

Nowcasting with Large Bayesian VARs

Jacopo Cimadomo, Domenico Giannone, Michele Lenza,
Francesca Monti, Andrej Sokol

Discussion by Minchul Shin¹ (FRB Philadelphia)

Conference on
Real-Time Data Analysis, Methods and Applications, 2020

¹**Disclaimer:** The views expressed here are my own and do not necessarily represent the views of the Federal Reserve Bank of Philadelphia or the Federal Reserve System.

This Paper ...

Modern Big Data analytics, often referred to as the three “Vs”

1. Volume
2. Variety
3. Velocity

Goal: Bring BVAR models up to these challenges

- ▶ Nowcasting/forecasting US real GDP
- ▶ Several policy exercises

This Paper ... More Precisely

The authors perform a systematic (pseudo) out-of-sample forecasting evaluation to compare performance of

- ▶ Three different mixed-frequency BVAR approaches
- ▶ Mixed frequency dynamic factor model currently used at NY-Fed

Point and density forecast evaluation for 2005Q1–2018Q1

This paper definitely adds a significant value to the literature; help others think about ...

- ▶ Available VAR modeling options for nowcasting U.S. GDP
- ▶ Their out-of-sample predictive performance (point and density)

One Comment/Question

In the conclusion ...

“This paper has shown that BVARs can be successfully used to handle Big Data – i.e., a large set of macroeconomic time-series with different frequencies, staggered release dates, and various other irregularities – for real-time nowcasting”

How large is large? How big do we need for good fore/now-casting?

Macroeconomic Forecasting in Big Data World



Search ALFRED

[Categories](#) [Sources](#) [Releases](#) [API](#) [Help](#)

St. Louis

Archival FRED

Economic Data Time Travel from the St. Louis Fed's Economic Research Division.

ALFRED® allows you to retrieve vintage versions of economic data that were available on specific dates in history. In general, economic data for past observation periods are revised as more accurate estimates become available. As a result, previous vintages of data can be superseded and may no longer be available from various data sources. Vintage or real-time economic data allows academics to reproduce others' research, build more accurate forecasting models, and analyze economic policy decisions using the data available at the time. For more information, read the [help page](#) and the presentation [ALFRED: Capturing data as it happens](#).

[Notify Me When New Series and Features Are Added](#)

Currently, vintage data are available for **766,518 series in 9 categories**

Money, Banking, & Finance (16,233)

Population, Employment, & Labor

Markets (54,478)

National Accounts (158,139)

Production & Business Activity

(74,460)

Prices (23,426)

International Data (200,084)

Greenbook Projections (26)

U.S. Regional Data (766,020)

Academic Data (19,236)

The earliest vintage available is 1927-01-26 for series Industrial Production Index.

Macroeconomic Forecasting in Big Data World

We are living in a world with “big” data. For example, we have 766,518 time-series in ALFRED

1. **Volume:**

- ▶ The number of series is large (766,518 variables)

2. **Variety:**

- ▶ Mixed-frequency
- ▶ Heterogenous trending behavior (Mixed-roots)
- ▶ Heterogenous seasonal behavior
- ▶ Heterogenous missing observations pattern

3. **Velocity:**

- ▶ Release dates are different (online estimation and forecasting)

This Paper ...

This paper builds BVAR models with **18** key macroeconomic variables (14 monthly and 4 quarterly)

NY-Fed's DFM is based on about **37** variables (monthly and quarterly)

How did the authors bring # of variables from 766,518 to 18?

- ▶ Indeed, models are of Bayesian, and there must have been some subjective choices reducing the number from 766,518 to 18
- ▶ Economic theory and experts' opinion play a role
- ▶ How large is large for (B)VAR?

How Large Is Large For VAR?

I did a quick Google scholar search with the term “large Bayesian vector autoregression” (as of 10/7/2020).

Top 8 list from Google scholar search and # of variables

1. (2010, JAE) – 131 variables
2. (2012, JBF) – 17 variables
3. (2019, JoE) – 125 variables
4. (2019, JAE) – 138 variables
5. (2009, IJF) – 33 variables
6. (2014, IJF) – 14 variables
7. (2020, JBES) – 20 variables
8. (2016, JBES) – 14 variables

of variables in their VARs ranges from 14 to 138

*Another Google scholar search with “macroeconomic forecasting high dimensional” leads to a similar number with the largest being about 250 variables

How Large Is Large For VAR?

“Large” oftentimes refers to large parameter/state space

For example, B-BVAR specification in the paper

- ▶ 18 variables (4 quarterly, 14 monthly), 5 lags
- ▶ # of VAR coefficients in the conditional mean

$$76^2 \times 5 + 76 = 28,956$$

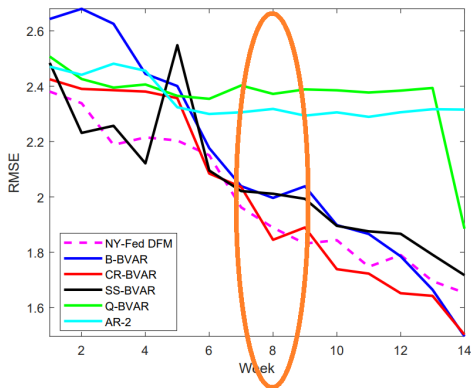
- ▶ # of VAR coefficients in the conditional variance

$$76 \times 77/2 = 2,926$$

- ▶ These numbers explode once we allow for time-varying coefficients (e.g., stochastic volatility)
- ▶ Regularization becomes important

Big-data versus Big-(parameter/state-space)

Experts' Sparse Model *versus* ...



