A New Approach to Nowcast Indian Gross Value Added*

Soumya Bhadury,[†] Robert C. M. Beyer,[§] Sanjib Pohit[‡]

Abstract

In India, quarterly growth of Gross Value Added (GVA) is published with a large lag and nowcasts are exacerbated by data challenges typically faced by emerging market economies, such as big data revisions, mixed frequencies data publication, small sample size, non-synchronous nature of data releases, and data releases with varying lags. In this paper, we present a new framework to nowcast India's GVA that incorporates information of mixed data frequencies and other data characteristics. In addition, we add evening-hour luminosity as a crucial high-frequency indicator. Changes in nightlight intensity contain information about economic activity, especially in countries with a large informal sector and significant data challenges, including in India. We illustrate our framework for the 'trade, hotels, transport, communication and services related to broadcasting' bloc of the Indian GVA.

JEL Classification: C32, C51, C53

Keywords: *Nowcast, India, Gross Value Added, Evening-hour Luminosity, Dynamic Factor Analysis, EM Algorithm*

---

* In this paper we discuss the general framework and present first results for the 'trade, hotels, transport, communication and services related to broadcasting' bloc of the Indian GVA. We would like to thank Apoorv Raj Singh for excellent research assistance and Virgilio Galdo for cleaning the nightlight data.

† Reserve Bank of India, soumya.bhadury18@gmail.com

§ The World Bank, rcmbeyer@worldbank.org

‡ National Council of Applied Economic Research, spohit@ncaer.org; spohit@gmail.com

1. Introduction

Most emerging market economies (EMEs) face considerable challenges with regard to observing and predicting economic activity. Data on Gross Value Added (GVA) and Gross Domestic Product (GDP) are published with a large lag and often revised considerably over time. For example, the first release of the quarterly GDP/GVA growth in India is published approximately 7–8 weeks after the end of the reference quarter. The growth estimates of the Central Statistics Office (CSO) avoid assessing the current state of the economy altogether. Nowcasting models are a crucial tool for many institutions worldwide, especially central banks, for monitoring the state of the economy in real time. Hitherto however, they were mainly employed in developed countries and nowcasting models for developing countries have started emerging only recently (Dahlhaus et al. 2017). We have designed a nowcasting framework suited for the Indian economy with the aim of predicting the very recent past, the present, and the very near future growth of real GVA in India.

We observe the following five data challenges in India, which are also faced by other EMEs:

(a) *Big data revisions*: According to Sapre and Sengupta (2017), the average revision of GDP estimates in India is + 0.5 percentage points.

(b) *Mixed frequencies data publication*: The index on mining in India, for example, is published monthly, whereas the foreign exchange (FOREX) assets data is published bi-weekly and the National Stock Exchange (NSE) data is published daily.

(c) *Small sample size*: The CSO has recently replaced the earlier 2004-05 base year with 2011-12, and updated the National Account Statistics (NAS) methodology to align it with more recent international guidelines. Using the data that has been revised in line with the updated methodology, we now have a shorter time series.

(d) *Non-synchronous data releases*: Hard data releases in India are non-synchronous. For example, data on the monthly production of coal and crude oil is typically released on the last working day of the month, on the monthly production of commercial vehicles during the middle of the month, and on railway freight traffic of major commodities during the first 10 days of the month.

(e) *Varying data lags*: Data on the monthly production of steel and fertilisers for the month of December is released in the month of January of the following year. However, data on the mining and quarrying index for the month of November is released in the month of January of the following year with a lag of more than a month. Together, (d) and (e) result in jagged-edge data.

Our framework builds on similar efforts by Bhattacharya et al. (2011) Bragoli and Fosten (2016), and Bhadury and Pohit (2017, 2018). Bhattacharya et al. (2011) compare the performance of univariate quarterly models with bridge models that are based on high-frequency monthly indicators for close-range forecasting. Their forecasting procedure is based on a real-time simulation, following official data releases in India. They suggest that bridge models perform moderately well in tracking current quarter GDP growth in India, and find mixed evidence about the incremental predictive power of Indian survey data. The works by Bhadury and Pohit (2017, 2018) follows the approach taken by Bhattacharya et al (2011) but differs from them on two counts: (a) nowcasting GDP at sectoral levels and (2) explore with more high frequency indicators than the former study. Bragoli and Fosten (2016) adopt a new approach to proxy for unavailable data on service exports in India, by including industrial production in the US and Eurozone. One of the important contributions of the paper by Bragoli and Fosten (2016) is to construct a real-time GDP series for India that allows for the analysis of the impact of data revisions by the CSO. Their study suggests that the best model for tracking real-time GDP growth is one that includes not just real and nominal variables and prices, but also an international series. The paper also finds that the results of the nowcasting model for India change when growth is tracked from the first data release rather than final release. This suggests that preliminary releases of GDP data might not be very useful for growth predictions.

Our work contributes to the literature in three important ways. First, it incorporates information of mixed data frequencies and other data characteristics, such as jagged edge data and successfully makes use of incremental information from mixed-frequency and non-synchronous data releases. Second, it employs separate high frequency indicators for tracking GVA growth in each of the six subcomponents of GVA. Dynamic factor models (similar to Forni et al. 2000; 2003; Forni and Lippi 2001; Stock and Watson 2002; Doz et al. 2007) have been employed to extract common factors from a set of monthly core indicators that are uniquely associated with a specific GVA bloc. Third, it adds data on evening-hour luminosity (World Bank 2017, Beyer et al. 2018). A strong relationship between economic activity and nightlight intensity is by now well established (e.g. Henderson et al. 2011). Since April 2012, monthly nightlight data from VIIRS satellites is publicly available (Elvidge et al. 2013). However, monthly nightlight data is not yet used systematically for nowcasting. Nightlight data is potentially an insightful source of information especially in countries with a large informal sector and other data limitations. We have integrated cleaned high-frequency geospatial satellite data on nightlight intensity (World Bank 2017, Beyer et al. 2018), which we expect will improve predictions, especially in the near future when few other data series would be available. In this paper, we present the general framework and provide empirical results for the *trade, hotels, transport, communication and services related to broadcasting* bloc of Indian GVA.

2. Methodology

We produce a series of three real-time GVA sector nowcasts prior to the official data release using four different methods: Ordinary Least Squares (OLS), Mixed Data Sampling (MIDAS), Principal Component Analysis (PCA), and Dynamic Factor Model (DFM). The last two methods make use of different data-shrinkage techniques, which may improve nowcasts as multiple macroeconomic variables exhibit co-movements adding to the convolution of the forecasting process.

2.1. Ordinary Least Squares (OLS)

The simple linear regression helps in investigating relationships between variables,

$$y_i = \alpha_1 + \alpha_2 * x_i + \gamma_i \tag{1}$$

The non-random or structural component is $\alpha_1 + \alpha_2 * x_i$; where x_i is the independent or explanatory variable with α_1 and α_2 as the model parameters, and γ_i is the random component or error term. We adopt the Ordinary Least Squares (OLS) method to optimally estimate the model parameters. This implies that the model parameters, $\hat{\alpha}_1$ and $\hat{\alpha}_2$ are estimated by minimising the sum of all squared deviations.

$$\min_{\hat{\alpha}_1,\hat{\alpha}_2} \sum_{i=1}^{n}(e_i)^2 = \min_{\hat{\alpha}_1,\hat{\alpha}_2} \sum_{i=1}^{n}(y_i - \hat{\alpha}_1 - \hat{\alpha}_2 * x_i)^2 = S(\hat{\alpha}_1, \hat{\alpha}_2) \tag{2}$$

2.2. Mixed Data Sampling (MIDAS)

Mixed Data Sampling (MIDAS), as the name suggests, involves regression models for datasets with processes of different frequencies. It is generally used when the frequency of the explanatory variables is higher than the frequency of the dependent variable (Ghysels et al. 2004, 2007; Marcellino and Schumacher 2007). The explanatory variables can also have frequencies different from each other. The basic equation for MIDAS is similar to that of distributed lag models, exhibiting a dynamic relationship between the dependent and independent variables. However, there are still significant differences between the two methods. The basic equation for MIDAS is (Ghysels et al. 2004):

$$Y_t = \beta_0 + B\left(L^{\frac{1}{m}}\right) X_t^{(m)} + \epsilon_t^{(m)} \tag{3}$$

where $B\left(L^{\frac{1}{m}}\right) = \sum_{j=0}^{j^{max}} B(j)L^{j/m}$ is a polynomial of length j^{max} in the $L^{1/m}$ operator and the $L^{j/m}$ operator lags $X_t^{(m)}$ by j/m periods. We can run the above non-linear regression to estimate the dependent variable. Extracting the maximum information from the dataset requires a suitable polynomial, which, in turn, may involve an increased number of lags of $X_{(t-j)}^{m}$ data. This necessitates estimation of many parameters and is one of the shortcomings of MIDAS (Ghysels et al. 2004).

2.3. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is extremely helpful if the number of predictor series is very large (Stock and Watson 2002). In such cases, the principal components of the N-dimensional multiple time series of candidate predictors can serve as consistent estimators of the factors. The factors can then form a time series, and serve as the dependent variable when regressed upon the variable to be forecast. To derive the principal component estimator, we first write the non-linear least squares objective function as follows (Stock and Watson 2002):

$$V(\widetilde{F}, \tilde{\Lambda}) = (NT)^{-1} \sum_{i} \sum_{t} (x_{it} - \tilde{\lambda}_i \widetilde{F}_t)^2 \tag{4}$$

We minimise the above equation to minimise the error term associated with the model. The classical principal components technique is used to solve this by equating $\widehat{\Lambda}$ to the Eigenvectors of X'X corresponding to its r largest Eigenvalues. This gives us the principal components estimator, $\widehat{F}$:

$$\hat{F} = \frac{X'\hat{\Lambda}}{N} \tag{5}$$

This estimate is point-wise consistent (for any date t) and has limiting mean squared error (MSE) over all values of t that converge to 0. The consistency can be derived from the function $F_t' - F_t$. At both N and T tending to infinity, the above function tends to zero, implying the consistency of the factor estimates.

PCA is based on primarily two properties, namely variance maximization and dimension reduction. However, maximization of variance in the p-dimensional data-space under a quadratic constraint, leads to many variable loadings on the first principal component. This is a major drawback of the PCA. One way to avoid this is to use the varimax rotation procedure.

2.4. Dynamic Factor Model (DFM)

The Dynamic Factor Model (DFM) exploits hidden trends that are dynamic, and filters out extra information by selecting the most relevant factors out of multiple time series variables. Forni, et al. (2000) suggest a generalised dynamic factor model which caters to most business cycle problems. It also takes up correlated idiosyncratic components, rejecting orthogonality at the same time, as in most cases it is an unrealistic assumption. An advantage of DFM is that it covers both autoregressive and moving average responses to common factors unlike the static factor model (Stock and Watson 2002) and thus provides the best model alternative for the nowcasting process.

