Forecasting with Real Time Information Flow

Anirban Sanyal

Department of Economics, UCSC

December 17, 2021

The Views expressed in this presentation, belongs to the presenter only and does not reflect the views of the organization to which they belong. The findings are drawn from published papers of the presenter.

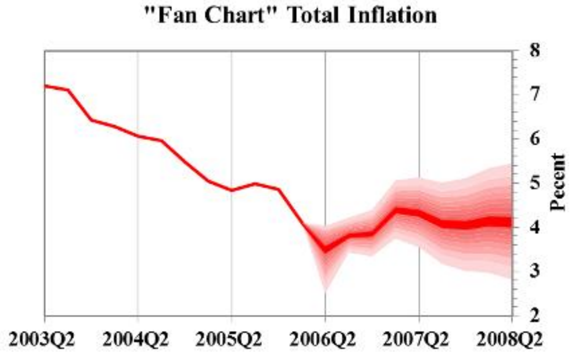
Agenda

- **Part I: Nowcasting models**
  - Nowcasting models
  - Real time information flow
    - High growth sectors
    - Corporate sentiments

Part - I: Nowcasting

Central Bank forward looking view

- Forward guidance by Central Banks
- Outlook on growth, inflation and interest rate path
- Short term vs long term forecast

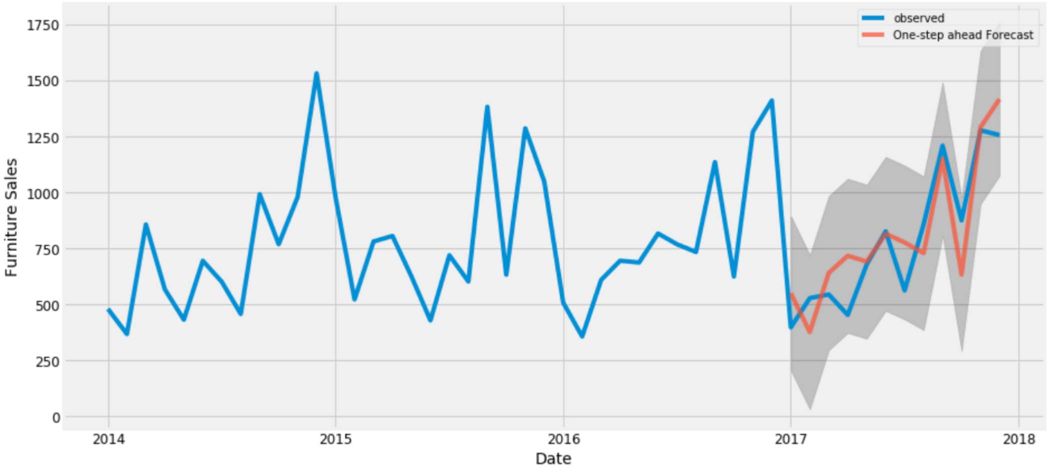


Forecasting - Short term vs Long term

- Long term forecast - driven by fundamentals
- Short term forecast - lacks fundamentals
- Mechanism behind short term forecast - data generating process and information impact
- Time series models used for forecasting
  - Univariate models - AR, MA and ARMA
  - Multivariate models - VAR, VECM

