	sa	1
50	nse_nifty	India: Stock Price Index: NSE: Nifty (AVG, Nov-3-95=1000)	stock	nsa	1
51	nse_nifty_junior	India: Stock Price Index: NSE: Nifty Junior (AVG, Nov-3-1996=1000)	stock	nsa	1
52	nsa_500	India: Stock Price Index: NSE 500 (AVG, 1994=1000)	stock	nsa	1
53	nse_it	India: Stock Price Index: NSE: Information Technologies (AVG, Jan-01-96=1000)	stock	nsa	1
54	nse_defty	India: Stock Price Index: NSE: Defty (AVG, Nov-03-95=1000)	stock	nsa	1
55	HPI	India: Industrial Workers Consumer Price Index (SA, Apr.16-Mar.17=100)	stock	sa	1
56	HVI	India: Eight Core Industry Infrastructure Index (SA, Apr.11-Mar.12=100)	stock	sa	1
57	NFGBR	India: Central Govt: Revenue Surplus/Deficit (NSA, 10 Mil.Rupee)	flow	nsa	1
58	NFGEB	India: Central Government: Expenditure (NSA, 10 Mil.Rupees)	flow	nsa	1
59	NFYFB	India: External Commercial Borrowings (NSA, Thous.US$)	flow	nsa	1
60	NIE1P	India: Exports: Jute Yarn (NSA, Mil.US$)	flow	nsa	1

No.	Series_name	Description	Frequency Conversion	Seasonal Adjustment	Coefficient Sign
61	NIE2P	India: Exports: Jute Hessian (NSA, Mil.US$)	flow	nsa	1
62	NIE3P	India: Exports: Other Jute Manufactures (NSA, Mil.US$)	flow	nsa	1
63	NIEAJ	India: Exports: Poultry Products (NSA, Mil.US$)	flow	nsa	1
64	NIEC8	India: Exports: Rice: Basmoti (NSA, Mil.US$)	flow	nsa	1
65	NIECB	India: Exports: Rice Other Than Basmoti (NSA, Mil.US$)	flow	nsa	1
66	NIECS	India: Exports: Groundnut (NSA, Mil.US$)	flow	nsa	1
67	NIECV	India: Exports: Seasame Seeds (NSA, Mil.US$)	flow	nsa	1
68	NIECW	India: Exports: Niger Seeds (NSA, Mil.US$)	flow	nsa	1
69	NIED1	India: Exports: Fresh Vegetables (NSA, Mil.US$)	flow	nsa	1
70	NIED3	India: Exports: Processed Vegetables (NSA, Mil.US$)	flow	nsa	1
71	NIED7	India: Exports: Castor Oil (NSA, Mil.US$)	flow	nsa	1
72	NIED8	India: Exports: Cashew Nut Shell Liquid (NSA, Mil.US$)	flow	nsa	1
73	NIEDM	India: Exports: Pulses (NSA, Mil.US$)	flow	nsa	1
74	NIEFA	India: Exports: Sugar (NSA, Mil.US$)	flow	nsa	1
75	NIEGN	India: Exports: Fruits and Vegetable Seeds (NSA, Mil.US$)	flow	nsa	1
76	NIEGQ	India: Exports: Floriculture Products (NSA, Mil.US$)	flow	nsa	1
77	NIECA	India: Exports: Wheat (NSA, Mil.US$)	flow	nsa	1
78	NIEGU	India: Exports: Guar Gum Meal (NSA, Mil.US$)	flow	nsa	1
79	NIEHE	India: Exports: Unmanufactured Tobacco (NSA, Mil.US$)	flow	nsa	1
80	NLFS	India: Foreign Currency Reserves: Securities (NSA, EOP, Mil.US$)	stock	nsa	1

Note: in the “**Frequency Conversion**” column, the Stock vs. Flow nature of each variable is used to convert monthly variables to quarterly frequency. We will apply “Sum observation” to “flow” variables and “Average observation” to “Stock/Index” variables. In the “**Seasonal Adjustment**” column, the seasonality of each variable is used to decide whether to apply the X-13 seasonal adjustment procedure. “NSA” means the series has not been seasonally adjusted and needs to be adjusted. In the “**Coefficient Sign**” column, the expected sign of the coefficient is entered based on an economic prior between the regressor and the target variable. “-1” means a negative coefficient is expected, “1” means a positive coefficient is expected.

Annex III. Other Country Examples

1. Argentina:

1) Data

No.	Series_name	Description	Frequency Conversion	Seasonal Adjustment	Coefficient Sign
1	XNE	Argentina: Multilateral Exchange Rate (NSA, Dec-17-15=100)	stock	nsa	1
2	XRE	Argentina: Real Multilateral Exchange Rate (NSA, Dec-17-15=100)	stock	nsa	1
3	XJRB	Argentina: JPMorgan Real Broad Effective Exchnge Rate Index, PPI Based(2010=100)	stock	nsa	-1
4	XJCB	Argentina: JPMorgan Real Broad Effective Exch Rate Index, CPI Based (2010=100)	stock	nsa	-1
5	HT	Argentina: G. Buenos Aires CCI: General Level (1993=100)	stock	nsa	1
6	AP	Argentina: Motor Vehicle Production (NSA, Units)	flow	nsa	1
7	APB	Argentina: Motor Vehicle Production: Automobiles (NSA, Units)	flow	nsa	1
8	IBD	Argentina: Goods Trade Balance (SA, Mil.US$)	flow	sa	1
9	IXD	Argentina: Goods Exports (SA, Mil.US$)	flow	sa	1
10	IMD	Argentina: Goods Imports (SA, Mil.US$)	flow	sa	-1
11	VLD	Argentina: Leading Indicator Index (2004=100)	stock	nsa	1
12	VCC	Argentina: National Consumer Confidence Index (SA, 50+=Growth)	stock	sa	1
13	GVI	Argentina: Economic Activity Indicator (SA, 2004=100)	stock	sa	1
14	GFMI	Argentina: Economic Activity: Financial Intermediation (NSA, 2004=100)	stock	nsa	1
15	GPRI	Argentina: Econ Act: Real Estate, Business & Rental Activities (NSA, 2004=100)	stock	nsa	1
16	GGI	Argentina: Econ Act: Public Admin & Defense, SocSec Plans (NSA, 2004=100)	stock	nsa	1
17	GDI	Argentina: Economic Activity: Education (NSA, 2004=100)	stock	nsa	1
18	GHSI	Argentina: Economic Activity: Social & Health Services (NSA, 2004=100)	stock	nsa	1
19	GSOI	Argentina: EconAct: Other Community, Social & Personal Srvc Act (NSA, 2004=100)	stock	nsa	1
20	HSCA	Argentina: Synthetic Indicator of Construction Activity (SA, 2004=100)	stock	sa	1
21	IFFB	Argentina: IP: FIEL: Food and Beverages (SA, 1993=100)	stock	sa	1
22	IFTB	Argentina: IP: FIEL: Cigarettes (SA, 1993=100)	stock	sa	1
23	IFTW	Argentina: IP: FIEL: Textiles (SA, 1993=100)	stock	sa	1
24	IFPP	Argentina: IP: FIEL: Paper & Pulp (SA, 1993=100)	stock	sa	1
25	IFFU	Argentina: IP: FIEL: Fuel (SA, 1993=100)	stock	sa	1
26	IFCH	Argentina: IP: FIEL: Chemicals & Plastics (SA, 1993=100)	stock	sa	1
27	IFMM	Argentina: IP: FIEL: Nonmetallic Minerals (SA, 1993=100)	stock	sa	1
28	IFQ	Argentina: IP: FIEL: Mining (SA, 1993=100)	stock	sa	1
29	IFAU	Argentina: IP: FIEL: Automotive (SA, 1993=100)	stock	sa	1
30	DU	Argentina: Capacity Utilization (NSA, %)	stock	nsa	1