A DFM assumes that the n-dimensional vector of stationary observed variables $(\lambda_{1,t},\cdots,\lambda_{n,t})$ is driven by a vector of r unobserved dynamic factors $(f_{1,t},\cdots,f_{r,t})$. The features that are specific to individual series, such as measurement errors, are captured by idiosyncratic errors $(\varepsilon_{1,t},\cdots,\varepsilon_{n,t})$. $(\gamma_1,\cdots,\gamma_r)$ is an r- dimensional vector and does not vary over time.

The empirical model can be summarised in the following equation:

$$\lambda_{i,t} = \gamma_i' F_t + \varepsilon_{i,t};\ i = 1,\ldots,n, t = 1,..,T \tag{6}$$

The two components $\zeta_{i,t} = \gamma_i' F_t$ and $\varepsilon_{i,t}$ are orthogonal unobserved stochastic processes. $\zeta_{i,t} = \gamma_i' F_t$ is the linear combination of the r unobserved common factors, F_t , reflecting the bulk of the co-movement in the economy. We allow the idiosyncratic component, $\varepsilon_{i,t}$, to follow an AR(1) process;

$$\begin{aligned} &\varepsilon_{i,t} = \alpha_i \varepsilon_{i,t-1} + e_{i,t}; e_{i,t} \sim iid\ \mathrm{N}(0,\sigma_i) \\ &E[\sigma_{i,t},\sigma_{j,s}] = 0,\ for\ i \neq j \end{aligned} \tag{7}$$

Equation (8) is the matrix notation representation of Equation (7), where

$$\begin{aligned} &X_t = \Gamma F_t + \Pi_t \\ &X_t = (\lambda_{1,t},\lambda_{2,t},\cdots,\lambda_{n,t})', \Pi_t = (\varepsilon_{1,t},\varepsilon_{2,t},\cdots,\varepsilon_{n,t})' \text{ and } \Gamma=(\gamma_1,\cdots,\gamma_r) \end{aligned} \tag{8}$$

There are different ways in which large DFMs can be estimated. Giannone et al. (2008) point out that DFMs are specially equipped to monitor macroeconomic conditions in real time or nowcasting. Doz et.al. (2012) show the consistency and robustness properties of the likelihood-based (for example, quasi-maximum likelihood estimator) methods when both the size of the samples and the cross-sections are large. Furthermore, the results remain robust to cross-sectional mis-specification, correlation between idiosyncratic components or non-Gaussian features in the dataset.

To address data irregularities, especially those associated with mixed frequencies and non-synchronicity of the data releases, we have adopted the Kalman filtering technique. The Kalman filter uses the expectation maximisation (EM) algorithm, which can handle both mixed frequencies and missing data (see Banbura and Modugno 2014). The algorithm is initialised by computing principal components, and the model parameters are estimated by OLS regression, treating the principal components as if they were the true common factors. This is a good initialisation, especially with big data, given that principal components are reliable estimates of the common factors. For example, the principal components come from the largest Eigenvalues of the sample correlation matrix of the series,

$$S = \frac{1}{T}\sum_{i=1}^{T} X_t X_t^{'} \tag{9}$$

The r largest principal components are extracted from the sample correlation matrix. D is the $r \times r$ diagonal matrix with diagonal elements given by the largest r Eigenvalues of S, and denoted by V the $n \times r$ matrix corresponding Eigenvectors s.t the normalisation gives $V^{'}V = I_r$. Following is the approximation of the common factors:

$$\tilde{F} = V^{'} X_t \tag{10}$$

Once we have estimated the common factors, $\tilde{F}$, we can estimate the factor loadings, Γ, and the covariance matrix of the idiosyncratic components, Π. This is done by regressing the data series on the estimated common factors, as follows:

$$\hat{\Gamma} = \sum_t X_t \tilde{F}_t^{'} (\tilde{F}_t \tilde{F}_t^{'})^{-1} = V \tag{11}$$

The estimated covariance matrix of the idiosyncratic components, $\hat{\Pi}$, is as follows:

$$\hat{\Pi} = diag(S - VDV) \tag{12}$$

The dynamic factor equation parameters, A and B, can be estimated from VAR, on the common factors, $\tilde{F}_t$, where $F_t = AF_{t-1} + Bu_t$. These estimates, $\hat{\Gamma}, \hat{\Pi}, \hat{A}, \hat{B}$, have been proven to be consistent as $n, T \rightarrow \infty$ by Forni et al. (2000). Given the estimated parameters, in the second step, an updated estimate of the common factors is obtained using the Kalman smoother. The Multivariate Autoregressive State-Space Modeling

(MARSS) package in R is capable of handling jagged-edge data. For example, we attempt to Nowcast (-1), that is, one month before the official data release for FY 2017–18, Q3 by CSO. Q3 comprises the months of October, November and December, 2017, and the official release date for Q3 is February 28, 2018. (A special DFM setup, as represented in the MARSS package in R, is shown in Appendix A.)

2.5. Real-time Nowcasts

The major data irregularities associated with the calendar release of Indian data are twofold: non-synchronicity of data releases, and data that are available in mixed frequencies. To address the former, we attempt to compare the relative performance of models that can make use of the incremental information as against models that are rendered ineffective to work with such data. For example, models that employ the principal component technique are unable to process a jagged-edge dataset. On the other hand, models that employ DFMs are naturally cast in a state-space framework, and inferences are performed using the Kalman filtering technique that can handle jagged-edge datasets. Similarly, to address the latter, we compare models that adopt the regular OLS method versus those that adopt the MIDAS method, which are specially equipped to handle mixed frequency data.

To the best of our knowledge, this paper represents the first attempt to add nightlight intensity to a systematic nowcasting framework. A strong correlation of nightlight intensity and economic activity is well established (Henderson et al. 2012) and has been confirmed for South Asian countries (World Bank 2017; Beyer et al. 2018). In India, nightlight intensity is especially strongly correlated with activity in the services sector. Since activity in the services sector is difficult to measure, for example, due to large informality, we expect information on nightlight intensity to improve nowcasts.

We produce a sequence of three real-time GVA sector nowcasts prior to the official data release. We label the three sequence releases as follows:

(i) **Nowcast (-2)**: This covers the period two months before the official data release. Nowcast (-2) is the biggest challenge, and we call it ***low-visibility nowcasting*** due to the unavailability of hard data covering the entire reference quarter.

(ii) **Nowcast (-1)** is performed one month before the official data release and we call it ***medium reach nowcast.***

(iii) **Nowcast (-0.5)** is a ***close range nowcast***, performed 15 days before the official release of the GVA numbers. All hard-data indicators are available till the end of a reference quarter during this period, which is expected to improve the precision of the nowcast.

3. Data

The quarterly time-series GVA is divided into seven major blocks: agriculture; mining and quarrying; manufacturing; electricity, gas and water supply, and construction; financial, real estate and professional services; public administration, defence, and other services; and trade, hotels, transport, communication and services related to broadcasting. In this paper, we present the nowcasting results for the latter sector, which we will hereafter refer to as 'trade GVA'.

The high-frequency monthly data series associated with the trade GVA are railway freight traffic of major commodities, cargo traffic through major ports, air cargo, foreign tourist arrivals in India, the telecommunication subscriber base, Purchasing Managers' Index (PMI)-services index, production of commercial vehicles, and nightlight intensity. For high-frequency indicators suitable for nowcast the other sectors, refer to Table B.1 in Appendix B. The release of the data is non-synchronous (Table 1). In addition, the lag in data releases with respect to the reference month varies between the railway freight and cargo traffic via port.

Table 1: Varying Data Releases for Trade GVA in India

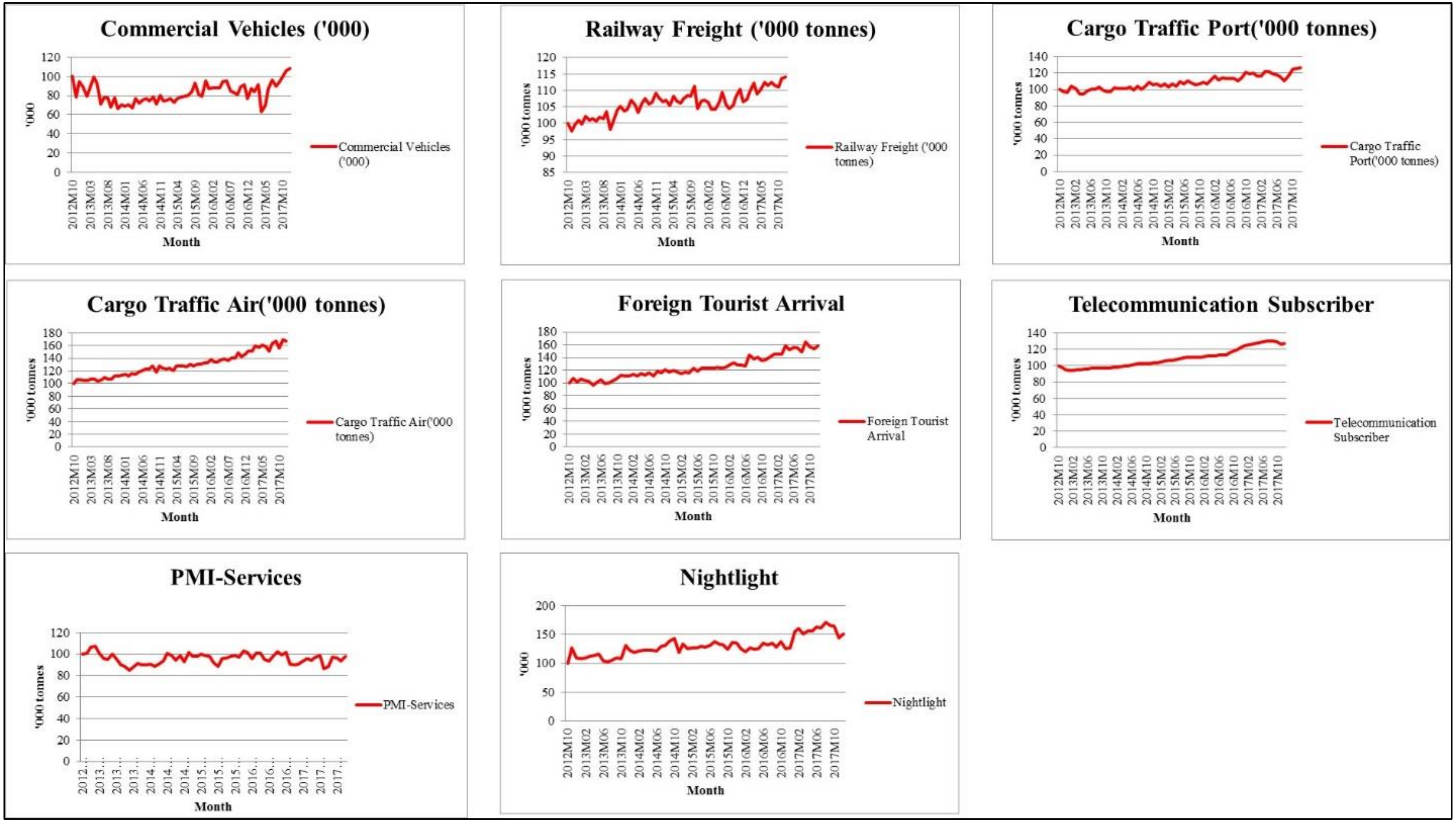
	INDICATORS	RELEASE DATE	REFERENCE PERIOD
Trade, Hotels, Transport, Communication and Services Related to Broadcasting	Railway freight traffic of major commodities	Between January 1 and January 10, 2018	December, 2017
	Cargo traffic – ports	January 15, 2018	November, 2017
	Cargo traffic – air	January 15, 2018	November, 2017
	Foreign Tourist Arrivals in India	Between January 10 and January 15, 2018	December, 2017
	Telecommunication subscriber base	January 11, 2018	November, 2017
	Nikkei PMI-Services Index	January 4, 2018	December, 2017
	Commercial Vehicles	15 January, 2018	November, 2017
	Nightlight	January 30, 2018	December, 2017

