

# Nowcasting Indian GDP <sup>1</sup>

Daniela Bragoli<sup>2</sup> Jack Fosten<sup>3</sup>

July 21, 2017

First Draft: March 2016

Word Count: Approx. 9000

**Abstract:** Nowcasting has become a useful tool for making timely predictions of GDP in a data-rich environment. However, in developing economies this is more challenging due to substantial revisions in GDP data and the limited availability of predictor variables. Taking India as an important leading case, we use a dynamic factor model nowcasting method to analyse these two issues. Firstly, we propose to compare nowcasts of the first release of GDP to those of the final release to assess differences in their predictability. Secondly, we expand a standard set of predictors typically used for nowcasting GDP with nominal and international series, in order to proxy the variation in missing employment and service sector variables in India. We find that the factor model improves over several benchmarks including bridge equations, but only for the final GDP release and not for the first release. Also, the nominal and international series improve predictions over and above real series. This suggests that future studies of nowcasting in developing economies, which have similar issues of data revisions and availability as India, should be careful in analysing first versus final release GDP data, and may find that predictions are improved when including additional variables from more timely international data sources.

**JEL Classification:** C38, C53, E37, O11, O47.

**Keywords:** Nowcasting, Emerging Markets, Data Revisions, Dynamic Factor Model, Economic Growth.

---

<sup>1</sup>We would like to thank participants of several research meetings held at Now-Casting Economics Ltd., London, for their feedback and advice. We also directly thank Now-Casting Economics Ltd. for access to data. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

<sup>2</sup>Università Cattolica, via Necchi 9, 29100 Milano (Italy), e-mail: daniela.bragoli@unicatt.it

<sup>3</sup>Corresponding Author. University of East Anglia, Norwich Research Park, Norwich, Norfolk, NR4 7TJ (UK), e-mail: j.fosten@uea.ac.uk

# 1 Introduction

In recent years, nowcasting has emerged as an important tool to produce timely predictions of economic activity variables such as Gross Domestic Product (GDP). Since GDP figures are only published on a quarterly basis and typically with a publication lag of more than a month, nowcasting has provided a solution to policymakers and other market participants by allowing them to obtain a timely snapshot of the current state of the economy required to set policy or to inform investment decisions. In developed economies such as the United States and the Euro Area, where rich and timely economic datasets are available, papers such as Evans (2005), Giannone et al. (2008) and Bańbura et al. (2013) have established that suitable nowcasting methods can provide predictions of GDP which are both more timely and at least as accurate as surveys of professional forecasters or naïve benchmark models.

However, in the context of developing economies, the problem of nowcasting is much less straightforward due to the issues of data timeliness, availability and quality. This is confirmed by quantitative measures such as the “Data Quality Index” (DQI) constructed by World Economics, where the best rankings are dominated by developed countries in Europe and North America.<sup>1</sup> In this paper we focus on the case of India, not least since it is one of the world’s largest and fastest-growing economies, but also because there has been recent media interest in the way the Indian Central Statistical Office (CSO) heavily revises its figures for GDP and periodically changes its methodology of calculating these figures. Furthermore, nowcasting Indian GDP is of particular relevance given its publication lag of almost 2 months after the end of the quarter; a delay which is far longer than most other developing countries such as China (3 weeks) and Indonesia (5 weeks), and developed countries such as the UK and US (4 weeks) and Japan (6 weeks), albeit similar to countries such as Brazil and Canada.

The first main contribution of this paper is to provide an analysis of the GDP data revisions process in the case of India. For this we propose to compare nowcasts of the first GDP release to the final release based on a nowcasting methodology which uses a dynamic factor model

---

<sup>1</sup>For more information, see <http://www.worldeconomics.com/pages/Data-Quality-Index.aspx> [Last accessed: 5th April 2016].

to accommodate a large number of predictors. Since GDP revisions data are not available for developing economies like India in databases such as the *OECD Real-Time Data and Revisions Database*, we firstly construct a series of first-release GDP figures using the available press releases from the CSO. As a starting point, we use statistical measures based on the “news” versus “noise” data revisions hypothesis developed by Mankiw et al. (1984) and Mankiw and Shapiro (1986). However, in developing countries such as India, where revisions data is only available for a around a decade, we note that these measures may not always give a robust conclusion about the revisions process. We therefore argue that our nowcasting methodology may be preferred for comparing predictions of first- and final-release GDP in developing economies, as it uses the co-variation of GDP with a large number of predictor variables. This acts like a nowcasting version of the out-of-sample test of Aruoba (2008). Although there are a few existing studies of nowcasting Indian GDP, such as Bhattacharya et al. (2011) and Dahlhaus et al. (2014), none of these studies look at the effect GDP revisions have on the nowcasting procedure.

The second contribution is that we provide a discussion and analysis of the effect of missing variables in the context of predicting Indian GDP. Specifically, we note that many variables which are typical in empirical nowcasting studies, for example employment series, are not available in a timely fashion from the Indian statistical authorities. Furthermore, there are variables which would be particularly relevant to use in a model for Indian GDP, such as trade in services, for which timely information is also not available. Starting with a small set of real variables used in studies such as Bańbura et al. (2013), we expand our study in two ways. Firstly, we include financial series such as stock prices and the exchange rate with the US dollar. This helps us to establish a dataset with a more complete representation of the Indian economy. Next, we introduce international series such as US and Euro-zone industrial production, motivated by the fact that these series correlate to a reasonable extent with exports from the Indian service sector to those economies. Our approach is able to pick up the effect of improving data richness, in contrast to existing studies for India which fix a single dataset and make comparisons either of different nowcasting methodologies (Bhattacharya et al., 2011) or different countries (Dahlhaus

et al., 2014).

The methodology we employ is the dynamic factor model approach to nowcasting suggested by Giannone et al. (2008). Since seasonally-adjusted GDP data are not available from the CSO, we target the year-on-year (YoY) real GDP growth rate and therefore use the same state space representation of the model as in Giannone et al. (2013) and Luciani et al. (2015). We consider three model variants based on a data structure of real; real and nominal; and real, nominal and international series, and compare our models to a variety of benchmarks including bridge equations and professional forecasters. Furthermore, we compare our results to those when using the first release of GDP as a target variable, thereby adding to the literature of data revisions in factor models, such as Bernanke and Boivin (2003) and, more recently, Clements (2015).

Our results uncover several important findings. Firstly, we discover that the ‘big data’ factor method improves significantly over all benchmarks in predicting the final release of real GDP growth. However, this result does not hold when predicting the first release of GDP as a target variable. The fact that the first release is difficult to predict, whereas the final release can be predicted using economic information, indicates that caution should be taken when making decisions based on the first release of GDP data. When generalising nowcasting studies to other developing economies with similarly large data revisions, this point must be taken into consideration. Secondly, we find that there are substantial gains in nowcast performance when adding in nominal series to a dataset of real variables. Moreover, adding in international series from the US, Euro-zone and Asia, which may at least partially proxy missing data on the Indian service sector, also improves the results, especially in the financial crisis period in 2008-2009. This indicates that, until developing countries have accurate and timely data available on a wide range of series, there may be benefits to searching for variables from other countries’ statistical authorities as these can improve predictions of real GDP.

This paper adds to an increasing body of empirical literature on nowcasting real GDP. There are now a large amount of studies looking at nowcasting GDP in the developed world, all of which are well-surveyed in the chapter of Bańbura et al. (2011). There is considerably less

literature on nowcasting in the developing world, presumably due to the previously mentioned issues of data availability and quality. Recent studies include nowcasting mainland China (Giannone et al., 2013), Brazil (Bragoli et al., 2015), Indonesia (Luciani et al., 2015), Mexico (Caruso, 2015) and the BRIC economies (Dahlhaus et al., 2014). There have been even fewer studies on nowcasting Indian GDP, such as the aforementioned examples of Bhattacharya et al. (2011) and a subset of the results in Dahlhaus et al. (2014).

The rest of the paper is organised as follows. Section 2 and 3 describe in detail the Indian real GDP series and the dataset of predictors we use. Section 4 describes the dynamic factor model methodology we employ and Section 5 the benchmarks we use. Section 6 presents the results of our out-of-sample forecasting exercise. Finally, Section 7 concludes the paper.

## **2 The Target Variable: Real GDP**

Indian Real GDP data is produced by the CSO at a quarterly frequency. It has been available since the late 1990s and is published with a delay of almost 2 months, which means that first quarter GDP (of the calendar year from January to March) is only released at the end of May. Together with total GDP, the CSO also publishes other measures of GDP which have been used in the literature, for example Bhattacharya et al. (2011) measure economic growth using variables for GDP excluding agriculture and GDP excluding agriculture and other services.<sup>2</sup>

We consider total real GDP as the target variable as in Dahlhaus et al. (2014). This is because our aim is to focus only on the variables monitored by the markets and the headline numbers identified by national statistical offices, central banks, local newspapers and other media.<sup>3</sup>

Given that the CSO does not seasonally adjust its GDP figures, we focus our attention on the year-on-year (YoY) growth rate of GDP rather than choosing some particular seasonal ad-

---

<sup>2</sup>The choice of Bhattacharya et al. (2011) is motivated by the fact that some of these sectors' developments are unrelated to the business cycle movements of the economy and display considerable volatility. In particular Indian agriculture is affected by strong seasonality which depend on the outcome of the monsoon and government services are affected by significant short run volatility due to the dynamics of public sector outlays.