2) Selected Baseline Model

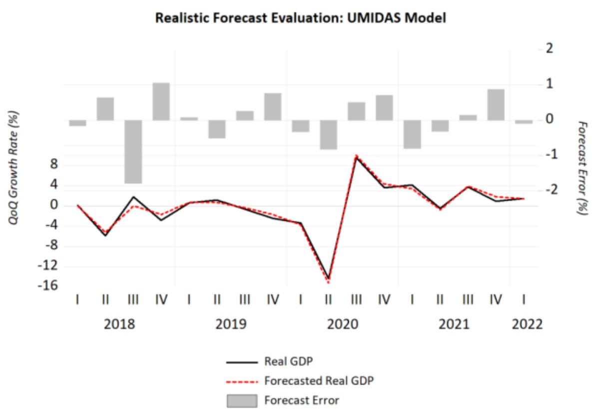
Dependent Variable: DLOG_RGDP
Method: ARMA Maximum Likelihood (BFGS)
Sample: 2004Q2 2022Q2
Included observations: 73
Convergence achieved after 14 iterations
Coefficient covariance computed using outer product of gradients

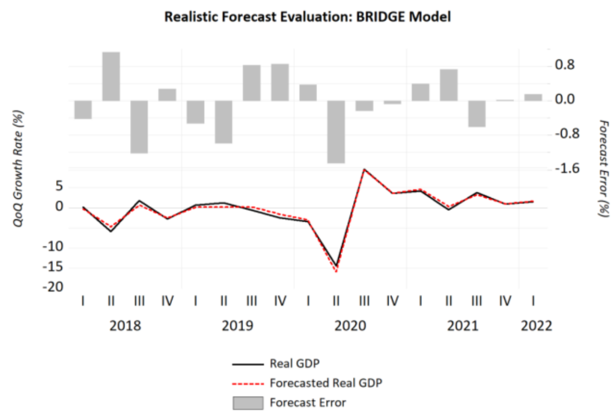
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.00E-05	0.000251	0.039982	0.9682
DLOG(GVI)	0.984779	0.012234	80.49502	0.0000
AR(1)	-0.841065	0.095844	-8.775390	0.0000
AR(2)	-0.950080	0.061174	-15.53072	0.0000
AR(3)	-0.715994	0.072175	-9.920233	0.0000
SIGMASQ	4.61E-05	8.02E-06	5.751150	0.0000

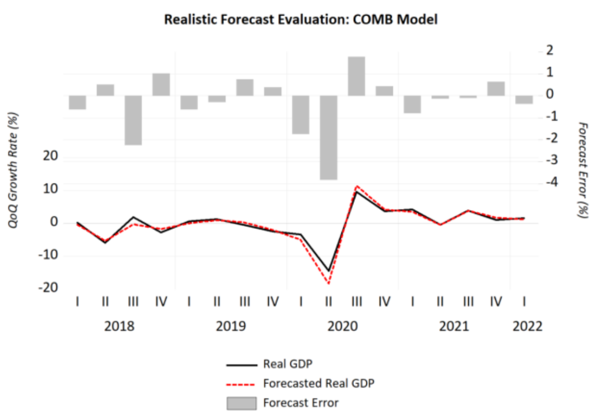
R-squared	0.953881	Mean dependent var	0.005410
Adjusted R-squared	0.950440	S.D. dependent var	0.031852
S.E. of regression	0.007091	Akaike info criterion	-6.931814
Sum squared resid	0.003369	Schwarz criterion	-6.743557
Log likelihood	259.0112	Hannan-Quinn criter.	-6.856790
F-statistic	277.1555	Durbin-Watson stat	2.428392
Prob(F-statistic)	0.000000		

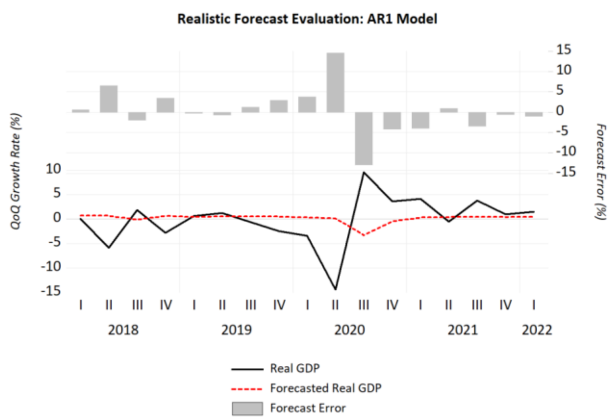
Note: **GVI** is Argentina: Economic Activity Indicator (SA, 2004=100).