Source: Summarised by authors from government data sites

Except for nightlight intensity, all data have been obtained from the Economic Outlook database of the Centre for Monitoring Indian Economy Pvt Ltd (CMIE). All series are available in monthly frequency from April 2012 to December 2017 and are measured in different units. All data have been converted into indices for inter-temporal comparability and intra-temporal consistency. The seasonally adjusted values were obtained by adopting Signal Extraction in ARIMA Time Series (SEATS)-based seasonal adjustment with automatic outlier detection and Time series Regression with ARIMA noise, Missing values and Outliers (TRAMO)-based automatic Autoregressive Integrated Moving Average (ARIMA). Finally, scale effects are eliminated. Figure 1 shows the data for the eight high-frequency indicators used in this paper.

The plots in Figure 1 give a preliminary idea about the co-movement of the different indicator series. Except PMI-Services, all the indicators reflect a positive trend. The commercial vehicles and PMI-Services are also the two variables exhibiting maximum deviation.

Figure 1: High-frequency Indicators



Source: Authors' computation

The monthly data on evening hour luminosity has been obtained from the Visible Infrared Imaging Radiometer Suite (VIIRS) of National Aeronautics and Space Administration (NASA) and National Oceanic and Atmospheric Administration (NOAA), USA. VIIRS data has a wider radiometric detection range than former generation of similar satellites, which solves the issue of over-saturation at bright core centres (Elvidge et al. 2013). However, the publicly available VIIRS data still requires processing before use, as some temporary lights and background noise remain. We follow the procedure discussed in Beyer et al. (2018) and remove all observations from areas categorized as background noise mask. After outlier removal, these areas are identified by clustering the remaining observations based on their intensity.

4. Results

4.1. *Correlations of High-frequency Indicators with Trade GVA and Nightlight*

In Table 2, we report the correlations of all high-frequency indicators and trade GVA in terms of both levels and y-o-y growth. We have converted all monthly series to quarterly frequency. The level correlations use 21 observations, from the 3rd quarter of 2012–13 to the 3rd quarter of 2017–18. Except for PMI-services and commercial vehicles, all the other indicators are correlated at the 5 per cent significance level with trade GVA. The cargo traffic air exhibits the highest correlation with GVA with a correlation coefficient of 0.93 followed by cargo traffic port. The level correlation between trade GVA and nightlight is 0.82 and is also statistically significant at the 5 per cent level. PMI-services is the only variable negatively correlated with GVA. The year-on-year (y-o-y) growth correlation calculations use 17 observations starting from the 3rd quarter of 2013–14 to the 3rd quarter of 2017–18. The variables railway freight, foreign tourist arrivals, PMI-services, and nightlight intensity exhibit a negative growth correlation with the trade GVA. None of the variables displays a statistically significant correlation. The y-o-y growth in cargo traffic port exhibits the highest correlation with the GVA growth with a correlation coefficient of 0.38 followed by PMI-services growth. As expected, the growth correlations are significantly smaller as compared to the level correlations.

Table 2: Correlations of High-frequency Indicators and Trade GVA

Indicator	Levels	Year-on-year Growth
Commercial Vehicles ('000)	0.4022(21)	0.0906(17)
Railway Freight ('000 tonnes)	0.8288(21)	-0.2073(17)
Cargo Traffic Port ('000 tonnes)	0.9266(21)	0.3769(17)
Cargo Traffic Air ('000 tonnes)	0.9313(21)	0.0392(17)
Foreign Tourist Arrivals	0.9074(21)	-0.0101(17)
Telecommunication Subscriber	0.9232(21)	0.1737(17)
PMI-Services	-0.0682(21)	-0.2618(17)
Nightlight	0.8205(21)	-0.1019(17)

Note: *The numbers in parentheses in the table denote the number of observations used for correlation calculation.
Source: Authors' computation

In Table 3, we report the correlations between the levels and y-o-y growth of all the high-frequency indicators and nightlight intensity. In terms of the levels, except for PMI-services and commercial vehicles, all the series have a strong correlation with nightlight intensity. This is evident from the fact that the correlation coefficient values for all the other indicators are statistically significant at the 5 per cent level. The indicator 'foreign tourist arrivals' has the highest correlation with GVA with a correlation coefficient of 0.92, followed by cargo traffic air. PMI-Services is the only variable with a level that is negatively correlated with nightlight. In terms of y-o-y growth, except for railway freight and cargo traffic port, all the series depict a positive correlation with nightlight intensity. However, the correlation coefficients for all the indicators except cargo traffic port are statistically insignificant at the 5 per cent level, implying a poor correlation between the growth patterns of the indicators and nightlight.

Table 3: Correlation of High-frequency Indicators with Nightlights

Indicator	Levels	Year-on-year Growth
Commercial Vehicles ('000)	0.3121(21)	0.0745(17)
Railway Freight ('000 tonnes)	0.8966(21)	-0.0332(17)
Cargo Traffic Port ('000 tonnes)	0.786(21)	-0.6011(17)
Cargo Traffic Air ('000 tonnes)	0.9219(21)	0.4583(17)
Foreign Tourist Arrivals	0.9236(21)	0.4742(17)
Telecommunication Subscriber	0.9116(21)	0.3902(17)
PMI-Services	-0.1353(21)	0.0507(17)

Note: *The numbers in parentheses in the table denote the number of observations used for the correlation calculation

Source: Authors' computation

4.2. *Data Shrinking Procedures*

4.2.1. Principal Component Analysis

The sequences of nowcasts using principal components for each reference quarter have been generated from six datasets (details of all the six datasets have been provided in Appendix C). As discussed, the principal component method, unlike the dynamic factor model method, is not equipped to handle jagged-edge data.

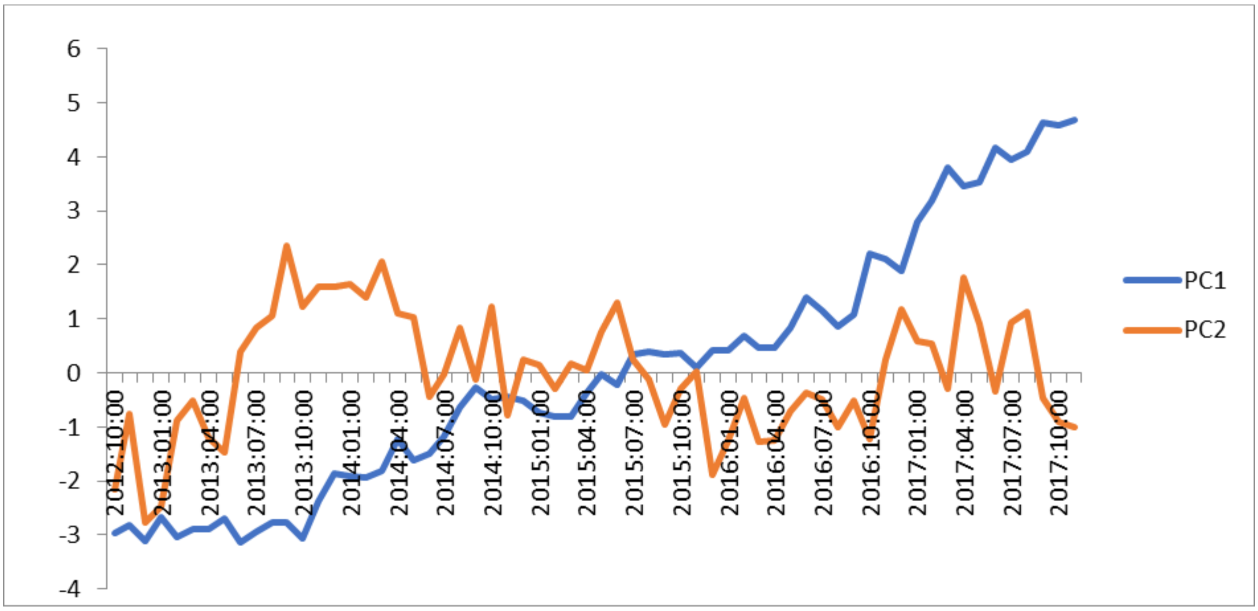
Table 4: Eigenvalues and Proportions Corresponding to Principal Components

Component	Eigenvalue	Proportion	Cumulative
1	5.4344	0.6793	0.6793
2	1.2785	0.1598	0.8391
3	0.6368	0.0796	0.9187
4	0.2961	0.037	0.9557
5	0.2503	0.0313	0.987
6	0.0544	0.0068	0.9938
7	0.0272	0.0034	0.9972
8	0.0222	0.0028	1