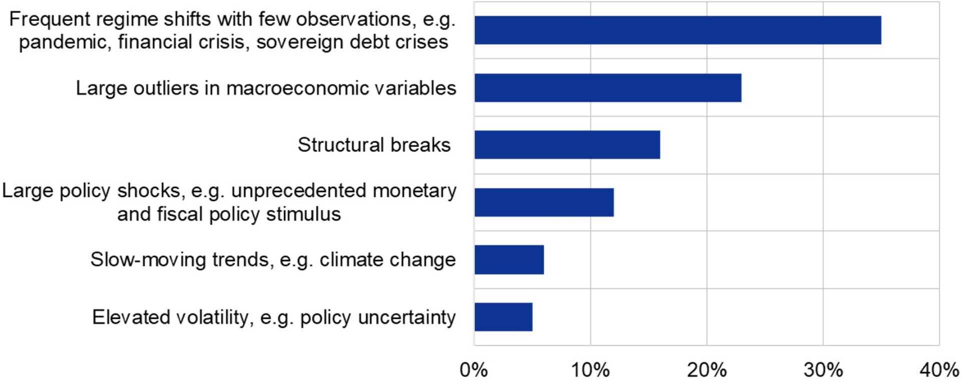
Why time series models fail

Figure: Actual vs Forecast



Why time series models fail

Figure: Challenges to forecasting



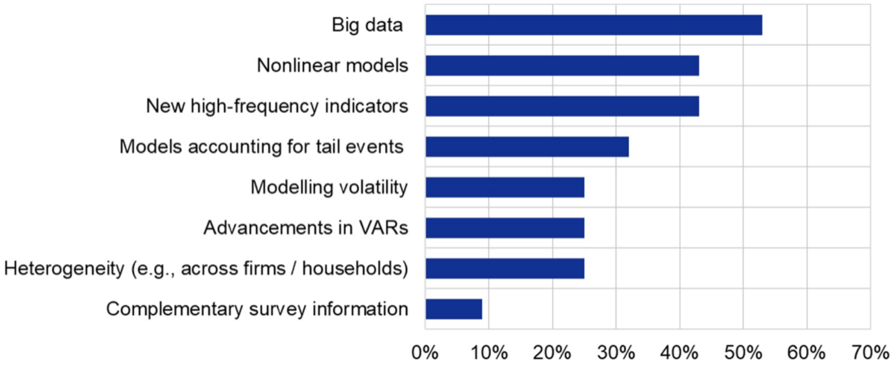
(Source: ECB's 11th Conference on Forecasting Techniques)

Other challenges

- High growth sectors like real estate
- High volatility - investment growth
- Lack of timely data
- Base year changes
- Aggregate forecast vis-a-vis component series forecast

Possible solution

Figure: Possible way out



(Source: ECB's 11th Conference on Forecasting Techniques)

Short term forecasting

Why time series models fail

- Time series models - based on lagged values
- Lagged values - persistence effect

Way out

- Time varying models

$$
\begin{aligned}
y_t &= \alpha_{0t} + \alpha_{1t} y_{t-1} + \alpha_{2t} y_{t-2} + ... + \epsilon_t \\
\alpha_t &= \alpha_{t-1} + \eta_t
\end{aligned} \tag{1}
$$

- Nowcasting - Incorporates news in forecasting

Nowcasting

- Forecasting in real time $\Rightarrow$ Nowcasting
- Gather information from high frequency indicators
- Curse of dimensionality (Hence factors)
- Combine high frequency indicators using dynamic factor model
- **Challenges**
  - Data release calendar differs
  - Not all data releases at same time

Nowcasting - Dynamic factor models

Framework by Giannone et. al. (2006)

$$\begin{aligned} X_t &= \alpha_0 + \alpha_1 F_t + \epsilon_t \\ F_t &= A F_{t-1} + B\eta_t \end{aligned} \tag{2}$$

Bridge equation

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \gamma_2 F_t + \zeta_t \tag{3}$$

- ϵ_t idiosyncratic error (noise)
- Assumption: $\epsilon_t \sim \mathbf{N}(0, \Sigma)$

$$\sigma_{ij}^2 = \begin{cases} \sigma_{ij}^2 & \text{if data available} \\ \infty & \text{if data not available} \end{cases} \tag{4}$$

High frequency indicators used

- Industrial production (IIP)
- Eight Core (EC)
- Consumer price index (CPI)
- Wholesale price index (WPI)
- Money and credit
- Payment system indicators
- PMI and forward looking survey

Variable selection

- Using all variables (Bai & Ng (2002))
- Use a subset of variables
- Variable selection - LASSO, RIDGE or Elastic Net

