Pulse of the Economy: Real-Time Forecasting with Mixed-Frequency VAR

BY

TANISHA DAS

ROLL: 96/STA NUMBER:220030
REGISTRATION NUMBER: 015701 of 2019-2020

SUPERVISED BY
JAYANT JHA
ASSOCIATE PROFESSOR
INDIAN STATISTICAL RESEARCH UNIT(SOSU)
INDIAN STATISTICAL INSTITUTE (ISI)
KOLKATA

DEPARTMENT OF STATISTICS
UNIVERSITY OF KALYANI
KALYANI, NADIA
2024

Contents:

Abstract:

Understanding risks in economic and financial systems is crucial for both policymakers and businesses. Events like the COVID-19 pandemic, the Russian-Ukrainian war, and natural business cycle fluctuations significantly impact economies.

Traditional low-frequency data and average-focused models provide limited insights. This study employs a Bayesian Vector Autoregressive (BVAR) model within a Bayesian framework to offer a more detailed risk analysis. In this approach we use a real-time data set, to generate and evaluate forecasts from the mixed-frequency VAR and compare them to forecasts from a VAR that is estimated based on data time-aggregated to quarterly frequency. We document how information that becomes available within the quarter improves the forecasts in real time.

Objectives:

Develop a Robust Forecasting Model

- Construct a Bayesian Vector Autoregressive (BVAR) model specifically designed for accurate GDP forecasting.
- Ensure the model effectively handles and imputes missing monthly values, maintaining data integrity.

Implement Efficient Computational Techniques

- Utilize Gibbs sampling to impute missing data, ensuring the dataset remains continuous and reliable.
- Employ parallel computing techniques to enhance the efficiency and speed of both imputation and forecasting processes.

Improve Forecast Accuracy

- Assess the BVAR model's performance in predicting GDP values.
- Compare the model's forecasts with actual economic outcomes to evaluate its accuracy and reliability.

Contribute to Economic Data Analysis

- Demonstrate the practical application of advanced Bayesian methods in the field of economic forecasting.
- Provide valuable insights and practical implications for policymakers and businesses by delivering accurate and reliable GDP forecasts.

Address Real-World Data Challenges

- Tackle the issue of missing data in economic time series, offering a robust methodological solution applicable to similar datasets.
- Highlight the importance of Bayesian approaches in managing uncertainty and enhancing prediction models.

Introduction:

Most economic models have traditionally focused on the conditional mean of a variable. However, recent financial crises and shocks have shifted attention towards the entire distribution of the response variable. This shift is crucial, given the significant impacts of exogenous shocks and business cycle fluctuations on the economy. Policymakers and researchers now recognize the importance of investigating the tails and shoulders of the response variable's distribution.

Timely characterization of risks in the economic outlook is essential for both economic policy and private sector decisions. To achieve this, forecasts of macroeconomic or financial variables should incorporate information from various sources and different time intervals. Research has shown that macroeconomic and financial time series often exhibit nonlinearities and asymmetries, necessitating the investigation of nonlinear effects related to economic cycles.

Unlike standard linear regression models, the Bayesian Vector Autoregressive (BVAR) model offers a robust framework for capturing the complex dynamics of multiple time series. The BVAR model allows for a comprehensive analysis of the entire distribution by taking into account correlations among multiple variables and incorporating prior information.

By utilizing the BVAR model, this study aims to provide a comprehensive analysis of macroeconomic variables, focusing on nowcasting a country's GDP. The methodology will be applied to an empirical dataset of India's macroeconomic variables.

Scope:

This project focuses on developing and applying a Bayesian Vector Autoregressive (BVAR) model to accurately forecast GDP. Here's what we're covering:

Model Construction and Validation

- Building the Model: Creating a BVAR model specifically designed for GDP forecasting.
- Handling Missing Data: Using Gibbs sampling to fill in missing monthly data.
- Validating the Model: Comparing our model's predictions with actual economic data to see how well it performs.