Source: Authors' computation

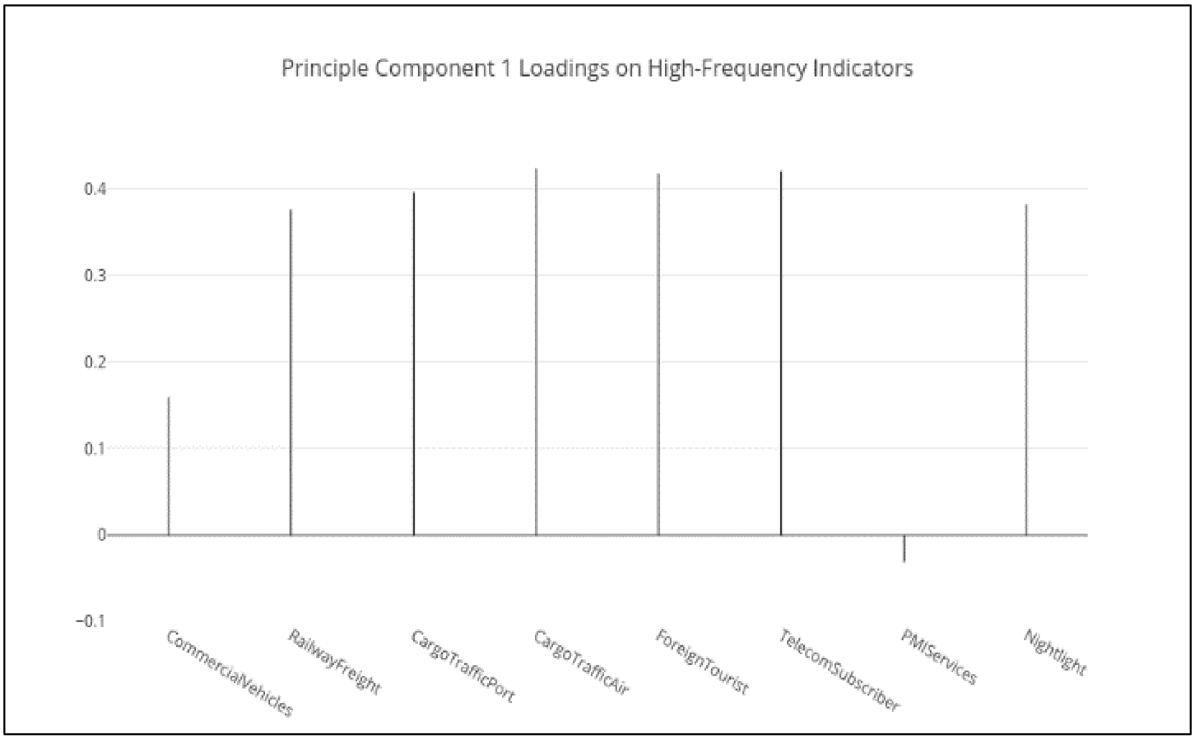
Table 4 reports Eigenvalues corresponding to the eight principal components. Based on the Eigenvalues, only the principal components with Eigenvalues greater than the mean are selected. This is because the first principal component already explains 68 per cent of the variance in the dataset and the second principal component explains around 16 per cent. Together they explain 84 per cent of the total variance in the dataset and only these two principal components are used for analysis. Figure 2 represents the time-series plot of the first two principal components, derived from eight high-frequency indicator variables using the data shrinkage method. Figure 3 represents the loading of the high-frequency monthly indicators on principal components. Variables such as cargo traffic via air, foreign tourist arrivals, and the telecommunication subscriber base are heavily loaded on the first principal component.

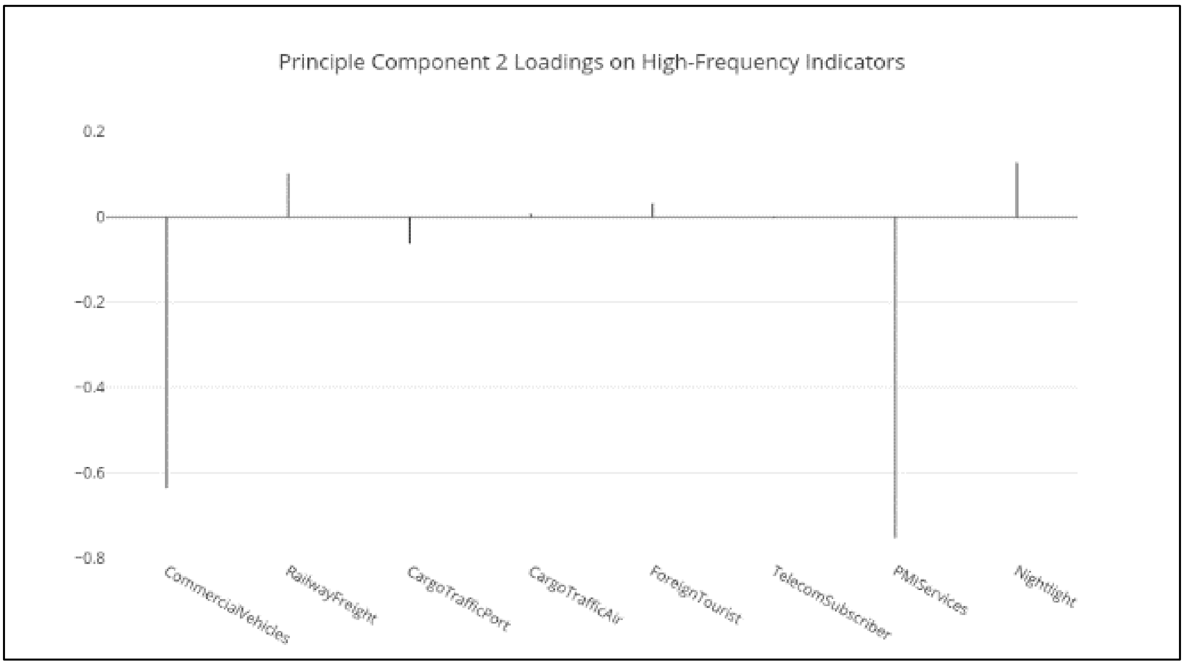
Figure 2: First Two Principal Components of Trade GVA Indicators



Source: Authors' computation

Figure 3: Loadings on Principal Components





Source: Authors' computation

4.2.2. DYNAMIC FACTOR MODEL

The sequences of nowcasts using dynamic factors (DFs) for each reference quarter have been generated from six datasets (more details about the datasets are provided in Appendix D). Note that the DFM is equipped to handle the non-synchronous data series.

Figure 4: Non-synchronous Nature of Data Releases in India and Jagged-edge Time-series Data

Month	Vehicle	Rail	PortCargo	AirCargo	Tourist	Telecom	PMI	Nightlight
Sep, 2012	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕
⋮	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕
Nov, 2017	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕
Dec, 2017	×	⊕	×	×	⊕	×	⊕	⊕
Jan, 2018 → N(-1)	×	×	×	×	×	×	×	×
Feb 28, 2018 → Official Release	×	×	×	×	×	×	×	×

Source: Authors' collation from government release calendar

The model selection results are shown in Table 5. Generally, the model with the lowest Akaike's Information Corrected Criterion (AICC) has three trends. It also appears that models with a diagonal and unequal R matrix fit the data much better than the other models with different complex structures for the observation errors (i.e. models with diagonal and unequal forms for R account for nearly all the AICC weight).

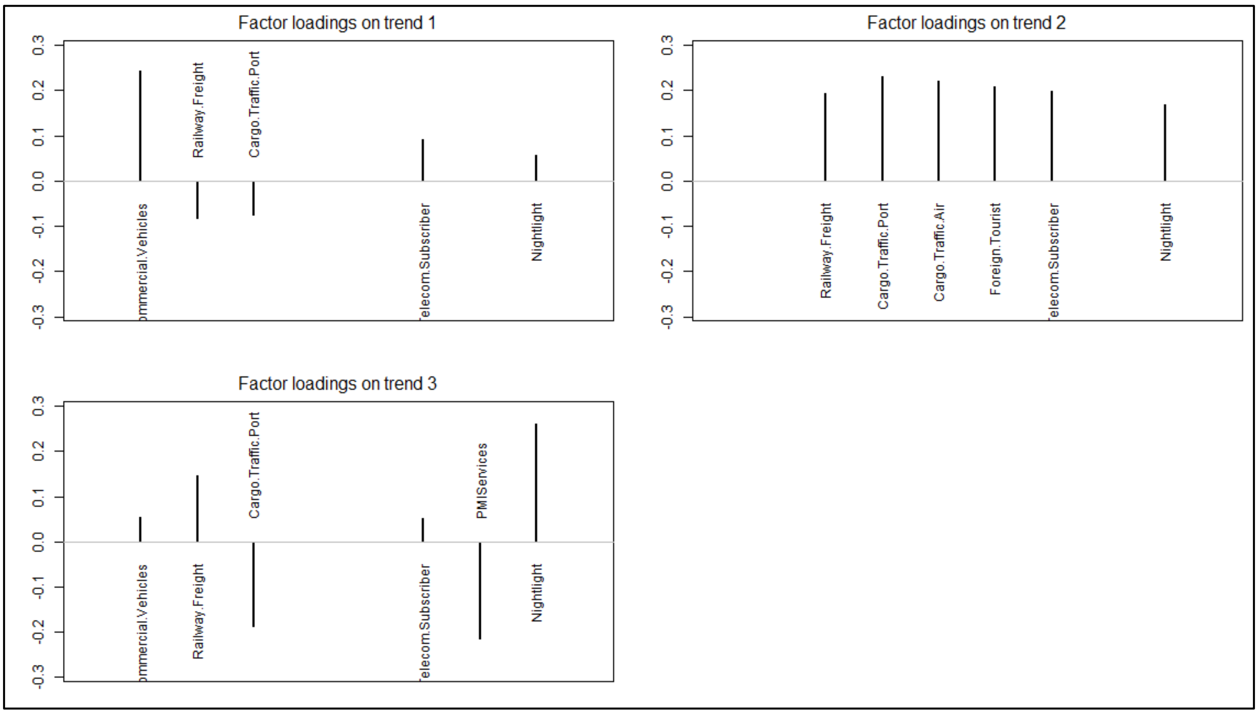
Table 5: Results of the Dynamic Factor Model Selection

S. No.	R	m	logLik	K	AICc	delta.AICc	Ak.wt
1	diagonal and equal	1	-453.6953	9	925.758	412.544727	2.26E-90
2	diagonal and equal	2	-367.4622	16	768.0507	254.83752	3.97E-56
3	diagonal and equal	3	-319.5883	22	685.2981	172.084875	3.70E-38
4	diagonal and equal	4	-283.7961	27	624.7956	111.582381	5.09E-25
5	diagonal and equal	5	-232.6614	31	531.562	18.348804	8.95E-05
6	diagonal and equal	6	-233.4859	34	540.09	26.876811	1.26E-06
7	diagonal and equal	7	-234.0485	36	545.8507	32.637487	7.06E-08
8	diagonal and unequal	1	-254.8969	16	542.9201	29.706831	3.06E-07
9	diagonal and unequal	2	-234.2965	23	516.9124	3.699199	1.36E-01
10	diagonal and unequal	3	-225.7556	29	513.2132	0	8.63E-01
11	diagonal and unequal	4	-227.0401	34	527.1985	13.985252	7.93E-04
12	diagonal and unequal	5	-237.0755	38	556.5805	43.367304	3.30E-10
13	diagonal and unequal	6	-255.2969	41	600.1135	86.900293	1.16E-19
14	diagonal and unequal	7	-267.973	43	630.2442	117.030923	3.34E-26
15	equalvarcov	1	-440.6627	10	901.7752	388.56197	3.64E-85
16	equalvarcov	2	-363.6825	17	762.6348	249.421539	5.96E-55
17	equalvarcov	3	-321.6211	23	691.5615	178.348241	1.62E-39
18	equalvarcov	4	-290.8041	28	641.0562	127.843009	1.50E-28
19	equalvarcov	5	-248.0404	32	564.6032	51.390005	5.98E-12
20	equalvarcov	6	-241.6439	35	558.7188	45.505523	1.13E-10
21	equalvarcov	7	-240.2278	37	560.5421	47.328877	4.56E-11
22	unconstrained	1	-252.6281	44	601.9595	88.74625	4.63E-20
23	unconstrained	2	-224.6458	51	563.1309	49.91772	1.25E-11
24	unconstrained	3	-207.2521	57	543.4634	30.250207	2.33E-07
25	unconstrained	4	-198.2857	62	538.4479	25.234624	2.86E-06
26	unconstrained	5	-194.3115	66	541.0479	27.834668	7.80E-07
27	unconstrained	6	-193.1692	69	546.8034	33.5902	4.39E-08
28	unconstrained	7	-190.8326	71	547.5531	34.339876	3.02E-08

Source: Authors' computation

There are many ways of doing factor rotations and we have adopted the varimax rotation, which seeks a rotation matrix H that creates the largest difference between loadings. The rotated factor loadings for the best model are shown in Figure 5.

