Reserve Bank of India Occasional Papers
Vol. 37, No. 1&2, 2016

Nowcasting Indian GVA Growth in a Mixed Frequency Setup

Indrajit Roy, Anirban Sanyal, Aloke Kumar Ghosh*

This paper attempts to nowcast quarterly real non-agricultural Gross Value Added (GVA) growth for India using a dynamic factor model (DFM), following two different approaches. Multi-level variable selection using turning-point analysis and elastic-net framework has been adopted to overcome the over-fitting problem while selecting of variables. The paper finds significant improvement in forecast accuracy of one-quarter ahead forecast using nowcasting framework as compared with the naïve models. The two-factor model is found to be the most precise when compared with other higher-order factor models and naïve models. The forecast performance improves marginally when stochastic volatility is introduced in the model.

JEL Classification : C51, C52, C53, C32, C38, E50, E17

Keywords : Nowcasting, Cross-correlation, Business cycle, LARS-EN, Dynamic factor model, Kalman filter, Principal component, Cross-section

Introduction

Central banks track various macroeconomic indicators for forward-looking assessment of the state of an economy. Gross Domestic Product (GDP) is an important component for policy analysis. However, data relating to GDP are released with a lag and the release calendar is often asynchronous with the monetary policy calendar. In the absence of any real-time information on GDP, 'central banks' adopt different strategies to deal with the data gaps. Forecasting is one option which can provide forward-looking guidance on economic growth. Model-based forecasting is often criticised for limited

* Indrajit Roy is a Director, Anirban Sanyal is a Research Officer in the Department of Statistics and Information Management and Aloke Ghosh is an Assistant Adviser in the Department of Currency Management, Reserve Bank of India, Mumbai. Authors are thankful to Domenico Giannone for sharing his matlab code for dynamic factor model and Marcellino for sharing his inputs that helped in refining the model used in the paper. Data used in the paper are sourced from the public domain. Views expressed in the paper are those of the authors and not necessarily of the organisation to which they belong.

information content, while incorporating a larger information set brings in the problem of dimensionality. The other approach, namely 'multiple indicator' approach, is governed by tracking high-frequency data of economic indicators. The challenge arises when different indicators depict diverging scenarios about the state of the economy. In such a situation, summarising the information content from a larger information set and incorporating only the relevant information in the forecasting model provides a viable solution. Such an approach, also called nowcasting, gained popularity among forecasters in the aftermath of the global financial crisis of 2007.

The background of nowcasting goes back to the introduction of bridge models, which were formulated to bridge the gap between target series using available indicators (Bessec, 2013). However, the 'curse of dimensionality' restricted the use of a large information set. Stock and Watson (1999, 2002) introduced the first static and dynamic factor model for incorporating a larger information set using factors. The nowcasting framework builds upon the dynamic factor model (DFM) in a mixed-frequency setup in such a manner that the factors are updated based on the latest available information. In this context, the nowcasting framework suggested by Giannone *et al*. (2008) used around 200 high-frequency economic indicators with a varying data release calendar for nowcasting the US GDP using DFM. Later on, other central banks started incorporating similar nowcasting frameworks for short-term forecasting (Angelini *et al.*, 2011; Altissimo *et al.*, 2001, 2007; Schumacher, 2007, 2010; D'Agostino *et al.*, 2008; Marcellino *et al.*, 2013). The emergence of nowcasting and its wide adaptability point towards its growing popularity among economic forecasters and policymakers. Over fifty-four different research papers on short-term forecasting have been published in different top-ranked international journals between 2014 and 2016, which highlights the sustained interest in the methodologies for improving short-term forecasting.

While nowcasting has been adopted across advanced economies in the aftermath of the financial crisis of 2007, the emerging market economies (EMEs), which face comparatively more severe data release lags and revisions, started exploring the potential of nowcasting only in recent times. In India, the Central Statistics Office (CSO) releases quarterly GDP/ GVA data after a lag of about 60 days. For instance, GDP/GVA estimates for Q3: 2016-17 were released on February 29, 2017 *i.e.*, after a lag of about two months. Hence, any assessment of growth in the current quarter has to rely on the information available on high-frequency indicators. Nowcasting can help estimating GVA using the latest available information on such economic indicators, before the release of first advance estimate by the CSO. It can also incorporate the periodic data revisions for updating the short-term forecast of GVA.

In this paper, an attempt has been made to apply two main available nowcasting approaches to Indian data. The short-term forecasts generated from these nowcasting approaches have been validated against the available short-term forecasting models (also called naïve models) using an expanding-window approach. The variable selection for factor model has been carried out by applying a multi-level approach to turning point analysis mimicking business cycle approach. The paper finds significant improvement in short-term forecast precision using nowcasting models as compared to naïve models. The rest of the paper is organised as follows. Section II presents a brief literature review on the different approaches to use DFM in nowcasting. Section III illustrates the methodology along with a separate discussion on selection of variables. Details of data used are presented in Section IV, followed by the empirical findings in Section V. Section VI concludes with a few remarks on the proposed approach for nowcasting Indian GDP/GVA.