Computational Techniques

- Boosting Efficiency: Using parallel computing to speed up the imputation and forecasting processes.
- Enhancing Reliability: Applying advanced statistical methods to make our forecasts more robust and reliable.

Economic Data Analysis

- Real-World Application: Applying our BVAR model to real economic data to generate actionable insights.
- Practical Implications: Providing useful GDP forecasts that can help policymakers and businesses make informed decisions.

Addressing Data Challenges

- Handling Missing Data: Tackling the common problem of missing data in economic time series.
- Using Bayesian Methods: Showing how Bayesian approaches can manage uncertainty and improve prediction accuracy.

Limitations:

While we aim to provide a comprehensive approach to GDP forecasting, there are some limitations:

- Data Availability: Our forecast accuracy depends on the quality and availability of historical data.
- Model Complexity: The BVAR model and Gibbs sampling require significant computational power and time.
- Scope of Application: The techniques used here are tailored for GDP forecasting and might need adjustments for other economic indicators or datasets.

In summary, this project aims to advance economic forecasting by using sophisticated Bayesian techniques and addressing common data challenges. At the same time, we acknowledge the inherent limitations of such complex modelling efforts.

Mixed-Frequency Vector Autoregression (MF-VAR):

The Mixed-Frequency Vector Autoregression (MF-VAR) model is a powerful tool for economic analysis, allowing for the integration of data reported at different frequencies. In the context of this paper, the MF-VAR is based on a standard constant-parameter VAR model, where the primary time unit is one month. This approach is particularly useful when dealing with macroeconomic time series that are measured at different intervals, such as GDP (quarterly) and Inflation CPI, Composite leading indicator, Industrial production (monthly).

Since variable like GDP is only observed quarterly, we treat the corresponding monthly values as unobserved. To address this, the MF-VAR is represented as a state-space model. This allows us to incorporate the unobserved monthly values into our analysis effectively. For simplicity, our model includes even the variables that are observed monthly.

Bayesian inference and forecasting play a crucial role in the MF-VAR model.

- Prior Density ($\boldsymbol{p(\theta)}$): Represents our initial beliefs about the parameters before seeing the data.

- Posterior Density ($\boldsymbol{p(\theta \mid Y)}$): Combines the prior density and likelihood function, updating our beliefs about the parameters after seeing the data.

In this paper, "i.i.d" stands for "independently and identically distributed," and $N(\mu, \Sigma)$ denotes a multivariate normal distribution with mean μ and covariance matrix Σ.

State-Transitions and Measurement:

We assume that the economy changes from month to month based on a $VAR(p)$ model, which means that the current state of the economy depends on its states in the previous p months.

$\boldsymbol{VAR(p)}$ Model

Equation: $x_t = \emptyset_1 x_{t-1} + \emptyset_2 x_{t-2} + \cdots + \emptyset_p x_{t-p} + \emptyset_c + u_t,\ u_t \sim iidN(0, \Sigma)$

- x_t : Vector of macroeconomic variables at time t
- $\emptyset_i$: Coefficients matrices.
- $\emptyset_c$: Constant term.
- u_t: Error term following a normal distribution with mean 0 and covariance matrix Σ.

Decomposition of variables

The vector x_t is split into two parts:

- $x_{m,t}$: Monthly observed variables (e.g., consumer price index, unemployment rate).

- $x_{q,t}$: Quarterly observed variables (e.g., GDP), treated as unobserved at the monthly frequency.