<sup>3</sup>In addition to this, Haver Analytics, the data provider from which we extract our data, has a shorter historic availability for the other two GDP series starting from 2011 onwards. This would not currently be enough observations to perform a meaningful nowcasting study.

justment routine in order to use quarter-on-quarter (QoQ) growth. Not only does this reduce the reliance on biases caused by specific seasonal adjustment routines (see Luciani et al., 2015) but it also makes our results comparable to virtually all institutions and data agencies which provide Indian GDP growth data in a YoY format.<sup>4</sup>

An important issue with developing countries' GDP data in general is regarding the revisions and periodic rebasing of real GDP. In India, not only does the CSO revise quarterly GDP figures twice, as is standard across all countries, it also changes the methodology of GDP calculations each time the base year for real GDP is changed. This has resulted in significant media attention. For example when GDP was re-based in early 2015, Bloomberg<sup>5</sup> reported confusion amongst economists that India received a “boost from changing the way it measures GDP” in spite of negative trends in other variables such as slowing credit growth. This regular revisions and rebasing process has caused India to perform relatively poorly in measures such as the “Data Quality Index” constructed by World Economics.<sup>6</sup>

For this reason we feel that it is important to consider not only a fully revised GDP data series, as is done in the majority of nowcasting studies including the aforementioned studies for India, but also the GDP data which was available in real time. Since a real-time dataset for Indian GDP is not available, we constructed a dataset from the press releases of the CSO website which were available from May 2009.<sup>7</sup> To the best of our knowledge we are the first to obtain real time dataset for Indian GDP, which is not currently available in databases such as the *OECD Real-Time Data and Revisions Database*. On the other hand, we are unable to locate a source for constructing a real time database for other series we consider, so our study will not be fully real-time as in Clements (2015).

Figure 1 shows a graph depicting the fully revised real GDP series against the real time data from 1996 to 2015.<sup>8</sup> Since the press releases only start in May 2009, only the last portion of

---

<sup>4</sup>The YoY growth rate is also used by Bhattacharya et al. (2011) whereas QoQ growth is used in the paper of Dahlhaus et al. (2014) as they compare India to 4 other countries.

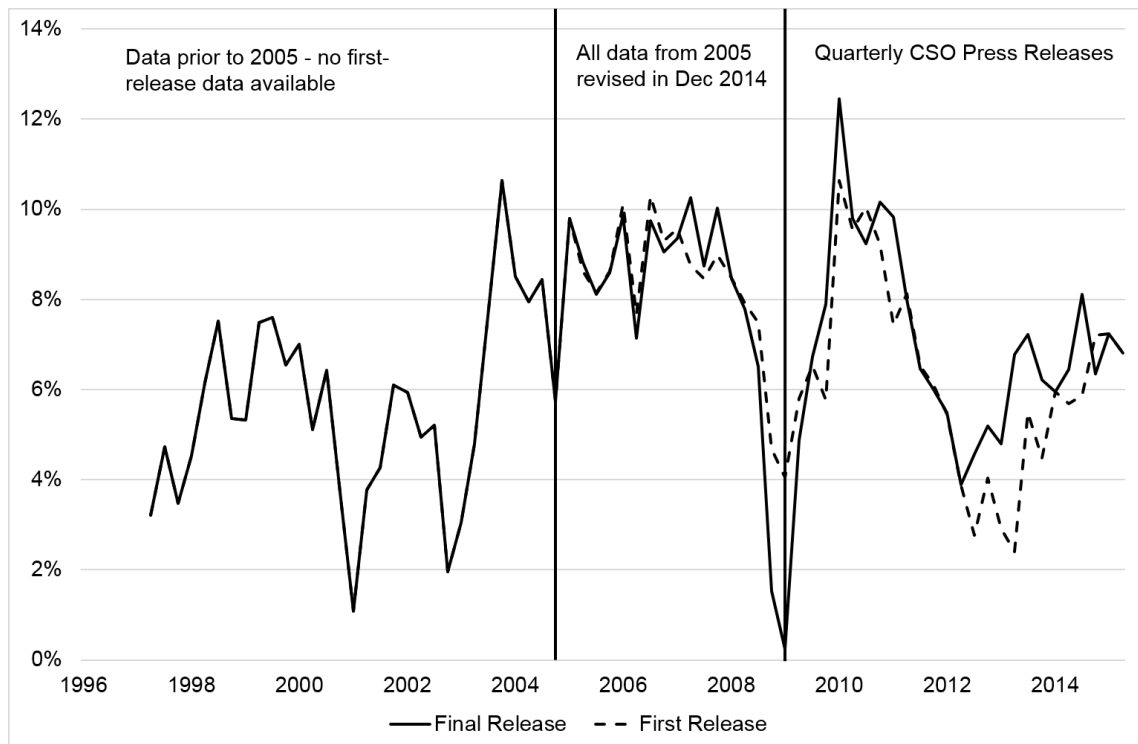
<sup>5</sup>See: <http://www.bloomberg.com/news/articles/2015-02-10/no-1-for-now-india-s-new-gdp-data-join-global-revisions> [Last accessed 2nd April 2016].

<sup>6</sup>As of April 2016, India was placed 47th out of 155 countries. For more information, see <http://www.worlddeconomics.com/pages/Data-Quality-Index.aspx> [Last accessed: 5th April 2016].

<sup>7</sup>These press releases were last accessed in April 2016.

<sup>8</sup>Fully revised data is taken from Haver Analytics. A detailed description of the method for constructing the

Figure 1: First Release vs. Fully Revised YoY Real GDP Growth



**Notes.** The Y-axis reports the YoY growth rate of the real GDP measure.

the data is truly “first release” data. The real time and final vintage data differ in the portion between 2005 and 2008 because all of the data since 2005 were revised in December 2014.

Figure 1 shows that differences between the first release and final release data are substantial in places, particularly during the financial crisis in 2008-2009, where the first release growth rate was far higher than the final out-turn. The final release YoY growth rate is also larger in the period after 2012; a point which was the focus of the aforementioned media attention.

Further analysis of the revisions process to GDP will be presented later in the main results section.

---

first release data is available upon request.

### 3 Input Variables

In predicting real GDP, standard nowcasting applications tend to use predictors which concentrate on the real side of the economy, not including prices and financial variables. The chapter of Bańbura et al. (2013) notes that real series, such as industrial production indices and employment series, while suffering from long publication lags, tend to provide precise signals for real GDP. On the other hand, survey-based series do not provide as precise a signal but are typically published in a very timely fashion, usually within days of the end of the reference month.

The construction of a representative dataset based on real variables is not a simple task for India, given the amount of missing variables relative to typical nowcasting studies. Almost all GDP nowcasting studies for other countries include, for instance, series on unemployment and employment. However, the Labour Bureau of India only produces quarterly series relating to employment and unemployment, and these series are only available from 2009, have periods of missing data, and are not published until around 2 quarters after the end of the reference quarter. This makes them impractical to use from a nowcasting perspective. Variables such as capacity utilisation and inventories are similarly only available at a quarterly or even annual basis and are therefore omitted from this study. Other commonly used variables such as housing starts, building permits and retail sales are, to the best of our knowledge, not measured by any statistical authorities in India.

Constrained by the availability of real variables, we will compare three models which incorporate different sets of input variables. The full set of variables we consider is described in Table 1. The first (Model A) includes real hard data and two different surveys: monthly PMI (manufacturing and services) and the Industrial Performance Assessment<sup>9</sup> which is rather timely, but released at a quarterly frequency.

The second set of variables (Model B) augments the previous set with financial indicators and prices. The benefit of using nominal and financial variables is that it gives us a wider

---

<sup>9</sup>This is a survey published within the Reserve Bank of India's Industrial Outlook Survey. We include the Overall Business Situation Assessment, which is based on positive, negative or neutral responses from over 1000 manufacturing firms.



coverage of variables for factor estimation, given the limited scope of the real variables, and also that, even when looking at monthly averages, these variables are available with no publication delay having been aggregated from daily series.

We aim to choose nominal variables similar to those used in past studies looking at nominal versus real factor estimates, such as those in Boivin and Ng (2006). We consider the monthly average variables for the rupee to US dollar (\$) exchange rate, the Bombay Stock Exchange Sensex 30 stock price index, the 3 month treasury bill spread over the Reserve Bank's rate and the M1 measure of the money supply. We also include several consumer and wholesale price index series. The consumer price index (CPI) is the timeliest of these price series, being released 2 weeks after the end of the reference month. The Wholesale Price indices (WPI) are included in our study as the CSO documents their use in the production of GDP figures. For a similar reason, we also include the CPI series for industrial workers, agricultural labourers and rural labourers.