Section II
Literature Review

Dynamic factor models emerged as an effective mechanism for reducing dimensionality in the forecasting domain. Unlike the first and second generation DFM, the third generation deals with the time invariant correlation structure, using a larger data set in state space framework (Stock, 2010). In this approach, a principal component analysis (PCA) is used to derive the initial estimates of the factors and subsequently the factors are updated based on the latest available observations of the constituent economic indicators in a state-space framework. The use of factor models for nowcasting in a mixed frequency setup was formally introduced by Giannone *et al.* (2008), where the nowcasting model was developed using monthly data on a large set of macroeconomic and other indicators. The use of DFM for nowcasting GDP was first used in 2005 by the Board of Governors of the US Federal Reserve. Angelini *et al.* (2011) introduced a similar framework for the European Central Bank (ECB). The research work of Altissimo *et al.* (2011), Schumacher (2007 and 2010), and D'Agostino *et al.* (2008) in the related area helped to introduce this framework in Banca d'Italia, Bundesbank and Central Bank of Ireland, respectively. Barhoumi *et al.* (2010) compared the forecast performance of dynamic factor models with other quarterly models (AR, VAR and bridge equations) for select European countries and they observed significant improvement in forecasting performance in case of dynamic factor models that used monthly data. As regards variable selection, soft data which are primarily obtained through different surveys and available in a timely manner are used in DFM. However, the forecasting precision using such soft indicators is often found to be weaker than that using hard data (Banbura and Runstler, 2011). With this background, Giannone *et al.* (2008) used both hard and soft indicators for nowcasting GDP. Camacho *et al.* (2010) used a similar approach for developing the Euro-STIN. Euro-STIN represents a model that combines short-run economic indicators along with GDP revisions for increasing the precision of short-term forecasts of GDP in the Euro area. Their approach incorporated the Kalman filter model of Mariano and Murasawa (2003) to deal with mixed frequency on top of the 'strict DFM' framework of Stock and Watson (1991). The information content of GDP data revisions was previously used by Evans (2005) and Coenen *et al.* (2005). However, the incorporation of revision history along with soft and hard data was first introduced in Euro-STIN. The framework used in Euro-STIN was later extended by Marcellino *et al.* (2013) for nowcasting the Euro area GDP. The novelty of this framework was that stochastic volatility was introduced using the time-varying variance parameter. The model was estimated using multi-move Gibbs sampling approach as suggested by Carter and Kohn (1995). While nowcasting GDP of Germany, Girardi *et al*. (2016) observed improvement in short-term forecast performance using the partial least square approach for deriving the factor estimate. Lamprou (2016) and Feldkircher (2016), however, observed that DFM using bridge model had better forecast precision compared to time series models. The forecast performance of these models, nevertheless, was found to be changing over time and that suggested for a continuous review of the modelling framework to improve forecast performance (Feldkircher, 2016).

Section III
Framework

India's GVA data are released by the CSO at a quarterly frequency with a lag of about 60 days. The quarterly growth rate of GVA (Y_t^q) for any current quarter, therefore, has to be forecast based on monthly available data. In this process, two different situations generally arise – firstly the data vintages change after the release of data and accordingly, the information base of nowcasting changes; secondly, as the new data become available, the old data get revised and the dynamics change. Also, the data release calendars of different indicators in any month vary across months and thus at any point of time, the chance of getting 'Jagged Heads' or 'unbalanced panel' is significantly high. This paper considers a dynamic factor model for nowcasting GDP using approaches suggested by Giannone *et al.* (2008) and Marcellino *et al.* (2013).

Considering Θ_t^n be the information set comprising of n indicators up to time t, the nowcasting problem boils down to predicting $(Y_t^q|\Theta_t^n)$. The information set Θ_t^n comprises two parts, namely soft information ($\Theta_t^{n_1}$) and hard information ($\Theta_t^{n_2}$). Soft information primarily incorporates survey-based sentiment. Further, hard information is segregated into two subsets – core data which have been used for estimating GVA growth in India and other hard data. Here, the data release calendar of $\Theta_t^{n_1}$ and $\Theta_t^{n_2}$ are not synchronous in nature and hence jagged edges develop in the information set infusing complexity in the modelling.

Let $\Theta_n^{r_j} = \{X_{it}|r_j, i = 1, 2, \dots n\}$ for t=1,2,3,4,……,T be the information available on n economic indicators at time t as per vintage r_j. Then the information set of each data vintage $\Theta_{r_j}^n$ and $\Theta_{r_{j-1}}^n$ differ from each other as new data get released and older data get revised.

Now, in order to forecast GVA, a mixed-frequency setup arises as GVA data are released at quarterly frequency and $\{X_{it}\}$is available at higher frequency (*i.e.*, monthly). For that, let us assume that quarterly GVA growth is tagged at the last month of the quarter which means that q=3m where (3m-2) and (3m-1) are two other months within the same quarter. Having assumed that, the next step is to consider the different data vintages $\Theta_n^{v_j}$ as the monthly data releases create multiple data vintages depending upon the date of release

Given these notations, the nowcasting exercise boils down to:

$Proj\left(\widetilde{Y_{r_j}^{3k}}\middle|\Theta_{r_j}^n\right) = E\left(Y_{r_j}^{3k}\middle|\Theta_{r_j}^n, Model\right) for\ r_j \in [(3m-2), 3m]$ where $(\widetilde{Y_{r_j}^{3k}}|\Theta_{r_j}^n)$ is the projected value of target variable and expectation is taken as the latest available vintage using suitable model.

