Journal of Quantitative Economics (2023) 21:213–234
https://doi.org/10.1007/s40953-022-00335-6

ORIGINAL ARTICLE

Nowcasting India's Quarterly GDP Growth: A Factor-Augmented Time-Varying Coefficient Regression Model (FA-TVCRM)

Rudrani Bhattacharya[1] · Bornali Bhandari[2] · Sudipto Mundle[3]

Accepted: 7 December 2022 / Published online: 11 January 2023

Abstract
Governments, central banks, private firms and others need high frequency information on the state of the economy for their decision making. However, a key indicator like GDP is only available quarterly and that too with a lag. Hence decision makers use high frequency daily, weekly or monthly information to project GDP growth in a given quarter. This method, known as nowcasting, started out in advanced country central banks using bridge models. Nowcasting is now based on more advanced techniques, mostly dynamic factor models. In this paper we use a novel approach, a Factor Augmented Time Varying Coefficient Regression (FA-TVCR) model, which allows us to extract information from a large number of high frequency indicators and at the same time inherently addresses the issue of frequent structural breaks encountered in Indian GDP growth. One specification of the FA-TVCR model is estimated using 19 variables available for a long period starting in 2007–08:Q1. Another specification estimates the model using a larger set of 28 indicators available for a shorter period starting in 2015–16:Q1. Comparing our model with two alternative models, we find that the FA-TVCR model outperforms a Dynamic Factor Model (DFM) model and a univariate Autoregressive Integrated Moving Average (ARIMA) model in terms of both in-sample and out-of-sample Root Mean Square Error (RMSE). Further, comparing the predictive power of the three models using the Diebold-Mariano test, we find that FA-TVCR model outperforms DFM consistently. In terms of out-of-sample forecast accuracy both the FA-TVCR model and the ARIMA model have the same predictive accuracy under normal conditions. However, the FA-TVCR model outperforms the ARIMA model when applied for nowcasting in periods of major shocks like the Covid–19 shock of 2020–21.

Keywords Nowcasting · Quarterly year-on-year GDP growth · State-space model · India

✉ Rudrani Bhattacharya
rudrani.bhattacharya@nipfp.org.in

Extended author information available on the last page of the article

JEL Classification C52 · C53 · O40

Introduction

Governments, central banks, private firms and others need high frequency data on the state of the economy for their decision making. However, data on Gross Domestic Product (GDP) growth, a key indicator of the state of the economy, is typically available only on a quarterly basis and that too with a lag. In India, for example, the quarterly GDP estimate is made available with a lag of two months. Consequently, decision makers use high frequency monthly, weekly or daily information to project GDP growth for a given quarter. This method of gauging the present state of the economy using information from high frequency indicators is known as 'nowcasting.' In this paper, we propose to employ a Factor Augmented Time Varying Coefficient Regression Model (FA-TVCRM) to nowcast quarterly year-on-year (y–o–y) growth in India.

Nowcasting was first introduced by central banks in the advanced economies from around 2000 onwards. The approach initially adopted was the Bridge Model (BM) where quarterly frequency national accounts variables were regressed on their lagged values and other high frequency indicators, converted to quarterly frequency (Baffigi et al. 2004). Subsequently, more advanced techniques, mainly Dynamic Factor Models (DFMs) were developed (Giannone et al. 2008; Banbura et al., 2010). In this class of models, quarter-on-quarter GDP growth and month-on-month growth of a large set of monthly indicators are assumed to be driven by a set of unobserved factors which follow a Vector Auto Regression (VAR) structure among themselves. DFMs have been successfully implemented to nowcast GDP growth in Euro Area (Giannone et al. 2008; Banbura et al., 2010), Japan (Urasawa 2014) and Canada (Chernis and Sekkel, 2017).

Nowcasting GDP growth has always been a challenge in emerging market economies because of the limitations of data availability, irregular release of high frequency indicators and frequent structural breaks in the data. Still, DFMs have been found to work satisfactorily for nowcasting GDP growth in countries like Brazil, Indonesia, Mexico, South Africa, Turkey, (Cepni et al., 2019; Cepni et al., 2020, Luciani et al., 2017). In India, the picture is mixed. While Bhattacharya et.al (2011) found that a bridge regression model performed better than a DFM, Bragoli and Fosten (2018) found that their DFM outperformed a bridge model. Bhadury et al. (2019) and Iyer & Sen Gupta (2020) also found DFMs to perform better compared to Random Walk and Auto Regressive Models.

In this paper, we have adopted a novel approach to address the frequent structural breaks in Indian GDP.[1] We employ a Factor Augmented Time Varying Coefficient

[1] Basu (2020) found four structural breaks in post-Independence India—1964–65, 1978, 1990–91 and in 2004–05. Much of the Indian empirical literature has examined structural breaks for India pre–2011–12 (the Great Financial Recession). However, Kar & Sen (2016) and Subramanian & Felman (2019) amongst others have presented evidence of a sharp economic slowdown post 2011–12.

Regression (FA-TVCR) model following Bhattacharya et al. (2019) and apply it using a large number of high frequency indicators to nowcast quarterly GDP growth. This model allows us to extract information from a large number of indicators and also inherently addresses the issue of frequent structural breaks in GDP growth. Comparing our model with two alternative models, we find that the FA-TVCR model outperforms a Dynamic Factor Model (DFM) and a univariate Autoregressive Integrated Moving Average (ARIMA) model in terms of both in-sample and out-of-sample Root Mean Square Error (RMSE). Further, comparing the predictive power of the three models using the Diebold-Mariano test, we find that FA-TVCR model outperforms DFM consistently. In terms of out-of-sample forecast accuracy, both the FA-TVCR model and the ARIMA model have the same predictive accuracy under normal conditions. However, the FA-TVCR model outperforms the ARIMA model when applied for nowcasting in periods of major shocks like the Covid-19 shock of 2020–21.

The rest of the paper is organised as follows: Sect. 2 details the model. Sect. "Model" describes the indicators that have been used. It should be pointed out that while monthly time series data is available for some indicators from 2004 to 05 onwards, some monthly indicators for some additional variables are available for shorter periods, including a few since 2014–15. Accordingly, the model has been estimated separately for two separate periods: Specification I estimates the model for the period 2007–08:Q1 to 2019–20:Q3 using only 19 indicators. Specification II for the period 2015–16:Q1 to 2019–20:Q3 includes a larger set of 28 indicators, i.e., those available since 2004–05 plus those that are available from 2014 to 2015 onwards. Sect. "The Data" reports the estimation results. Sect. "Model estimation" and "Performance of the models for the period including Covid-19 pandemic" discuss the forecast performance of the model for the pre-pandemic period Jan–Mar, 2019 to Oct–Dec, 2019, and the pandemic period from Jan–Mar, 2020 to Jan–Mar, 2021 respectively. Finally, Sect. "Concluding remarks" concludes the paper.