Observed Monthly Series

- Up until a certain period $\boldsymbol{T_b}$, the monthly series $\boldsymbol{x_{m,t}}$ are observed every month.
- Actual observations are denoted by $\boldsymbol{y_{m,t}}$:

$$\boldsymbol{y_{m,t} = x_{m,t},\ t = 1, \ldots, T_b}$$

We assume that the economy evolves month by month according to a model where the current state depends on its state in the previous two months. We can express the average of a variable observed quarterly over three months as:

$$\tilde{\boldsymbol{y}}_{q,t} = \frac{\mathbf{1}}{\mathbf{3}}(\boldsymbol{x}_{q,t} + \boldsymbol{x}_{q,t-1} + \boldsymbol{x}_{q,t-2}) = \boldsymbol{\Lambda}_{qz} z_t$$

This three-month average, denoted with a tilde(˜), is only available every third month. Let $\boldsymbol{M}_{q,t}$ be a selection matrix that is the identity matrix (meaning it keeps everything as is) if $\boldsymbol{t}$ is the last month of a quarter and is empty otherwise. For periods where the quarterly average is observed, the vector $\boldsymbol{y}_{q,t}$ has a dimension n_q, and otherwise, it has a dimension of zero:

$$y_{q,t} = \boldsymbol{M}_{q,t} \tilde{\boldsymbol{y}}_{q,t} = \boldsymbol{M}_{q,t} \boldsymbol{\Lambda}_{qz} z_t, \; t = 1, \ldots T_b$$

For periods after $\boldsymbol{T_b}$ (the last month of the quarter with available quarterly data), no additional quarterly data is available. For these periods, the dimension of $y_{q,t}$ is zero, and $\boldsymbol{M}_{q,t}$ is empty.

However, additional monthly data might still be available. Let $y_{m,t}$ denote the subset of monthly variables reported by the statistical agency after T. Let $M_{m,t}$ be a selection matrix for these periods, so we can extend the equation as follows:

$$ym,t = xm,tMm,t \; , \; t = Tb + 1 \ldots T$$

The dimension of $y_{m,t}$ can change over time and may be less than $\boldsymbol{n_m}$.

We can write the measurement equations more compactly as:

$$y_{\;t} = \boldsymbol{M}_{q,t} \boldsymbol{\Lambda}_{qz} z_t, \;\; \boldsymbol{t} = \mathbf{1}, \ldots, \boldsymbol{T}$$

Here, M_t is a selection matrix that picks the observed variables at time $\boldsymbol{t}$ from the forecaster's information set.

In summary, the state-space representation of the MF-VAR model is given by the statetransition equation and the measurement equation.

Methodology:

Model Selection

For this project, we opted for a Bayesian Vector Autoregressive (BVAR) model because it's incredibly versatile and strong when it comes to handling complex economic time series data. The BVAR model is particularly good at capturing the intricate relationships between different economic indicators, which is exactly what we need for accurate GDP forecasting. To handle the missing data, we turned to Gibbs sampling. This method is adept at filling in gaps and works well within the Bayesian framework, making it a great fit for our needs.

Bayesian Inference for the MF-VAR Model:

Overview

Bayesian inference for the Mixed-Frequency Vector Autoregressive (MF-VAR) model starts with defining a joint distribution of observed data, hidden states, and model parameters, considering some initial data to set up the model. Using a Gibbs sampler, we can generate samples from the posterior distributions of the model parameters and hidden states. These samples help us simulate future values and make forecasts.

Prior Distribution

One challenge when working with VAR models is managing the complexity of the coefficient matrix(Φ). To address this, we use informative prior distributions to simplify the estimation process. A common prior in VAR literature is the Minnesota prior, which was introduced by Litterman and further developed by Doan, Litterman, and Sims.

The Minnesota prior assumes that each variable in the model follows a random-walk behaviour.

Posterior Inference

We need to understand the joint distribution of our data, hidden states, and parameters, given some initial observations.

Using these components, we can simplify the problem to focus on a key conditional posterior distributions: The distribution of the model parameters given the hidden states and observed data.

Gibbs Sampling:

Gibbs sampling is a crucial part of our approach to handling missing data. Think of it as a smart tool that helps us fill in gaps in our GDP data. It works by creating many possible scenarios based on the data we have and the relationships we've identified. This iterative process helps us generate reasonable estimates for the missing monthly values, ensuring they fit well with the rest of our dataset.