Finally, the third set of variables (Model C) augments the previous set with timely variables from other countries, namely the US and the Euro Area Industrial Production and PMI and the Asian Sentix overall index. As mentioned above, the reasoning behind this selection is that trade in services, which is particularly important in India, is not well-represented by the available data. While an alternative could be to find bilateral trade data produced by other developed economies' statistical authorities,<sup>10</sup> we note that this data is also not readily available in a sufficiently timely manner to be of use in a nowcasting study. The US Census Bureau does produce a quarterly series for US imports of services from India which, while not of use in the model, has a relatively high correlation with the US Industrial Production Index, which gives us motivation for using this series and other similar series from partner trading zones. These series could also be viewed as a proxy for global demand which, in turn, will be related to demand for Indian service sector output.

Regarding the timeliness of the series for other countries, we note that soft data is available within the first days after the reference month: the Euro Zone PMI is released a week before

---

<sup>10</sup>We thank an anonymous referee for this suggestion.

Table 1: *The dataset*

Name	Model A	Model B	Model C	Publishing lag	Frequency	Source	Units/ Transf.
India Industrial Production	x	x	x	40 days	M	CSO	INDEX/YoY
India Exports	x	x	x	15 days	M	MoCI	US\$ /YoY
India Imports	x	x	x	15 days	M	MoCI	US\$ /YoY
India PMI Services	x	x	x	5 days	M	NK/MKT	INDEX /Level
India PMI Manufacturing	x	x	x	1/2 days	M	NK/MKT	INDEX /Level
India Crude Oil Production	x	x	x	28/30 days	M	MoCI	Units /YoY
India Steel Production	x	x	x	28/30 days	M	MoCI	Units /YoY
India Electricity Generation	x	x	x	28/30 days	M	MoCI	Units /YoY
India Industrial Performance Assessment	x	x	x	7/8 days	Q	RBI	% /Level
India Money Supply (M1)		x	x	40 days	M	RBI	LocCur /YoY
India Exchange Rate (Rupee/US\$, EOP)		x	x	1/5 days	M	RBI	RATIO
India 91-Day Treasury Bill		x	x	1/5 days	M	RBI	%
India Stock Prices Sensex/BSE 30		x	x	1/5 days	M	BSE	INDEX /YoY
India Industrial Workers Consumer Price		x	x	28/30 days	M	LB	INDEX /YoY
India Agricultural Laborers Consumer Price		x	x	20 days	M	LB	INDEX /YoY
India Rural Laborers Consumer Price		x	x	20 days	M	LB	INDEX /YoY
India Consumer Price		x	x	15 days	M	MOSPI	INDEX /YoY
India Wholesale Price Index (All Items)		x	x	15 days	M	MoCI	INDEX /YoY
India Wholesale Price Index (Fuel and Power)		x	x	15 days	M	MoCI	INDEX /YoY
US Industrial Production Index			x	15 days	M	FRB	INDEX /YoY
US ISM Manufacturing PMI			x	5 days	M	ISM	INDEX/Level
Euro Area 19 Industrial Production			x	50 days	M	EUROSTAT	INDEX /YoY
Euro-zone Manufacturing PMI			x	(-)-8 days	M	MKT	INDEX/Level
Asia (ex Japan) Sentix Overall Index			x	(-)-23 days	M	SENTIX	%/Level
India GDP (Billions Apr.11-Mar.12 Rupees)	x	x	x	60 days	Q	CSO	LocCur /YoY

**Notes. Models A,B,C:** report the variables considered in each model; **Publishing Lag:** is approximately the number of days from the end of the reference period; **Frequency:** indicates whether the series is monthly (M) or quarterly (Q); **Sources:** CSO (Central Statistics Office), MoCI (Ministry of Commerce and Industry), NK/MKT (Nikkei/Markit), RBI (Reserve Bank of India), BSE (Bombay Stock Exchange), LB (Labour Bureau), MOSPI (Ministry of Statistics & Programme Implementation), ISM (Institute for Supply Management), FRB (Federal Reserve Board), EUROSTAT (Statistical Office of European Communities), SENTIX (SENTIX GmbH).

the reference month ends, whereas the Asian Sentix Overall Index within the first week of the current month. Exchange rates, interest rates and stock prices are very timely and released at the beginning of the following month. Trade variables together with the Consumer Price Index and the Wholesale Price Index are released with a delay of 15 days, whereas Industrial Workers, Agricultural Laborers and Rural Laborers Consumer Price Indexes are released 5/10 days later. Indian Industrial Production is released after 40 days, but the production of the single sectors (Crude Oil, Steel and Electricity Generation) are released almost 10 days earlier. Also money supply is released with a delay of 40 days. As we can see from the timeliness of the variables the introduction of financial variables and prices, which are relatively more timely than the real set, might help improve the results of the nowcasting model.

Due to the lack of seasonally adjusted official data for many series, as was also the case in Giannone et al. (2013) for nowcasting mainland Chinese GDP, we consider the YoY transformation for many input variables except for the surveys which we report in levels and the 3 month treasury bill which is a spread over the Bank rate. The use of YoY transformations for the monthly series is also consistent with our target variable which is the YoY growth rate of real GDP, as mentioned earlier.

## 4 Methodology

We nowcast real GDP growth using a dynamic factor model, which is capable of producing a good representation of the dataset, while still allowing parsimony. The model takes advantage of the fact that there is considerable co-movement among the majority of macroeconomic data series, so that relatively few factors are required to explain the dynamics of many variables (see Sargent and Sims, 1977; Stock and Watson, 2002a,b, 2005; Giannone et al., 2005).

We are interested in a data span of  $T_M$  monthly observations on  $N_M$  stationary monthly predictors which we denote  $X_t^M = [X_{1,t}^M, \dots, X_{N_M,t}^M]$ .<sup>11</sup> We will use an estimation method which allows for certain variables to have missing data, as is common at the beginning and end of the

---

<sup>11</sup>We standardise the variables to have mean 0 and variance 1.

sample, so for some particular variable  $i$  it is possible that  $T_i^M < T^M$ . The variables are assumed to have the following factor structure:

$$X_t^M = \Lambda^M F_t + \varepsilon_t^M \quad (1)$$

where  $F_t$  is an  $r \times 1$  vector of unobserved factors at the monthly frequency,  $\Lambda^M$  is an  $N_M \times r$  matrix of factor loadings and  $\varepsilon_t$  is an  $N_M \times 1$  vector of idiosyncratic error components. We allow the individual idiosyncratic components to follow the AR(1) process  $\varepsilon_{i,t}^M = \alpha_i^M \varepsilon_{i,t-1} + e_{i,t}$  where  $e_{i,t} \sim iidN(0, \sigma_i^2)$  for all  $i = 1, \dots, N_M$ . Furthermore, we allow the factors to follow the stationary  $VAR(p)$  process:

$$F_t = A_1 F_{t-1} + \dots + A_p F_{t-p} + v_t \quad (2)$$

where the common shocks  $v_t \sim iidN(0, Q)$  and  $A_1, \dots, A_p$  are  $r \times r$  matrices of  $VAR$  coefficients. In the results we will let  $p = 2$ , in line with previous studies which have shown that including a small number of lags is sufficient for forecasting,<sup>12</sup> though the results do not change substantially when altering this parameter. The common and idiosyncratic shocks are assumed to be uncorrelated at all leads and lags, whereas the idiosyncratic shocks are allowed to be cross-sectionally correlated, but only by a limited amount (approximate factor structure).<sup>13</sup>

Since we use the YoY growth rate in quarterly GDP, in order to relate this YoY growth rate to the monthly factor, we follow the approach used by Giannone et al. (2013) and subsequently Luciani et al. (2015). We observe the quarterly levels of log real GDP, which we denote  $Y_t^Q$ , and obtain a YoY GDP growth series from these quarterly levels. These are understood as a ‘partially observed’ monthly series with quarterly observations attributed to the third month of the quarter. We call this growth rate  $y_t^Q = (1 - L^{12})Y_t^Q$ , where  $L$  is the lag operator at a monthly frequency. The quarterly levels of log GDP are also linked to unobserved monthly levels through the equation  $Y_t^Q \simeq (1 + L + L^2)Y_t^M$ , and so  $y_t^Q \simeq (1 - L^{12})(1 + L + L^2)Y_t^M = (1 + L + L^2)y_t^M$ ,

<sup>12</sup>See Forni et al. (2000) and Stock and Watson (2002b).

