## School of Economics Working Paper 2016-06

University of East Anglia

SCHOOL OF ECONOMICS

# Nowcasting Indian GDP

**by Daniela Bragoli***
**Jack Fosten****

*** Università Cattolica**
****School of Economics, University of East Anglia**

**Abstract**

We propose a nowcasting model for the Indian real GDP growth rate which uses the flow of relevant information to update predictions on a daily basis and can serve as a timely barometer to track the Indian development process. There are several challenges faced when nowcasting GDP in developing economies such as India. The first challenge is to proxy important missing variables such as international trade in the service sector. Our novel solution augments a baseline model with series on US and Euro-area output which improves predictions, particularly during the 2008-2009 global crisis. The second challenge is the impact of sizeable revisions to the GDP data. We construct a new series for real-time Indian GDP using press releases from the Central Statistics Office (CSO), finding that data revisions have a non-trivial influence on our results. Therefore, caution should be taken when evaluating predictions using the preliminary GDP release.

**JEL classification codes**
C38, C53, E37, O11, O47

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

School of Economics
University of East Anglia
Norwich Research Park
Norwich NR4 7TJ
United Kingdom
www.uea.ac.uk/economics

# Nowcasting Indian GDP

Daniela Bragoli[1] Jack Fosten[2]

June 6, 2016

First Draft: March 2016

**Abstract:** We propose a nowcasting model for the Indian real GDP growth rate which uses the flow of relevant information to update predictions on a daily basis and can serve as a timely barometer to track the Indian development process. There are several challenges faced when nowcasting GDP in developing economies such as India. The first challenge is to proxy important missing variables such as international trade in the service sector. Our novel solution augments a baseline model with series on US and Euro-area output which improves predictions, particularly during the 2008-2009 global crisis. The second challenge is the impact of sizeable revisions to the GDP data. We construct a new series for real-time Indian GDP using press releases from the Central Statistics Office (CSO), finding that data revisions have a non-trivial influence on our results. Therefore, caution should be taken when evaluating predictions using the preliminary GDP release.

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

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

[1]Università Cattolica, via Necchi 9, 29100 Milano (Italy), e-mail:daniela.bragoli@unicatt.it

[2]Corresponding Author. University of East Anglia, Norwich Research Park, Norwich, Norfolk, NR4 7TJ (UK), e-mail: j.fosten@uea.ac.uk

# 1 Introduction

Although Gross Domestic Product only provides a limited measure of people's material living standards, it remains the single most important indicator to capture economic activities in a given country and for this reason it is the most monitored variable by policy institutions and investors. The main problem with real Gross Domestic Product (GDP) is the fact that it is released on a quarterly basis and with a delay that ranges from $4$ weeks, as in the US or the UK, to $8$ weeks as in the case of Canada and India.

Long publication lags in economic output data have forced policy institutions and market participants to set their policies and to allocate their investments without having timely knowledge on current state of the economy and sometimes, even without knowing the recent past. Institutions had previously solved this problem by producing predictions of the current or even previous quarter using their own judgement. For developing countries, such as India, the task becomes even more challenging given the frequent revisions in the data.[3]

The purpose of this work is to propose a nowcasting model for the Indian real GDP growth rate which is capable of giving updated predictions on a daily basis as new relevant information arrives following the literature pioneered by Evans (2005) and Giannone et al. (2008). This method can be used both as a timely barometer to measure the Indian development process, or as a tool for interested investors to track short-term movements in the Indian economy. Furthermore, we also analyse the impact of data revisions in nowcasting Indian GDP which we do not believe has been addressed in the literature to date.

There are several reasons which lend motivation to nowcasting Indian GDP, and several challenges which we must overcome in this paper. The purpose of nowcasting is to make use of 'big data' in obtaining timely predictions of variables which are measured at a low frequency and with a delay in publication. This is particularly relevant in the case of the quarterly Indian GDP series, which is published around two months after the end of the quarter to which it refers.

The first main challenge is that many variables which are typical in empirical studies of

[3] In the first half of 2015 the Indian Central Statistics Office (CSO) changed their methodology of calculating real GDP, which altered existing numbers for year-on-year growth by up to 4% in a single quarter (see the cited newspaper articles and explanation in Section 2.1, below).

nowcasting GDP, such as 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. Our first main contribution is to construct a dataset of predictor variables, which has two main differences relative to existing studies. Firstly, we include financial (stock price) series, thereby moving away from the idea that only real variables tend to have strong short-term signals for GDP, as found by Bańbura et al. (2013) and others. This helps us to establish a dataset with a more complete representation of the Indian economy. Secondly, a novel contribution of our paper is that we proxy missing variables for the trade in services by introducing international series such as US and Euro-zone industrial production. In our results, we assess whether the addition of these variables is useful for predicting Indian GDP.