3) *Realistic Forecast Evaluation (out-of-sample): Nowcasting Model vs. Benchmarks*









2. Australia:

1) Data

No.	Series_name	Description	Frequency Conversion	Seasonal Adjustment	Coefficient Sign
1	NRTAR	Australia: Exchange Rate (Avg, US$/Australian$)	stock	nsa	-1
2	NXUSV	Australia: Exchange Rate (Avg, A$/Euro)	stock	nsa	1
3	NXEUV	Australia: JPMorgan Real Broad Effective Exchnge Rate Index, PPI Based(2010=100)	stock	nsa	-1
4	NXJRB	Australia: JPMorgan Real Broad Effective Exch Rate Index, CPI Based (2010=100)	stock	nsa	-1
5	NXJCB	Australia: Stock Price Index: All Ordinaries (EOP, Jan-1-80=500)	stock	nsa	1
6	SFM1	Australia: Labor Force: Unemployment Rate (SA, %)	stock	sa	1
7	SFM3	Australia: Underemployment (NSA, Thous)	stock	nsa	1
8	NFKAO	Australia: New Motor Vehicle Sales (SA, Units)	stock	sa	1
9	SHP	Australia: Official Reserve Assets (EOP, NSA, Mil.US$)	stock	nsa	1
10	SELUR	Australia: Performance of Manufacturing Index (SA, 50+=Expansion)	stock	sa	1
11	NELUD	Australia: News-Based Economic Policy Uncertainty Index (Mean=100)	stock	nsa	1
12	STRS	AU: New Home Sales: Detached Houses (SA, Units)	stock	sa	1
13	STRA	Australia: Average Rainfall (NSA, mm)	stock	nsa	1
14	HIB	Australia: Money Supply: M1 (EOP, SA, Bil.A$)	stock	sa	1
15	HIM	Australia: Money Supply: M3 (EOP, SA, Bil.A$)	stock	sa	1
16	NLFRG	Australia: Dwelling Units Approved (SA, Units)	stock	sa	1
17	SVCC	Australia: Retail Turnover (SA, Mil.A$)	stock	sa	1
18	SVM	Australia: Retail Turnover: Food (SA, Mil.A$)	stock	sa	1
19	NVKCO	Australia: Retail Turnover: Clothing, Footwear & Personal Accessory (SA, Mil.A$)	stock	sa	1
20	NVIUC	Australia: Trade Balance in Goods (SA, Mil.A$)	stock	sa	1
21	NTA	Australia: Imports of Goods, fob (SA, Mil.A$)	stock	nsa	-1
22	AUSHPT	Australia: Tourist Arrival (NSA, Persons)	stock	sa	1
23	AUSHPTPB	AU: Dwelling Units Approved: Private Sector (SA, Units)	stock	sa	1
24	AUSHPOT	AU: Dwelling Units Approved: Public Sector (SA, Units)	stock	sa	1
25	AUNHNS	AU: Building Approvals(SA, Mil.A$)	stock	sa	1
26	AUNAWS	Australia: Short Term Visitor Arrivals (NSA, Number)	stock	nsa	1
27	AUNTMPM	Australia: Official Cash Rate (EOP, %)	stock	sa	-1
28	AUNRAIN	Australia: Westpac-Melbourne Inst Consumer Sentiment Index (SA, 100+=Favorable)	stock	sa	1
29	CVRT	Australia: New Motor Vehicle Sales: Passenger Vehicles (SA, Units)	stock	sa	1
30	VPKI	Australia: Performance of Construction Index (SA, 50+=Expansion)	stock	sa	1

2) Selected Baseline Model

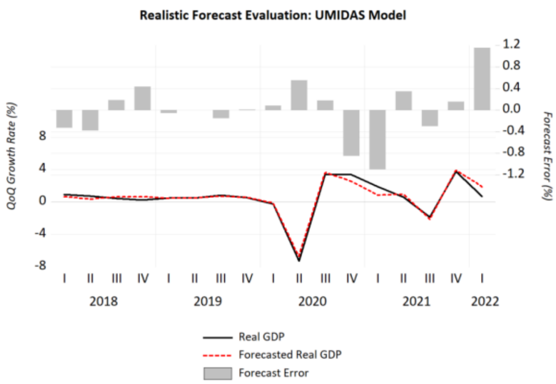
Dependent Variable: DLOG_RGDP
Method: Least Squares
Sample (adjusted): 2000Q2 2022Q2
Included observations: 89 after adjustments

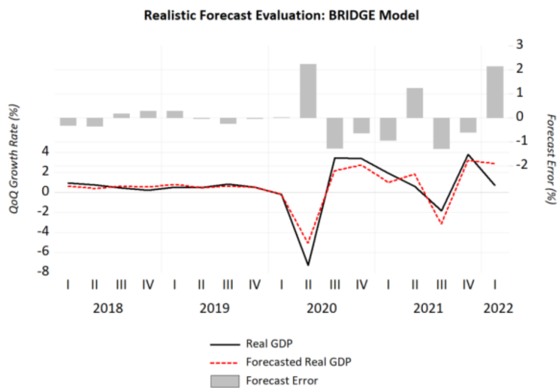
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.006215	0.000560	11.09542	0.0000
DLOG(AUSHPT)	0.009057	0.001341	6.754041	0.0000
DLOG(NVKCO)	0.075749	0.008676	8.731061	0.0000
DLOG(CVRT)	0.032111	0.009224	3.481380	0.0008

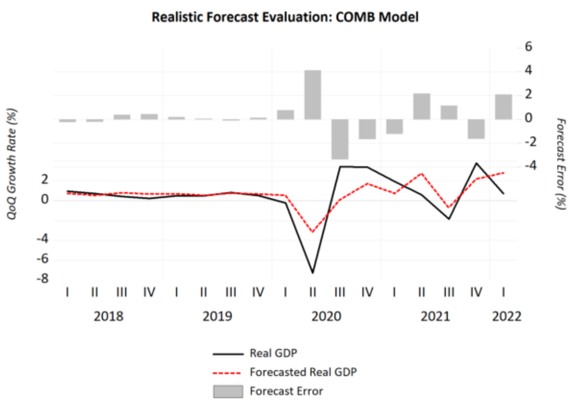
R-squared	0.800186	Mean dependent var	0.006804
Adjusted R-squared	0.793134	S.D. dependent var	0.011223
S.E. of regression	0.005105	Akaike info criterion	-7.673413
Sum squared resid	0.002215	Schwarz criterion	-7.561564
Log likelihood	345.4669	Hannan-Quinn criter.	-7.628330
F-statistic	113.4654	Durbin-Watson stat	2.097271
Prob(F-statistic)	0.000000		