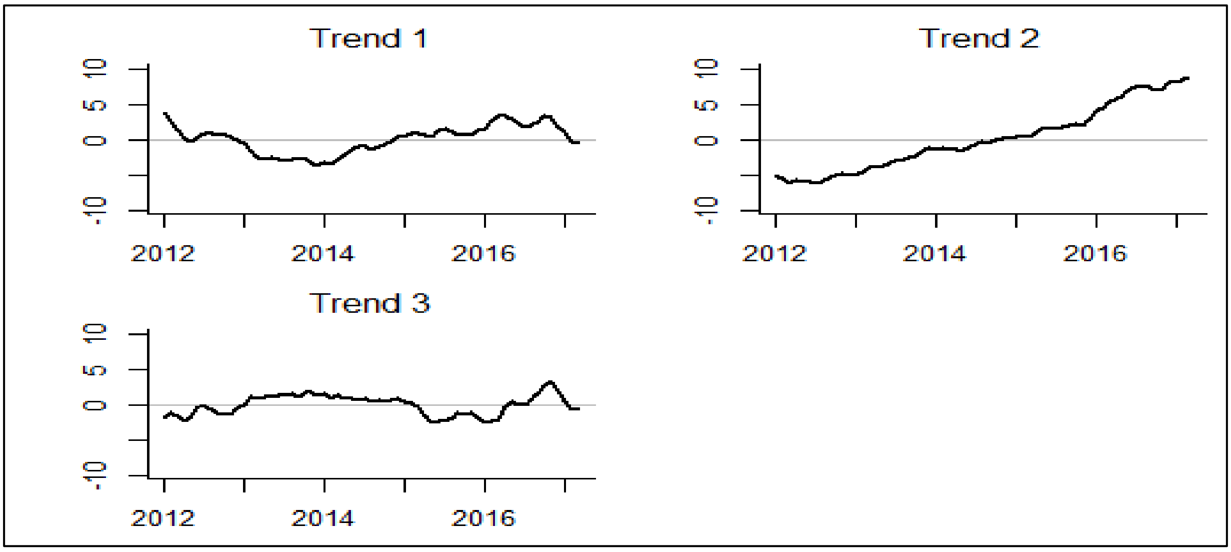
Figure 5: The Rotated Factor Loading from the Best Fitting Model



Source: Authors' computation

The plot of the unobserved trends (following varimax rotation) from the best model fit to the economic indicator data is shown in Figure 6.

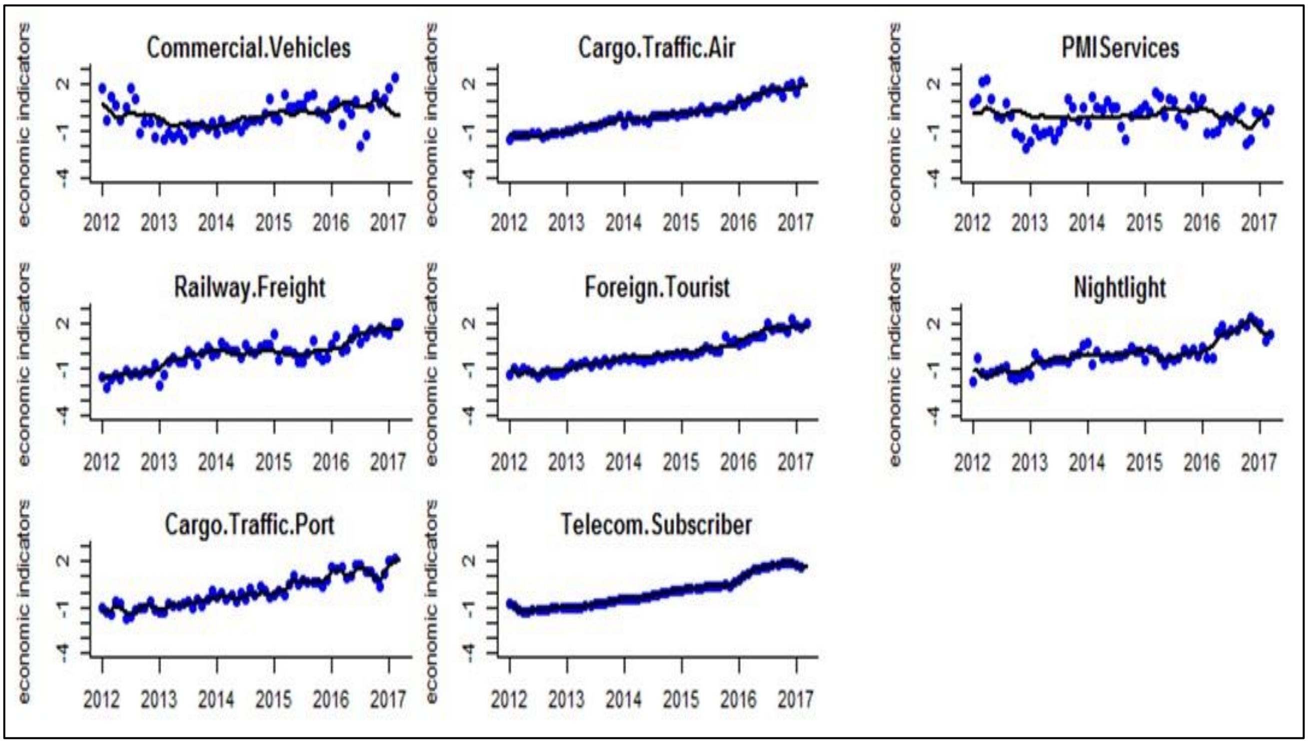
Figure 6: Plot of the Unobserved Trend Following Varimax Rotation



Source: Authors' computation

Finally, the results are shown for the best model with the appropriate factor and trend rotations. The plots of the model fits are shown in Figure 7. This model was run to convergence by using a very high number of maximum iterations (maxit=25000). The model does an adequate job of capturing some of the high-frequency variation in the time series, though the overall fit is much better for some than for other time series.

Figure 7: Plots of the Indicators with Model Fits (Dark Lines) from a Model with Trends



Source: Authors' computation

4.3. Nowcasting

Table 6 shows the root mean squared error (RMSE) for the nowcasts for the third and fourth quarters of the calendar year 2017.[1] Growth in the third quarter, at 9.1 per cent, has been slightly higher than the average over the last few years. Growth in the fourth quarter moderated slightly to 8.6 per cent. As benchmarks for the different nowcast models, we employ a naïve and an autoregressive nowcast. The former simply assumes that growth in an ongoing quarter will be the same as in the last quarter. The latter is based on an autoregressive model of order one and a constant, and does not rely on any other data. The average nowcast error over the two quarters of the naïve forecast is 0.66; and the average forecast error of the second benchmark is 0.53. For both benchmarks, the nowcast is closer to the actual growth in the fourth quarter. Since no high-frequency indicators are employed here, the nowcast remains the same throughout the quarter for which trade GVA is nowcast.

[1] The months of July, August, and September correspond to Q3 of a calendar year and Q2 of a fiscal year. Similarly, the months of October, November, and December correspond to Q4 of a calendar year and Q3 of a fiscal year. The actual nowcasts are shown in Table E1 and Table E2 in the Appendix.

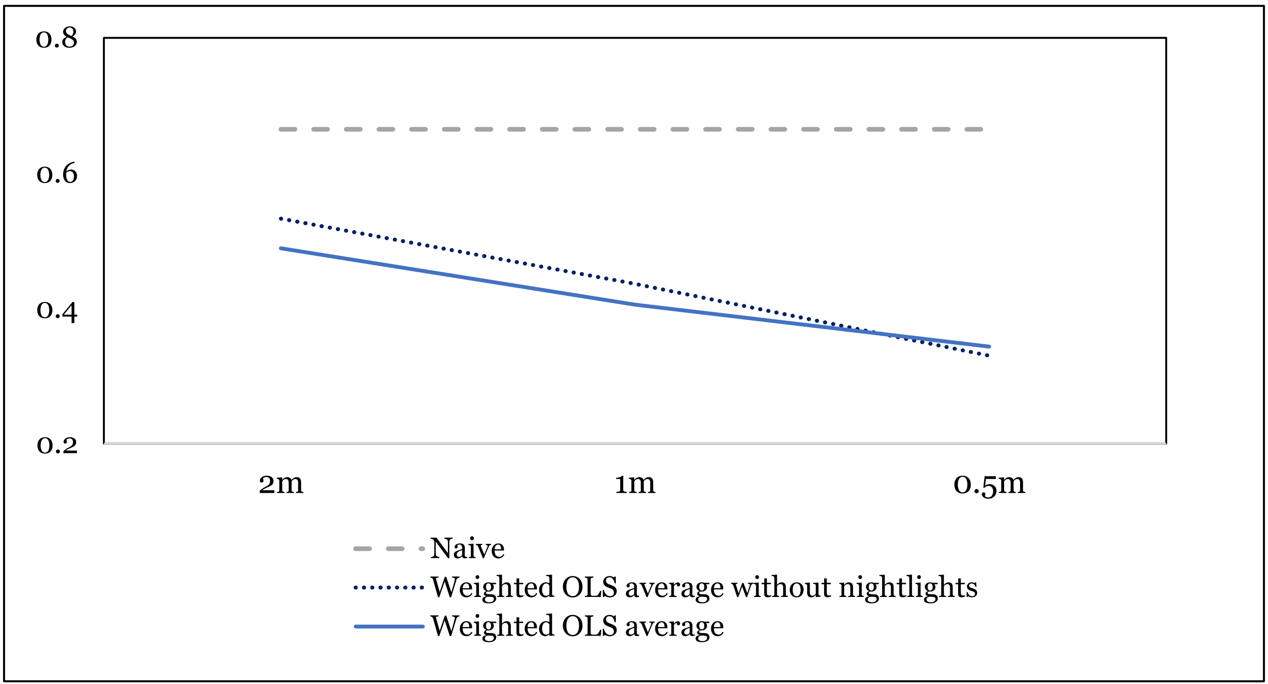
Table 6: Out-of-sample Root Mean Squared Error for Nowcasts Based on High-frequency Indicators