<sup>13</sup>This is the model studied by Doz et al. (2012) and Doz et al. (2011), which is a special case of Forni et al. (2009). See Luciani (2014) for a comprehensive survey on DFM.

where  $y_t^M$  is the unobserved monthly YoY GDP growth.<sup>14</sup> Finally by specifying a factor model for the monthly variable  $y_t^M = \Lambda^Q F_t + \varepsilon_t^Q$  we get:

$$\begin{aligned} y_t^Q &\simeq (1 + L + L^2) y_t^M = (1 + L + L^2) (\Lambda^Q F_t + \varepsilon_t^Q) \\ &= \Lambda^Q F_t + \Lambda^Q F_{t-1} + \Lambda^Q F_{t-2} + \varepsilon_t^Q + \varepsilon_{t-1}^Q + \varepsilon_{t-2}^Q \end{aligned} \quad (3)$$

Equations (1), (2) and (3) together are cast into a state space form, and can be estimated, at a monthly frequency using quasi maximum likelihood with the expectation-maximization (EM) algorithm approach following Doz et al. (2012).<sup>15</sup> The use of the EM algorithm enables us to address the issues of mixed frequencies and the ragged edge problem. Alternative approaches include Bayesian estimation, as used by Luciani and Ricci (2014) in nowcasting Norwegian GDP.

In order to provide timely nowcasts of the target GDP growth variable, we use a data release calendar which allows us to track the evolution of the information set. We denote the information set  $\Omega_v$ , where the index  $v$  denotes the ‘vintage’ of the data; the date at which one or more data points are released, in contrast to the index  $t$  which indexes the time period of the observation. In performing a pseudo out-of-sample experiment, we expand the information set one vintage at a time from  $\Omega_v$  to  $\Omega_{v+1}$  which mimics the way data is released in real time. At every point of release we can obtain a new estimate of the factors  $\mathbb{E}[F_t|\Omega_v]$  and therefore a new prediction of GDP growth  $\mathbb{E}[y_t^Q|\Omega_v]$ . Full details of this procedure is explained in the chapter of Bańbura et al. (2011).

## 5 Benchmarks

In order to assess the forecasting accuracy of our results we compare them with different benchmarks: two univariate models, the survey conducted by Bloomberg and the professional

---

<sup>14</sup>This is a modified version of the relation used in Mariano and Murasawa (2003) who link QoQ growth to the unobserved monthly equivalent.

<sup>15</sup>Bańbura and Modugno (2014) showed how the EM algorithm could be adapted for estimating the model when there are general patterns of missing data.

forecasts collected from the Reserve Bank of India.

## 5.1 Univariate Autoregressive Model

The first model we use as a benchmark is the first-order autoregressive AR(1) model. Given the quarterly YoY growth rate of GDP,  $y_t^Q$ , the AR(1) model is given by the following equation:

$$y_t^Q = \rho y_{t-1}^Q + \epsilon_t^Q \quad (4)$$

where  $\epsilon_t^Q$  is a zero-mean quarterly error term and the autoregressive parameter satisfies  $|\rho| < 1$ .

## 5.2 Univariate Bridge Models

The second benchmark we consider is the average prediction taken over a large number of univariate bridge models. We use regressions of the quarterly YoY growth rate of GDP,  $y_t^Q$ , onto quarterly aggregates of each of the  $N_M$  monthly input variables,  $X_t^M$ . For these models, we use all 23 of the available monthly predictors listed in Table 3 which corresponds to the set of monthly predictors used in Model C for the dynamic factor model. Predictions of the target GDP growth series are obtained in two steps.

In the first step, the monthly indicators are forecast over the remainder of the quarter to obtain forecasts of their quarterly aggregates ( $x_{i,t}^Q$ ). Each variable may require a different number of ‘bridging’ predictions as each series may have a different number of missing values at the end of the sample. These forecasts are made using individual autoregressive models. The number of autoregressive lags is determined using the BIC information criterion with a maximum number of  $p^{max} = 12$  lags.

In the second step, a simple bivariate bridge equation of quarterly GDP is run onto each predictor. Then the GDP prediction is made using these regression coefficients and the quarterly forecast obtained in the previous step.<sup>16</sup> Having obtained separate bridge equation predictions

---

<sup>16</sup>For further details on bridge equation methods see Parigi and Schlitzer (1995), Kitchen and Monaco (2003) and Baffigi et al. (2004).

at a given point in time, rather than reporting results for each of the  $N_M$  bridge predictions, we use the simple average of all of the bridge predictions.<sup>17</sup>

### 5.3 Bloomberg and Professional Forecasters

Bloomberg conducts a survey and collects forecasts from analysts and economists in order to produce predictions for GDP and other market-relevant variables before their release dates. These predictions are published as soon as they have at least three respondents to their questionnaire, which is generally around two weeks before the release of the relevant data series. Thereafter the prediction is continually revised up to 24 hours before the release.

The Reserve Bank of India (RBI) has been conducting a Survey of Professional Forecasters since September 2007. In our study we will use the median of around 26 quarterly professional forecasts of the YoY real GDP growth rate, which was available prior to 2015. It is worth noting that, since 2015, the RBI has been producing a more regular (bi-monthly) survey of professional forecasters, but these forecasts are only available for the growth in Gross Value Added. This is not a relevant consideration here as it is outside our sample window, but is useful for future forecast evaluation studies for Indian output growth.

## 6 Model Evaluation

In order to evaluate the performance of the model we report a ‘pseudo real time’ historical reconstruction from 2007:Q1 to 2014:Q4. We recursively estimate the model described in Equations (1), (2) and (3), first estimating using data from 2000:Q1 up to 2007:Q1, and then adding data points as the information set increases in vintage from, say,  $\Omega_v$  to  $\Omega_{v+1}$ . As mentioned earlier, this is done by constructing a full release calendar for all of the observations of all of the variables in our dataset, which allows us to model the data flow. For each of the predictions, we can compute the forecast error by subtracting the prediction from the actual level of quarterly GDP growth. The autoregressive and bridge equation models are also re-estimated

---

<sup>17</sup>We thank an anonymous referee for suggesting we explore this benchmark modelling procedure.

in each vintage, using the same available data as in the dynamic factor model approach. For the bridge equation approach this entails re-estimating the monthly forecasting equations including determining the number of lags by the BIC, and also the quarterly bridge equations.<sup>18</sup>

For a given reference quarter in the out-of-sample period, we begin making predictions 90 days before the start of the quarter (the ‘forecast period’), we continue making predictions throughout the reference quarter (the ‘nowcast period’), and continue into the period after the reference quarter up until the day before the release date of GDP (the ‘backcast period’). Therefore for each calendar quarter there is a period of around 240 days (the ‘prediction period’) over which predictions are continuously updated. In obtaining a measure of nowcast accuracy, for each point in the prediction period we compute the root mean square forecast error (RMSFE) by taking the square root of the mean of the quarterly squared forecast errors over the out-of-sample period from from 2007Q1 to 2014Q4.

## 6.1 Results: Final Release GDP Target

The results of the historical evaluation using the YoY growth of the fully revised real GDP series as the target variable are reported in the figures below. Figure 2 compares the YoY GDP nowcast based on three different models (Panel A), we also include the median prediction from the RBI survey of professional forecasters, an AR(1) model and the mean bridge equation model (“Bridge”) as benchmarks (Panel B).

Panel A shows that the introduction of financial indicators and prices (model B) as well as international variables (model C) seem to help in improving the forecasting performance. The lagging in the predictions, which characterize model A, disappears in models B and C both in capturing the beginning of the global recession in 2008 and also the pick-up in 2009; a notable improvement. It would appear that international variables help a lot in capturing the timing and the depth of the global recession. This success potentially arises from mitigating the problem that timely service sector trade data is not available in nowcasting Indian GDP. Figure 4 in

---

<sup>18</sup>To be clear, each time a new data point is released, the bridge prediction for that variable is re-evaluated. The average bridge prediction is then re-calculated using this new prediction and the remaining  $N_M - 1$  unchanged prediction.



Appendix A displays a measure of individual variables' impact on predictions throughout the nowcast period. This figure confirms the assertion that the impact of the international series is large, and also the financial series such as the BSE 30 stock price index.

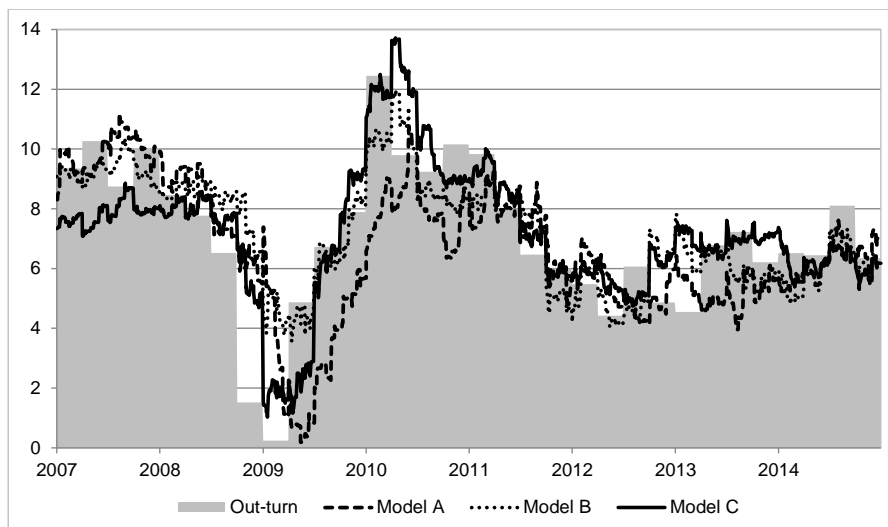
Panel B of Figure 2 compares model C (which is visually the best performing model) with several benchmarks. The AR model tracks YoY GDP well, but with a clear delay in timing. The bridge model and the Reserve Bank of India survey of professional forecasters, which is the only benchmark available that produces a nowcast of Indian Real GDP, seem to follow the AR model without being able to capture at all the timing and depth of the 2008-2009 crisis. Panel B clearly shows the improvement in performance of the nowcasting model compared to the benchmarks we are considering.