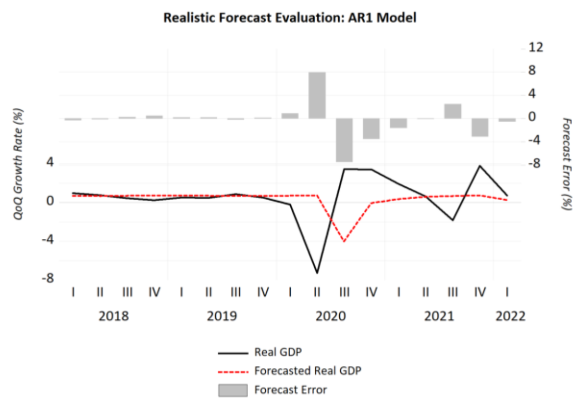
Note: **AUSHPT** is Australia: Tourist Arrival (NSA, Persons); **NVKCO** is Australia: Retail Turnover: Clothing, Footwear & Personal Accessory (SA, Mil. A$); **CVRT** is Australia: New Motor Vehicle Sales: Passenger Vehicles (SA, Units)

3) Realistic Forecast Evaluation (out-of-sample): Nowcasting Model vs. Benchmarks









3. South Africa:

1) Data

No.	Series_name	Description	Frequency Conversion	Seasonal Adjustment	Coefficient Sign
1	RR	South Africa: Interest Rates: Average Repo Rate (%)	stock	sa	-1
2	RLV	South Africa: Prime Lending Rate, Predominant Rate (Avg, %)	stock	sa	-1
3	XJRB	South Africa: JPMorgan Real Broad Effective Exch Rate Index, PPI Based(2010=100)	stock	nsa	-1
4	XJCB	South Africa: JPMorgan Real Broad Effective Exch Rate Index, CPI Based(2010=100)	stock	nsa	-1
5	KMDV	South Africa: MSCI Share Price Index, US$ (AVG, DEC-31-92=100)	stock	nsa	1
6	FVX	South Africa: Volatility Index (AVG, Index)	stock	nsa	-1
7	LFGR	South Africa: Gross Gold and Other Foreign Reserves (EOP, NSA, Mil.Rand)	stock	nsa	1
8	CVL	South Africa: Domestic Vehicle Sales (NSA, Units)	flow	nsa	1
9	CVLI	South Africa: New Vehicles Sold (SA, 2019=100)	flow	sa	1
10	VP	South Africa: Manufacturing PMI (SA, 50+=expansion)	stock	sa	1
11	HAC	S.Africa: Building Plans Passed (SA, Thous.2015.Rand)	flow	sa	1
12	DM	South Africa: Manufacturing Production: Volume (SA, 2019=100)	flow	sa	1
13	DV	South Africa: Electricity Production (SA, 2019=100)	flow	sa	1
14	DN	South Africa: Mining Production (SA, 2019=100)	flow	sa	1
15	VMDG	South Africa: Volume of Production: Gold Mining (SA, 2019=100)	flow	sa	1
16	VMDM	South Africa: Volume of Production: Manufacturing (SA, 2019=100)	flow	sa	1
17	CE	South Africa: Electricity Consumed (SA, Gigawatt Hours)	flow	sa	1
18	TRS	South Africa: Retail Sales: Current Prices (SA, Mil.Rand)	flow	sa	1
19	H199SRG	South Africa: Retail Sales: General Dealers (SA, Mil.Rand)	flow	sa	1
20	H199RFN	South Africa: Retail Sales: Food/Beverages/Tobacco in Spec Stores (SA, Mil.Rand)	flow	sa	1
21	H199TR2	South Africa: Ret Sales: Textiles/Clothing/Footwear/Leather Goods (SA, Mil.Rand)	flow	sa	1
22	H199TR3	South Africa: Retail Sales: HH Furniture/Appliance/Equipment (SA, Mil.Rand)	flow	sa	1
23	H199SUO	South Africa: Retail Sales: Hardware/Paint/Glass Retailers (SA, Mil.Rand)	flow	sa	1
24	H199SRO	South Africa: Retail Sales: All Other Retailers (SA, Mil.Rand)	flow	sa	1
25	VLD	South Africa: Business Cycles: Composite Leading Indicator (SA, 2015=100)	stock	sa	1
26	RC	South Africa: Interest Rates: Interbank Call Money Rate (%)	stock	sa	-1
27	VLC	South Africa: Business Cycles: Coincident Indicator (SA, 2015=100)	stock	sa	1
28	VLAG	South Africa: Business Cycles: Lagging Indicator (SA, 2015=100)	stock	sa	1
29	TA	South Africa: Tourist Arrivals (NSA, Persons)	flow	nsa	1
30	POGI	South Africa: Fuel Prices: Inland Leaded Replacement Petrol 93 (RSA c/litre)	stock	nsa	1

2) Selected Baseline Model

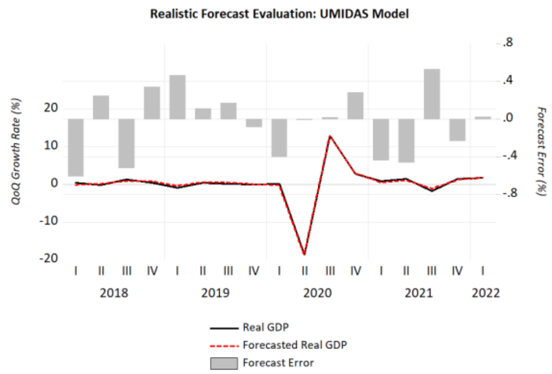
Dependent Variable: DLOG(RGDP)
Method: Least Squares
Sample (adjusted): 2005Q2 2022Q2
Included observations: 69 after adjustments