The second key issue is the size of the revisions made to Indian GDP by the CSO. Not only do revisions take place on a quarterly basis, but the methodology of calculating real GDP changes periodically each time the series is re-based with a new price level. In order to analyse the effect of revisions, our second main contribution is in constructing a real-time series of Indian GDP using the available press releases from the CSO. To the best of our knowledge, this is the first attempt to analyse such real-time data in the case of India, whilst real-time databases are becoming more widely available in countries like the United States (*Philadelphia Fed: Real-time Dataset for Macroeconomists*) and other OECD countries (*OECD Real-Time Data and Revisions Database*). This allows us to perform historical re-constructions of competing nowcasting models, keeping in mind the data truly available to forecasters at the time.

This paper adds to an increasing body of empirical literature on nowcasting real GDP. There is 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. (2013). There is considerably less literature on nowcasting in the developing world, presumably due to issues of data availability and quality. Recent studies include nowcasting mainland China (Giannone et al., 2013), Brazil (Bragoli et al., 2015), Mexico (Caruso, 2015) and the BRIC economies (Dahlhaus et al., 2014). There have been even fewer studies on nowcasting Indian GDP, examples include Bhattacharya et al. (2011) and a subset of the results in Dahlhaus et al. (2014).

The methodology we employ is the dynamic factor model approach of Bańbura and Modugno (2014). The use of a factor model, which dates back to the seminal works of Stock and Watson (2002a,b), allows us to estimate a couple of common factors which capture the bulk of the co-movement amongst the Indian economic variables in our dataset. The factors are estimated by maximum likelihood as in Doz et al. (2012), while the mixed frequency of monthly and quarterly variables and the 'ragged-edge' are accounted for in the estimation procedure as in Bańbura and Modugno (2014). Since seasonally-adjusted data are not available, 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). We consider three model variants based on a data structure of real; real and nominal; and real, nominal and international series. Furthermore, we compare our results to those when using the real-time target GDP variable. In doing so, our results add to the literature of data revisions in factor models, such as Bernanke and Boivin (2003) and, more recently, Clements (2015).

To briefly preview the results, our key findings are that there are significant gains to considering nominal as well as just real series in nowcasting Indian GDP. Furthermore, the idea of proxying the missing service sector trade variables with international series from the US, Euro-zone and Asia has a positive impact, particularly improving the results in the financial crisis period in 2008-2009. Turning to the effect of data revisions, we find that there are large differences to the results when considering real-time GDP. This runs contrary to existing studies, including the aforementioned studies of Bernanke and Boivin (2003) and Clements (2015), which tend to find that data revisions do not have much impact on factor models. This is an indication that care should be taken when evaluating predictions of the Indian economy due to the sizeable nature of the revisions to the preliminary estimate.

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

## 2 Data

In this section we describe the data used in constructing a nowcasting model for Indian GDP. All data are extracted from the Haver Analytics database, and the sample of monthly and quarterly variables runs from January 2000 to December 2014. We first describe in detail the Indian GDP series before turning to the remainder of the dataset.

### 2.1 Indian GDP Data

Indian Real GDP data is produced by the Central Statistical Office (CSO) at a quarterly frequency. It has been available since the late 1990s and is published with a delay of almost 2 months. This delay is somewhat longer than in the US or the UK (4 weeks), European Countries or Japan (6 weeks) and China (3 weeks), although comparable to that of some other countries such as Canada and Brazil. It is also notable that the CSO does not seasonally adjust its GDP figures. We therefore focus attention on the year-on-year growth in GDP, which is the measure widely followed by market participants.

The aim of the statistical model we propose in this article is to predict GDP before the official figures are published by taking advantage of the information in the flow of economic data releases that precede them and updating our prediction with each successive release. Given the sizeable publication delay of Indian GDP there is strong motivation for producing a reliable nowcasting model.

In addition to this large publication lag, there is also an issue regarding the revisions and periodic rebasing of the real GDP data, which may have an impact on nowcasting studies. 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 rebased in early 2015, Bloomberg[4] reported confusion amongst economists that India received a "boost from changing the way it measures GDP" in spite of negative trends in other variables

[4]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 2015].

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.[5]

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, but also the GDP data which was available in real time. Since a real-time dataset for Indian GDP is not available, we construct such a dataset from the press releases available at the Central Statistical Office website since May 2009. 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.[6] Since the press releases only start in May 2009, only the last portion of 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 was revised in December 2014.

The differences between these two series are substantial in places. Particularly during the financial crisis in 2008-2009, we see that the data which was originally published was revised substantially, indicating a deeper slowdown than had previously been thought. The other main finding is that the final vintage has a larger YoY growth rate in the period after 2012; a point which was the focus of the aforementioned media attention.