Figure 3 (Panel A) compares the RMSFE of the three models, on average for all the calendar quarters in the historical reconstruction period. Panel B compares the RMSFE of the best nowcasting model (model C) with the AR(1) predictions, which only change when GDP is released, the bridge model, the RBI professional forecasts and the Bloomberg survey of independent forecasters, which is published the day before the GDP release.

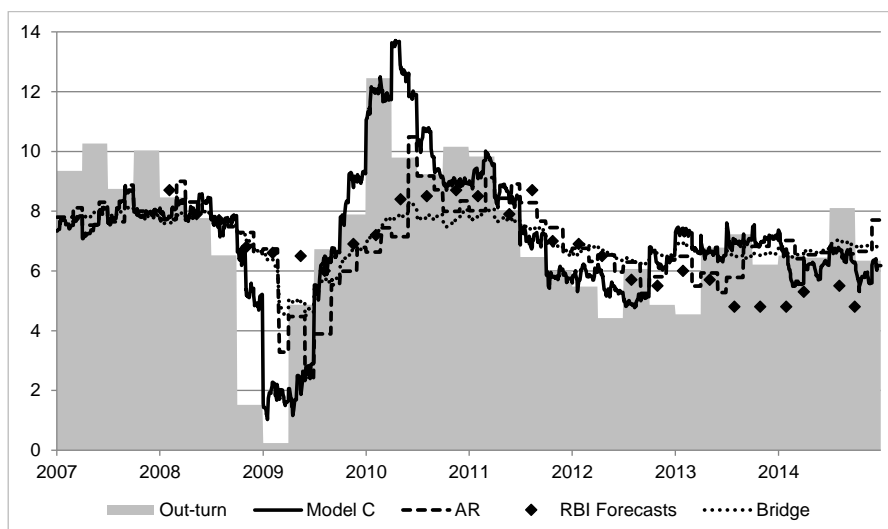
The results show that, as we move through the prediction period from forecast to nowcast to backcast, all models seem to improve as we approach the publication date of GDP. We also see that by adding in nominal, financial and international series, models B and C appear to show some significant improvement over the AR model, although their results appear to be similar to one another. Finally the bridge model does not improve over the DFM, which is a different outcome compared to the literature on nowcasting India (see Bhattacharya et al., 2011 and Matheson, 2011, who find that factor models perform poorly for India). This result is particularly true with the inclusion of financial and international series; a conclusion which was not reached by the existing literature.

Table 2 reports the test of Diebold and Mariano (1995) for equal predictive accuracy to assess whether these differences in forecasting performance between models are significant. We take model C as the benchmark with which we compare all the remaining models (AR, Bridge, model A and model B). For each month of the forecast, nowcast and backcast period we report

Figure 2: YoY GDP nowcast: model comparisons



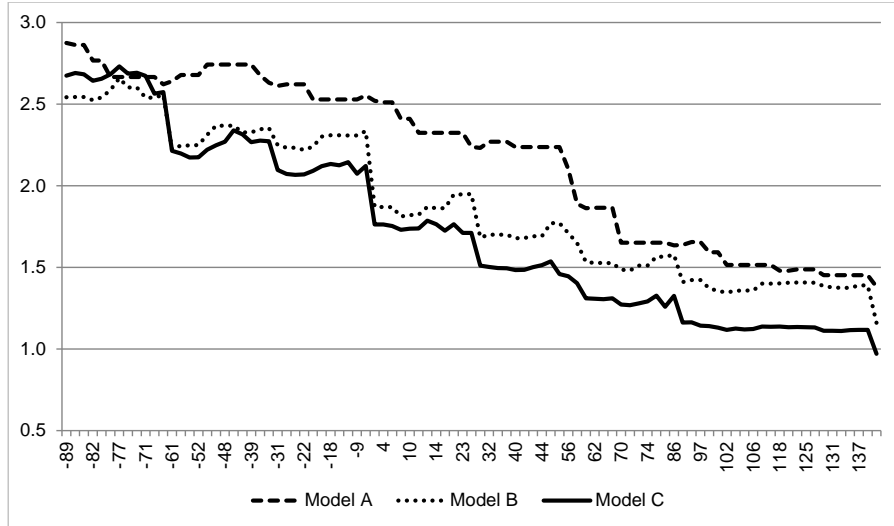
Panel A: Comparison of Nowcasting models A, B and C



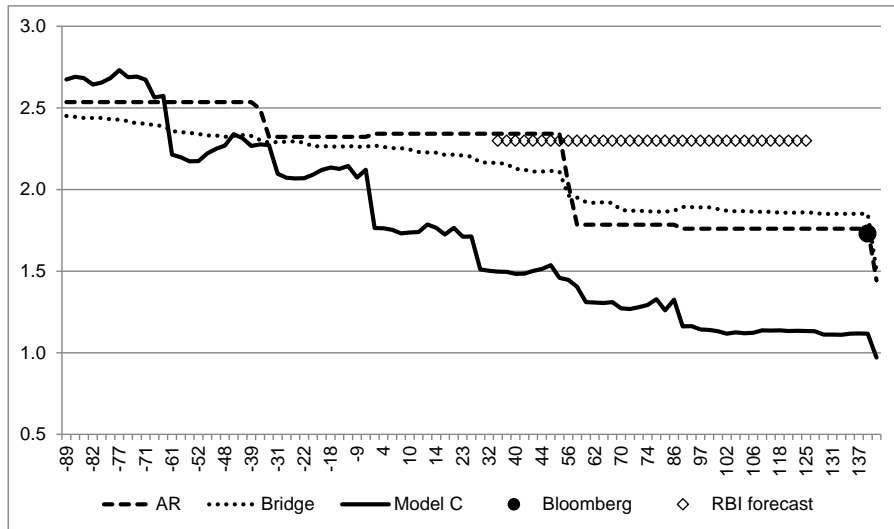
Panel B: Comparison of Nowcasting model C and benchmarks

**Notes:** The Y-axis represents YoY growth rates. Panel A reports the comparison between YoY GDP nowcast using only real variables (model A), using real financial and nominal variables (model B), using real financial, nominal and international variables (model C) and the GDP actual value (“Out-turn”). Panel B reports model C with the bridge model, the AR model, the Reserve Bank of India professional forecasts and the GDP actual value.

Figure 3: RMSFE



Panel A: Comparison of RMSFE for models A, B and C



Panel B: Comparison of RMSFE of model C and benchmarks

**Notes:** The Y-axis reports the RMSFE over the period 2007:Q1 to 2014:Q. The forecast accuracy is evaluated from the first month of the previous quarter to the time when GDP is released. The X-axis reports the distance in terms of days from the beginning of the current quarter.

the sample average of the difference between the squared errors of the AR, bridge, model A and model B forecasts with respect to the benchmark model (model C), which coincides with the first release of the month: the Manufacturing Purchasing Managers Index. A positive value of the DM test statistic implies that model C has lower RMSFE than the competing model, whereas a negative value means model C has a higher RMSFE. We report the value of the DM test and its standard deviation estimated using heteroskedasticity and autocorrelation robust (HAC) standard errors. These can be used to perform the DM *t*-test of the null hypothesis that RMSFE is the same between model C and the competitor model.

Table 2: Diebold-Mariano test of equal forecasting accuracy

	Forecast				Nowcast				Backcast			
	AR	Bridge	Model A	Model B	AR	Bridge	Model A	Model B	AR	Bridge	Model A	Model B
1m	-0.66 (1.28)	-1.01 (1.24)	1.17 (2.27)	-0.62 (1.42)	2.41 (1.64)	1.99 (1.56)	3.28* (2.02)	0.41 (0.89)	1.75* (1.04)	2.14** (1.04)	1.34 (0.83)	0.65 (0.56)
2m	1.54 (1.44)	0.77 (1.29)	2.09 (1.98)	0.07 (1.28)	3.21** (1.65)	2.34* (1.41)	2.71** (1.25)	0.53 (0.92)	1.81* (1.04)	2.07** (1.00)	0.90 (0.60)	0.69 (0.52)
3m	1.04 (1.01)	0.89 (1.04)	2.47 (1.98)	0.67 (1.05)	1.47 (0.91)	1.97** (1.00)	1.76** (0.84)	0.64 (0.66)	-	-	-	-

**Notes:** The table reports the mean of the loss differential series and the HAC estimator of its standard error in the first, second, and third month of the forecast, nowcast, and backcast, respectively. A negative sign on the estimated constant indicates that the model outperforms the benchmark (model C). There are no predictions in Month 3 of the backcast period as GDP has already been released by this point. Significance levels: \* 10%; \*\* 5%.