Model

Nowcasting of quarterly y–o-y GDP growth is essentially predicting the GDP growth for the quarter Q_t, using information from high frequency indicators (we use monthly indicators for our analysis) spanning that particular quarter Q_t. The estimation process consists of the following steps:

(i) Depending on the flow of information for the set of monthly indicators for months i, where i = 1,2,3 span quarter Q_t, the nowcasting is conducted for "2 months ahead", "1 month ahead" and "zero month ahead" of GDP data released by the statistical agency of the country. Since high frequency indicators are released with different lags on different dates in a month, addressing the "ragged-edge data" problem at the end of the sample period is a major challenge in the nowcasting methodology. Converting monthly indicators into quarterly frequency by forecasting the observation/s unavailable for the month/s in a particular quarter is a commonly used method of handling the ragged-edge data problem.[2]

(ii) When a monthly indicator Y_i is available till month i = 1 in quarter Q_t, we forecast the values for i = 2 and 3 in quarter t using Seasonal ARIMA (p,d,q)(P, D, Q) model:

$$\phi(L)\Phi(L^s)(1-L)^d(1-L^s)^D Y_i = \theta(L)\theta(L^s)\in_i \tag{2.1}$$

where L is the lag operator ($LY_i = Y_{i-1}$); s is the seasonal period and hence s = 12 for monthly data; $\phi(L) = 1 - \phi_1 L\text{-}\phi_2 L^2 - \dots - \phi_p L^p$ is the non-seasonal autoregressive (AR) operator; $\Phi(L) = 1 - \Phi_1 L\text{-}\Phi_2 L^{2s} - \dots - \Phi_P L^{Ps}$ is the seasonal AR operator; $\theta(L) = 1 - \theta_1 L\text{-}\theta_2 L^2 - \dots - \theta_q L^q$ is the non-seasonal moving average (MA) operator; and $\theta(L) = 1 - \theta_1 L\text{-}\theta_2 L^2 - \dots - \theta_Q L^Q$ is the seasonal MA operator. Here D stands for the number of differencing required to remove seasonal unit root. Similarly, d represents the number of differencing required to remove the non-seasonal unit root. Here $\in_i$ is the *i.i.d* error with zero mean and variance σ^2. The Seasonal ARIMA structure is optimally chosen using X13-SEATS Seasonal Adjustment Programme of U.S. Census Bureau. When a monthly indicator Y_i is available till month i = 1 and 2 in quarter Q_t, we forecast the values for i = 3 in quarter Q_t using the same method. Forecasting is not conducted when information on a monthly indicator is available for all three months spanning quarter $Q_{t.}$

(iii) Once information for all the three months spanning the quarter Q_t are obtained, the monthly series is converted to quarterly frequency. The quarterly y–o-y growth of the indicator is then derived.

(iv) Assuming that a set of unobserved factors determines performance of the observed monthly economic indicators, the static factors are estimated from y–o-y growth in the set of monthly indicators converted to quarterly frequency using Principal Component Analysis (Stock and Watson, 2002). The first k numbers of factors that explain at least 80% of the variation in the data are chosen. The weighted sum of estimated factors provides a single composite indicator where weights are the shares of variance of each factor in total variation.

(v) Next, we regress quarterly y–o-y growth in GDP available till quarter Q_{t-1} on the k number of factors till quarter Q_{t-1} and one period lagged GDP growth where the regression coefficients are assumed to vary over time. Finally, using the estimated coefficient and the values of k factors obtained for the quarter Q_t from the set of monthly indicators, the nowcast of the GDP growth for Q_t is obtained.

The details of the regression model are as follows:
Measurement equation:

[2] Since this method attaches equal weights to all the monthly observations in a quarter, more complex weighting schemes, widely known as "Mixed Data Sampling (MIDAS)" method (Marcellino and Schumacher, 2010; Forni and Marcellino, 2014) have been applied to both regression and DFM structure. We are unable to apply this method for nowcasting Indian GDP growth because of the paucity of data.

$$y_t = X_t' \beta_t + \in_t \tag{2.2}$$

where X_t s is a $(k+1 \times 1)$ vector consisting of k number of chosen factors F_t and one quarter lagged GDP growth.

Transition equation.

$$(\beta_{t+1} - \beta)\mathbf{G}(\beta_t \ - \ \beta) + \mathbf{v}_{t+1} \tag{2.3}$$

If the eigen values of the (k + 1 × k + 1) matrix G are all inside the unit circle, then β is interpreted as the average or steady-state value for the coefficient vector.

Assuming that,

$$\begin{pmatrix} V_{t+1} \\ \in_t \end{pmatrix} | X_{t,Z_{t-1}} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} Q & 0 \\ 0 & \sigma^2 \end{pmatrix}\right] \tag{2.4}$$

where $z_{t-1} \equiv (y'_{t-1}, y'_{t-2}, \ldots\ldots, y'_1, X'_{t-1}, X'_{t-2,} \ldots\ldots\ldots.., X'_1).'$

Here the regression coefficients β are not unknown constants but latent, stochastic variables that follow random walks, estimated by Kalman Filter (Hamilton 1994; Kim and Nelson 1999). Equations (2.2), (2.3) and (2.4) represent the state-space form of the time-varying parameter model, with state vector $\mathbf{s}_t = \beta_t - \bar{\beta}$.