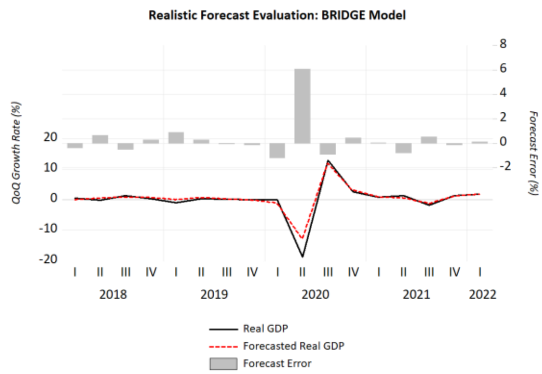
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000130	0.000847	0.153981	0.8781
DLOG(H199SRO)	0.060186	0.014486	4.154901	0.0001
DLOG(VLC)	0.237951	0.033333	7.138577	0.0000
DLOG(TRS)	0.155229	0.046086	3.368227	0.0013
DLOG(CE)	0.089775	0.033575	2.673835	0.0095

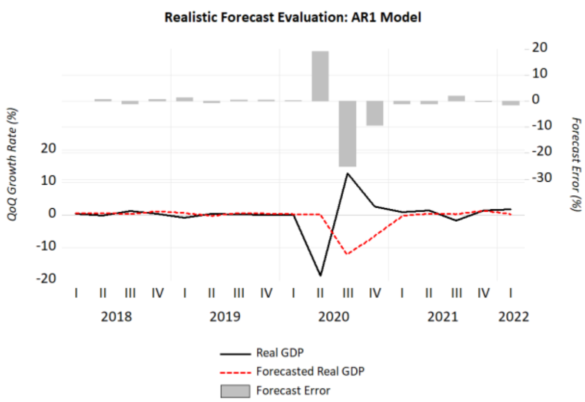
R-squared	0.980105	Mean dependent var	0.004554
Adjusted R-squared	0.978861	S.D. dependent var	0.028766
S.E. of regression	0.004182	Akaike info criterion	-8.046173
Sum squared resid	0.001120	Schwarz criterion	-7.884281
Log likelihood	282.5930	Hannan-Quinn criter.	-7.981945
F-statistic	788.2084	Durbin-Watson stat	2.247614
Prob(F-statistic)	0.000000		

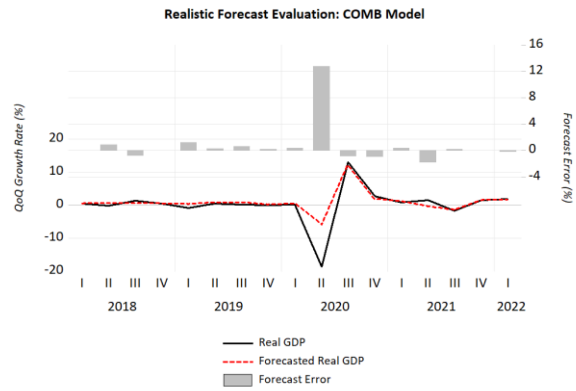
Note: **H199SRO** is South Africa: Retail Sales: All Other Retailers (SA, Mil. Rand); **VLC** is South Africa: Business Cycles: Coincident Indicator (SA, 2015=100); **TRS** is South Africa: Retail Sales: Current Prices (SA, Mil. Rand); **CE** is South Africa: Electricity Consumed (SA, Gigawatt Hours).

3) *Realistic Forecast Evaluation (out-of-sample): Nowcasting Model vs. Benchmarks*









4. United Kingdom:

1) Data

No.	Series_name	Description	Frequency Conversion	Seasonal Adjustment	Coefficient Sign
1	NXUSV	U.K.: Exchange Rate (Avg, US$/Pound)	stock	nsa	-1
2	NXEUV	U.K.: Exchange Rate (Avg, Pounds/Euro)	stock	nsa	1
3	NXJRB	U.K.: JPMorgan Real Broad Effective Exchange Rate Index, PPI Based (2010=100)	stock	nsa	-1
4	NXJCB	UK: JPMorgan Real Broad Effective Exchange Rate Index, CPI Based (2010=100)	stock	nsa	-1
5	NFKFT	U.K.: London Stock Exchange: FTSE 100 (AVG, Jan-2-84=1000)	stock	nsa	1
6	HPHPC	U.K.: Harmonized Index of Consumer Prices [HICP] (SA, 2015=100)	stock	sa	1
7	HPPM	U.K.: PPI: Gross Output Prices: Manufactured Products (SA, 2015=100)	stock	sa	1
8	NPOIL	European Brent Spot Price FOB ($/Barrel)	stock	nsa	1
9	SHG	U.K.: Nationwide Building Society House Price Index (SA, Q1-93=100)	stock	sa	1
10	STRSC	Great Britain: Retail Sales Volume Index (SA, 2019=100)	stock	sa	1
11	STRS	Great Britain: Retail Sales Value Index (SA, 2019=100)	stock	sa	1
12	HCVRT	U.K.: New Passenger Car Registrations (SA, Units)	flow	sa	1
13	SFCB	U.K.: M4 Lending to Private Sector (SA, EOP, Mil.GBP)	flow	sa	1
14	SPFTT	UK: Terms of Trade (SA, 2010=100)	stock	sa	1
15	SD	UK: Industrial Production (SA, 2019=100)	stock	sa	1
16	SDM	UK: Industrial Production: Manufacturing (SA, 2019=100)	stock	sa	1
17	SDMF	UK: Industrial Production: Food, Beverages and Tobacco (SA, 2019=100)	stock	sa	1
18	SDMT	UK: Industrial Production: Textiles, Apparel and Leather Goods (SA, 2019=100)	stock	sa	1
19	SDMSM	UK: Industrial Production: Motor Vehicles/Trailers/Semitrailers(SA, 2019=100)	stock	sa	1
20	SDN	UK: Industrial Production: Mining & Quarrying (SA, 2019=100)	stock	sa	1
21	SDV	UK: IP: Electricity, Gas, Steam & Air Conditioning Supply (SA, 2019=100)	stock	sa	1
22	SDVW	UK: Industrial Production: Water, Waste Mgmt, Remediation(SA, 2019=100)	stock	sa	1
23	HOAU	U.K.: Motor Vehicle Production: Cars (SA, Units)	flow	sa	1
24	HOTU	U.K.: Motor Vehicle Production: Commercial Vehicles (SA, Units)	flow	sa	1
25	SEWA	Great Britain: Average Weekly Earnings [incl Bonus]: Whole Economy (SA, GBP)	stock	sa	1
26	HTS	U.K.: Manufacturing Turnover (SA, Mil.Pounds)	stock	sa	1
27	SIBT	U.K.: Trade Balance in Goods and Services (SA, Mil.Pounds)	flow	sa	1
28	SIXT	U.K.: Exports of Goods and Services (SA, Mil.Pounds)	flow	sa	1
29	SIMT	U.K.: Imports of Goods and Services (SA, Mil.Pounds)	stock	sa	-1
30	NRTAR	U.K.: Official Bank Rate (EOP, %)	stock	sa	-1