Elastic Net framework

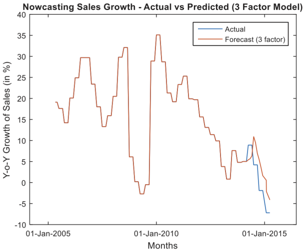
$$Y_t = \gamma_0 + \gamma_1 X_t + \epsilon_t \tag{5}$$

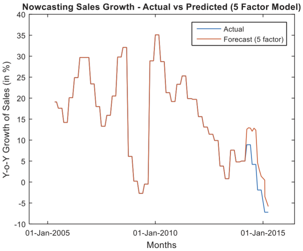
$$min \sum (Y_t - \hat{Y}_t)^2 + \lambda_1 \sum |\gamma_{1i}| + \lambda_2 \sum |\gamma_{1i}|^2 \tag{6}$$

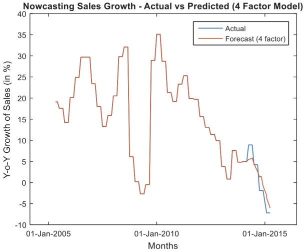
References

- Sanyal, Anirban and Abhiman Das, "Nowcasting sales growth of manufacturing companies in India", Applied Economics, 2018
- Roy, Indrajit, Anirban Sanyal and Aloke Ghosh, "Nowcasting Indian GVA Growth in a Mixed Frequency Setup", RBI Occasional Paper, 2016

Corporate sales: Forecast performance







Corporate sales: Forecast performance

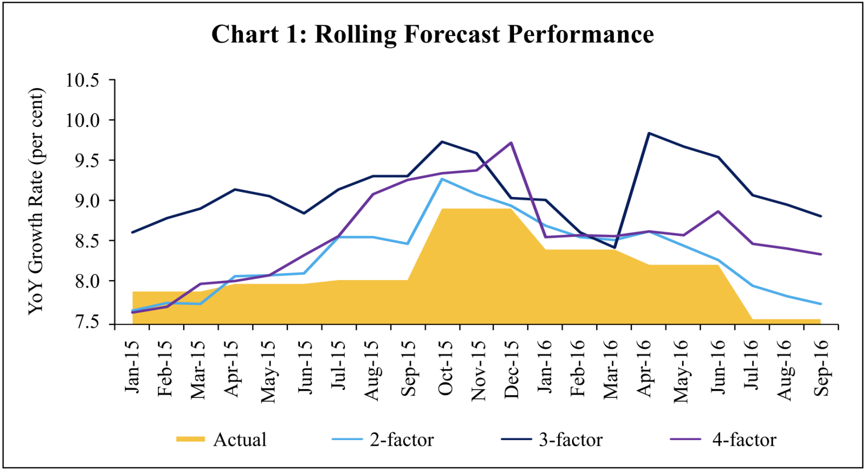
Table 5. Rolling RMSE of nowcasting model and ARIMA.

	Rolling RMSE			
	4 quarters	6 quarter	8 quarter	10 quarter
ARIMA	14.4	13.2	13.7	13.7
3-factor model	5.0	4.3	4.9	6.0
4-factor model	2.7	4.9	5.8	6.0

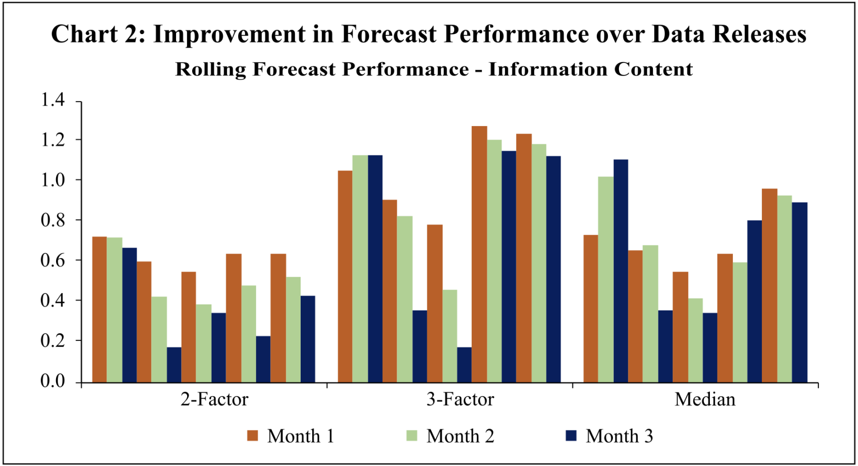
Table 6. Rolling RMSE of combination forecast.