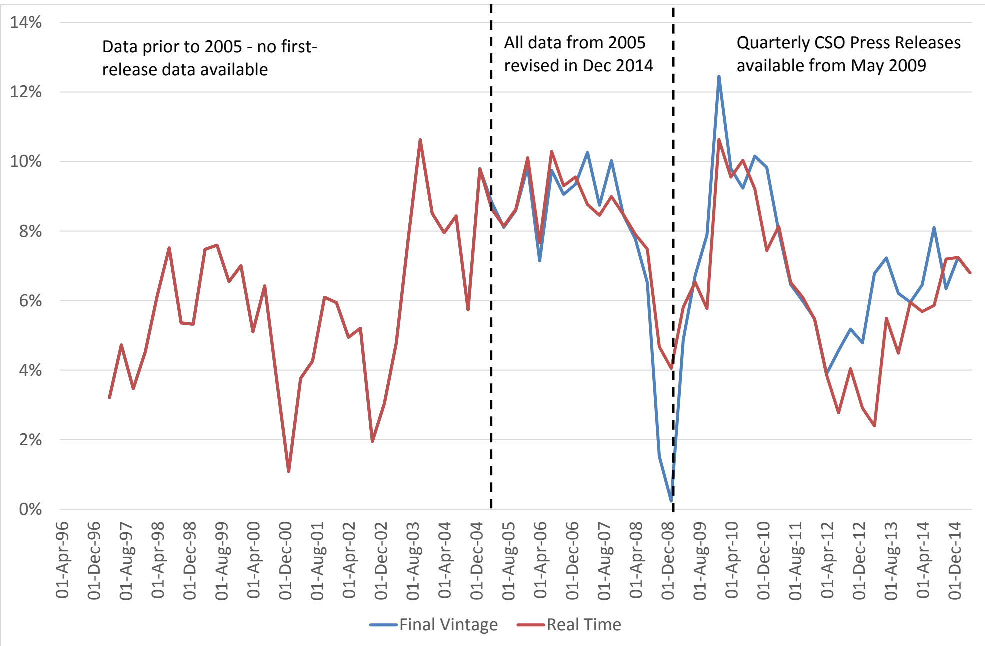
In the results, we will compare the findings of our nowcasting models when tracking both the real-time and final vintage GDP data. Any major differences in the results will indicate that care should be taken when considering which GDP series to use.

---

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

[6]Fully revised data is taken from Haver Analytics. A detailed description of the method for constructing the first release data is available upon request.

Figure 1: Real Time vs. Fully Revised YoY Real GDP Growth




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

## 2.2 The Dataset

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 that market participants, which also try to infer the current economic condition by monitoring the behaviour of other indicators that are linked to GDP, also face a data constraint.

Another important consideration is the amount of missing variables for India. 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[7] 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 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

[7]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.

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. We feel that using international series from Europe, the US and Asia may provide a reasonable and timely proxy for Indian service sector exports.

Regarding the timeliness of the series as for other countries soft data is available within the first days after the reference month, the Euro Zone PMI is released a week before 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 is 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 lack of seasonally adjusted official data for many series, as Giannone et al. (2013) in nowcasting mainland Chinese GDP, we target the year-on-year growth rate of real GDP at a quarterly frequency and we consider year-on-year transformation for many input variables excluding surveys that we report in levels and the 3 month treasury bill which is a spread over the Bank rate.

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 at Market Prices | 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).

## 2.3 Benchmark Comparisons

In order to assess the forecasting accuracy of our results we compare them against the survey conducted by Bloomberg and the professional forecasts collected from the Reserve Bank of India.

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. Bloomberg publishes predictions 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 final number is usually close to the actual release value.

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 year-on-year 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.

# 3 Methodology

In this paper we employ a dynamic factor model approach to nowcasting which is based on the work of Giannone et al. (2008) and Bańbura and Modugno (2014). We have $T_M$ monthly observations on $N_M$ stationary monthly predictors which we denote $X_t^M = [X_{1,t}^M, ..., X_{N_M,t}^M]$.[8] These variables are assumed to have the following factor structure:

$$X_t^M = \Lambda^M F_t + \varepsilon_t^M \tag{1}$$

[8] We standardise the variables to have mean 0 and variance 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 $\varepsilon^M_{i,t} = \alpha^M_i \varepsilon_{i,t-1} + e_{i,t}$ where $e_{i,t} \sim iidN(0, \sigma^2_i)$ for all $i = 1, ..., N_M$. Furthermore, we allow the factors to follow the stationary $VAR(p)$ process:

$$F_t = A_1 F_{t-1} + ... + A_p F_{t-p} + v_t \tag{2}$$

where $v_t \sim iidN(0, Q)$ and $A_1, ..., A_p$ are $r \times r$ matrices of $VAR$ coefficients. In the results we will let $p = 2$, in line with previous studies though the results do not change substantially when altering this parameter.