The test confirms the following results. Model C provides statistically significant gains over the AR and bridge models from the second month of the nowcast period through to the end of the backcast period. This result implies that the factor model is capable of predicting Indian GDP over and above naïve models. Furthermore, model C improves upon model A, which only uses real series, in the nowcast period and so financial variables and prices help considerably in improving the nowcasting model. Since models B and C are not significantly different from each other, we conclude that adding international variables does not have a significant effect over the whole evaluation period, although Figure 2 (Panel A) suggests that they did help to predict Indian GDP during the global recession in 2008-2009.

In comparing the nowcasts to other sources such as the Bloomberg survey, the fact that we

use fully revised data can in principle distort the results in favour of the model, given that the Bloomberg survey relies on real time information. In the next part, we repeat this exercise using the real time data shown in Figure 1.

## **6.2 Results: First Release GDP Target**

Following the earlier introduction to the revisions process of Indian GDP data, in this section we repeat the results of the previous section changing the target variable from the YoY growth in the final release GDP series to the growth in the first release GDP series. The results are summarised in Table 3, and additional graphs are displayed in Figures 5 and 6 in Appendix B.

We can see from the results that the picture changes somewhat relative to the previous section. From Figure 5 (Panel A) model A still has a lagging behaviour during the 2008-2009 crisis period, but seems to perform better than the other 2 models during the second GDP downturn in 2012. Models B and C seem to have similar behaviour, even if model B is more suitable than model C in capturing the depth and recovery from the great recession. Panel B shows on the other hand that model B's performance does not improve that much compared to the other benchmarks. It is better in grasping the great recession, whereas the AR model better mimics GDP realizations in the 2012 downturn.

In terms of RMSFE results Figure 6 (Panel A) shows that the three models are on average similar in term of performance, while model B is visually the best model. If we compare the best model (model B) with other benchmarks there seems to be no difference at all. These results are confirmed by the Diebold-Mariano tests in Table 3. From the tests we can conclude that model B outperforms the bridge model and model C, whereas it has the same performance as model A and the AR.

In order to better understand these results, it is helpful to provide some more in-depth analysis of the revisions process to GDP. This additional analysis can be found in Appendix C. Following the “news” versus “noise” data revisions literature proposed by Mankiw et al. (1984) and Mankiw and Shapiro (1986) we produce measures such as bias and variance of revisions, displayed in Tables 4 and 5. Our results show on the one hand that Indian GDP revisions are

Table 3: Diebold-Mariano test of equal forecasting accuracy with first release GDP target

	Forecast				Nowcast				Backcast			
	AR	Bridge	Model A	Model C	AR	Bridge	Model A	Model C	AR	Bridge	Model A	Model C
1m	-1.05	-0.62	0.60	1.20	0.09	0.86	0.77	0.96*	0.01	1.20**	0.17	0.36
	1.24	0.72	1.39	0.98	0.95	0.57	1.02	0.53	0.87	0.60	0.76	0.24
2m	0.17	0.36	0.83	1.07	0.56	1.10**	0.75	1.01**	0.13	1.26*	0.26	0.39*
	0.96	0.45	1.18	0.71	0.84	0.47	0.96	0.45	0.86	0.66	0.72	0.23
3m	-0.29	0.55	0.83	0.89	-0.08	1.16**	0.28	0.34	-	-	-	-
	0.93	0.50	1.23	0.55	0.81	0.53	0.76	0.32	-	-	-	-

**Notes:** Same as for Table 2 but with the first release GDP variable as the target.

subject to large fluctuations with high variance, but on the other hand it is not clear whether the first release is a predictor of the final release or vice versa, and so there is no clear evidence for the “news” or “noise” hypothesis. This may partly be due to the small sample size we have available in constructing these statistics.

Unlike the regression approach of Mankiw et al. (1984) and Mankiw and Shapiro (1986), Aruoba (2008) uses an out-of-sample forecasting approach to determine whether revisions are predictable or not. The results we have presented in this section can be viewed as an out-of-sample nowcasting approach to determine whether the first release data is as predictable as the fully revised data, when we predict these series using multiple different variables. Therefore, to some extent, our results can be considered as a nowcasting variant of the test of Aruoba (2008).

In light of both the analysis of the revisions process and the out-of-sample nowcast evaluation study, the conclusion we draw is that our model tracks the fully revised GDP figure much better than it does the first-release data. The fact that the model does not track the first release as well seems to be a symptom of the sizeable revisions which, in turn, has a strong impact on the conclusions between different predictive models.

This result is interesting relative to the existing literature as it contrasts with the results of Bernanke and Boivin (2003) and Clements (2015) who find that data revisions do not really matter in predicting GDP in the US when using factor models. This indicates that the nature of revisions in Indian GDP can cause problems for those wishing to make predictions. However, we should also note that our results cannot be interpreted as ‘fully real-time’ as in Bernanke and Boivin (2003) and Clements (2015). A fully real-time study would require us to have real-time

information on all variables in the dataset, and also to properly account for the revisions process by using the actual real-time data vintages as inputs rather than the first-release series.

## 7 Conclusion

This paper produces a nowcasting model for tracking Indian GDP in real time. We use a factor model approach which allows regular updates of predictions to be made each time a new piece of information arrives. The first main contribution is that we analyse data revisions by comparing the nowcasts of the first release of GDP data to the final release. This requires the construction of a real-time Indian GDP series, not analysed in previous studies. We then overcome the limitations of data availability in India by constructing datasets with nominal series, and by using external series from other countries such as the US, Euro Area and Asia. Not only does this help to pick up some of the variation in missing variables such as employment, but the international series may to some degree proxy the variation in Indian service sector exports. We start with a small set of variables from the real side of the economy which is standard in nowcasting studies (Bańbura et al., 2013) and build up our study with the new series. This allows us to arrive at conclusions which were not possible in existing studies on nowcasting India and other developing countries which use a single dataset only.

We make several important findings. Firstly we find that the factor model improves over benchmark models and professional forecasters only for the final release GDP series, and not for the first release. Since the first release is not as well-predicted by available economic information as the final release, caution must be taken when making decisions based on the first release. We suggest that future studies on nowcasting developing economies, with sizeable GDP revisions as in India, also make the distinction between predicting first versus final release data. Secondly, we find that adding in nominal series and, to a lesser extent, international series, yields substantial gains in nowcast accuracy. The addition of international series seems to improve the nowcasts of Indian GDP mostly during the global crisis period in 2008-2009. We find that the factor method improves over a benchmark bridge equation method, differently

to Bhattacharya et al. (2011), potentially as the factors from our expanded dataset are more representative of the Indian economy. Our results imply that searching for series outside of the standard set of variables used in nowcasting studies can improve nowcasts in developing countries such as India. This may also include finding series from other countries' statistical authorities which are more readily available than data in the domestic country, and might be able to proxy the variation of missing series.



## References

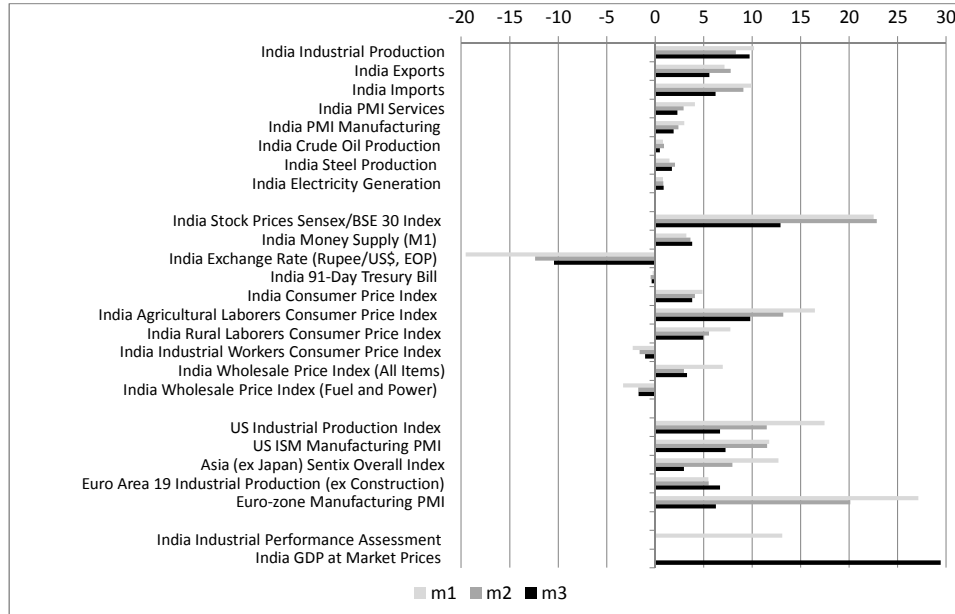
- Aruoba, S. B. (2008). Data Revisions are not Well Behaved. *Journal of Money, Credit and Banking* 40(2-3), 319–340.
- Baffigi, A., R. Golinelli, and G. Parigi (2004). Bridge models to forecast the euro area GDP. *International Journal of Forecasting* 20(3), 447–460.
- Bañbura, M., D. Giannone, M. Modugno, and L. Reichlin (2013). Now-Casting and the Real-Time Data Flow. In G. Elliott and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 2A. North-Holland.
- Bañbura, M., D. Giannone, and L. Reichlin (2011). Nowcasting. In M. Clements and D. Hendry (Eds.), *Oxford Handbook on Economic Forecasting*. Oxford University Press.
- Bañbura, 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.
- Bernanke, B. S. and J. Boivin (2003). Monetary Policy in a Data-rich Environment. *Journal of Monetary Economics* 50(3), 525–546.
- Bhattacharya, R., R. Pandey, and G. Veronese (2011). Tracking India Growth in Real Time. *National Institute of Public Finance and Policy Working Paper No. 2011-90*.
- Boivin, J. and S. Ng (2006). Are More Data Always Better for Factor Analysis? *Journal of Econometrics* 132(1), 169–194.
- Bragoli, D., L. Metelli, and M. Modugno (2015). The Importance of Updating: Evidence from a Brazilian Nowcasting Model. *OECD Journal: Journal of Business Cycle Measurement and Analysis* 2015(1), 5–22.
- Caruso, A. (2015). Nowcasting Mexican GDP. *ECARES Working Papers*.
- Clements, M. P. (2015). Real-time Factor Model Forecasting and the Effects of Instability. *Computational Statistics & Data Analysis*.
- Dahlhaus, T., J.-D. Gunette, and G. Vasishtha (2014). Nowcasting BRIC+M in Real Time. *Bank of Canada Working Paper 2015-38*.
- Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business*