The measurement equation can then be re-written as.

$$y_t = X_t' \chi + X' s_t + \in_t \tag{2.5}$$

which is an observation equation with $\mathbf{a}(\mathbf{X_t}) = X_t' \overline{\beta}$, $\mathbf{H}(\mathbf{X_t}) = \mathbf{X_t}$, and $\mathbf{R}(\mathbf{X_t}) = \sigma^2$. These values are then used in the following Kalman Filter iterations (see Hamilton(1994) for details):

$$\hat{s}_{t|t} = \hat{s}_{t|t-1} + \left\{P_{t|t-1} H(X_t) \left[H(X_t)'\right] P_{t|t-1} H(X_t) + R(X_t)\right]^{-1} \times [y_t - a(X_t) - H(X_t)'] \hat{s}_{t|t-1} \right\} \tag{2.6}$$

$$P_{t|t} = P_{t|t-1} - \left\{P_{t|t-1} H(X_t)\right\} \times \left[H(X_t)'\right] P_{t|t-1} H(X_t) + R(F_t)]^{-1} H(X_t)' P_{t|t-1} \quad \} \tag{2.7}$$

$$s_{t+1} | X, z_{t-1} \sim N(\hat{s}_{t+1|t}, P_{t+1|t}) \tag{2.8}$$

$$\hat{s}_{t+1|t} = G\hat{s}_{t|t} \tag{2.9}$$

$$P_{t+1|t} = GP_{t|t} G' + Q \tag{2.10}$$

The Data

The target variable in our analysis is the quarterly y–o-y growth rate of aggregate GDP in India. The GDP data are sourced from the Central Statistical Organisation, Ministry of Statistics and Programme Implementation (CSO, MOSPI) for the period

2004–05: Q1, to 2020–21:Q4.[3] In its quarterly GDP estimates, MoSPI regularly publishes the indicators which are used to estimate it. We have used the same set of indicators plus some additional indicators, mostly drawn from the Centre for Monitoring Indian Economy (CMIE). The high frequency dataset consists of 29 monthly indicators which have been listed in Appendix A.

The monthly indicators are taken for the period April, 2004 to February, 2021. The data sources, along with their date of periodic release are given in Appendix A, Table 8.

We test for stationarity of quarterly y–o–y growth rates of the high frequency indicators using Augmented Dickey-Fuller (ADF) test and Phillips-Perron (PP) test with the null hypothesis of presence of unit root in the series. We also employ Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test with the null hypothesis that the series is stationary around a constant or a deterministic trend against the alternative hypothesis that the series contains unit root. All variables are found to be stationary by either one or both the tests, except for CPI inflation and growth in cargo movement by air.[4]

Model Estimation

Since India experienced a massive contraction shock in the 2020–21 due to the Covid–19 pandemic, we first estimate our model till the period ending at Oct–Dec, 2019, i.e., before the outbreak of the pandemic, and evaluate the model performance till that period., We then assess the performance of the model for the pandemic period i.e., Jan–Mar, 2020 to Jan–Mar, 2021.

As a first step, we apply static factor analysis to summarise the information about the performance of the economy from quarterly y–o–y growth rates of monthly indicators converted to quarterly frequency. The full sample or long period analysis in Specification I includes 19 high frequency indicators.

The factors extracting information from these 19 indicators are then estimated using Maximum Likelihood Method. The number of factors estimated are 3.[5] Table 1 reports factor loadings, i.e., the correlation of each of the indicators with the estimated latent factors. The factor loadings give the variance explained by the data associated with each factor. As a rule of thumb, in our analysis a factor loading with value 0.6 or more is taken to indicate that the factor extracts sufficient variation from that variable. Table 1 suggests that the factor F1 extracts variations from growth in aggregate deposits, food and non-food credit, exports of goods and GST revenue.

The factor F2 extracts variations from the growth in cargo handled in major sea ports, IIP, National Stock Exchange turnover and the indicator for tourism. Finally, the factor F3 extracts variations from the growth in car sales, production of two wheelers and the production of commercial vehicles.

[3] In India, the financial year calendar starts from 1st April of a particular calendar year to 31st March of the following year. Thus 2004–05: Q1 refers to the April–June quarter in the year 2004. In this paper we have followed the Indian financial year calendar.

[4] The unit root test results are available from the authors on request.

[5] We choose three factors as only 78 percent of the variation is explained by the first two factors.

Table 1 Loadings of indicators in estimated factors

Indicator (YoY growth, %)	*F1*	*F2*	*F3*
Passenger Car Sales	0.31	0.26	0.60
Cargo Handled in Ports	0.03	0.87	0.04
CPI	0.45	(–)0.30	0.46
Aggregate Deposits	0.77	0.18	(–)0.04
Electricity Demand	0.24	0.13	0.20
Exports of Goods	0.58	0.05	0.37
Food Credit	0.59	–0.38	(–)0.07
GST	0.62	0.31	0.53
IIP	0.43	0.85	0.22
Non-food Credit	0.98	0.13	0.08
Non-oil Imports of Goods	0.47	(–)0.04	0.23
NSE Turnover	(–)0.16	0.61	0.004
Deviation of Rain from Normal Level	0.10	0.08	(–)0.14
Revenue Expenditure (Net of Interest Payments)	0.31	(–)0.14	–0.20
Rice Production	0.15	0.12	0.07
Net Tax Revenue	0.15	0.36	0.23
No. of tourists arrival in India	(–)0.04	0.52	0.24
Production of Two Wheelers	(–)0.07	(–)0.002	0.91
Production of Commercial Vehicles	(–)0.01	0.36	0.76

Source: Authors' estimates

Next, Eq. (2) is estimated where GDP growth is regressed on F1, F2, F3 and one period lagged GDP growth. The coefficients are assumed to vary over time and are estimated in a state-space framework, using Kalman filtration technique. Figure 1 depicts actual GDP growth along with the estimated GDP growth for the sample period Jan–Mar, 2008–Oct–Dec, 2019. The figure shows that predicted GDP growth using FA-TVCR model captures most of the turning points in Indian GDP growth series for the full sample period quite well.

In the second exercise, Specification II, we incorporate a few additional variables which are available for a shorter sample period. Information from a total of 28 high-frequency indicators is used in this exercise. The number of factors estimated is 3. However, given the limited number of observations in this exercise, only the first factor explaining 71.5 percent variation in the data is considered for the time-varying coefficient regression model (see Table 2).[6]

The factor loadings reported in Table 2 suggests that the first factor extracts variations from transport services indicators such as cargo movements via sea, air and rail, and passengers travelling by air; production indicators such as IIP, production of coal and cement, production of two wheelers and commercial vehicles; trade indicators such as exports of goods and services, and non-oil imports of goods and

[6] Given that the data on consumption of finished steel products are available from December, 2013, the quarterly y-o-y growth of this indicator is available from the quarter Jan-Mar, 2015. Consequently, we