(a) Root mean squared error

In week 8, the best model's RMSE is about 1.9% using 18–37 key macroeconomic variables

Is this the best possible RMSE we can obtain? Even when we started from 766,518 time-series from the ALFRED?

Monitoring U.S. Economy, Past

Burns and Mitchell (1946)'s "Dating Specific and Business Cycles"

- ▶ They use **44 (pre-selected) time-series** to analyze turning points

Shiskin (1961)'s "Statistics for Short-Term Economic Forecasting"

- ▶ *"Groups of different series representing many economic activities – **well over 300** – have already been used, and more will probably be added ... "*
- ▶ *"... These studies have been and will continue to be supplemented by different kinds of statistical analyses of as yet uncovered areas of economic activity and **by seasonally adjusted key weekly series**"*

What have changed since then?

- ▶ Statistical and computational accuracy/efficiency
- ▶ A new set of variables
- ▶ Better measurement

Are we using all possible available data?

- ▶ What is the best possible attainable RMSE by using the set of variables available today relative to the set of variables available 50 years ago?

Monitoring U.S. Economy, Present

Recent monitoring tools that are used and released by some Feds

1. Chicago Fed National Activity Index (CFNAI)
 - ▶ 85 macroeconomic variables (Monthly, Quarterly)
 - ▶ Evans, Liu, and Pham-Kanter (2002)
2. Aruoba-Diebold-Scotti (ADS) Index (PHIL-Fed)
 - ▶ 6 variables (Weekly, Monthly, Quarterly)
 - ▶ Aruoba, Diebold, Scotti (2009)
3. GDPNow (ATL-Fed)
 - ▶ 100–200 variables (Monthly, Quarterly)
 - ▶ Higgins (2014)
4. Nowcasting Report (NY-Fed)
 - ▶ 37 variables (Monthly, Quarterly)
 - ▶ Bok, Caratelli, Giannone, Sbordone, Tambalotti (2017)

Monitoring U.S. Economy, Present

Recent monitoring tools that are used and released by some Feds

1. Chicago Fed National Activity Index (CFNAI)
 - ▶ 85 macroeconomic variables (Monthly, Quarterly)
 - ▶ Evans, Liu, and Pham-Kanter (2002)
2. Aruoba-Diebold-Scotti (ADS) Index (PHIL-Fed)
 - ▶ 6 variables (Weekly, Monthly, Quarterly)
 - ▶ Aruoba, Diebold, Scotti (2009)
3. GDPNow (ATL-Fed)
 - ▶ 100–200 variables (Monthly, Quarterly)
 - ▶ Higgins (2014)
4. Nowcasting Report (NY-Fed)
 - ▶ 37 variables (Monthly, Quarterly)
 - ▶ Bok, Caratelli, Giannone, Sbordone, Tambalotti (2017)
5. **Brave-Butters-Kelley Index (CHI-Fed)**
 - ▶ 500 variables (Monthly, Quarterly)
 - ▶ Brave, Butters, Kelley (2019)
6. **Weekly Economic Index (NY-Fed and DAL-Fed)**
 - ▶ 10 variables (Daily, Weekly)
 - ▶ Lewis, Mertens, Stock (2020)
7. **Daily News Sentiment Index (SF-Fed)**
 - ▶ 16 major U.S. newspapers
 - ▶ Buckman, Shapiro, Sudhof, Wilson (2020)

Moving toward... More data (500+), Higher frequency (daily), Complex data type (text)

A large number of macroeconomic variables are available

- ▶ Mixed-frequency (Daily, Weekly, Monthly, Quarterly)
- ▶ Heterogenous trending behavior
- ▶ Heterogenous seasonal behavior
- ▶ Heterogenous missing pattern
- ▶ Heterogenous aggregation level
- ▶ Text-data, transaction data, etc.

Several decisions to make to feed the data into standard econometric models (BVARs, DFMs, etc.). Experts' opinion is always valuable:

- ▶ Stock-Watson and McCracken-Ng datasets usually come with data transformation suggestion
- ▶ A smaller set of key macroeconomic variables “judiciously” pre-selected by a researcher
- ▶ Regularization becomes very important: various priors available

It would be useful to have an automatic and less subjective way to screen, process, and transform raw “Big” data, and test previous experts' opinion and convention in real-time

Conclusion

This paper provides convincing evidence that Bayesian VAR models are able to produce accurate probabilistic predictions about the US real GDP in real-time

- ▶ Point and density prediction evaluation for three different MF-VAR modeling strategies and the MF-DFM
- ▶ BVARs and DFMs are all useful for nowcasting/forecasting
- ▶ They are useful for other policy exercises (IRF, conditional Forecasting, etc.)

Final remarks

- ▶ Developing a way to evaluate and improve “Expert’s sparse model” in real-time may be valuable
- ▶ Modeling time-varying higher moments (SD, Skew, Tail) may improve density prediction performance
- ▶ Documenting computational costs may be useful