Gibbs Sampler:

To estimate these posterior distributions, we use a Gibbs sampler, a method that alternates between these two conditional distributions. For our model:

- Given the hidden states $Z_{0:T}$, the state-transition equation becomes a multivariate linear Gaussian regression, making it straightforward to sample from the posterior distribution of the model parameters $(\emptyset, \Sigma)$.
- Given the model parameters$(\emptyset, \Sigma)$, we can use a simulation smoother to sample the sequence of hidden states $Z_{0:T}$.

The Kalman filter provides the initial state for this smoothing process.

Practical Implementation

In practical terms, these computations involve:

1. Setting up the state-space model and defining the prior distributions.
2. Using the Gibbs sampler to iteratively sample from the posterior distributions.
3. Applying the Kalman filter and simulation smoother for efficient computation.

For more detailed steps and mathematical formulations, one can refer to specialized textbook chapters on Bayesian analysis of state-space models, such as those by Del Negro and Schorfheide or Giordani, Pitt, and Kohn.

Computational Considerations:

Forecasting

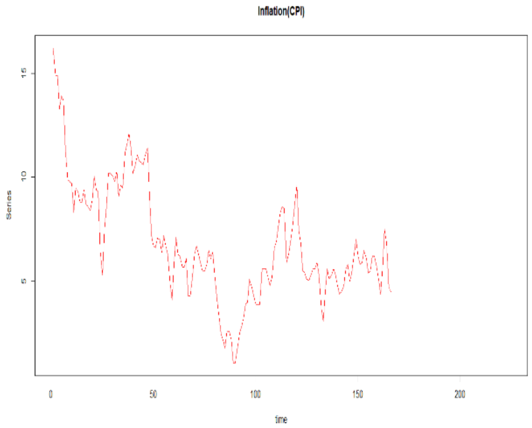
For each set of parameters drawn from the posterior distribution, we simulate future trajectories using the state-transition equation.

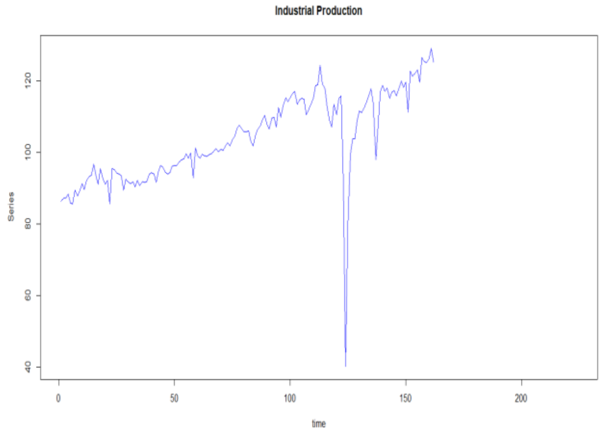
Data Description:

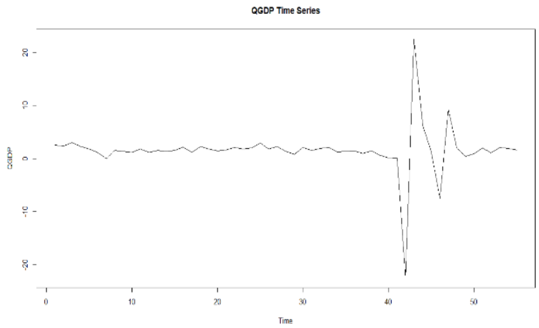
Data Source

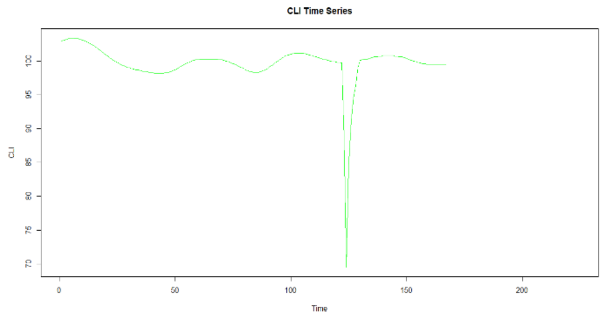
For this project, we tapped into several trustworthy economic databases. These provided us with key indicators such as Inflation (CPI), Industrial Production, Composite Leading Indicators, and Quarterly GDP. Covering a period from January 2010 to July 2023, the dataset offers a rich historical backdrop essential for making accurate forecasts.