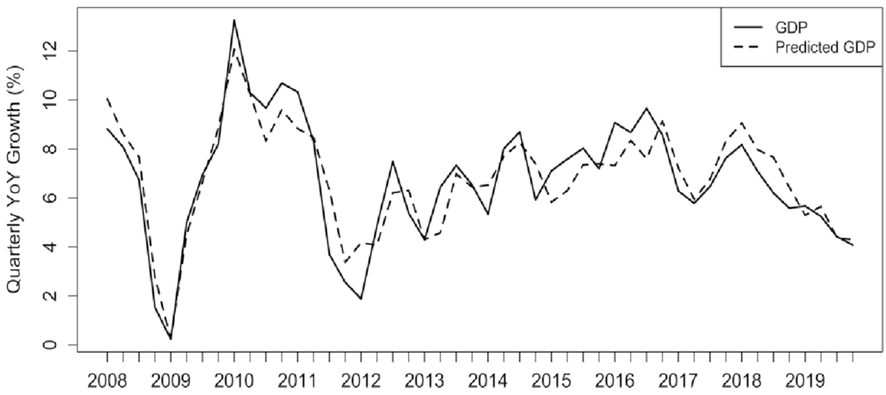

Fig. 1 Actual and Predicted GDP Growth (Jan–Mar, 2008–Oct–Dec, 2019). *Source*: Authors' estimates

services; electricity demand and supply; other demand indicators such as car sales, fiscal indicators such as net tax revenue and consolidated GST revenue.

Forecast Performance of FA-TVCR Model

We compare forecasting performance of our model with two alternating modelling strategy, namely, a univariate ARIMA model and a Dynamic Factor Model (DFM).

Univariate Quarterly Model

Univariate Auto-Regressive Integrated Moving Average (ARIMA) models is the simplest method of projecting growth rate of an economy. Following this method, quarterly y-o-y growth rate of GDP (g_t^Q) can be modelled as

$$(1-L)^d g_t^Q = A + A_1 g_{t-1}^Q + A_2 g_{t-2}^Q + \ldots . + A_p g_{t-p}^Q + B_1 \varepsilon_t + B_2 \varepsilon_{t-1} + B_3 \varepsilon_{t-2} + \ldots . + B_q \varepsilon_{t-q} \tag{5.1}$$

Here, L is the lag operator ($Lg_t^Q = g_{t-1}^Q$); d is the order of integration; $A_{1,}\ A_{2,\ \ldots..,}A_{\mathrm{p}}$ are the coefficients of autoregressive components; $B_{1,}\ B_{2,}\ \ldots\ldots., B_{\mathrm{q}}$ are the coefficients of moving average components; and ε_t is the *i.i.d* error term. Given information available at period t, GDP growth rate for period $t+1$ can be projected as

$$\widehat{g_{t+1|t}^{Q}} = \hat{A} + \sum_{i=0}^{p} \hat{A}_i \left(\widehat{g_{t-i}^{Q}} - \hat{A} \right) + \hat{\varepsilon}_t + \sum_{j=1}^{q} \hat{B}_j \widehat{\varepsilon_{t-j}},$$

Table 2 Loadings of indicators in estimated factors: Jan–Mar, 2015 onwards

Indicator (YoY growth, %)	*F1*
Cargo Movement by Air	0.78
Passengers Travelled by Air	0.58
Car Sales	0.83
Cargo Handled in Ports	0.75
Production of Coal	0.74
Production of Cement	0.83
CPI	(–)0.31
Aggregate Deposits	(–)0.20
Electricity Supply	0.82
Electricity Demand	0.85
Exports of Goods and Services	0.57
Food Credit	(–)0.39
GST	0.89
IIP	0.90
Non-food Credit	0.40
Non-oil Imports of Goods and Services	0.78
NSE Turnover	(–)0.09
Production of Crude Oil	0.39
Deviation of Rain from Normal Level	(–)0.46
Revenue Expenditure (Net of Interest Payments)	(–)0.07
Rice Production	0.13
Goods Movement vial Rail	0.90
Passengers Travelled by Rail	0.47
Net Tax Revenue	0.57
Telephone/Mobile Subscribers	0.13
Domestic Sale of Tractors	0.42
Production of Two Wheelers	0.96
Production of Commercial Vehicles	0.89

Source: Authors' estimates

$$where \quad \varepsilon_{t-j} = \frac{\left(1 - A_1L - A_2L^2 - \ldots . - A_pL^p\right)}{\left(1 + B_1 + B_2 + \cdots + B_q\right)}(g_t^Q - A) \tag{5.2}$$

The order of integration d and the order of AR and MA process, i.e., p and q are optimally chosen using Box-Jenkins methodology in an iterative process and using

Footnote 6 (continued)
have 24 observations for each of 28 indicators. Since the number of observations is less than the number of variables, the factor matrices are rank deficient and ML Estimator technique is not applicable (Robertson and Sumons, 2007). Hence we apply Iterated Principal Factor method in this stage.

AIC, BIC and SC criteria. The optimally chosen univariate model for quarterly YoY GDP growth follows an ARIMA(1,0,1) specification.

Dynamic Factor Model

The Dynamic Factor Model (DFM) assumes that a common unobservable state variable s_t drives N number of macroeconomic indicators y_t. The framework of Dynamic Factor Model (DFM) is outlined as follows:

$$y_t = As_t + By_{t-1} + e_t \tag{5.3}$$

$$s_t = C + \varphi s_{t-1} + u_t \tag{5.4}$$

where y_t is $(N \times 1)$, s_t is $(K \times 1)$, A is $(N \times K)$, B is $(N \times N)$ and φ is $(K \times K)$. Here A, B, C are parameters to be estimated and e_t and u_t are modelled as Gaussian error terms $e_t \sim iid\, N\,(0, R)$, $u_t \sim iid\, N\,(0, Q)$, and $E(e_t u_t) = 0$.

The DFM specification is a state-space model where the first equation, the measurement equation, describes the relation between the observed variable y_t and the unobserved state variable s_t. Eq. (B.1) is the transition equation which describes the dynamics of unobserved variables. All the variables in the model are required to be stationary.

The model estimation aims at estimating the parameters A, B, C and φ to recover the unobserved state space variable s_t. The model is estimated using Kalman filtering technique which is a recursive algorithm that provides an optimal estimate of s_t conditional on information up to time $t-1$ and knowledge of the state space parameters A, B, C, φ, R and Q. The estimation results from DFM specifications are reported in Table 9 in Appendix B.