	Rolling RMSE			
Nowcasting models	4 quarter	6 quarter	8 quarter	10 quarter
3-factor model	5.0	4.3	4.9	6.0
4-factor model	2.7	4.9	5.8	6.0
Forecast combination	3.0	4.0	4.7	5.5

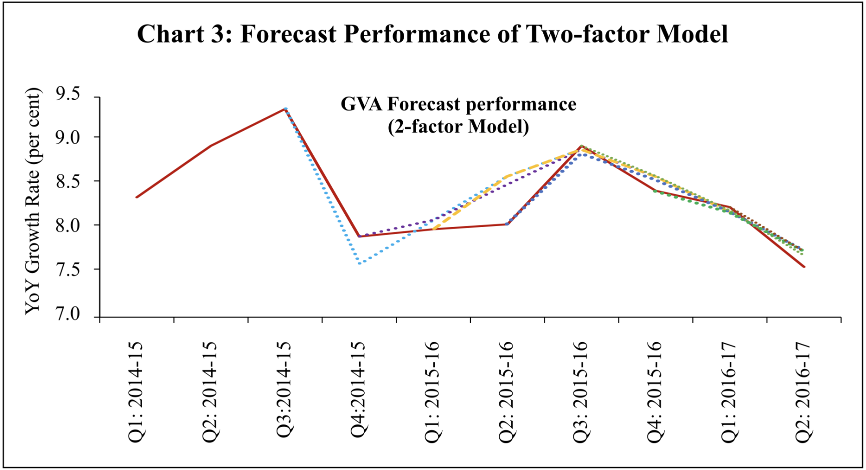
Non Agri GVA: Forecast performance



Non Agri GVA: Forecast performance



Non Agri GVA: Forecast performance



Non Agri GVA: Forecast performance

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

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

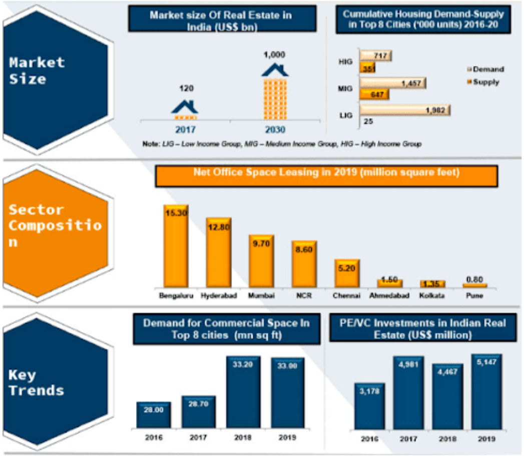
Note: 1-Q = 1 quater ahed forecast; similarly 2-Q, 3-Q, 4-Q.

Forecasting high growth sectors

Reference

- Mitra, Pratik, Anirban Sanyal and Sohini Chowdhury, "Nowcasting real estate activity in India using Google trend data", RBI Occasional Paper, 2017