To model the quarterly GDP variable, since the CSO does not produce seasonally adjusted figures, and because the market follows the year-on-year growth in quarterly GDP, we also use the quarterly year-on-year real GDP growth rate as an input. In order to relate this year-on-year growth rate to the monthly factor, we follow the approach used by Giannone et al. (2013).

We observe the quarterly levels of real GDP, which we denote $Y^Q_t$. We obtain a year-on-year GDP growth series from these quarterly levels, which as in Bańbura and Modugno (2014) are understood as a 'partially observed' monthly series with quarterly observations attributed to the third month of the quarter. We call this growth rate $y^Q_t = (1 - L^{12})Y^Q_t$, where $L$ is the lag operator at a monthly frequency. The quarterly levels of GDP are also linked to unobserved monthly levels of GDP through the equation $Y^Q_t = (1 + L + L^2)Y^M_t$, and so $y^Q_t = (1 - L^{12})(1 + L + L^2)Y^M_t = (1 + L + L^2)y^M_t$, where $y^M_t$ is the unobserved monthly year-on-year GDP growth.[9] Finally by specifying a factor model for the monthly variable $y^M_t = \Lambda^Q F_t + \varepsilon^Q_t$ we get:

$$\begin{aligned} y^Q_t = \left(1 + L + L^2\right) y^M_t &= \left(1 + L + L^2\right) \left(\Lambda^Q F_t + \varepsilon^Q_t\right) \\ &= \Lambda^Q F_t + \Lambda^Q F_{t-1} + \Lambda^Q F_{t-2} + \varepsilon^Q_t + \varepsilon^Q_{t-1} + \varepsilon^Q_{t-2} \end{aligned} \tag{3}$$

[9]This is a modified version of the relation used in Mariano and Murasawa (2003) who link quarter-on-quarter growth to the unobserved monthly equivalent.

Equations (1), (2) and (3) together are cast into a state space form, and can be estimated using the EM algorithm approach of Doz et al. (2012), as described in Bańbura and Modugno (2014). The use of the EM algorithm enables us to address the issues of mixed frequencies and the ragged edge problem.

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 date at which some data point is released, in contrast to the index $t$, above, which indexes the period of the observation. In performing a pseudo out-of-sample experiment, we expand the information set one data point at a time from $\Omega_v$ to $\Omega_{v+1}$ and so on. 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. (2013).

# 4 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 a single data point at a time, as the information set expands from, say, $\Omega_v$ to $\Omega_{v+1}$. We do this by constructing a full release calendar for all of the observations of all of the variables in our dataset, which allows us to model this 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.

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 is 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 averaging the root of the quarterly squared forecast errors over the out-of-sample period from from 2007Q1 to 2014Q4.

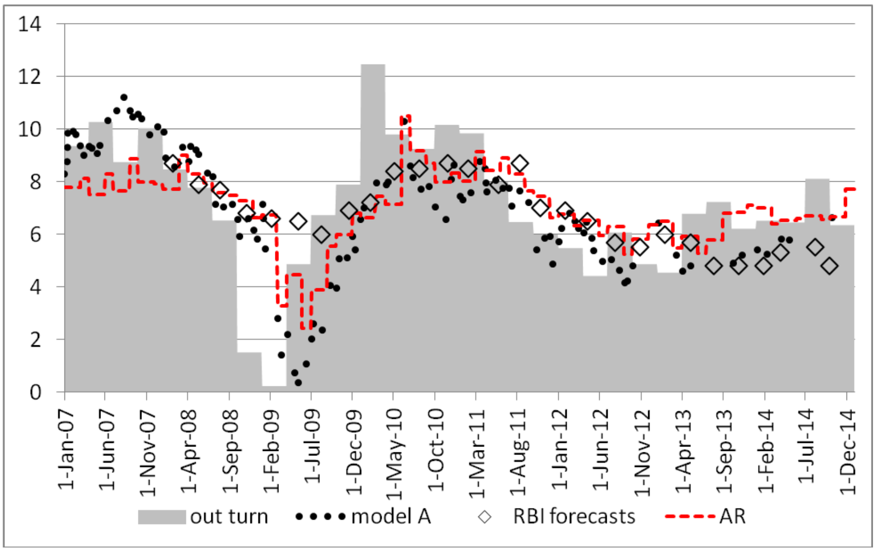
## 4.1 Results: Fully Revised GDP Target

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 year-on-year GDP nowcast based on three different models, we also include the median prediction from the Reserve Bank of India (RBI) survey of professional forecasters and a 1-lag autoregressive (AR) model as benchmarks. Panel A shows that the AR model tracks YoY GDP well but with a clear delay in timing. Our baseline specification which estimates the factors using only real variables, Model A, does not appear to deviate strongly from the AR predictions. Also, the Reserve Bank of India survey of professional forecasters, which is the only benchmark available that produces a nowcast of Indian Real GDP, seems to follow the AR model without being able to capture at all the timing and depth of the 2008-2009 crisis.