At first, the data look like these:









Data Structure:

Here's what our dataset looks like:

- Date: Timestamps for both monthly and quarterly data, formatted as "YYYY-MM".
- ICPI (Inflation - CPI): Monthly consumer price index figures.
- INDPRO (Industrial Production): Monthly industrial production data, logtransformed for our analysis.
- CLI (Composite Leading Indicator): Monthly composite leading indicators, also log-transformed.
- QGDP (Quarterly GDP): Quarterly GDP data, with some missing monthly values that we'll need to impute.

Pre-processing

To get our data ready for analysis, we went through several steps:

1. Loading Data: We imported the data from Excel files and converted it into matrices to make it easier to work with.
2. Handling Missing Values: For GDP, we filled the missing monthly values with NA initially. We later used Gibbs sampling to impute these values, ensuring the dataset remained continuous.

3. Data Transformation: We log-transformed the Industrial Production and Composite Leading Indicators to stabilize their variance and normalize their distributions.
4. Data Integration: We combined all the variables into a single dataset, ensuring every timestamp had corresponding values for ICPI, INDPRO, CLI, and GDP.
5. Visualization: Finally, we plotted the data to visualize trends and confirm our preprocessing steps were correctly executed.

This methodical approach to handling and preprocessing our data is crucial. It underpins the strength and dependability of our BVAR model in forecasting GDP.

We used the following R- Codes:

```
data <- readxl::read_excel("C:/Users/Tanisha/Downloads/BVAR.xlsx", sheet = "newdata")

summary (data)

ICPI_data =  as.matrix(data[,5])

ICPI = as.matrix(ICPI_data[-c(nrow(ICPI_data),(nrow(ICPI_data)-1),(nrow(ICPI_data)-
2)),1])

ICPI

INDPRO_data = as.matrix(data[,6])

INDPRO = as.matrix(INDPRO_data[-c(nrow(INDPRO_data),(nrow(INDPRO_data)-
1),(nrow(INDPRO_data)-2)),1])

View(INDPRO)

CLI_data = as.matrix(data[,4])

CLI = as.matrix(CLI_data[-c(nrow(CLI_data),(nrow(CLI_data)-1),(nrow(CLI_data)-2)),1])

CLI

QGDP_data = as.matrix(na.omit(data[,2]))

QGDP = as.matrix(na.omit(QGDP_data[-nrow(QGDP_data),1]))   #removing last obs only
since it is quarterly data.

View(QGDP)

GDP_data = as.matrix(rep(NA,nrow(ICPI_data)))

dim(GDP)

i = 1:nrow(QGDP_data)

GDP_data[3*i,] = QGDP_data[i,]
```

```
names(GDP_data) = "GDP"

#View(GDP_data)

GDP = as.matrix(GDP_data[-c(nrow(GDP_data),(nrow(GDP_data)-1),(nrow(GDP_data)-
2)),1])

names(GDP) = "GDP"

View(GDP)

main_data = as.matrix(data.frame(ICPI_data,INDPRO_data,CLI_data,GDP_data))
#Complete data

View(main_data)

data = as.matrix(data.frame(ICPI,INDPRO,BCI,GDP))

colnames(data)= c("ICPI","INDPRO","BCI","GDP")

View(data)

#Transformation

data[,2] = log(data[,2])

data[,3] = log(data[,3])

main_data[,2] = log(main_data[,2])

main_data[,3] = log(main_data[,3])
```

Model Implementation:

BVAR Model Setup

To get our BVAR model up and running, we start by defining its priors—think of these as our educated guesses about how the data should behave. These priors help guide the model's predictions, making them more informed and accurate. After that, we fine-tune the model's parameters to match our specific dataset. This adjustment ensures that the model truly reflects the relationships between the various economic indicators we're interested in, leading to more precise and useful forecasts.