Forecast Performance of FA-TVCRM for Specification I: Using Indicators Available from 2004–05

In order to evaluate the in-sample and out-of-sample performance of the model, we divide the sample period into train data and test data periods respectively. The train data period is Apr–Jun, 2007 to Oct–Dec, 2018. The test data period is Jan–Mar, 2019 to Oct–Dec 2019. We estimate the model for the train data period using the FA–TVCR model Specification I and the in-sample Root Mean Square Error (RMSE) is found to be 0.35.[7]

We then obtain the out-of-sample period nowcasts for the four quarters of 2019 in the following way. With the model estimated till Oct–Dec, 2018, the nowcast of GDP growth for Jan–Mar, 2019 is obtained using the estimated coefficients and the factors summarizing information from high frequency indicators available for the quarter Jan–Mar, 2019. The model is then re-estimated with data till Jan–Mar, 2019

[7] The model is estimated with the indicators standardized using their respective mean and standard deviation which is a standard practice in the estimation of forecasting models.

Table 3 Comparing in-sample forecast performance of FA-TVCR model Specification I and alternative models

Model	In-sample RMSE	DM Test H0: Two forecasts have same predictive accuracy H1: Two forecasts have different predictive accuracy		DM Test H0: Two forecasts have same predictive accuracy H1: Forecast 1 is more accurate than forecast 2	
		Test Statistic	p-value	Test statistic	p-value
FA-TVCRM Specification I	0.35				
DFM	1.00				
ARIMA	0.70				
FA-TVCRM Specification I vs. DFM		(–)3.018	0.003	(–)3.018	0.001
FA-TVCRM Specification I vs. ARIMA		(–)2.775	0.006	(–)2.775	0.003

Source: Authors' estimates

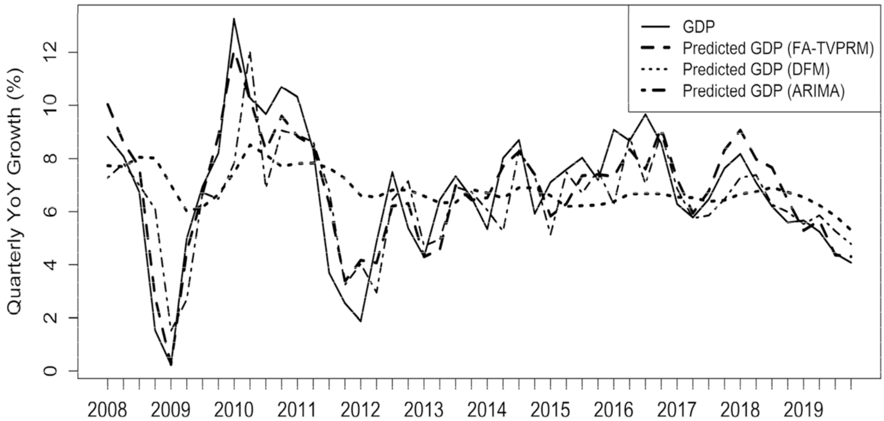

Fig. 2 In-sample fit from alternative models. *Source*: Authors' estimates

and the nowcast for Apr–Jun, 2019 is obtained using the re-estimated parameters and the factors estimated using monthly indicators for Apr–Jun, 2019. We repeat this procedure to obtain nowcast of GDP growth for the quarter Oct–Dec, 2019.

Table 3 and Fig. 2 compare the in-sample nowcast performance of the FA-TVCR model Specification I with two alternative models, namely a Dynamic Factor Model (DFM) and a univariate ARIMA model.

In terms of in-sample RMSE, FA-TVCR model Specification I performs best with the lowest RMSE, followed by the ARIMA model and DFM (Table 3). Additionally, when we compare the predictive power of FA-TVCR Specification I and ARIMA models using the Diebold-Mariano test (Diebold and Mariano 1995), the

Table 4 Comparing out-of-sample forecast performance of alternative models

Model	Out-of-sample RMSE	DM Test H0: Two forecasts have same predictive accuracy H1: Two forecasts have different predictive accuracy		DM Test H0: Two forecasts have same predictive accuracy H1: Forecast 1 is more accurate than forecast 2	
		Test Statistic	p-value	Test statistic	p-value
FA-TVCRM Specification I	0.33				
DFM	0.82				
ARIMA	0.51				
FA-TVCRM Specification I vs. DFM		(–)2.720	0.007	(–)2.720	0.003
FA-TVCRM Specification I vs. ARIMA		(–)0.958	0.338	(–)0.958	0.169

Source: Authors' estimates

Table 5 Out of sample projection of GDP growth rates for Q4 2018–19 to Q3: 2019–20

Period	Quarterly y–o-y GDP growth (Actual)	Quarterly y–o-y GDP growth (Flash estimates)	Quarterly y–o-y GDP growth (Projection from Specification I till pre-covid period)
Jan-Mar 2019	5.8	5.7	3.7
Apr-Jun 2019	5.4	5.2	5.2
Jul-Sep 2019	4.6	4.4	4.3
Oct-Dec 2019	3.3	4.1	3.8

Source: Authors' estimates

null hypothesis that the two models have the same predictive accuracy is rejected at 5% level of significance against the alternative hypothesis that the two models have different accuracy as well as against the alternative hypothesis that nowcast from FA-TVCR model is more accurate than the nowcast from the ARIMA model (Table 3 row 5, columns 2–5).

Further, when we compare the predictive power of FA-TVCR model with DFM using Diebold-Mariano test, the null hypothesis that the two models have the same predictive accuracy is rejected at a 5% level of significance against the alternative hypothesis that the two models have different accuracy as well as against the alternative hypothesis that nowcast from FA-TVCR model is more accurate than the nowcast from DFM (row 4, columns 2–5 in Table 3).

In terms of out-of-sample RMSE, FA-TVCR model performs best followed by ARIMA model and DFM (Table 4). Using the Diebold-Mariano test, we reject the null hypothesis that FA-TVCRM and DFM have same predictive accuracy against the alternative hypothesis that the out-of-sample nowcast from FA-TVCRM is more accurate than the out-of-sample nowcast from DFM (row 4, columns 2 to 5 in Table 4).

However, we cannot reject the null hypothesis that the predictive accuracy of the FA-TVCR and the ARIMA models are the same.

Table 5 compares out of sample projections of quarterly y–o-y GDP growth with the actual growth outcome for the period Jan–Mar 2019 to Oct–Dec 2019 using Specification 1. Since a subset of indicators used in Specification II are available from Jan-Mar 2015, we do not have sufficient number of observations to perform out-of-sample projection for the pre-covid period. Hence we compare performance of out of sample nowcasts for the covid period using Specification II, along with Specification I in Table 7.