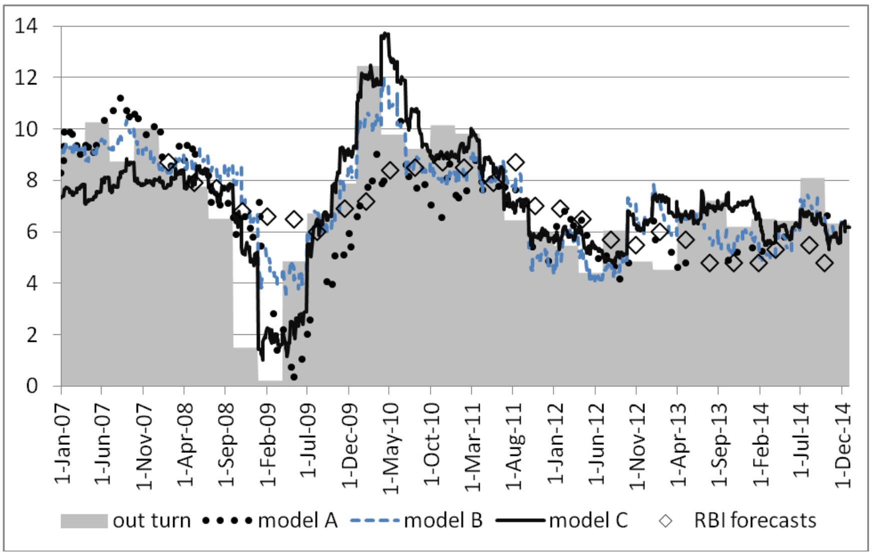
Panel B of Figure 2 compares the three different model specifications we consider. The performance of Models B and C shows that the introduction of financial indicators, prices and international variables seem to help in improving the forecasting performance. The delay in the prediction disappears both 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 is in mitigating the problem that timely service sector trade data is not available in nowcasting Indian GDP. Figure 6 in the Appendix 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.

Figure 3 compares the root mean squared forecast error (RMSFE) of the three models, on average for all the calendar quarters in the historical reconstruction period. The performance

Figure 2: YoY GDP nowcast: model comparisons



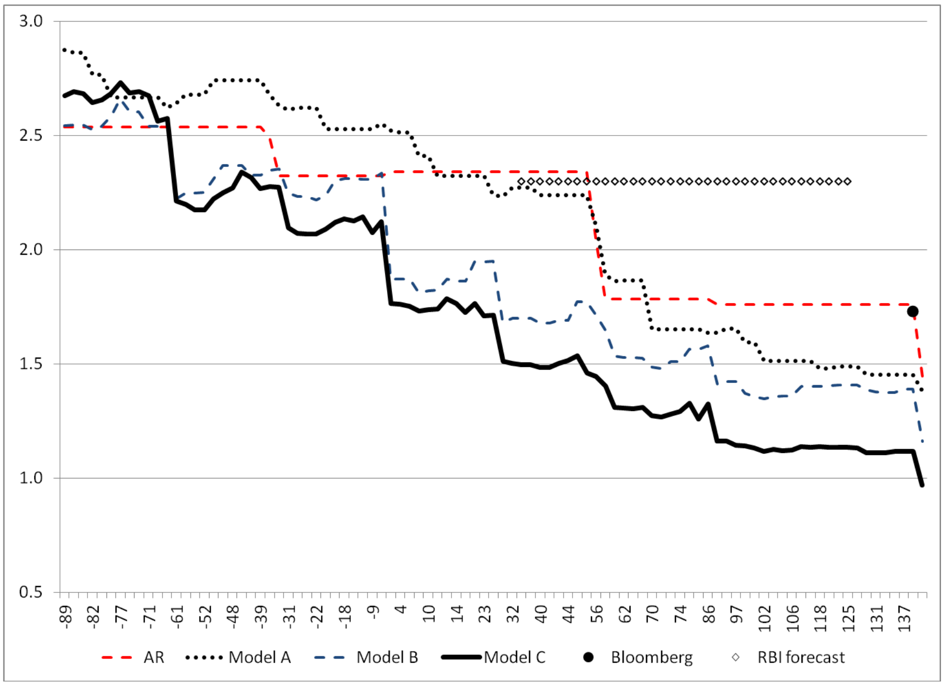

Panel A




Panel B

**Notes:** Panel A reports the comparison between YoY GDP nowcast using only real variables (model A) the AR model, the Reserve Bank of India professional forecasts and the GDP actual value. Panel B reports the three nowcasting models 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.