The forecast precision is calculated as an inverse of uncertainty which is defined as

$$Uncertainty_{r_j}^{3k} = E\left[\left(\widetilde{Y_{r_j}^{3k}} - Y^{3k}\right)^2 \middle| Model\right]$$

As more and more monthly data are released, $\Theta_n^{r_j}$ expands and throws more information for the forecast. Hence, we can expect the precision of the forecast to improve with more data releases, *i.e.*;

$$Uncertainty_{v_j}^{3k} \leq Uncertainty_{v_{j-1}}^{3k}$$

Framework used by Giannone

Giannone *et al.* (2008) used a dynamic factor model encompassing a large set of indicators. As the number of indicators increases, the number of unknown parameters also increases, and the curse of dimensionality crops up which limits degrees of freedom for residual estimates. The factor models are used for overcoming the dimensionality problem while capturing the information to a major extent.

The factor models are typically expressed as follows:

$$X_{it|r_j} = \mu_i + \lambda_i F_t + \epsilon_{it|r_j} \quad for\ i = 1(1)n \tag{1}$$

where μ_i is the intercept part indicating common level and F_t (Order: k x 1) relating to the factors spanning the information set. Also λ_i F_t and $\epsilon_{it|v_j}$ are assumed to be independent.

Eq(1) can be written as:

$$X_{t|r} = \mu + \Lambda F_t + E_{t|r_j} \tag{2}$$

The significance of F_t lies in the fact that the components of F_t span the information set $E_{t|r_j}$ but reduce dimensionality problem. The idiosyncratic component $X_{t|r_j}$ is the unexplained part and consists of variable specific shocks primarily attributed to the exogenous impact and possible revisions in macroeconomic variables.

Due to changing dynamics among different macroeconomic variables, the dynamics of common factors play an important role. For the sake of simplicity, the dynamics of the common factors F_t can be expressed as an AR(1) process as below:

$$F_t = AF_{t-1} + B\eta_t \tag{3}$$

where A is AR(1) coefficient matrix of order $k \times k$ and B (*Order:* $r \times q$) represents structural relationship between common factors. The shock to common factors is a white noise process. Forni *et al.* (2005) preferred using a larger set of common factors over idiosyncratic shocks in order to capture the lead-lag relationship among the variables $\{X_t\}$ along the business cycle movements.

Since the data release calendar of different variables within the information set $\{X_t\}$ differs, the chance of getting an unbalanced panel cannot be ruled out. For that, we assume that:

$$E\left(\epsilon^2_{it|v_j}\right) = \widetilde{\phi_i} = \begin{cases} \phi_i & where \quad Y_{it|r_j} \; is \; available \\ \infty & where \quad Y_{it|r_j} \; is \; not \; available \; (NA) \end{cases} \tag{4}$$

Here $E\left(\epsilon^2_{it|v_j}\right) = \infty$ ensures that no weightage would be given to variables having missing data at information vintage Θ_{r_j}.

Thus from (4), we get

$$\text{L } E\left(\epsilon_{t|v_j}\, \epsilon'_{s|v_j}\right) = \begin{cases} diag\left(\widetilde{\phi_i}, i = 1(1)N\right) \; if \; t = s \\ \infty \; if \; t \neq s \end{cases} \tag{5}$$

Also $E(\epsilon_{t|v_j}\, \eta_{s|v_j}) = 0$ for all s indicates independence between idiosyncratic shocks and shocks to common factor. Once the coefficients of equation (2) and (3) are estimated, the factors are estimated based on the latest available vintage Θ_{r_j} and estimated coefficients.

Equations (2) and (3) correspond to state-space representation. Assuming the errors follow a Gaussian process, the factors can be estimated by Kalman filter and the precision of the estimate is assessed as follows:

$$Precision_{s|v_j} = E[(F_t - \widehat{F_t})(F_{t-s} - \widehat{F_{t-s}})'|\Theta_{v_j}]$$

The news content of vintage v_j is represented by

$$News_{v_j} = Proj\left[Y_{t|v_j}\right] - Proj[Y_{t|v_{j-1}}]$$

which indicates the incremental information content in the latest available vintage over the previous data release.

Framework used by Marcellino

Marcellino *et al.* (2013) assumed that the quarterly estimate of GVA is assumed to be a geometric average of monthly unobserved GVA figures. Such assumption is viable in case the month on month changes in GVA are expected to be minimal. Under this assumption,

$$\ln(Y^L_{3M}) = \frac{1}{3} \times (\ln(Y^*_{3M-2}) + \ln(Y^*_{3M-1}) + \ln(Y^*_{3M}))$$

$$\Rightarrow Y^{3M} = \frac{1}{3} \times (y_{3M} + 2y_{3M-1} + 3y_{3M} + 2y_{3M-4} + y_{3M-5})$$

Here $y^{3M} = \Delta\ln(\mathrm{Y}^{\mathrm{L}}_{3\mathrm{M}})$ and $y_i = \Delta\ln(Y^*_i)\, , i = 1,2,3, \ldots.$