	2017Q3			2017Q4		
	2m	**1m**	**0.5m**	**2m**	**1m**	**0.5m**
Benchmarks						
Naïve	0.80	0.80	0.80	0.53	0.53	0.53
AR1	0.83	0.83	0.83	0.22	0.22	0.22
Ordinary Least Squares						
Weighted Average	0.88	0.79	0.66	0.10	0.02	0.03
Weighted Average without Nightlights	0.90	0.80	0.65	0.17	0.07	0.01
Cargo Traffic Air	0.81	0.82	0.64	0.33	0.31	0.26
Cargo Traffic Port	0.74	0.99	0.93	0.23	0.33	0.38
Nightlight Intensity	0.76	0.70	0.70	0.39	0.32	0.32
Services PMI	1.19	0.83	0.83	0.34	0.13	0.13
Railway Freight	0.47	0.19	0.19	0.63	0.91	0.91
Telecom Subscribers	0.98	0.84	0.69	0.33	0.64	0.43
Foreign Tourist Arrivals	0.68	0.71	0.71	0.33	0.32	0.32
Commercial Vehicles	1.43	1.22	0.56	0.10	0.19	0.74
Mixed Data Sampling						
Weighted Average	1.09	0.49	0.68	0.33	0.10	0.25
Weighted Average without Nightlights	1.20	0.50	0.72	0.23	0.23	0.40
Cargo Traffic Air	0.77	2.99	0.62	1.87	1.12	0.65
Cargo Traffic Port	1.78	1.32	1.09	1.51	1.39	1.79
Nightlight Intensity	0.25	0.46	0.46	1.02	0.79	0.79
Services PMI	1.55	0.29	0.29	0.22	0.85	0.85
Railway Freight	1.21	0.78	0.78	0.47	0.92	0.92
Telecom Subscribers	0.17	1.39	2.19	0.13	0.48	0.16
Foreign Tourist Arrivals	1.21	1.37	1.37	0.05	0.03	0.03
Commercial Vehicles	1.74	1.91	0.49	0.37	0.33	0.61

Source: Authors' computation

Next, we compute nowcasts based on the individual high-frequency indicators and then combine the different nowcasts by weighting the in-sample-fit of the different high-frequency indicators. The average weighted OLS forecast error is lower than that of both benchmarks and lower than any nowcast based on only one high-frequency indicator. It improves the nowcast performance, especially for the fourth quarter, in which the nowcasts are very close to the actual growth, and as more data becomes available. This can be seen in Figure 8, which shows the average forecast error of different nowcasting models. Shortly before the growth estimates are officially announced, the forecast error using the weighted OLS model is only half as large as for the naïve model. In line with expectations, nightlight intensity improves the nowcast early in the quarter. Shortly before the official growth is announced, however, they do not provide any additional value for the two quarters analysed here.

Figure 8: Average Root Mean Squared Error from Different Nowcast Models



Source: Authors' computation

The nowcasts based on mixed data sampling are a bit of a mixed bag—sometimes closer to the actual growth (e.g. 2017Q3 - 1m) or very similar (e.g., 2017Q3 – 0.5m) and worse in some cases (e.g. 2017Q4). Whether mixed data sampling improves forecasts or not depends on whether the relative weight of the monthly observations in explaining the quarterly GVA in the in-sample estimation period is the same as in the out-of-sample forecast period. If one monthly observation of a series with a good fit in the past is off during the nowcast period, the nowcast will perform worse than in the OLS case. More out-of-sample forecasts will be needed to determine whether mixed data sampling is improving the overall nowcasts for this sub-sector.

Table 7: Out-of-sample Root Mean Squared Error for Nowcasts Based on Shrinking Methods

	2017Q3			2017Q4		
	2m	**1m**	**0.5m**	**2m**	**1m**	**0.5m**
Principal Components						
OLS with nightlights	0.95	0.90	0.88	0.16	0.03	0.08
OLS without nightlights	0.81	0.80	0.66	0.07	0.31	0.02
MIDAS with nightlights	0.95	0.90	0.88	0.16	0.03	0.08
MIDAS without nightlights	0.81	0.66	0.66	0.07	0.02	0.02
Dynamic Factor						
OLS with nightlights	1.43	0.83	0.81	0.34	0.30	0.19
OLS without nightlights	0.84	1.01	0.73	0.31	0.28	1.04
MIDAS with nightlights	0.95	0.75	0.21	0.16	1.21	1.05
MIDAS without nightlights	0.81	1.07	1.15	0.07	0.04	0.53

Source: Authors' computation

As regards the nowcasts presented so far, the nowcast performance based on principal components and dynamic factors is much better for the fourth quarter than for the third. The nowcasts based on the first two principal components are similar to those using the weighted average of all the high-frequency indicators. While in general the nowcast error becomes smaller over time, there are a few exceptions in which the nowcasts worsen as more data becomes available. In contrast to the weighted OLS nowcasts, including nightlight intensity does not improve the nowcasts here. In a few cases, dynamic factors improve the nowcast as compared to principal components, but somewhat surprisingly, the forecast performance deteriorates in some cases. More out-of-sample nowcasts are needed to establish whether dynamic factors are improving over the other models or not.

4.4. Forecasting

In addition to nowcasting trade GVA, all the models discussed in the paper can be used to compute forecasts. To do so, first individual series are forecast using a simple autoregressive model. Based on the nowcasting models described above, one can then provide a forecast for quarters in the future. In Table 8, we present the one quarter ahead forecasting errors for the fourth quarter of calendar year 2017 and the first quarter of 2018.[2]

[2] The actual forecasts are shown in Table F1 and Table F2 in the Appendix.

Table 8: One Quarter Ahead Forecast RMSE Based on High-frequency Indicators

	2017Q4			2018Q1		
	2m	**1m**	**0.5m**	**2m**	**1m**	**0.5m**
Benchmarks						
Naïve	0.27	0.27	0.27	2.80	2.80	2.80
AR1	0.30	0.30	0.30	2.01	2.01	2.01
Ordinary Least Squares						
Weighted Average	0.20	0.22	0.21	2.12	2.21	2.09
Weighted Average without nightlights	0.20	0.21	0.21	2.13	2.23	2.09
Cargo Traffic Air	0.22	0.19	0.29	2.05	2.08	2.06
Cargo Traffic Port	0.17	0.00	0.13	2.17	2.11	1.80
Nightlight Intensity	0.20	0.23	0.23	2.01	2.07	2.07
Services PMI	0.16	0.28	0.28	2.08	2.00	2.00
Railway Freight	0.30	0.49	0.49	2.10	2.29	2.29
Telecom Subscribers	0.58	0.49	0.01	2.31	2.59	2.20
Foreign Tourist Arrivals	0.32	0.24	0.24	2.09	2.08	2.08
Commercial Vehicles	0.32	0.19	0.27	2.13	2.42	2.20
Mixed Data Sampling						
Weighted Average	0.27	0.24	0.48	2.13	2.50	2.25
Weighted Average without nightlights	0.28	0.32	0.49	2.16	2.57	2.29
Cargo Traffic Air	0.18	2.42	2.00	2.48	4.78	3.97
Cargo Traffic Port	0.32	0.20	0.31	2.00	1.82	0.95
Nightlight Intensity	0.22	0.36	0.36	1.94	2.00	2.00
Services PMI	1.26	0.39	0.39	1.91	1.84	1.84
Railway Freight	0.16	0.13	0.13	2.40	2.38	2.38
Telecom Subscribers	0.10	0.20	0.59	2.25	2.51	2.11
Foreign Tourist Arrivals	0.27	0.43	0.43	2.10	2.11	2.11
Commercial Vehicles	0.11	0.39	0.23	1.96	2.52	2.66

Source: Authors' computation

The one quarter ahead forecast error for the fourth quarter of 2017 is much smaller than the forecast error for the first quarter of 2018. It is around 0.2 for the former, but close to 2.2 for the latter. For the first forecast, the one step ahead forecasts based on the weighted OLS model beats both benchmarks and mixed data sampling does not improve over OLS. The OLS forecast is roughly a third closer to actual growth than the two benchmarks. For the second, all high-frequency indicators point to higher than announced growth. The actual growth rate for this quarter has just been released and may still be revised. Based on our model, we expect a (strong) upward revision.

Table 9: One Quarter Ahead Forecast RMSE Based on Shrinking Methods

	2017Q4			2018Q1		
	2m	**1m**	**0.5m**	**2m**	**1m**	**0.5m**
Principal Components						
OLS with nightlights	0.31	0.18	0.18	2.22	2.28	2.19
OLS without nightlights	0.58	0.56	0.48	1.94	1.73	2.09
MIDAS with nightlights	0.31	0.18	0.18	2.22	2.28	2.19
MIDAS without nightlights	0.58	0.48	0.48	1.94	2.09	2.09
Dynamic Factor						
OLS with nightlights	0.52	0.51	0.39	2.23	2.05	2.25
OLS without nightlights	0.03	0.70	0.65	1.65	1.73	2.20
MIDAS with nightlights	0.31	0.18	0.18	2.22	2.28	2.19
MIDAS without nightlights	0.58	0.56	0.48	1.94	1.73	2.09

Source: Authors' computation

The results for the forecasts based on shrinking methods are again mixed (Table 9). The forecasts from the principal component model with nightlight intensity and the dynamic factor MIDAS model are the lowest in some cases, for example, in the fourth quarter of 2017. As expected, the methods cannot correct for the large forecast error in the first quarter of 2018.

5. Conclusion

In this paper, we have presented a new framework to nowcast India's GVA that incorporates information of mixed data frequencies and other data characteristics. In addition, we have added evening-hour luminosity as a crucial high-frequency indicator. The first results suggest that nightlights may improve nowcasts, especially early in a quarter, before other data becomes available (Table 9). More out-of-sample nowcasts are needed to determine whether differences in the forecast accuracy are statistically significant and accurately compare the forecasting performance of different models. This is a necessary condition to be able to judge whether nightlight intensity adds important information to now- and forecasts of trade GVA, and to determine the models with the best performance.

The information content of nightlight intensity can likely be strengthened. We used the same nightlight data that has been used in Beyer et al. (2018), but this cleaning procedure does not necessarily result in the best data for nowcasting, for which the cleaning procedure could be optimised.