Figure 3: RMSFE




**Notes:** The Y-axis reports the root-mean-squared forecast error (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.

of the three model versions is also compared to that of the autoregressive forecast, which only changes when GDP is released, 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 the 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 Model A seems to be have RMSFE very close to that of the AR model, as expected on inspecting Figure 2. On the other hand, 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.

Table 2 reports the test of Diebold and Mariano (1995) for equal predictive accuracy to assess whether these differences in forecasting performance between the are significant. We take Model C as the benchmark with which we compare all the remaining models (AR, Model A and Model B), as this model includes the largest set of variables. For each month of the forecast, nowcast and backcast period we report the sample average of the difference between the squared errors of the AR, 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. We report the value of the DM test and its standard deviation estimated using heteroskedasticity and autocorrelation robust (HAC) standard errors.

The test confirms the following results. Model C provides statistically significant gains over the AR model from the second month of the nowcast period through to the end of the backcast period. This is a strong result as it means that using timely information can help to predict Indian GDP over and above naïve models. Furthermore, model C improves upon Model A, which only uses real series, the nowcast and to a lesser extent in the first month of the backcast 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 suggests that they did help to predict Indian GDP during the global recession in 2008-2009.

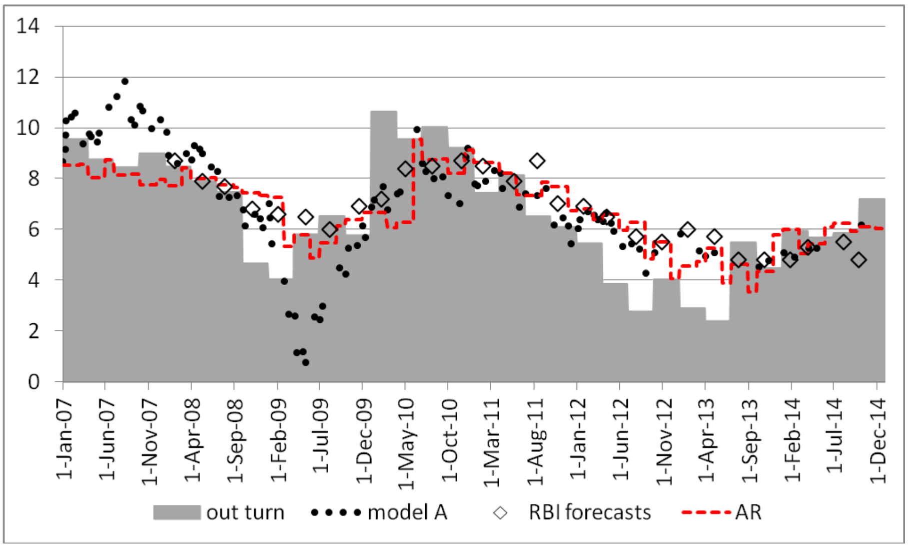
In comparing the nowcasts to other sources such as the Bloomberg survey, the fact that we use fully revised data can in principle distort the results in favour of the model, given that the Bloomberg survey relies on real time information. In the next part, we repeat this exercise using the real time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

## 4.2 Results: Real-Time GDP Target

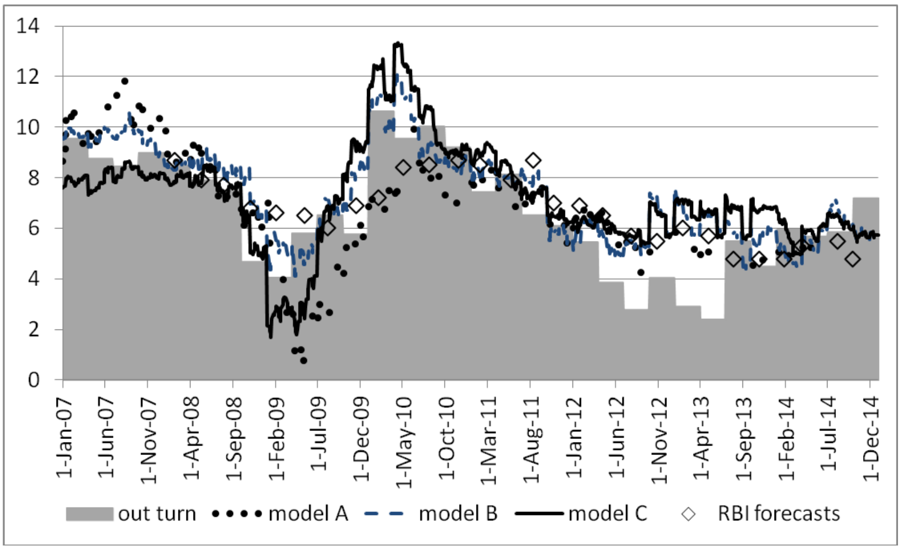
In this section we repeat the results of the previous section, changing the target variable from the YoY growth in the fully-revised GDP series to the growth in the real-time GDP series, which was described in Section 2.1. The results are displayed in Figures 4 and 5 and Table 3.