The dynamic factor model framework can be written as:

$$\begin{pmatrix} y_t \\ X_t \end{pmatrix} = \begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix} + \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} F_t + \begin{pmatrix} \epsilon_t^q \\ \epsilon_t^m \end{pmatrix} \quad (6)$$

Factor estimates are assumed to follow random walk process with lag=1 whereas the components of stochastic volatility follow AR process. Accordingly, the transition equations can be represented as:

$$\begin{aligned} \Phi^1(L)F_t &= \eta_t e^{\lambda_{f,t}/2} \\ \Phi^2(L)\epsilon_t^q &= \eta_t^1 \sigma_q e^{\lambda_{q,t}/2} \\ \Phi^3(L)\epsilon_t^{m_j} &= \eta_t^2 \sigma_m e^{\lambda_{m_j,t}/2} \end{aligned} \quad (7)$$

where are independent N(0,1) random variables and $\Phi^{\mathrm{i}}(L)$ are polynomials of order a_i:

$$\Phi_i(L) = 1 - \sum_{j=1}^{P_i} \phi_i L^i$$

The stochastic volatility is induced in the system through transition equation of λ_{it} which is assumed to follow a drift-less random walk, *i.e.*,

$$\lambda_{it} = \lambda_{it-1} + \eta_t^4, where\; \eta_t^4 \sim N(0, \sigma_\lambda^2)$$

Following Kim and Nelson (1999), the model was estimated using the multi move Gibbs Sampling approach with Metropolis Hastings algorithm for selecting the draws.

Naïve Models

The forecast performance of the nowcasting models has been analysed using the rolling forecast approach with expansionary window size. The benchmark models considered in this paper fall under time series and structural models. A list of naïve models used in this paper is given below:

Time Series Models	Structural Models
ARIMA Holt-Winters SETAR 2 regime SETAR 3-regime LSTAR AAR Artificial Neural Network	VAR Models (Interest Rate, CPI Inflation and GVA) TVP-VAR Model

The lag selection of VAR models has been decided using SIC criteria. Further, the time varying models of VAR have been framed in line with Primeceri (2005) and estimated using multi move Gibbs Sampling following Cater and Cohn (1994). In-sample residual diagnostics and test of convergence[1] have been validated for goodness of fit and stability.

Section IV
Selection of Variables

The success of a dynamic factor models lies in extracting factors from a pool of economic variables which contain information on the target variable. Thus, the information content of each variable should be cross checked before selecting a final pool. Stock and Watson (1999 and 2002) advocated the use of very large information sets, which have also been considered in a majority of the papers that use DFM. Giannone *et al.* (2008) incorporated 200 odd economic indicators for nowcasting US GDP growth. However, a different stance was adopted by Bai and Ng (2008) and Boivin and Ng (2006), who suggested the use of a relatively smaller set of variables to start with. Following that methodology, Angelini *et al.* (2011) and Camacho *et al.* (2010) used a smaller set of variables (also called core variables) and then incrementally included more variables one by one, based on improvement in forecast performance.

One most commonly accepted way to check the nature of dependence among the variables is the cross-correlation test which not only provides the significance of cross-correlation at various lags but also suggests the nature of dependence using the sign of correlation. In this context, the variables showing significant cross correlation at lag = 0 and having appropriate signs comprise the first pool of variables. However, while high cross-correlation affirms a co-movement among the series, it does not address one of the major aspects of a business cycle, that is switching of processes between which is identified as the most critical property of any business cycle indicator by Burns and Mitchell (1946). In view of this, Lahiri and Yao (2006) used turning point analysis by applying Bry-Boschan (1971) algorithm. The turning point analysis of the target variable and other economic indicators provide sufficient

[1] The authors are grateful to Primiceri for sharing his views on the assessment of convergence criteria in a time varying setup.

insight about the different phases of business cycle movement observable in each series. Though Lahiri and Yao (2006) proposed statistical coherence tests for identifying regime changes, the scrutiny of turning point analysis of any business cycle indicator requires sufficiently large number of observations, which is not suitable for our study due to the limitation of adequate data points. So, we rely on visual inspection of regime switches in the target variable and the explanatory variables. Comparing the regimes of target series, any variable (or regressor) having recession and boom regimes during different periods would be suspected for the fact that the concerned regressor would not be able to predict the turnaround points of the target variable efficiently. Hence, the cross-correlation test along with regime switching behaviour would provide sufficient screening of regressors for nowcasting. Adding up variables based on the above criteria could be one solution to define a larger pool of variables.

On the contrary, as identified by Boivin and Ng (2006), adding more and more variables into dataset may not result in improvement of forecasting performance, as some of the variables may have influences from other variables which do not impact the target variable. Also, if the idiosyncratic components are large and correlated with each other, adding further variables may not result in better accuracy in forecasting. Thus, pre-selection of variables poses a crucial challenge to forecasting. Marie (2013) used elastic net framework using Least Angle Regression - ElasticNet (LARS-EN) algorithm for selecting the regressor. LARS-EN algorithm typically uses sequential backward selection of variables using ARDL model for checking the explanatory power of regressors and penalising L1 and L2 norm of regression coefficients.