Our nowcasting framework can be applied in the same manner to all other sub-components of the GVA. The only exception may be agriculture, for which we have so far been unable to identify suitable correlated high-frequency indicators, though increased availability of daytime satellite imagery of agricultural land and better weather data may already allow for nowcasting agricultural GVA in the same way. Once all the sub-components of GVA are nowcast, they can then be easily aggregated into an overall GVA nowcast. We plan to do this in future work.

While we will apply the framework to now- and forecast Indian GVA, the model setup captures data characteristics common to the EMEs and can hence be applied to other countries as well.

References

Bańbura, M. and M. Modugno. (2014). "Maximum Likelihood Estimation of Factor Models on Datasets with Arbitrary Pattern of Missing Data," *Journal of Applied Econometrics*, 29(1): 133–160.

Beyer, R., E. Chhabra, V. Galdo, and M. Rama. (2018). "Measuring Districts' Monthly Economic Activity from Outer Space," World Bank Policy Research Working Paper, No 8503.

Bhadury, Soumya, and S. Pohit. (2017). "NCAER's nowcasting Model of India: A New approach to know the Now-GDP," in *The Macolm S Adiseshiah 2017-18 Mid-Year Reviewof the Economy, NCAER, 11 November, 2017.*

Bhadury, Soumya, and S. Pohit. (2018). "Nowcasting' the Indian economy: A new approach to know the now-GDP," (with S. Bhadury), *Ideas for India,* 12 Jan,

, accessed June 18, 2018.

Bhattacharya, R., R. Pandey, and G. Veronese. (2011). "Tracking India Growth in Real-time," *National Institute of Public Finance and Policy Working Paper 11/90*, https://ideas.repec.org/p/npf/wpaper/11-90.html, accessed June 25, 2018.

Bragoli, D. and J. Fosten. (2016). "Nowcasting Indian GDP (No. 2016-06)," *School of Economics Working Paper*, University of East Anglia, Norwich, UK.

Dahlhaus, T., J.-D. Guénette, and G. Vasishtha. (2017). „Nowcasting BRIC+M in Real-time," *International Journal of Forecasting*, 33(4): 915–935.

Doz, C., D. Giannone, and L. Reichlin. (2007). "A Two-Step Estimator for Large Approximate Dynamic Factor Models Based on Kalman Filtering," CEPR Discussion Paper No. 6043.

Doz, C., D. Giannone, and L. Reichlin. (2012). "A Quasi–maximum Likelihood Approach for Large, Approximate Dynamic Factor Models," *The Review of Economics and Statistics*, 94(4): 1014–1024.

Elvidge, C.D., K.E. Baugh, M. Zhizhin, and F.C. Hsu. (2013). "Why VIIRS Data Are Superior to DMSP for Mapping Night-time Lights," *Proceedings of the Asia-Pacific Advanced Network*, 35: 62–69.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin. (2000). "The Generalized Dynamic Factor Model: Identification and Estimation," *The Review of Economics and Statistics*, 82(4): 540–554.

Forni, M. and M. Lippi. (2001). "The Generalized Dynamic Factor Model: Representation Theory," *Econometric Theory*, 17(06): 1113–1141.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin. (2003). "Do Financial Variables Help Forecasting Inflation and Real Activity in the Euro Area?" *Journal of Monetary Economics*, 50(6): 1243–1255.

Giannone, D., L. Reichlin, and D. Small. (2008). "Nowcasting: The Real-time Informational Content of Macroeconomic Data," *Journal of Monetary Economics,* 55(4): 665–676.

Ghysels, E., P. Santa-Clara, and R. Valkanov. (2004). *The MIDAS Touch: Mixed Data Sampling Regression Models*. University of North Carolina. Mimeo, http://docentes.fe.unl.pt/~psc/MIDAS.pdf, accessed June 25, 2018.

Ghysels, E., A. Sinko, and R. Valkanov (2007). “MIDAS Regressions: Further Results and New Directions,” *Econometric Reviews*, 26(1): 53–90. https://doi.org/10.1080/07474930600972467

Henderson, V., A. Storeygard, and D.N. Weil. (2011). “A Bright Idea for Measuring Economic Growth,” *The American Economic Review*, 101(3): 194–199.

Marcellino, M. and C. Schumacher. (2007). *Factor-MIDAS for Nowcasting and Forecasting with Ragged-edge Data: A Model Comparison for German GDP*. Frankfurt am Main: Dt. Bundesbank.

Sapre, A. and R. Sengupta. (2017). “An Analysis of Revisions in Indian GDP Data”, *NIPFP Working Paper Series WP-213-2017*, November, http://macrofinance.nipfp.org.in/PDF/SapreSengupta_WP_2017_213.pdf, accessed June 25, 2018.

Stock, J.H. M.W. Watson. (2002). “Forecasting Using Principal Components from a Large Number of Predictors,” *Journal of the American Statistical Association*, *97*(460): 1167–1179.

World Bank. 2017. *South Asia Economic Focus, Fall 2017: Growth Out of the Blue*. Washington, DC: World Bank.

Zuur, A., I. Tuck, and N. Bailey. (2003).   "Dynamic Factor Analysis to Estimate Common Trends in Fisheries Time Series,"Canadian Journal of Fisheries and Aquatic Sciences, 60(5): 542–552.

Appendices

Appendix A

The DFA in the MARSS package has the following structure:

- Observation (y) are modeled as linear combination of hidden trends (x) and factor loadings (Z) plus some offsets a

  $x_t = x_{t-1} + w_t$ where $w_t \sim MVN(0,Q)$

  $y_t = Zx_t + a + v_t$ where $v_t \sim MVN(0,R)$

  $x_0 \sim MVN(\Pi,\Lambda)$
- It is important to write the DFA model in MARSS form. Let's say there is a data set with six observed time series i.e. n=6.
- And it requires to fit a model with three hidden trends, m=3.
- Writing the DFA model in MARSS matrix form (ignoring the error structure and initial conditions for now).

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}_t = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}_{t-1} + \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}_t$$

- Notice the process error of the hidden trend, $w_t \sim MVN(0,Q)$ can be written as follows:

$$\begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}_t \sim MVN(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{bmatrix})$$

- The matrix form representation of the equation between (y), hidden trend (x) and factor loading (Z) is as follows:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{bmatrix}_t = \begin{bmatrix} z_{11} & z_{12} & z_{13} \\ z_{21} & z_{22} & z_{23} \\ z_{31} & z_{32} & z_{33} \\ z_{41} & z_{42} & z_{43} \\ z_{51} & z_{52} & z_{53} \\ z_{61} & z_{62} & z_{63} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}_t + \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{bmatrix}_t$$

- The observation error can be written as: $v_t \sim MVN(0,R)$

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{bmatrix}_t \sim MVN(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} r_{11} & r_{12} & r_{13} & r_{14} & r_{15} & r_{16} \\ r_{21} & r_{22} & r_{23} & r_{24} & r_{25} & r_{26} \\ r_{31} & r_{32} & r_{33} & r_{34} & r_{35} & r_{36} \\ r_{41} & r_{42} & r_{43} & r_{44} & r_{45} & r_{46} \\ r_{51} & r_{52} & r_{53} & r_{54} & r_{55} & r_{56} \\ r_{61} & r_{62} & r_{63} & r_{64} & r_{65} & r_{66} \end{bmatrix})$$

- Harvey's identifiability constraints are written below:
- If Z, a, Q are not constrained, then the DFA model is unidentifiable.
- In the first $m-1$ rows of Z, the $z-value$ in the $j-th$ column and the $i-th$ row set to zero, if $j > i$.

- a is constrained so that first m values are set to zero.
- Q is set equal to the identity matrix (I_m).
- Zuur et al. (2003) found that Harvey's second constraint, the EM algorithm, is not particularly robust and takes time to converge.
- Zuur et al. (2003) found that EM algorithm behaves better if you constrain each of the time series in x to have a mean of zero across $t=1$ to T
- Zuur et al. (2003) replaced the estimates of hidden state, x_t^T , coming out of the Kalman smoother with $x_t^T - \bar{x}$ for $t=1$ to T ; where $\bar{x}$ is mean of x_t across t .
- With this approach, you estimate all of the a elements, which represents the average level of y_t relative to $Z(x_t - \bar{x})$.
- However, it was found out that demeaning x_t^T in this way can cause the EM algorithm to have errors (decline in log-likelihood).
- Instead, demeaning data is followed by fixing all elements of a to zero.
- Using this revised constraints, DFA will become:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}_t = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}_{t-1} + \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}_t ,$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{bmatrix}_t = \begin{bmatrix} z_{11} & 0 & 0 \\ z_{21} & z_{22} & 0 \\ z_{31} & z_{32} & z_{33} \\ z_{41} & z_{42} & z_{43} \\ z_{51} & z_{52} & z_{53} \\ z_{61} & z_{62} & z_{63} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}_t + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{bmatrix}_t$$

$$\begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}_t \sim MVN(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix})$$

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{bmatrix}_t \sim MVN(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} r_{11} & r_{12} & r_{13} & r_{14} & r_{15} & r_{16} \\ r_{21} & r_{22} & r_{23} & r_{24} & r_{25} & r_{26} \\ r_{31} & r_{32} & r_{33} & r_{34} & r_{35} & r_{36} \\ r_{41} & r_{42} & r_{43} & r_{44} & r_{45} & r_{46} \\ r_{51} & r_{52} & r_{53} & r_{54} & r_{55} & r_{56} \\ r_{61} & r_{62} & r_{63} & r_{64} & r_{65} & r_{66} \end{bmatrix})$$

- To complete our model, it is required to set the initial condition of the state.
- Following Zuur et al. (2003), the initial state vector (x_0) is set to have zero mean and diagonal variance-covariance matrix with large variance.

$$x_0 \sim MVN(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{bmatrix})$$

We assured parameter convergence by using enough iterations.