2) Selected Baseline Model

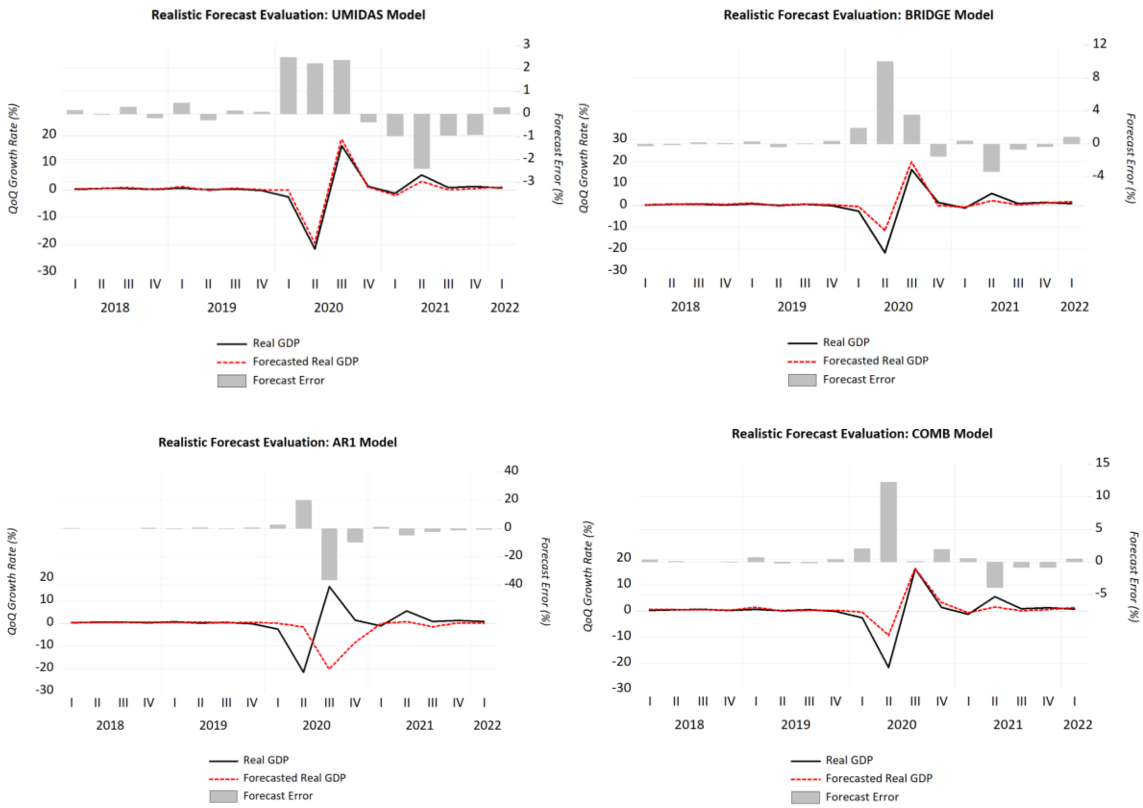
Dependent Variable: DLOG(RGDP)
Method: ARMA Maximum Likelihood (BFGS)
Sample: 2000Q2 2022Q2
Included observations: 89
Convergence achieved after 20 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.60E-05	0.000852	0.018751	0.9851
DLOG(SDM)	0.592887	0.025737	23.03623	0.0000
DLOG(HCVRT)	0.014111	0.004880	2.891404	0.0050
DLOG(STRS)	0.217246	0.023982	9.058781	0.0000
DLOG(NXUSV)	-0.065803	0.021851	-3.011516	0.0035
AR(1)	-0.407451	0.215217	-1.893208	0.0620
MA(1)	0.726362	0.202095	3.594157	0.0006
MA(2)	0.340511	0.140864	2.417300	0.0179
MA(4)	-0.378619	0.077984	-4.855080	0.0000
SIGMASQ	3.49E-05	5.89E-06	5.925658	0.0000

R-squared	0.961080	Mean dependent var	0.003807
Adjusted R-squared	0.956646	S.D. dependent var	0.030126
S.E. of regression	0.006273	Akaike info criterion	-7.186770
Sum squared resid	0.003108	Schwarz criterion	-6.907148
Log likelihood	329.8112	Hannan-Quinn criter.	-7.074062
F-statistic	216.7543	Durbin-Watson stat	1.954775
Prob(F-statistic)	0.000000		

Note: **SDM** is UK: Industrial Production: Manufacturing (SA, 2019=100); **HCVRT** is U.K.: New Passenger Car Registrations (SA, Units); **STRS** is Great Britain: Retail Sales Volume Index (SA, 2019=100); **NXUSV** is U.K.: Exchange Rate (Avg, US$/Pound).

3) *Realistic Forecast Evaluation (out-of-sample): Nowcasting Model vs. Benchmarks*



5. United States:

1) Data:

No.	Series_name	Description	Frequency Conversion	Seasonal Adjustment	Coefficient Sign
1	CEXP	University of Michigan: Consumer Expectations (NSA, Q1-66=100)	stock	nsa	1
2	USPHPIM	FHFA House Price Index: Purchase Only, United States (SA, Jan-91=100)	stock	sa	1
3	TSTH	Real Manufacturing & Trade Sales: All Industries (SA, Mil.Chn.2012$)	stock	sa	1
4	YPM	Personal Income (SAAR, Bil.$)	stock	sa	1
5	PRFA	Prices Received by Farmers: All Farm Products (2011=100)	stock	nsa	1
6	JQI	US Private Sector Job Quality Index (<100=Greater Share of Low Quality Jobs)	stock	nsa	1
7	RST	Retail Sales: Retail & Food Services (SA Mil.$)	flow	sa	1
8	NTS	Sales: Total Business (SA, Mil.$)	flow	sa	1
9	TABCA	Advance Trade Balance, Customs Value (SA, Mil.$)	flow	sa	1
10	TAXEAVA	Advance Exports: Automotive Vehicles, Parts and Engines (SA, Mil.$)	flow	sa	1
11	TAXECNA	Advance Exports: Consumer Goods (SA, Mil.$)	flow	sa	1
12	TAXEOMA	Advance Exports: Other Goods (SA, Mil.$)	flow	sa	1
13	NMONC	Manufacturers' New Orders: Nondefense Capital Goods (SA, Mil.1982$)	stock	sa	1
14	FM2C	Real Money Stock: M2 (SA, Bil.Chn.2012$)	stock	sa	1
15	YPLTPMH	Real Personal Income less Transfer Payments (SAAR, Bil.Chn.2012$)	flow	sa	1
16	FFEDTAR	Federal Open Market Committee: Fed Funds Target Rate (%)	stock	sa	-1
17	FXWSJ	Wall Street Journal Dollar Index (AVG, Jun-6-01=100)	stock	nsa	1
18	SPNA	Stock Price Index: NASDAQ Composite (Feb-5-71=100)	stock	nsa	1
19	PZDJAF	Bloomberg Commodity Index (Jan-2-91=100)	stock	nsa	1
20	PZTEXP	Spot Oil Price: West Texas Intermediate [Prior'82=Posted Price] ($/Barrel)	stock	nsa	1
21	HST	Housing Starts (SAAR, Thous.Units)	flow	sa	1
22	HPT	New Pvt Housing Units Authorized by Building Permit (SAAR, Thous.Units)	flow	sa	1
23	CUTC	Capacity: Industry (SA, Percent of 2017 Output)	stock	sa	1
24	PCOAL	Production: U.S. Coal Production (NSA, Short Tons)	flow	nsa	1
25	NRST	Retail Sales & Food Services (SA, Mil.$)	flow	sa	1
26	NAPMVDI	ISM Mfg: Supplier Deliveries Index (SA, 50+ = Slower)	stock	sa	1
27	HPT	New Pvt Housing Units Authorized by Building Permit (SAAR, Thous.Units)	stock	sa	1
28	SP500	Stock Price Index: Standard & Poor's 500 Composite (1941-43=10)	stock	nsa	1
29	LANAGRA	All Employees: Total Nonfarm (SA, Thous)	stock	sa	1
30	IP	Industrial Production Index (SA, 2017=100)	stock	sa	1

2) Selected Baseline Model

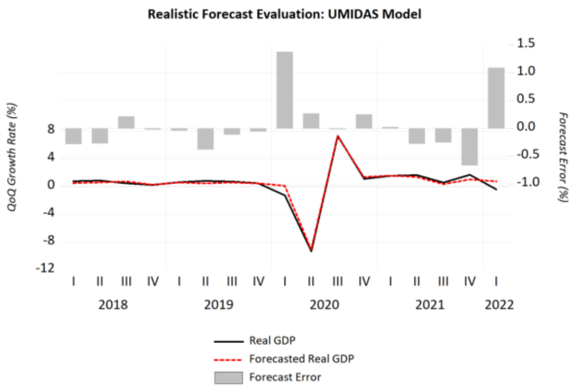
Dependent Variable: DLOG_RGDP
Method: Least Squares
Sample (adjusted): 2000Q2 2022Q2
Included observations: 89 after adjustments

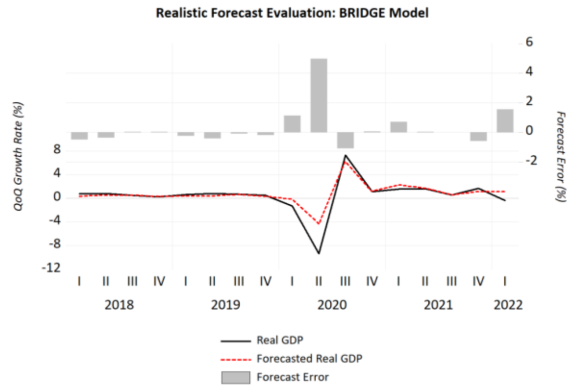
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.002219	0.000522	4.247760	0.0001
DLOG(LANAGRA)	0.499324	0.043424	11.49887	0.0000
DLOG(TSTH)	0.278859	0.050729	5.497072	0.0000
DLOG(NRST)	0.080174	0.034608	2.316639	0.0230
DLOG(CEXP)	0.016958	0.006155	2.755139	0.0072

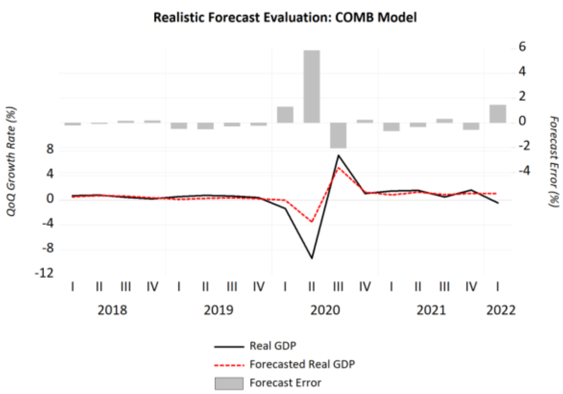
R-squared	0.911379	Mean dependent var	0.004726
Adjusted R-squared	0.907159	S.D. dependent var	0.014187
S.E. of regression	0.004323	Akaike info criterion	-7.995258
Sum squared resid	0.001570	Schwarz criterion	-7.855447
Log likelihood	360.7890	Hannan-Quinn criter.	-7.938904
F-statistic	215.9652	Durbin-Watson stat	1.981167
Prob(F-statistic)	0.000000		

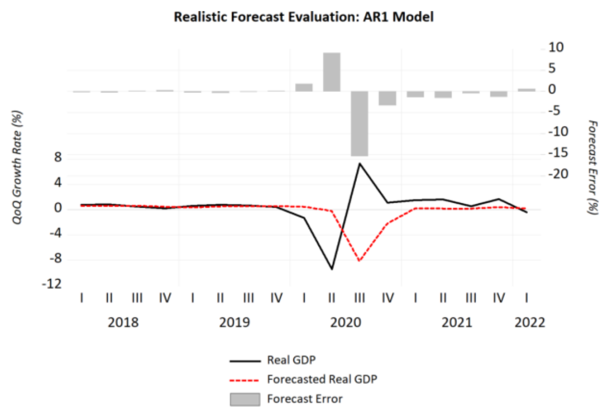
Note: **LANAGRA** is All Employees: Total Nonfarm (SA, Thous)**; TSTH** is Real Manufacturing & Trade Sales: All Industries (SA, Mil.Chn.2012$)**; NRST** is Retail Sales & Food Services (SA, Mil.$); **CEXP** is University of Michigan: Consumer Expectations (NSA, Q1-66=100)

4) Realistic Forecast Evaluation (out-of-sample): Nowcasting Model vs. Benchmarks









Bibliography

Armesto, M. T., Engemann, K. M., & Owyan, M. T. (2010). Forecasting with Mixed Frequencies. *Federal Reserve Bank of St. Louis Review*, 521-36.