The ARDL framework can be written as:

$$S_t = \alpha + \beta_1 S_{t-1} + \beta_2 S_{t-2} + \cdots + \beta_k S_{t-k} + \gamma X_t + \zeta_t \qquad (A.1)$$

EN criteria (suggested by Zou and Hastie, 2005) is represented as:

$$\min_{\beta} \sum_{t=1}^{T} \left(S_t - \widehat{S_t}\right)^2 + \lambda_1 \sum_{i=1}^{N} |\gamma_i| + \lambda_2 \sum_{i=1}^{N} \gamma_i^2$$

where λ_1 and λ_2 are the penalty parameters of L^1 and L^2 norms of regression coefficients. Basically, EN criteria is a combination of LASSO and Ridge regression which Zoe and Hastie (2005) suggested as more efficient than LASSO and Ridge. So we resort to LARS-EN algorithm for the final selection of pooled regressor.

Section V
Data Used

In this paper, GVA at basic price has been considered for nowcasting. As a supply-side measure, GVA at basic prices (real) comprises three major sectors namely agriculture (around 14-15 per cent), industry (around 22-23 per cent) and services (around 57-62 per cent). Among these three sectors, agriculture growth remains highly seasonal in nature and depends upon exogenous variables like rainfall, reservoir status and sowing pattern, the information on some of which are not available at higher frequency. Hence, GVA excluding the agriculture sector (also called non-agriculture GVA) has been considered as the target variable for the nowcasting exercise in this paper. GVA data are available from Q1: 2011-12, only after the latest rebasing exercise carried out by the CSO. The back series of GVA at basic prices has been constructed using a bottom-up approach. Using this approach, the linking factor has been estimated separately for agriculture, industry and services based on the common overlapping period of 2004-05 base and 2011-12 base. The back series of this sectoral GVA were first estimated by applying the linking factor and then aggregated to derive GVA at basic prices.

As indicated in the previous section, the indicators pool used in this paper can be segregated into three major parts:

$$\Theta_t = (\Theta_t^C, \Theta_t^H, \Theta_t^S)$$

where $\Theta_t^C, \Theta_t^H, \Theta_t^S$ indicate core indicators, hard indicators and soft indicators, respectively.

The quarterly estimate of GVA is obtained using the benchmark indicator approach, where selected high-frequency variables are tracked to extrapolate the YoY growth rate of different sectors of GVA. Finally, the overall GVA estimate is obtained by aggregating the sectoral GVA estimates. The list of indicators used by CSO for estimating GVA is provided in Table 1.

Table 1: List of Core Indicators Used

Sectors	Indicators Used	Frequency
Industry	IIP Mining	Monthly
	IIP Electricity	Monthly
	GVA of Manufacturing Companies	Quarterly
	GVA of Petroleum Companies	Quarterly
Services	Cement Production	Monthly
	Steel Consumption	Monthly
	Production of Commercial Motor Vehicles	Monthly
	Sales of Commercial Motor Vehicles	Monthly
	Cargo Handled at Major Ports	Monthly
	Air Traffic (Passenger & Freight)	Monthly
	Foreign Tourist Arrival	Monthly
	Hotel Occupancy Rate	Monthly
	Sales Tax	Monthly
	Service Tax	Monthly
	GVA of Wholesale Trade Companies	Quarterly
	GVA of Hotel & Restaurants	Quarterly
	Aggregate Deposits	Monthly
	Bank Credit	Monthly
	Insurance Premium	Monthly
	GVA of Real Estate Companies	Quarterly
	Profitability of IT Companies	Quarterly
	Central Government Non-plan Expenditure	Monthly

Hard indicators considered in the paper are provided in Table 2.

As far as soft variables are concerned, both the PMI-manufacturing and services-along with their components have been considered in this paper. Since PMI data are available from April 2005 for manufacturing and from December 2005 for services, the data for the target variable and other economic indicators have been considered for the period January 2006 to September 2016.

Table 2: List of Hard Indicators Used

IIP and its Components	IIP Basic Goods IIP Capital Goods IIP Consumer Goods IIP Consumer Durables IIP Consumer Non-durables IIP Intermediate Goods IIP Manufacturing (NIC 2 digit)
Eight Core	Overall Eight Core (EC) Index
Interest Rate	Weighted Average Call Money Rate 10-years G-sec Yield 91-days T-Bills Rate
Demand Condition	Passenger Car Sales Three-wheeler Production Three-wheeler Sales Two-wheeler Production Two-wheeler Sales
Inflation	WPI Headline WPI Core Inflation WPI Manufacturing
External Sector	Exports (USD) Non-oil Imports (USD) Non-oil, Non-gold Imports (USD)
Money & Banking	Currency in Circulation Currency with the Public Reserve Money Narrow Money Broad Money
Global Variables	IMF Commodity Prices IMF Metal Prices Crude Oil Prices (Indian Basket) Baltic Dry Index

Section VI
Empirical Findings

The business cycle dating algorithm proposed by Bry and Boschan (1971) has been adopted to track the turnaround points of the target variable (*i.e.*, non-agriculture GVA or NAGVA). Similar analysis was carried out on quarterly transformed data for the select economic indicators (hard and soft). The final set of hard and soft indicators used for the nowcasting exercise is provided in Table 3.