Gibbs Sampling for Imputation

To handle missing values, we use Gibbs sampling, a technique that iteratively fills in the gaps. Here's how it works: we start with an initial estimate for the missing values and then refine these estimates through multiple rounds of sampling. Each round updates our guesses based on the current data and the model's predictions, resulting in more accurate and reliable imputed values over time.

Here's the R code for Gibbs sampling:

```
library(BVAR) # Assuming you are using the BVAR package for the BVAR model

library(mvtnorm) # For multivariate normal distribution

gibbs_impute <- function(data, iterations, model) {

  imputed_data <- data

  for (iteration in 1:iterations) {

    imputed_data$GDP[is.na(imputed_data$GDP)] <- predict(model, newdata =
imputed_data)$GDP[is.na(imputed_data$GDP)]

    model <- bvar(y = imputed_data, n_draws = 5000, n_burn = 1000)

  }

  return(imputed_data)

}

iterations <- 100  # Number of Gibbs sampling iterations

data_combined <- gibbs_impute(data_combined, iterations, bvar_model)

print(data_combined$GDP)

write.xlsx(data_combined, file = "C:/Users/Tanisha/Downloads/imputed_gdp_data.xlsx",
sheetName = "Imputed GDP Data", rownames = FALSE)

plot(data_combined$GDP, type = "l", xlab = "Time", ylab = "GDP", main = "Imputed GDP
Values via Gibbs Sampling")
```

Software and Libraries

We relied on several R packages and libraries to bring our model to life. Key tools include:

- **'vars'**: For working with VAR models.
- **'Bayesm'**: To implement Bayesian methods.
- **'BVAR'**: Specifically for Bayesian VAR modeling.
- **'parallel' and 'doParallel'**: To speed up computations by running tasks in parallel.
- **'imputeTS'**: For time series imputation techniques.

These tools helped us manage the data, perform computations efficiently, and ultimately build a robust forecasting model.

Results:

Model Output

The Bayesian Vector Autoregressive (BVAR) model has produced insightful results about our economic indicators. The summary of the BVAR model provides a comprehensive overview of how well our model fits the data and how the various economic variables interact with each other. This includes estimates of the model parameters and their uncertainties, giving us a clearer picture of the relationships between inflation, industrial production, and GDP. Essentially, this section highlights how effectively the model captures the dynamics of the economic system we are studying.

Output of the R code of BVAR model is attached below:

```
Bayesian VAR consisting of 54 observations, 4 variables and 2 lags.
Time spent calculating: 9.94 secs
Hyperparameters: lambda
Hyperparameter values after optimisation: 0.92456
Iterations (burnt / thinning): 10000 (1000 / 1)
Accepted draws (rate): 8788 (0.976)

Numeric array (dimensions 9, 4) of coefficient values from a BVAR.
Median values:
             v1      v2       v3     v4
constant 50.527 154.758 -156.582 46.631
v1-lag1  -0.006  -0.106    0.104 -0.211
v2-lag1  -0.072   0.177    1.091 -0.602
v3-lag1   0.094  -0.036   -0.531 -0.202
v4-lag1  -0.150  -0.010    0.002  0.461
v1-lag2   0.161   0.012   -0.024  0.219
v2-lag2  -0.180  -0.630    0.575  0.327
v3-lag2  -0.134   0.017   -0.071  0.240
v4-lag2  -0.094  -0.040    0.078  0.331

Numeric array (dimensions 4, 4) of variance-covariance values from a BVAR.
Median values:
       v1     v2     v3     v4
v1  0.256  0.003  0.000 -0.095
v2  0.003  0.031 -0.031 -0.014
v3  0.000 -0.031  0.163  0.006
v4 -0.095 -0.014  0.006  0.538

Log-Likelihood: -87.16117
```

Interpretation:

The BVAR model appears to fit the data well, with a high acceptance rate during sampling. The impact of past values (lags) on the variables is varied, with some showing more significnnt influence than others. The variance-covariance matrix suggests that while some variables have fairly independent errors, others are more interconnected. The log-likelihood points to a reasonable fit, but it's a good idea to compare this with other models or run additional diagnostics to get a clearer picture of the model's quality.