We can see from the results that the picture changes somewhat relative to the previous

Figure 4: YoY GDP nowcast: model comparisons with real time GDP target




Panel A




Panel B

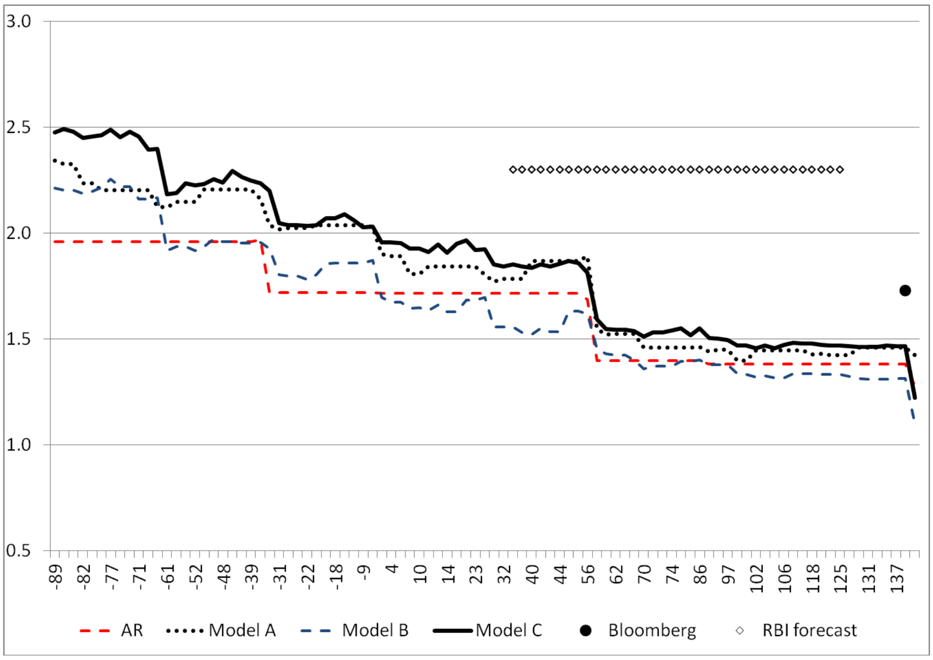
**Notes:** Charts are the same as for Figure 2 with the real-time GDP variable as the target.

Table 2: Diebold-Mariano test of equal forecasting accuracy

| | Forecast | | | Nowcast | | | Backcast | | |
|---|---|---|---|---|---|---|---|---|---|
| | AR | Model A | Model B | AR | Model A | Model B | AR | Model A | Model B |
| 1m | -0.66 | 1.17 | -0.62 | 2.41 | 3.28 | 0.41 | 1.75 | 1.34 | 0.65 |
| | (1.28) | (2.27) | (1.42) | (1.64) | (2.02) | (0.89) | (1.04) | (0.83) | (0.56) |
| 2m | 1.54 | 2.09 | 0.07 | 3.21 | 2.71 | 0.53 | 1.81 | 0.90 | 0.69 |
| | (1.44) | (1.98) | (1.28) | (1.65) | (1.25) | (0.92) | (1.04) | (0.60) | (0.52) |
| 3m | 1.04 | 2.47 | 0.67 | 1.47 | 1.76 | 0.64 | - | - | - |
| | (1.01) | (1.98) | (1.05) | (0.91) | (0.84) | (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. There are no predictions in Month 3 of the backcast period as GDP has already been released by this point. The AR, Model A and Model B are compared against Model C.

Figure 5: RMSFE for real-time GDP target




**Notes:** Same as for Figure 3 but with the real-time GDP variable as the target.

section. The performance of the autoregressive model is similar to that in the previous section, and appears to suffer significant predictive delays. When looking at the predictions from Models

Table 3: Diebold-Mariano test of equal forecasting accuracy with real-time GDP target

| | Forecast | | | Nowcast | | | Backcast | | |
|---|---|---|---|---|---|---|---|---|---|
| | AR | Model A | Model B | AR | Model A | Model B | AR | Model A | Model B |
| 1m | -2.24 | -0.60 | -1.20 | -0.87 | -0.19 | -0.96 | -0.35 | -0.18 | -0.36 |
| | (1.40) | (1.61) | (0.98) | (0.94) | (0.99) | (0.53) | (0.95) | (0.84) | (0.24) |
| 2m | 1.54 | 2.09 | 0.07 | 3.21 | 2.71 | 0.53 | 1.81 | 0.90 | 0.69 |
| | (1.44) | (1.98) | (1.28) | (1.65) | (1.25) | (0.92) | (1.04) | (0.60) | (0.52) |
| 3m | 1.04 | 2.47 | 0.67 | 1.47 | 1.76 | 0.64 | - | - | - |
| | (1.01) | (1.98) | (1.05) | (0.91) | (0.84) | (0.66) | - | - | - |

**Notes:** Same as for Table 2 but with the real-time GDP variable as the target.

A, B and C in Figure 4, we find that the predictions seem to track the shape of the real-time GDP variable slightly better, but tend to suffer from over-predictions at peaks and under-predictions at troughs, particularly in the 2008-2009 crisis period.