- and *Economic Statistics* 13(3), 253–263.
- Doz, C., D. Giannone, and L. Reichlin (2011). A two-step estimator for large approximate dynamic factor models based on Kalman filtering. *Journal of Econometrics* 164(1), 188–205.
- 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.
- Evans, M. (2005). Where are we now? Real-Time estimates of the Macroeconomy . *International Journal of Central Banking* 1(2).
- Forni, M., D. Giannone, M. Lippi, and L. Reichlin (2009). Opening the black box: Structural factor models with large cross sections. *Econometric Theory* 25(05), 1319–1347.
- Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000). The generalized dynamic-factor model: Identification and estimation. *Review of Economics and statistics* 82(4), 540–554.
- Giannone, D., S. M. Agrippino, and M. Modugno (2013). Nowcasting China Real GDP. *Working Paper*.
- Giannone, D., L. Reichlin, and L. Sala (2005). Monetary policy in real time. In M. Gertler and K. Rogoff (Eds.), *NBER Macroeconomics Annual*. MIT Press.
- 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.
- Kitchen, J. and R. Monaco (2003). Real-time forecasting in practice. *Business Economics* 38(4), 10.
- Luciani, M. (2014). Large-dimensional dynamic factor models in real-time: A survey. Available at SSRN: <https://ssrn.com/abstract=2511872>.
- Luciani, M., M. Pundit, A. Ramayandi, and G. Veronese (2015). Nowcasting Indonesia. *Finance and Economics Discussion Series 2015-100*. Washington: Board of Governors of the Federal Reserve System.
- Luciani, M. and L. Ricci (2014). Nowcasting Norway. *International Journal of Central Banking* 10(4), 215–248.

- Mankiw, N. G., D. E. Runkle, and M. D. Shapiro (1984). Are Preliminary Announcements of the Money Stock Rational Forecasts? *Journal of Monetary economics* 14(1), 15–27.
- Mankiw, N. G. and M. D. Shapiro (1986). News or noise? An Analysis of GNP Revisions. *NBER Working Paper*.
- Mariano, R. S. and Y. Murasawa (2003). A New Coincident Index of Business Cycles Based on Monthly and Quarterly Series. *Journal of Applied Econometrics* 18(4), 427–443.
- Matheson, T. (2011). New indicators for tracking growth in real time. Working paper No. WP/11/43, International Monetary Fund.
- Parigi, G. and G. Schlitzler (1995). Quarterly forecasts of the Italian business cycle by means of monthly economic indicators. *Journal of Forecasting* 14(2), 117–141.
- Sargent, T. and C. Sims (1977). Business cycle modeling without pretending to have too much a priori economic theory. In C. Sims (Ed.), *New Methods in Business Cycle Research*. Federal Reserve Bank of Minneapolis.
- Stock, J. H. and M. W. Watson (2002a). Forecasting Using Principal Components from a Large Number of Predictors. *Journal of the American Statistical Association* 97(460), 1167–1179.
- Stock, J. H. and M. W. Watson (2002b). Macroeconomic Forecasting Using Diffusion Indexes. *Journal of Business and Economic Statistics* 20(2), 147–162.
- Stock, J. H. and M. W. Watson (2005). Understanding changes in international business cycle dynamics. *Journal of the European Economic Association* 3(5), 968–1006.

## Appendix A: Variables' Impact

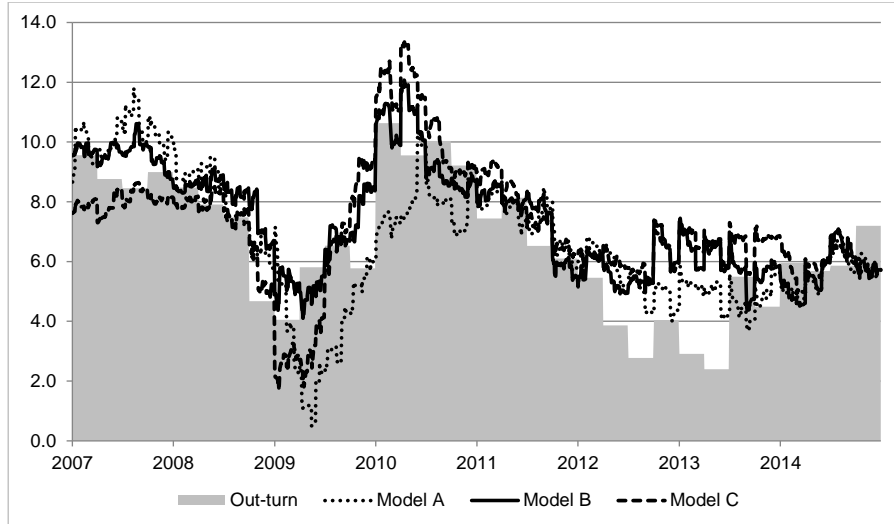
Figure 4: Variables' relevance



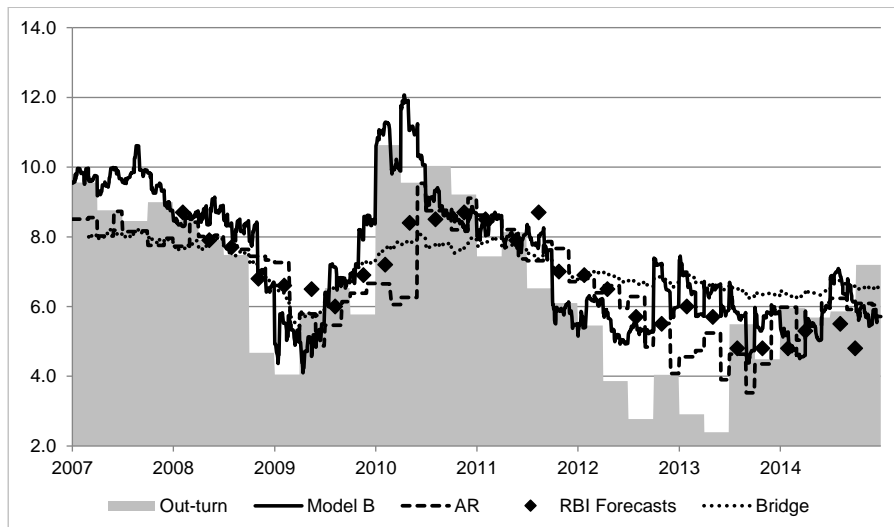
**Notes.** Variables' average impact in the first (m1), second (m2), and third (m3) month of the nowcast period.