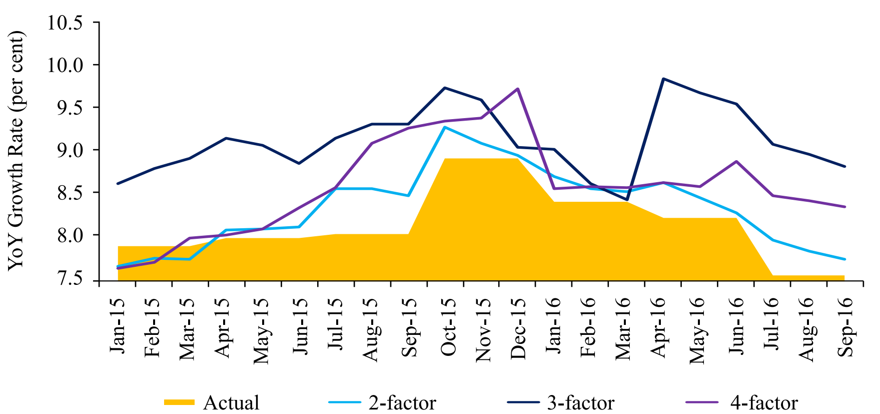
Table 3: Final List of Indicators Used

Broad Category	Indicators
IIP and its Components	IIP Basic Goods
	IIP Consumer Goods
	IIP Consumer Durables
	IIP Consumer Non-durables
	IIP Intermediate Goods
	IIP Manufacturing (NIC 2 digit)
Eight Core	Overall EC Index
Interest Rate	Weighted Average Call Money Rate
	10-years G-sec Yield
	91-days T-bills Rate
Demand Condition	Passenger Car Sales
	Three-wheeler Sales
	Two-wheeler Sales
Inflation	WPI Headline
	WPI Core Inflation
	WPI Manufacturing
External Sector	Exports (USD)
	Non-oil Imports (USD)
Money & Banking	Currency in Circulation
	Currency with the Public
	Reserve Money
	Narrow Money
	Broad Money

Broad Category	Indicators
Global Variables	IMF Commodity Price IMF Metal Price Baltic Dry Index
PMI Manufacturing	Overall Index Output New Orders Output Price Input Cost
PMI Services	Overall Index New Business Price Charged Input Price

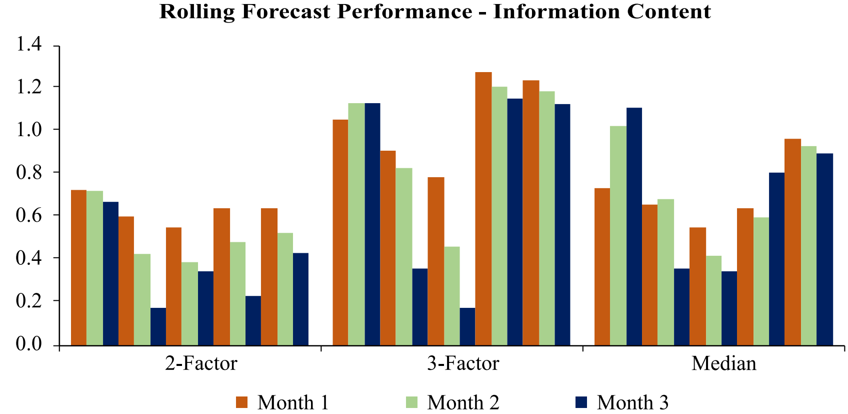
Using these indicators, the dynamic factor model was developed following Giannone *et al.* (2008) and Marcellino *et al.* (2013). Initially, the forecast precision of two-factor, three-factor and four factor models[2] was analysed using the rolling forecast mechanism in an incremental (or expansionary) window approach. The forecast performance of the two-factor model was found to be most precise among these models for immediate one-quarter ahead forecast (Chart 1).

Chart 1: Rolling Forecast Performance



[2] 2-factor model explains 76 per cent of total variability, 3-factor model explains 85 per cent variability whereas 4-factor model explains 90 per cent variability.

Chart 2: Improvement in Forecast Performance over Data Releases



The forecast performance was found to gradually improve over time during a quarter as more and more data became available and old data got revised (Chart 2). Beyond one-quarter also the forecast performance of two-factor model was found to be comparatively more precise (Chart 3).

Finally, the rolling forecast errors summarised in terms of root mean square error (RMSE) indicate significant improvement in the performance of forecast provided by the nowcasting models *vis-à-vis* that provided by Naïve Models. Among the two different approaches, the approach followed by Marcellino *et al.* (2013) was found to be marginally more precise than Giannone *et al.* (2008) approach (Table 4).