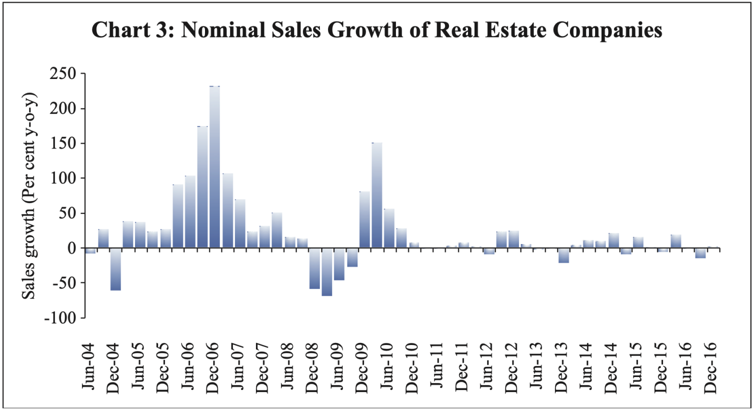
Real estate sector



Source: IBEF

Figure 1: India's Real Estate Sector Infographic

Real estate sector



Challenges in Forecasting

- Lack of high frequency indicators from NSSO/ CSO
- Landscape changes frequently
- Other indicators lacks tracking property
- **Solution**: Google Trend, Social media feeds and news coverage

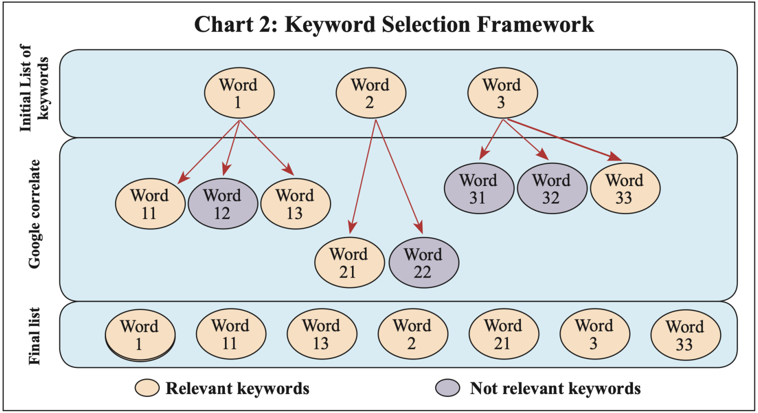
Using Google trends data

- **Google trend**: Search intensity data from google across locations
- Using Google Correlate (Choi and Varian (2012), Kholodin et al., (2010))
- Define search intensity

$$
\begin{aligned}
R_t^i &= \frac{N_t^i}{\sum N_t^i} \times 100 \\
S_t^i &= \frac{R_t^i}{\max R_t^i}
\end{aligned} \tag{7}
$$

- Bootstrapping to remove sampling variation (1000 instances of search frequency)

Selection of keywords



Combining search intensity

- Simple average
- weighted average (weights being inverse of variance)
- Principal component analysis
- Dynamic factor estimates

Nowcasting

$$\begin{aligned} X_t &= \alpha_0 + \alpha_1 F_t + \epsilon_t \\ F_t &= A F_{t-1} + B\eta_t \end{aligned} \tag{8}$$

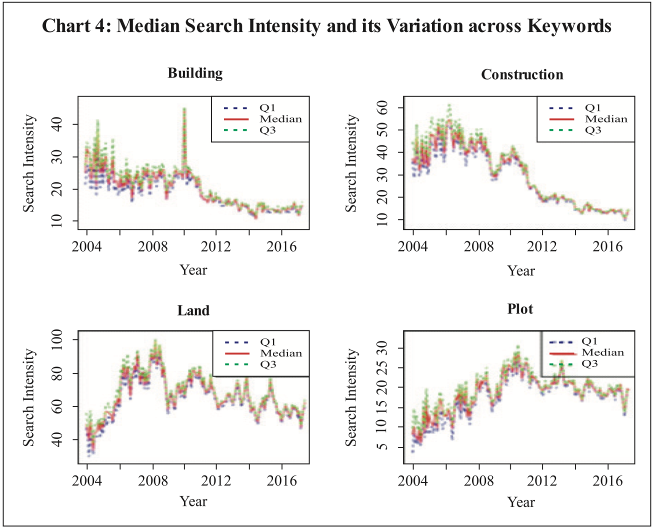
Bridge equation

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \gamma_2 F_t + \zeta_t \tag{9}$$