Performance of the Models for the Period Including Covid-19 Pandemic

We next investigate the nowcast performance of the model for the period including the Covid-19 pandemic when the economy was hit by a massive negative shocks. We estimate both the specifications of FA-TVCR model and the ARIMA model. We then test the in-sample and out-of-sample nowcast performance of all the three (Table 6).

The in-sample RMSE for the FA-TVCR model Specification I is the lowest of the three at 0.66, followed by FA-TVCR model Specification II and ARIMA. The RMSE of the FA-TVCR model Specification I is also the lowest of the three for the out-of-sample period. Both specifications of our FA-TVCR model also perform better than the ARIMA model in terms of the D-M test.

Comparing between the two specifications of FA-TVCR model, we find that both the in-sample RMSE and out-of-sample RMSE for Specification I is lower than that of Specification II. Further, Specification I better predicts the contraction of Apr–Jun, 2020, as well as the recovery pattern (Table 7). However, the DM test suggests that both the specifications have the same predictive accuracy, implying that the model is robust and its predictive power is invariant with respect to the length of the time series or the number of indicators used.

Concluding Remarks

Governments, central banks, private firms and others need high frequency data on the state of the economy for their decision making. However, a key indicator like GDP is only available quarterly and that too with a lag. Decision makers have therefore adopted the technique of nowcasting, projecting quarterly GDP growth based on high frequency daily, weekly or monthly indicators, mostly based on DFM models. In this paper we have presented an alternative model, the FA-TVCR nowcasting model which allows us to extract information from a large number of indicators and also inherently addresses the issue of frequent structural breaks in GDP growth. This model has been estimated for India for a full sample period from January–March 2007 to October– December 2018 using 19 high frequency indicators (Specification I) and for a shorter sample period from April– June 2015 to October– December 2018 using a larger set of 28 indicators which are available for this shorter period (Specification II).

Comparing forecast performance of the two specifications, we find that the FA-TVCR model is robust in the sense that Specification I uses a fewer set of indicators for a longer

Table 6 Comparing out-of-sample forecast performance of alternative models for the period including the Covid-19 pandemic

Model	In-sample RMSE for period upto Jan-Mar, 2021	Out-of-sample RMSE (Jan–Mar, 2020 to Jan–Mar, 2021)	Out-of-sample RMSE upto Jan-Mar, 2021, excluding Apr–Jun, 2020	DM Test H0: Two forecasts have same predictive accuracy H1: Two forecasts have different predictive accuracy		DM Test H0: Two forecasts have same predictive accuracy H1: Forecast 1 is more accurate than forecast 2	
				Test Statistic	p-value	Test statistic	p-value
FA-TVCRM Spc I	0.66	1.99	0.50				
FA-TVCRM Spc II	0.77	4.23	1.10				
ARIMA	1.92	6.14					
FA-TVCRM Spc I vs. ARIMA (Q1 2020–Q1 2021)				(–)1.948	0.051	(–)1.948	0.026
FA-TVCRM Spc II vs. ARIMA (Q1 2020–Q1 2021)				(–)2.498	0.012	(–)2.498	0.006
FA_TVCRM Spc I vs. SPC II				0.890 (–)1.1979	0.423 0.231	0.890 (–)1.1979	0.116

Source: Authors' estimates

Table 7 Out of sample projection of GDP growth rates for Q4 2019–20 to Q4 2020–21

Period	Quarterly y–o-y GDP growth (Actual)	Quarterly y–o-y GDP growth (Flash estimates)	Quarterly y–o-y GDP growth (Projection from Specification I)	Quarterly y–o-y GDP growth (Projection from Specification II)
Jan-Mar 2020	3.0	3.1	2.7	3.5
Apr-Jun 2020	– 24.4	– 23.9	– 13.3	– 1.2
Jul-Sep 2020	– 7.4	– 7.5	– 5.4	– 5.9
Oct-Dec 2020	0.5	– 0.6	– 0.8	– 1.5
Jan-Mar 2021	1.6	2.4	2.1	– 3.0

Source: Authors' estimates

period and Specification II which uses a larger set of indicators for a shorter period are equally efficient in terms of predictive power. We also find that our model outperforms a DFM and a univariate ARIMA model in terms of both in-sample and out-of-sample RMSE. Comparing the predictive power of the three models using the Diebold-Mariano test, we find that the FA-TVCR model outperforms DFM consistently. Both our model and the ARIMA model have the same predictive accuracy in terms of out-of-sample forecast accuracy under normal conditions. However, our model outperforms the ARIMA model when applied for nowcasting during a period including the Covid-19 pandemic shock. It suggests that the FA-TVCR model is a more appropriate tool for nowcasting GDP in countries characterised by frequent structural breaks and large shocks.

With recent advances in computer science, machine learning (ML) algorithms are used in forecasting and nowcasting research to capture the complex dynamics of macroeconomic processes (Pratap and Sengupta 2019). Several studies have found that ML methods outperform time series models in nowcasting GDP growth in both advanced and emerging economies (Babii et al. 2021; Marcellino and Sivec 2021; Richardson et al. 2021; Muchisha et al. 2021). Further, recent developments in forecasting research highlight the importance of news and the google trend data (Bortoli et al. 2018; Woloszko 2020). These studies are primarily for data-rich advanced economies. Application of ML technique to forecast GDP growth in India, with its data constrained environment, and comparing its predictive power with the present state-space model can be an interesting extension of the present exercise.

Disclaimer

The views expressed in study are those of the authors and do not necessarily reflect the views and policies of the organisations they are affiliated in.

Appendix A

See appendix Tables 8, 9.