Appendix B

Table B.1: Data Release Calendar for All Blocs Contributing to GVA

AREA	INDICATORS	CALENDER RELEASE DATE	REFERENCE PERIOD
Mining and Quarrying	Mining and quarrying Index	January 12, 2018	November, 2017
	Monthly Production of Coal	January 31, 2018	December, 2017
	Monthly Production of Crude Oil	January 31, 2018	December, 2017
Manufacturing	Manufacturing Index	January 12, 2018	November, 2017
	Monthly Production of Steel and Fertilisers	January 31, 2018	December, 2017
	Monthly Production of Fertilisers	January 31, 2018	December, 2017
	Nikkei PMI Manufacturing Index	January 2, 2018	December, 2017
	Commercial Vehicle Production	January 15, 2018	November, 2017
	Commercial of two wheeler production	January 15, 2018	November, 2017
	Passenger Car Production	January 15, 2018	November, 2017
	Merchandise non-oil imports	January 15, 2018	December, 2017
Electricity, Gas, Water Supply and Construction	Electricity Index	January 31, 2018	December, 2017
	Monthly Crude Oil Production	January 31, 2018	December, 2017
	Monthly Cement Production	January 31, 2018	December, 2017
	Oil Imports	January 15, 2018	December, 2017

Trade, Hotels, Transport, Communication and Services Related to Broadcasting	Railway Freight Traffic of Major Commodities	Between January 1 and January 10, 2018	December, 2017
	Cargo Traffic – Ports	January 15, 2018	November, 2017
	Cargo Traffic – Air	January 15, 2018	November, 2017
	Foreign Tourist Arrivals in India	Between 10th January 10 and January 15, 2018	December, 2017
	Telecommunication Subscriber Base	January 11, 2018	November, 2017
	Nikkei PMI-Services Index	January 4, 2018	December, 2017
	Commercial Vehicles Production	January 15, 2018	November, 2017
	Nightlight	January 30, 2018	December, 2017
Financial, Real Estate and Professional Services	Bank Credit to Commercial Sectors	January 5, January 12, January 19, January 26, 2018	22nd December onwards(respectively)
	Nikkei PMIServices Index	January 4, 2018	December, 2017
	Foreign Institutional Investment	January 31, 2018	January, 2018
	FOREX (Foreign Exchange) Assets	January 5, January 12, January 19, January 26, 2018	December 22 onwards(respectively)
	NSE Trading Volume	Daily data in January, 2018	January, 2018
Public Administration, Defence, and Other Services	Expenditure of the Central Government Net of Interest Payments	Between January 24 and 31, 2018	November, 2017
	Reserve Bank of India(RBI) Net Credit to Government	5th January 5, January 12, January 19, and January 26, 2018	22nd December onwards(respectively)
	Receipts of Central Government	Between January 24 and January 31, 2018	November, 2017

Appendix C

- Datasets used for Principal Component Analysis (PCA)

Table C.1: Principal Components Dataset: Reference Quarter and Description

Dataset	Reference Quarter	Nowcast Horizon and Description
PC_WL_REF17Q3_N(-2)	2017Q3 (October, November, December)	Nowcast at the end of December, 2017, using Principal Components (PCs), i.e. two months prior to the official data release. The nowcast has been produced using nightlight data.
PC_WL_REF17Q3_N(-1)	2017Q3 (October, November, December)	Nowcast at the end of January, 2018, using Principal Components (PCs), i.e. one month prior to the official data release. The nowcast has been produced using nightlight data.
PC_WL_REF17Q3_N(-0.5)	2017Q3 (October, November, December)	Nowcast during mid-February using Principal Components (PCs), i.e. 15 days prior to the official data release. The nowcast has been produced using nightlight data.
PC_WoL_REF17Q3_N(-2)	2017Q3 (October, November, December)	Nowcast at the end of December, 2017, using Principal Components (PCs), i.e. two months prior to the official data release. The nowcast has been produced without using nightlight data.
PC_WoL_REF17Q3_N(-1)	2017Q3 (October, November, December)	Nowcast at the end of January, 2018, using Principal Components (PCs), i.e. one month prior to the official data release. The nowcast has been produced without using nightlight data.
PC_WoL_REF17Q3_N(-0.5)	2017Q3 (October, November, December)	Nowcast during mid-February using Principal Components (PCs), i.e. 15 days prior to the official data release. The nowcast has been produced without using nightlight data.

Appendix D

- Datasets used for Dynamic Factor Analysis (DFA).

Table D.1: Dynamic Factors Dataset: Reference Quarter and Description

Dataset	Reference Quarter	Nowcast Horizon and Description
DF_WL_REF17Q3_N(-2)	2017Q3 (October, November, December)	Nowcast at the end of December, 2017 using Dynamic Factors (DFs), i.e. two months prior to the official data release. The nowcast has been produced using nightlight data.
DF_WL_REF17Q3_N(-1)	2017Q3 (October, November, December)	Nowcast at the end of January, 2018 using Dynamic Factors (DFs), i.e. one month prior to the official data release. The nowcast has been produced using nightlight data.
DF_WL_REF17Q3_N(-0.5)	2017Q3 (October, November, December)	Nowcast during mid-February using Dynamic Factors (DFs), i.e. 15 days prior to the Official Data release. The nowcast has been produced using nightlight data.
DF_WoL_REF17Q3_N(-2)	2017Q3 (October, November, December)	Nowcast at the end of December, 2017 using Dynamic Factors (DFs), i.e. two months prior to the official data release. The nowcast has been produced without using nightlight data.
DF_WoL_REF17Q3_N(-1)	2017Q3 (October, November, December)	Nowcast at the end of January, 2018 using Dynamic Factors (DFs), i.e. one month prior to the official data release. The nowcast has been produced without using nightlight data.
DF_WoL_REF17Q3_N(-0.5)	2017Q3 (October, November, December)	Nowcast during mid-February using Dynamic Factors (DFs), i.e. 15 days prior to the official data release. The nowcast has been produced without using nightlight data.

Appendix E

Table E1: Out-of-sample Nowcasts Based on High-frequency Indicators

	2017Q3			2017Q4		
	2m	**1m**	**0.5m**	**2m**	**1m**	**0.5m**
Benchmarks						
Naïve	8.30	8.30	8.30	9.10	9.10	9.10
AR1	8.27	8.27	8.27	8.35	8.35	8.35
Ordinary Least Squares						
Weighted Average	8.22	8.31	8.44	8.47	8.55	8.60
Weighted Average without Nightlights	8.20	8.30	8.45	8.40	8.50	8.56
Cargo Traffic Air	8.29	8.28	8.46	8.24	8.26	8.31
Cargo Traffic Port	8.36	8.11	8.17	8.80	8.90	8.95
Nightlight Intensity	8.34	8.40	8.40	8.96	8.89	8.89
Services PMI	7.91	8.27	8.27	8.23	8.44	8.44
Railway Freight	8.63	8.91	8.91	7.94	7.66	7.66
Telecom Subscribers	8.12	8.26	8.41	8.90	9.21	9.00
Foreign Tourist Arrivals	8.42	8.39	8.39	8.24	8.25	8.25
Commercial Vehicles	7.67	7.88	8.54	8.47	8.76	9.31
Mixed Data Sampling						
Weighted Average	8.01	8.61	8.42	8.24	8.67	8.82
Weighted Average without Nightlights	7.90	8.60	8.38	8.34	8.80	8.97
Cargo Traffic Air	8.33	12.09	9.72	6.70	9.69	9.22
Cargo Traffic Port	7.32	7.78	8.01	10.08	9.96	10.36
Nightlight Intensity	8.85	8.64	8.64	7.55	7.78	7.78
Services PMI	7.55	9.39	9.39	8.35	9.42	9.42
Railway Freight	7.89	8.32	8.32	8.10	7.65	7.65
Telecom Subscribers	8.93	7.71	6.91	8.44	8.09	8.41
Foreign Tourist Arrivals	7.89	7.73	7.73	8.52	8.54	8.54
Commercial Vehicles	7.36	7.19	8.61	8.20	8.24	9.18

Source: Authors' computation

Table E2: Out-of-sample Nowcasts Based on Shrinking Methods

	2017Q3			2017Q4		
	2m	**1m**	**0.5m**	**2m**	**1m**	**0.5m**
Principal Components						
OLS with nightlights	8.15	8.20	8.22	8.41	8.60	8.65
OLS without nightlights	8.29	8.30	8.44	8.64	8.26	8.59
MIDAS with nightlights	8.15	8.20	8.22	8.41	8.60	8.65
MIDAS without nightlights	8.29	8.44	8.44	8.64	8.59	8.59
Dynamic Factor						
OLS with nightlights	7.67	8.27	8.29	8.91	8.27	8.38
OLS without nightlights	8.26	8.09	8.37	8.26	8.29	7.53
MIDAS with nightlights	8.15	9.85	8.89	8.41	7.36	7.52
MIDAS without nightlights	8.29	8.03	7.95	8.64	8.61	8.04

Source: Authors' computation

Appendix F

Table F1: One Quarter Ahead Forecasts Based on High-frequency Indicators

	2017Q4			2018Q1		
	2m	**1m**	**0.5m**	**2m**	**1m**	**0.5m**
Benchmarks						
Naïve	8.30	8.30	8.30	9.10	9.10	9.10
AR1	8.27	8.27	8.27	8.31	8.31	8.31
Ordinary Least Squares						
Weighted Average	8.37	8.35	8.36	8.42	8.51	8.39
Weighted Average without nightlights	8.37	8.36	8.36	8.43	8.53	8.39
Cargo Traffic Air	8.35	8.38	8.28	8.35	8.38	8.36
Cargo Traffic Port	8.40	8.57	8.70	8.47	8.41	8.10
Nightlight Intensity	8.37	8.34	8.34	8.31	8.37	8.37
Services PMI	8.41	8.29	8.29	8.38	8.30	8.30
Railway Freight	8.27	8.08	8.08	8.40	8.59	8.59
Telecom Subscribers	7.99	8.08	8.56	8.61	8.89	8.50
Foreign Tourist Arrivals	8.25	8.33	8.33	8.39	8.38	8.38
Commercial Vehicles	8.89	8.76	8.30	8.43	8.72	8.50
Mixed Data Sampling						
Weighted Average	8.30	8.81	8.09	8.43	8.80	8.55
Weighted Average without nightlights	8.29	8.89	8.08	8.46	8.87	8.59
Cargo Traffic Air	8.39	10.99	6.57	8.78	11.08	10.27
Cargo Traffic Port	8.25	8.77	8.88	8.30	8.12	7.25
Nightlight Intensity	8.35	8.21	8.21	8.24	8.30	8.30
Services PMI	7.31	8.18	8.18	8.21	8.14	8.14
Railway Freight	8.41	8.44	8.44	8.70	8.68	8.68
Telecom Subscribers	8.67	8.77	7.98	8.55	8.81	8.41
Foreign Tourist Arrivals	8.30	8.14	8.14	8.40	8.41	8.41
Commercial Vehicles	8.68	8.96	8.34	8.26	8.82	8.96

Source: Authors' computation

Table F2: One Quarter Ahead Forecasts Based on Shrinking Methods

	2017Q4			2018Q1		
	2m	**1m**	**0.5m**	**2m**	**1m**	**0.5m**
Principal Components						
OLS with nightlights	8.26	8.39	8.39	8.52	8.58	8.49
OLS without nightlights	7.99	8.01	8.09	8.24	8.03	8.39
MIDAS with nightlights	8.26	8.39	8.39	8.52	8.58	8.49
MIDAS without nightlights	7.99	8.09	8.09	8.24	8.39	8.39
Dynamic Factor						
OLS with nightlights	8.05	8.06	8.18	8.53	8.35	8.55
OLS without nightlights	8.54	7.87	7.92	7.95	8.03	8.50
MIDAS with nightlights	8.26	8.39	8.39	8.52	8.58	8.49
MIDAS without nightlights	7.99	8.01	8.09	8.24	8.03	8.39

Source: Authors' computation