Baffigia, A., Golinellib, R., & Parigia, G. (2004). Bridge Models to Forecast the euro area GDP. *International Journal of Forecasting*, Volume 20, Issue 3, July–September 2004, Pages 447-460.

Bańbura, M., Giannone, D., Modugno, M., & Reichlin, L. (2013). Nowcasting and the Real-Time Data Flow. *European Central Bank Working Paper*, No 1564.

Bok, B., Caratelli, D., Giannone, D., Sbordone, A., & Tambalotti, A. (2017). *Macroeconomic Nowcasting and Forecasting with Big Data.* New York City: Federal Reserve Bank of New York.

Bragoli, D., & Fosten, J. (2017). Nowcasting Indian GDP. *Oxford Bulletin of Economics and Statistics*, 260-282.

Breitung, J., & Hafner, C. (2016). A Simple Model for Now-casting Volatility Series. *LIDAM Discussion Papers ISBA 2016035 Université catholique de Louvain, Institute of Statistics, Biostatistics and Actuarial Sciences (ISBA).*

C.W.J.Granger, & P.Newbold. (1974). Spurious Regressions in Econometrics. *Journal of Econometrics*, Volume 2, Issue 2, July 1974, Pages 111-120.

Camacho, M., & Perez-Quiros, G. (2010). Introducing the Euro-sting: Short-term Indicator of Euro Area Growth. *Journal of Applied Econometrics*, Volume 25, Issue 4.

Chipman, J. S. (2014). Gauss-Markov Theorem. *International Encyclopedia of Statistical Science*, pp 577-582.

Chowdhury, M. Z., & Turin, T. C. (2020). Variable Selection Strategies and its Iimportance in Clinical Prediction Modelling. *Family Medicine and Community Health*, e000262. doi:10.1136/.

Cimadomo J, G. D. (2020). Nowcasting with Large Bayesian Vector Autoregressions. *Working Paper Series No. 2453. European Central Bank*.

Claudia, F., Massimiliano, M., & Christian, S. (2011). U-MIDAS: MIDAS regressions with Unrestricted Lag Polynomials. *Discussion Paper Series 1: Economic Studies*, 2011,35, Deutsche Bundesbank.

Duarte, P., & Sussmuth, B. (2014). Robust Implementation of a Parsimonious Dynamic Factor Model to Nowcast GDP. *CESifo Working Paper*, No. 4574.

Rueben Ellul & Germano Ruisi. (2022). "Nowcasting the Maltese economy with a dynamic factor model," *CBM Working Papers WP/02/2022, Central Bank of Malta*.

Foroni, C., & Marcellino, M. (2013). A Survey of Econometric Methods for Mixed Frequency Data . *Norges Bank Research Working Paper*.

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

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

Giannone, D., Agrippino, S. M., & Modugno, M. (2013). Nowcasting China Real GDP. *CIRANO*.

Giannone, D., Reichlin, L., & Small, D. (2005). *Nowcasting GDP and Inflation: The Real-Time Informational Contest of Macroeconomic Data Releases.* Washington, DC: Federal Reserve Board.

Higgins, P. (2014). GDPNow: A Model for GDP "Nowcasting". *Federal Reserve Bank of Altanta Working Paper Series*, 2014 - 7.

Hopp, D. (2022). Benchmarking Econometric and Machine Learning Methodologies in Nowcasting. *UNCTAD Research Paper No. 83*.

Hyndman, R. J., & Khandakar, Y. (2008). Automatic Time Series Forecasting: The Forecast Package for R. *Journal of Statistical Software*, 27(3), 1–22. https://doi.org/10.18637/jss.v027.i03.

IHS Markit - EViews. (2020). EViews 12 User's Guide I: Chapter 11. Series. Seal Beach, CA.

IHS Markit - EViews. (2022). *EViews 12 User's Guide II: Chapter 22. Regression Variable Selection.*

Ingenito, R., & Trehan, B. (1996). Using Monthly Data to Predict Quarterly Output. *Federal Reserve Bank of San Francisco Economic Review 3*, 3-11.

Lewis, D. J., Mertens, K., Stock, J. H., & Trivedi, M. (2020). *Measuring Real Activity Using a Weekly Economic Index.* New York City: Federal Reserve Bank of New York.

MathWorks. (2022). Retrieved from Seasonal Adjustment: https://www.mathworks.com/help/econ/seasonal-adjustment-1.html#:~:text=Deseasonalized%20data%20is%20useful%20for,data%20for%20future%20modeling%20purposes.

Preis, T., & Moat, H. S. (2014). Adaptive Nowcasting of Influenza Outbreaks using Google Searches. *R Soc Open Sci*, doi: 10.1098/rsos.140095.

Smith, G. (2018). Step Away from Stepwise. *Journal of Big Data*, 5, 32.

Stock, J. a. (2002). Macroeconomic Forecasting Using Diffusion Indexes. *Journal of Business and Economic Statistics*, 20(2), I47-I62.

Stock, J. a. (2002a). Forecasting using Principal Components from a Large Number of Predictors. *Journal of the American Statistical Association*, 97, II67-II79.

United States Census Bureau. (2022, July 11). Retrieved from X-13ARIMA-SEATS Seasonal Adjustment Program: https://www.census.gov/data/software/x13as.html

Wallis, K. F. (1986). Forecasting with an econometric model: The ‘Ragged Edge’ Problem. *Journal of Forecasting*, 1-13.

PUBLICATIONS

Identifying Optimal Indicators and Lag Terms for Nowcasting Models

Working Paper No. WP/23/45