Turning to the RMSFE results in the graph of Figure 5 and the Diebold-Mariano tests in Table 3, we see that the performance of Models A, B and C are very similar to that of the AR model; confirmed by the results of the DM test. The conclusion we draw from this exercise is that it appears that our model tracks the fully revised GDP figure much better than it does the first-release data. There is some ambiguity about whether professional forecasters and institutions are tracking the first release or the final release of data. However, the difference between the sets of results indicates that the size of the revisions to Indian GDP has a strong impact on the conclusions between different predictive models.

This result is interesting because 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.

## 5 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 when a new piece of information arrives. The first main contribution of the paper is in building a representative database of the Indian economy, taking a novel approach to proxy unavailable service exports by including a set of international variables such a industrial production in the United States and the Euro Area. Another contribution of the paper is in constructing a real-time Indian GDP series which allows us to analyse the impact of data revisions made by the Central Statistics Office.

The main conclusions of the paper are threefold. Firstly, the best nowcasting model is the one which includes real, nominal, price and international series. When continuously updated from around 6 months before the release of a quarterly GDP figure, the model's predictions are constantly improving, and perform well compared to a survey of professional forecasters. Secondly, the addition of international series improves the nowcasts, especially in predicting Indian GDP during the global crisis period in 2008-2009. Finally, the results from our nowcasting models change when tracking the first release data for GDP rather than the final release. This is in contrast to previous studies using real time data in factor models for developed economies, which typically find that data revisions are not important. We therefore suggest that caution should be exercised when evaluating predictions of the Indian economy based on the preliminary release of GDP.

## 6 Acknowledgements

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.

## References

Bańbura, M., G. D., M. M., and R. L. (2013). Now-Casting and the Real-Time Data Flow. In G. Elliott and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 2A. North-Holland.

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

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 (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)*.

Giannone, D., S. M. Agrippino, and M. Modugno (2013). Nowcasting China Real GDP. *Working Paper*.

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.

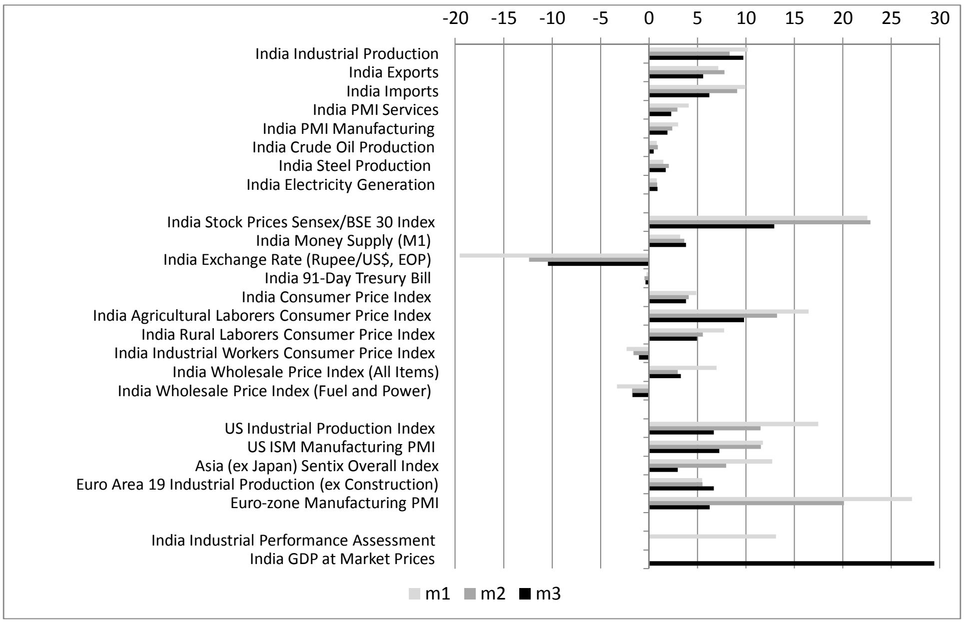
Mariano, R. S. and Y. Murasawa (2003). A New Coincident Index of Business Cycles Based on Monthly and Quarterly Series. *Journal of Applied Econometrics 18*(4), 427–443.

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.

# 7 Appendix

Figure 6: Variables' relevance




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