Table 8 Data Sources

Sector	Series	Source	Web link	Date of release	Notes
Agriculture	Rainfall	CMIE		1st of every month	Deviation from long period average rainfall is more important determinant of agricultural output than absolute rainfall level
	Domestic Sale of Tractors	Tractors Manufacturers Association	https://www.tmaindia.in/consolidated-monthly-reports-2021.php	Mid-month	Used only in Specification II of FA-TVCR Model
	Production of Rice	Department of Agriculture	https://eands.dacnet.nic.in/	No fixed date	
Industry	IIP (2011–12 base)	CSO, MOSPI	http://mospi.nic.in/iip	12th of every month	Data are published with a two months lag
	IIP (2004–05 base)	CSO, MOSPI	http://mospi.nic.in/iip		Used for Specification I of FA-TVCR Model and DFM
	Production of two wheelers	MoSPI Micro data	http://microdata.gov.in/nada43/index.php/catalog/148	12th of every month	Data are published with a two months lag
	Production of commercial vehicles	MoSPI Micro data	http://microdata.gov.in/nada43/index.php/catalog/148	12th of every month	Data are published with a two months lag
	Passenger car sales	CMIE		11^{th} of every month	Used in DFM and FT-TVCR Model
	Production of Coal	Office of Economic Adviser	https://eaindustry.nic.in/	30th of every month	Used only in Specification II of FA-TVCR Model
	Production of Crude Oil	Office of Economic Adviser	https://eaindustry.nic.in/		”
	Production of Cement	Office of Economic Adviser	https://eaindustry.nic.in/		”
	Consumption of Steel	CMIE		Mid-month	”
	Electricity Generation	CMIE		End-of-month	”

Table 8 (continued)

Sector	Series	Source	Web link	Date of release	Notes
	Imports of oil (Rs.)	Press release of Ministry of Commerce and Industry	https://pib.gov.in/PressReleaseIframePage.aspx?PRID=1704910	15th of a month	Used in DFM and FT-TVCR Model
Services	Total Telephone Subscribers	Telephone Regulatory Authority of India	https://trai.gov.in/release-publication/reports/telecom-subscriptions-reports	No fixed date	Published with 2 months lag Used only in Specification II of FA-TVCR Model
	Foreign tourists arrival in India	Ministry of Tourism Press Release	https://tourism.gov.in/market-research-and-statistics	No fixed date	Used for Specification I of FA-TVCR Model and DFM
	Cargo handled at major sea ports	Ministry of Ports, Shipping and Waterways	http://www.ipa.nic.in/	1st week of every month	Used in DFM and FT-TVCR Model
	Cargo handled at airports ('000 tonnes)	CMIE		Third week of every month	Used only in Specification II of FA-TVCR Model
	Number of air travel passengers (lakh)	CMIE		Third week of every month	Used only in Specification II of FA-TVCR Model
	Number of rail travel passengers (millions)	CMIE		First week of every month	Used only in Specification II of FA-TVCR Model
	Cargo handled at railways ('000 tonnes)	CMIE		First week of every month	Used only in Specification II of FA-TVCR Model
	Average Daily Turnover at NSE	National Stock Exchange	https://www1.nseindia.com/products/content/equities/equities/historical_equity_businessgrowth.htm	Weekly	Used in DFM and FT-TVCR Model

Table 8 (continued)

Sector	Series	Source	Web link	Date of release	Notes
Prices	CPI-IW (2001 base)	Labour Bureau/	http://labourbureau.gov.in/LBO_indexes.htm	Last week of every month	Used for Specification I of FA-TVCR Model and DFM
	CPI-IW (2016 base)	Labour Bureau/CMIE	http://labourbureau.gov.in/LBO_indexes.htm		Used for Specification I of FA-TVCR Model and DFM
	CPI (base 2011–12)	CSO, MOSPI	http://164.100.34.62:8080/Default1.aspx	12th of every month	Used in DFM and FT-TVCR Model
External Sector	Exports of goods (Rs crore)	Department of Commerce, Ministry of Commerce & Industry	https://commerce.gov.in/trade-statistics/	15th of every month	Used for Specification I of FA-TVCR Model and DFM
	Imports of non-oil goods (Rs crore)	”	”	”	”
	Exports of Goods and Services (Rs. crore)	”	”	”	Used only in Specification II of FA-TVCR Model
	Imports of non-oil goods and Services (Rs. crore)	”	”	”	”
Fiscal Indicators	Net tax revenue (Rs crore)	Controller General of Accounts	http://www.cga.nic.in/	30th of every month	
	Revenue Expenditure Net of Interest Payments (Rs crore)				
	GST Collection (Rs crore)	PIB, Ministry of Finance	https://pib.gov.in/SearchResults.aspx?q=GST&cx=003919640075425102515%3a4lg1hrnhj_k&cof=FORID%3a9#gsc.tab=0&gsc.q=GST&gsc.page=1	1st of every month	Backcast as proxy for indirect taxes Used in DFM and FT-TVCR Model

Table 8 (continued)

Sector	Series	Source	Web link	Date of release	Notes
Monetary Indicators	Food Credit (Rs crore)	RBI Weekly Statistical Supplement	https://rbi.org.in/scripts/BS_ViewWSS.aspx	2nd Friday of a month	Used in DFM and FT-TVCR Model
	Non-food Credit (Rs crore)	,,	,,	,,	,,
	Aggregate Bank Deposits (Rs crore)	,,	,,	,,	,,
Energy Demand	Electricity Requirements	CMIE		Daily	Used in DFM and FT-TVCR Model

Table 9 Estimation results of the DFM model

Indicator	Factor	Coefficient	p-value
$F_{\_t}$	$F_{\_(t-1)}$	0.883	0.000
GDP	F	0.182	0.000
Car sales	F	0.281	0.000
Cargo handled at major ports	F	0.215	0.003
CPI	F	0.241	0.007
Aggregate bank deposit	F	0.316	0.000
Electricity requirement	F	0.135	0.032
Exports	F	0.317	0.000
Food credit	F	0.071	0.431
GST	F	0.305	0.000
IIP	F	0.281	0.000
Non-food credit	F	0.255	0.000
Non-oil imports	F	0.217	0.003
NSE turnover	F	0.092	0.266
Rainfall deviation from normal level	F	0.013	0.874
Revenue expenditure net of interest payments	F	0.033	0.669
Rice production	F	0.089	0.240
Net tax revenue	F	0.169	0.007
Production of two wheelers	F	0.212	0.000
Production of commercial vehicles	F	0.282	0.000
Tourists arrival	F	0.154	0.021

Source: Authors' estimates

Appendix B

The number of factor optimally estimated by DFM is 1. Growth rate in all the macroeconomic indicators, except for food credit, NSE turnover, rainfall deviation, revenue expenditure, and rice production are explained by the estimated common unobserved factor.

Data availability Data used in the analysis can be available upon request to the authors.

References

Andrii Babii, Eric Ghysels, and Jonas Striaukas (April, 2021), "Machine Learning Time Series Regressions with an Application to Nowcasting," *Journal of Business and Economic Statistics*, (url: https://doi.org/10.1080/07350015.2021.1899933), Arvind Subramanian and Josh Felman, (2019), "India's Great Slowdown: What Happened? What's the Way Out?", *CID Faculty Working Paper No. 370*, Harvard University, https://www.hks.harvard.edu/sites/default/files/centers/cid/files/publications/faculty-working-papers/2019-12-cid-wp-369-indian-growth-diagnosis-remedies-final.pdf, December.