## Appendix B: Additional Results

Figure 5: YoY GDP nowcast: model comparisons with first release GDP target



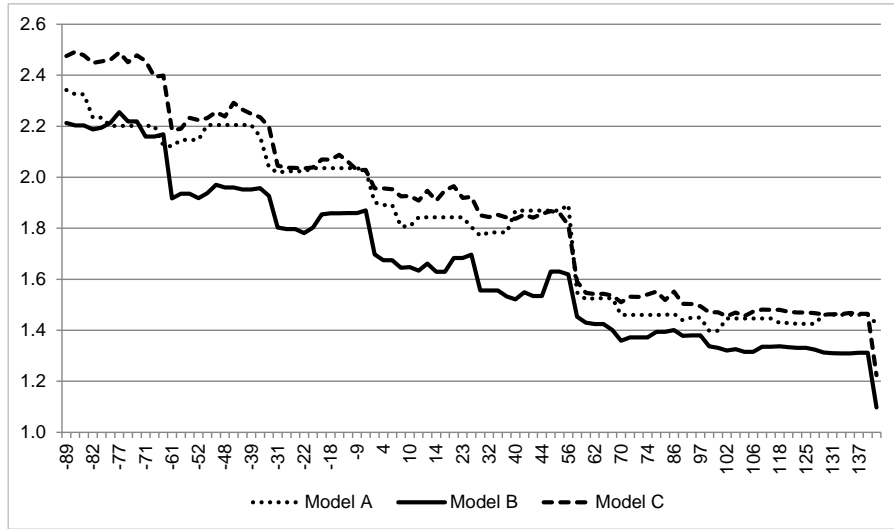
Panel A: Comparison of Nowcasting models A, B and C



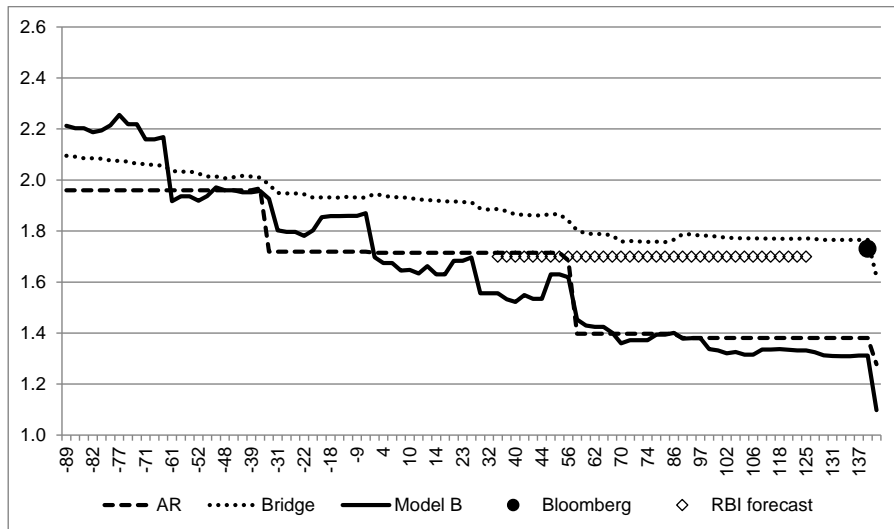
Panel B: Comparison of Nowcasting model B and benchmarks

**Notes:** Charts are the same as for Figure 2 with the first release GDP variable as the target.

Figure 6: RMSFE for first release GDP target



Panel A: Comparison of Nowcasting models A, B and C



Panel B: Comparison of Nowcasting model B and benchmarks

**Notes:** Same as for Figure 3 but with the first release GDP variable as the target.

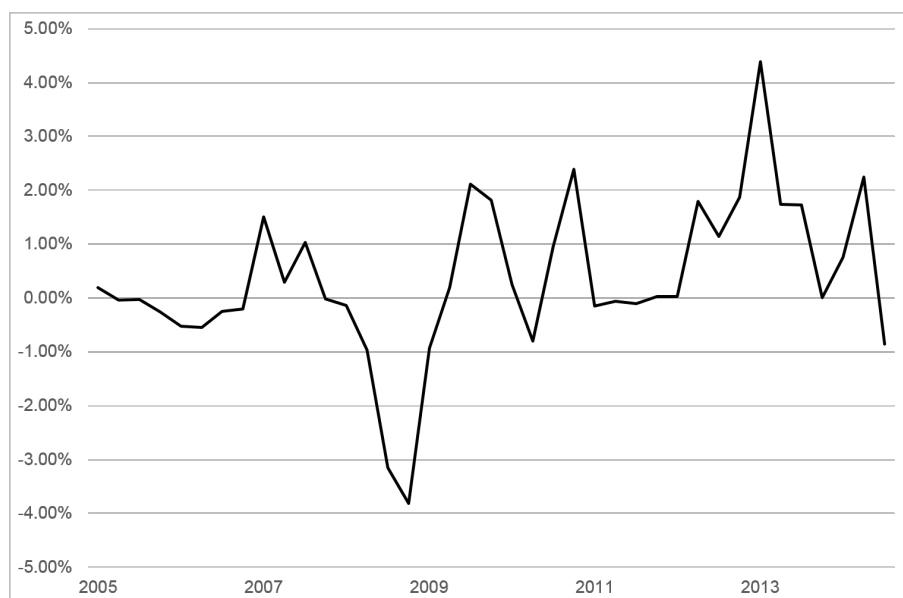
## Appendix C: The Nature of Indian GDP Revisions

In order to look further into the nature of data revisions in the Indian real GDP series, we follow the approach taken by Aruoba (2008).<sup>19</sup> Denoting the first-release real GDP growth series as  $y_t^1$  and the final release as  $y_t^f$ , the revisions process is defined to be:<sup>20</sup>

$$r_t \equiv y_t^f - y_t^1 \quad (5)$$

Figure 7 plots this series over time. Aruoba (2008) defines three properties for revisions to be well-behaved: 1) they must be unbiased:  $E[r_t] = 0$ ; 2) the revision variance  $Var[r_t]$  should be small; and 3) they should be unpredictable given the information set:  $E[r_t|I_t] = 0$ . Table 7 provides a set of summary statistics very similar to those of Aruoba (2008) which attempts to shed some further light on the revisions process.

Figure 7: Final Revisions to YoY Real GDP Growth



**Notes.** The Y-axis reports the revisions in YoY growth rates.

The mean of revisions to real GDP growth is only 0.35% YoY, which is not statistically

<sup>19</sup>We thank an anonymous referee for suggesting that we explore this avenue.

<sup>20</sup>As in Aruoba (2008), we define revisions as the difference in the YoY growth rate of the first- and final-release data rather than the growth of the difference in the raw series.

Table 4: Summary Statistics for  $r_t$ 

T	Mean	Min	Max	St. Dev.	Noise/Signal	Cor. With Initial	AR(1)
39	0.349	-3.815	4.386	1.455	0.595	-0.191	0.545

**Notes:** See discussion in the text for a full description of these numbers. A  $t$ -test was carried out separately that each of Mean, Cor. with Initial and AR(1) were equal to zero. The null hypothesis was only rejected AR(1), at the 1% level, and not at any conventional significance level for the others.

significantly different from zero using a HAC estimator of the standard error of the sample mean. Note that this is based only on the  $T = 39$  quarters for which we have revisions data available. The next statistics show very extreme minimum and maximum values of revisions with a seemingly large standard deviation. The statistic “Noise/Signal” takes the ratio of the standard deviation of  $r_t$  to the final release series  $y_t^f$ . The value of 0.595 is somewhat large, particularly in relation to the US results of Aruoba (2008) where this statistic is 0.28. This indicates that there is a large variance of the revisions process in India. The statistic “Cor. with Initial”, the correlation of  $r_t$  with the first release  $y_t^1$ , shows little correlation of the revisions with the first release, although the AR(1) coefficient of  $r_t$  is large and statistically significant showing some persistent behaviour in revisions.

We can also follow the procedures used in Mankiw et al. (1984) and Mankiw and Shapiro (1986) in attempting to determine whether the revisions to Indian GDP are “news” or “noise”. Both of these use regression-based procedures which we note might suffer from low power given the small quarterly sample of revisions data we have available. For the noise hypothesis to hold, the first release is equal to the final release plus a noise term, which can be tested using a regression of  $y_t^1$  onto  $y_t^f$  and testing whether the slope is equal to unity, and the constant equal to zero. The test of the news hypothesis reverses this regression to test whether the final release is predicted by the initial release. Aruoba (2008) augment this news regression with the lagged revisions variable, noting that past revisions should not be able to predict future releases. The results are displayed in Table 5.

The results from Table 5 are somewhat inconclusive for Indian GDP revisions. According to the regression results following Mankiw and Shapiro (1986) in Models 1 and 2, we reject the



Table 5: Testing the News vs. Noise Hypothesis

News Hypothesis		
Model 1: $y_t^1 = \alpha_1 + \beta_1 y_t^f + v_t^1$		
$H_0 : \alpha_1 = 0, \beta_1 = 1$	$W = 6.965$	$p = 0.031$
Noise Hypothesis		
Model 2: $y_t^f = \alpha_2 + \beta_2 y_t^1 + v_t^2$		
$H_0 : \alpha_2 = 0, \beta_2 = 1$	$W = 1.092$	$p = 0.579$
Model 3: $y_t^f = \alpha_3 + \beta_3 y_t^1 + \gamma r_{t-1} + v_t^3$		
$H_0 : \alpha_3 = 0, \beta_3 = 1, \gamma = 0$	$W = 13.881$	$p = 0.003$
<b>Notes:</b> $W$ denotes the standard Wald statistic for testing $H_0$ , using HAC standard errors.		

noise hypothesis and do not reject the news hypothesis. However, the Aruoba (2008) version of the test also rejects the news hypothesis, which gives us an ambiguous conclusion which was also the case when looking at US GDP revisions.

In the main text, we study the first release and final release series in more detail by looking at their respective predictability using nowcasting methods.