Chart 3: Forecast Performance of Two-factor Model

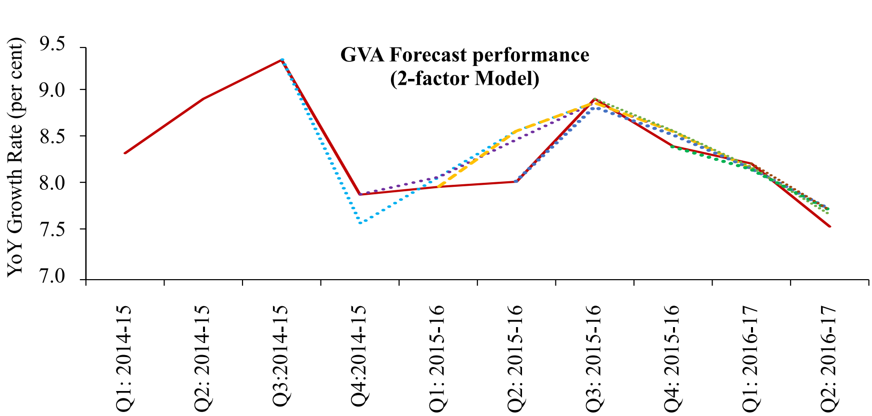


Table 4: Rolling RMSE – Nowcasting Model *vs* Naïve Model

Model	1-Q	2-Q	3-Q	4-Q
Naïve Models				
ARIMA	1.6	3.0	1.8	2.0
Holt Winters	1.7	2.2	2.3	2.4
SETAR - 3 Regime	1.1	1.2	1.5	2.2
SETAR - 2 Regime	1.3	2.9	2.6	2.0
LSTAR	1.2	1.9	1.9	1.4
AAR	1.6	3.2	3.0	3.2
Neural Network	1.5	2.5	2.4	2.8
Time Varying VAR	0.8	1.1	1.2	1.7
Nowcasting Model				
DFM-1	0.3	0.9	1.2	1.3
DFM-2	0.2	0.7	1.1	1.2

Note: 1-Q = 1 quater ahed forecast; similarly 2-Q, 3-Q, 4-Q.

Section VI
Conclusion

Nowcasting of GDP has become popular in the aftermath of the global financial crisis in 2007. The limited information problem cropped up prominently in the post-crisis scenario as most of the economic forecasting models failed to predict the crisis with significant probability. The nowcasting framework of Giannone *et al.* (2008) provided a convenient way to include a larger information set without facing the curse of dimensionality. Another novelty of the approach was to use the mixed frequency setup, given the asynchronous data release calendar. The emergence of nowcasting in an information overloaded environment helped in devising an alternative to the limited information approach of forecasting. This paper provide an assessment of the nowcasting experience in India using two different approaches.

The paper contributes to the growing literature of nowcasting and tries to implement the available frameworks to Indian high-frequency data. Following the estimation methodology of the CSO, an attempt has been made to use the available information sets for forecasting non-agricultural GVA. The paper finds significant improvement in forecast precision using nowcasting framework over naïve models. Also, it was observed that the nowcasting models are capable of forecasting NAGVA growth beyond one-quarter with a reasonable degree of precision. Further, the stochastic volatility approach suggested by Marcellino *et al.* (2013) is found to improve nowcast precision only marginally for NAGVA, compared to the approach suggested by Giannone *et al.* (2008).

References:

Altissimo, F., Bassanetti, A., Cristadoro, R., Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2001), "EuroCOIN: A Real Time Coincident Indicator of the Euro Area Business Cycle", CEPR Discussion Papers, 3108.

Baffigi, A., Golinelli, R., and Parigi, G. (2004), "Bridge Models to Forecast the Euro Area GDP", *International Journal of Forecasting,* 20 (3), 447–460.

Bai, J. (2003), "Inferential Theory for Factor Models of Large Dimensions", *Econometrica,* 71 (1), 135–171.

Bai, J. and Ng, S., (2002), "Determining the Number of Factors in Approximate Factor Models", *Econometrica,* 70 (1), 191–221.

Banbura, Marta, Giannone, Domenico, Modugno, Michele and Reichlin, Lucrezia (2013), "Now-casting and the Real-time Data Flow", Working Paper Series, European Central Bank, 1564.

Barhoumi, Karim et al. (2008), "Short-Term Forecasting of GDP using Large Monthly Datasets – A Pseudo Real-time Forecast Evaluation Exercise", Occasional Paper Series, European Central Bank, 84.

Bates , Brandon J., Plagborg-Møller, Mikkel, Stock, J. H. Stock and Watsonc, M. W. (2013), "Consistent Factor Estimation in Dynamic Factor Models with Structural Instability", *Journal of Econometrics*.

Bessec, Marie, (2013), "Short-Term Forecasts of French GDP: A Dynamic Factor Model with Targeted Predictors", *Journal of Forecasting,* 32, 500–511.

Bhattacharya, Rudrani, Pandey, Radhika and Veronese, Giovanni (2011), "Tracking India Growth in Real Time", Working Paper, *National Institute of Public Finance and Policy,* 2011-90.

Boivin, J. and Ng, S. (2005), "Understanding and Comparing Factor-based Forecasts", *International Journal of Central Banking,* 3, 117–151.

Camacho, Maximo and Perez-Quiros, Gabriel (2010), "Introducing the Euro-sting: Short-term Indicator of Euro Area Growth", *Journal of Applied Econometrics,* 25, 663–694.