Baffigi, Alberto, Roberto Golinelli, and Giuseppe Parigia. 2004. Bridge models to forecast euro area GDP. *International Journal of Forecasting* 20: 447–460.

Bortoli, Clément., Stéphanie. Combes, and Thomas Renault. 2018. Nowcasting GDP growth by reading newspapers. *Economie Et Statistique* 505–506: 17–33.

Charnis, Tony, and Rodrigo Sekkel. 2017. A dynamic factor model for nowcasting Canadian GDP growth. *Empirical Economics* 53: 217–234.

Daniela Bragoli and Jack Fosten. 2018. Nowcasting Indian GDP. *Oxford Bulletin of Economics and Statistics* 80 (2): 259–282.

Deepankar Basu, (2020), "Revisiting India's Growth Transitions" (2020). UMass Amherst Economics Working Papers. 284. https://doi.org/10.7275/16097049

Diebold, F.X., and R.S. Mariano. 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* 13: 253–263.

Foroni, Claudia, and Massimiliano Marcellino. 2014. A Comparison of Mixed Frequency Approaches for Nowcasting Euro Area Macroeconomic Aggregates. *International Journal of Forecasting* 30: 554–568.

Francis Diebold and Robert Mariano. 2002. Comparative predictive accuracy. *Journal of Business & Economics Statistics* 20 (1): 134–144.

Giannone, Domenico, Lucrezia Reichlin, and David Small. 2008. Nowcasting: the real time informational content of macroeconomic data. *Journal of Monetary Economics* 55: 665–676.

Hamilton, J.D. 1994. *Time Series Analysis*. Princeton University Press.

Stock, J.H., Watson, M.W. (2002), "Forecasting Using Principal Components from a Large Number of Predictors," *Journal of American Statistical Association*, 97 (2).

Kar, S., and K. Sen. 2016. *The political economy of india's growth episodes*. London: Palgrave Pivot.

Kim, C.J., and C.R. Nelson. 1999. *State space models with regime switching: classical and gibbs sampling approaches with applications*. MIT Press.

Luciani, Matteo, Madhavi Pundit, Arief Ramayandi, and Giovanni Veronese. 2018. Nowcasting Indonesia. *Empirical Econ* 55: 597–619.

Marcellino, Massimiliano, and Vasja Sivec. 2021. Nowcasting GDP growth in a small open economy. *National Institute Economic Review* 256: 127–161. https://doi.org/10.1017/nie.2021.13.

Massimiliano Marcellino and Christian Schumacher, (2010), "Factor MIDAS for Nowcasting and Forecasting with Ragged-Edge Data: A Model Comparison for German GDP**,"** *Oxford Bulletin of Economics and Statistics*, 72 (4), 518-550

Marta Bańbura, Domenico Giannone, and Lucrezia Reichlin, (2010), "Nowcasting," European Central Bank, Working Paper Series No. 1275.

Muchisha ND, Tamara N, Andriansyah, Soleh AM (2021) Nowcasting Indonesia's GDP Growth Using Machine Learning Algorithms, Stat Appl5 (2).

Oguzhan Cepni, I. Ethem Güney, and Norman R. Swanson, (January, 2020), "Forecasting and Nowcasting Emerging Market GDP Growth Rates: The Role of Latent Global Economic Policy Uncertainty and Macroeconomic Data Surprise Factors," Journal of Forecasting, 39 (1), 18-36

Oguzhan Cepni, I. Ethem Güney, and Norman R. Swanson, (April-June, 2019), "Nowcasting and Forecasting GDP in Emerging Markets using Global Financial and Macroeconomic Diffusion Indexes," *International Journal of Forecasting*, 35 (2), 555-572

Pratap, B., Sengupta, S., 2019. Macroeconomic Forecasting in India: Does Machine Learning Hold the Key to Better Forecasts? Working Paper WPS (DEPR): 04/ 2019, Reserve Bank of India.

Richardson A, van Florenstein Mulder T, Vehbi T, (2021) Nowcasting GDP using machine learning algorithms: a real time assessment, Int J Forecasting, 37 (2), pp: 941–948 Donald Robertson and James Symons, (2007), "Maximum Likelihood Factor Analysis with Rank-Deficient Sample Covariance Matrices, *Journal of Multivariate Analysis*, 98, pp: 813–828.

Robertson D, Symons J, (2007) Maximum likelihood factor analysis with rank-deficient sample covariance matrices, J Multivariate Anal 98: 813–828.

Rudrani Bhattacharya, Radhika Pandey and Giovanni Veronese, (2011), "Tracking India Growth in Real Time," National Institute of Public Finance and Policy, Working Paper No. 2011–90.

Rudrani Bhattacharya, Parma Chakravartti and Sudipto Mundle, (2019), “Forecasting India’s Economic Growth: A Time-Varying Parameter Regression Approach,” *Macroeconomics and Finance in Emerging Market Economies*, 12 (3).

Soumya Bhadury and Saurabh Ghosh and Pankaj Kumar, (2019), “Nowcasting GDP Growth Using A Coincident Economic for India”, MPRA Paper No. 96007.

Tara Iyer and Abhijit Sen Gupta, (May 2020), “Nowcasting Economic Growth in India: The Role of Rainfall,” Asian Development Bank Economics Working Paper Series No. 593.

Urasawa, Satoshi. 2014. Real-time GDP forecasting for japan: a dynamic factor model approach. *J Jpn Int Econ* 34: 116–134.

Woloszko N (2020) Tracking Activity in Real Time with Google Trends, Organisation for Economic Co-operation and Development, Working Paper No. ECO/WKP(2020)42.

Authors and Affiliations

Rudrani Bhattacharya[1] · Bornali Bhandari[2] · Sudipto Mundle[3]

Bornali Bhandari
bbhandari@ncaer.org

Sudipto Mundle
sudipto.mundle@gmail.com

[1] National Institute of Public Finance and Policy, 18/2, Satsang Vihar Marg, New Delhi 110067, India

[2] National Council of Applied Economic Research, 11, Indraprastha Estate, New Delhi 110002, India

[3] Centre for Development Studies, Medical College P.O, Prasanth Nagar, Ulloor, Thiruvananthapuram 695011, Kerala, India