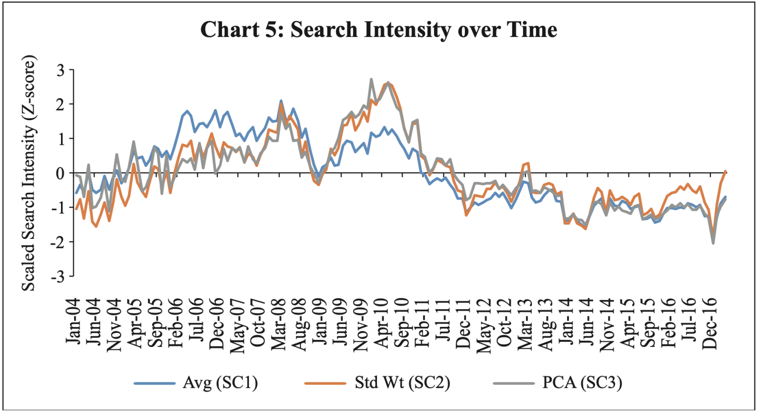
- ϵ_t idiosyncratic error (noise)
- Assumption: $\epsilon_t \sim \mathbf{N}(0, \Sigma)$

$$\sigma_{ij}^2 = \begin{cases} \sigma_{ij}^2 & \text{if data available} \\ \infty & \text{if data not available} \end{cases} \tag{10}$$

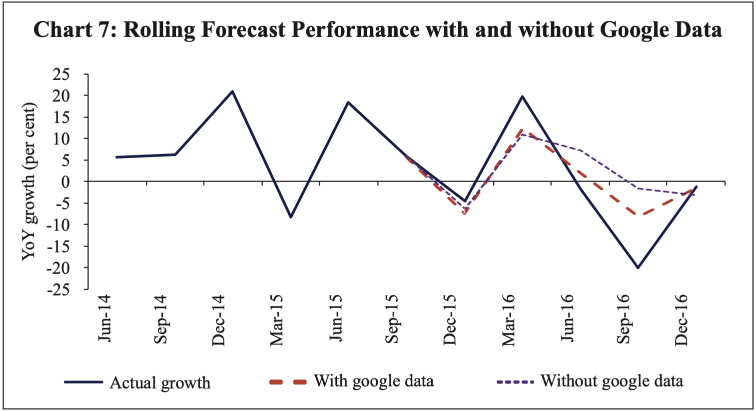
Search intensity



Search intensity



Search intensity



Corporate sentiment and nowcasting

Why nowcasting

- Lack of high frequency data
- Corporates report quarterly financial statement with lag of 45 days
- No timely data during Monetary policy strategy meeting
- Nowcasting for current state assessment

Sentiment analysis

- Sentiment analysis using newspaper reports
- Introduced in 2015-16
- Positive and negative keywords from different dictionary
  - Liu and Hu opinion lexicon: 60K keywords
  - SentiWordNet - 155K keywords (3 point scale)
  - WordStat - more than 9164 negative and 4847 positive word patterns
- News API used: Business Standard (since 2000 onward)
- Sentiment index using relative occurrence frequency

Nowcasting - new frontier after COVID

$$
\begin{aligned}
X_t &= \alpha_0 + \alpha_1 F_t + \epsilon_t \\
F_t &= A F_{t-1} + B\eta_t
\end{aligned} \tag{11}
$$

Bridge equation

$$
\begin{aligned}
Y_t &= \beta_{0t} + \beta_{1t} Y_{t-1} + \gamma_{2t} F_t + \zeta_t \\
&= \alpha_t Z_t + \epsilon_t
\end{aligned} \tag{12}
$$

Coefficient dynamics

$$
\begin{aligned}
\alpha_t &= \alpha_{t-1} + \eta_t^1 \\
h_t &= log(\sigma_t) \\
h_t &= \theta h_{t-1} + \eta_t^2
\end{aligned} \tag{13}
$$

Thank you