Carter, C., & Kohn, R. (1994). On Gibbs Sampling for State Space Models. Biometrika, 81(3), 541-553. doi:10.2307/2337125.

Chow, G.C. and Lin, A. (1971), "Best Linear Unbiased Interpolation, Distribution, and Extrapolation of Time Series by Related Series", The Review of Economics and Statistics 53 (4), 372–375.

D'Agostino, A. and Giannone, D. (2006), "Comparing Alternative Predictors based on Large-panel Dynamic Factor Models", Working Paper Series, European Central Bank, 680.

D'Agostino, A., Giannone, D. and Surico, P. (2006), "(Un) Predictability and Macroeconomic Stability", Working Paper Series, *European Central Bank,* 605,

Doz, Catherine Giannone, Domenico and Reichlin, Lucrezia (2011) "A step-step Estimator for Large Approximate Dynamic Factor Models Based on Kalman Filtering", *Journal of Econometrics,* 164, 188-205.

Eichler, Michael Motta, Giovanni and Sachs, Rainer von (2011), "Fitting Dynamic Factor Models to Non-stationary Time Series", *Journal of Econometrics,* 163, 51-70.

Evans, M.D.D., (2005), "Where Are We now? Real-time Estimates of the Macroeconomy", NBER Working Paper, 11064, *International Journal of Central Banking,* 1 (2), 127–175.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2005), "The Generalised Dynamic Factor Model: One-sided Estimation and Forecasting", *Journal of the American Statistical Association,* 100 (471), 830–840.

Giannone, D., Reichlin, L. and Sala, L. (2004), "Monetary Policy in Real Time", *NBER Macroeconomics Annual,* MIT Press, Cambridge, 161–200.

Giannone, Domenico Reichlin, Lucrezia and Small, David (2008), "Nowcasting: The Real Time Informational Content of Macroeconomic Data", *Journal of Monetary Economics,* 55, 665-676.

Giannone, Domenico, Reichlin, Lucrezia and Small, David (2005), "Nowcasting GDP and Inflation: The Real-Time Informational Content of Macroeconomic Data Releases", Finance and Economics Discussion Series, Divisions of Research & Statistics and Monetary Affairs, Federal Reserve Board, Washington, D.C.

Huang, Jih-Jeng Tzeng, Gwo-Hshiung and Ong, Chorng-Shyong (2006), "A Novel Algorithm for Dynamic Factor Analysis", *Applied Mathematics and Computation,* 175, 1288–1297.

Kitchen, J. and Monaco, R. M. (2003), "Real-time Forecasting in Practice: The U.S. Treasury Staff's Real-time GDP Forecast System", *Business Economics,* 10–19.

Koenig, E.F. Dolmas, S. and Piger, J. (2003), "The Use and Abuse of Real-time Data in Economic Forecasting", *The Review of Economics and Statistics,* 85 (3), 618–628.

Lahiri , Kajal and Monokroussos, George (2011), "Nowcasting US GDP: The Role of ISM Business Surveys", *International Journal of Forecasting*.

Lahiri , Kajal and Yao, Vincent Wenxiong (2006), "Economic Indicators for the US Transportation Sector", *Transportation Research Part A,* 40, 872–887.

Marcellino, M., Stock, J. H. and Watson, M. W. (2003), "Macroeconomic Forecasting in the Euro Area: Country Specific versus Area-wide Information", *European Economic Review,* 47 (1), 1–18.

Marcellino, Massimiliano Porqueddu, Mario and Venditti, Fabrizio (2013), "Short-term GDP Forecasting with a Mixed Frequency Dynamic Factor Model with Stochastic Volatility", Working Papers Number, Banca D'Italia, 896.

Mario Forni et al., (2007), "Opening the Black Box – Structural Factor Models with Large Cross Sections", Working Paper Series, *European Central Bank,* 712.

Negro, Marco Del, and Otrok, Christopher (2008), "Dynamic Factor Models with Time-Varying Parameters: Measuring Changes in International Business Cycles", Staff Report, *Federal Reserve Bank of New York.*

Orphanides, A., (2002), "Monetary-policy Rules and the Great Inflation", *American Economic Review,* 92 (2), 115–120.

Primiceri, Giorgio (2005), 'Time Varying Structural Vector Autoregressions and Monetary Policy', *Review of Economic Studies,* 72, 821–852.

Runstler, G. and Se´dillot, F. (2003), "Short-term Estimates of Euro Area Real GDP by Means of Monthly Data", Working Paper Series, 276, *European Central Bank*.

Silvia, John E, and Lahiri, Kajal (2011), "Transportation Indicators and Business Cycles", *Business Economics,* Vol. 46, Issue 4, 260-261.

Stock, J.H. and Watson, M. W. (2002), "Forecasting Using Principal Components from a Large Number of Predictors", *Journal of the American Statistical Association,* 97 (460), 147–162.

Stock, J.H. and Watson, M. W. (2002), "Macroeconomic Forecasting Using Diffusion Indexes", *Journal of Business and Economics Statistics,* 20 (2), 147–162.

Zou, Hui and Hastie, Trevor. (2005), "Regularization and Variable Selection via the Elastic Net", J.R. Stat Soc. B 67, Part 2, 301 – 320.