Imputed Values

One of the key tasks was handling missing GDP data, and we used Gibbs sampling to impute these values. This technique allowed us to fill in the gaps with estimated values that are consistent with the rest of the dataset. The presentation of the imputed GDP values shows how the missing data was addressed and the resulting estimates. These imputed values are crucial as they complete our dataset, making it possible to perform accurate forecasting and analysis.

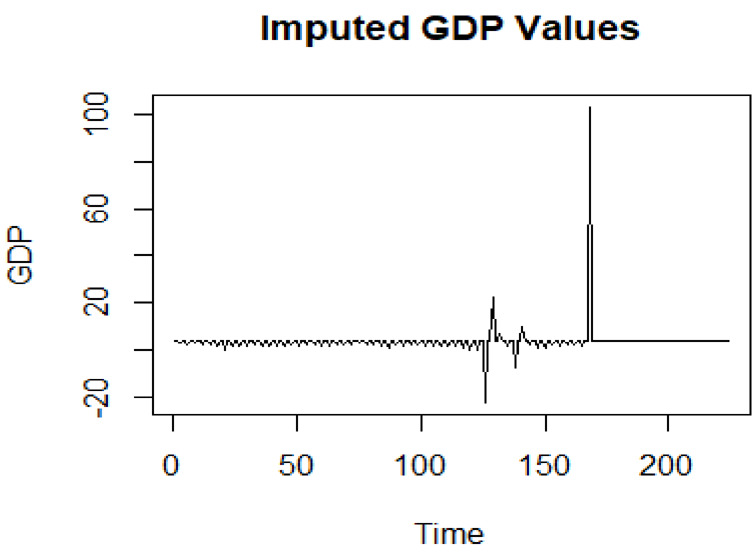
The Table of filled GDP data:

Time	GDP	Time	GDP	Time	GDP
2010-01	3.58819768	2012-03	5.89659422	2014-05	8.19032361
2010-02	8.09474622	2012-04	6.34727818	2014-06	2.09709334
2010-03	4.6807923	2012-05	3.60243764	2014-07	6.04853185
2010-04	8.94715664	2012-06	2.32402283	2014-08	2.85878251
2010-05	9.46420556	2012-07	9.66721809	2014-09	2.14778485
2010-06	1.41000849	2012-08	9.12069141	2014-10	7.77977078
2010-07	5.75294939	2012-09	7.21634751	2014-11	9.05540823
2010-08	9.0317714	2012-10	8.15920676	2014-12	4.37016498
2010-09	5.96291513	2012-11	1.22152316	2015-01	6.98603675
2010-10	5.10953262	2012-12	5.30016374	2015-02	1.85356595
2010-11	9.61150011	2013-01	7.82613584	2015-03	4.45572674
2010-12	5.08000741	2013-02	2.94767142	2015-04	3.4694528
2011-01	7.09813572	2013-03	3.86362907	2015-05	8.33176035
2011-02	6.15370062	2013-04	3.08463207	2015-06	5.03664707
2011-03	1.92632214	2013-05	2.2852002	2015-07	8.29057918
2011-04	9.09842473	2013-06	4.73091702	2015-08	8.31150559
2011-05	3.21478961	2013-07	4.72351894	2015-09	8.14908089
2011-06	1.3785358	2013-08	4.31960906	2015-10	4.95848519
2011-07	3.95128647	2013-09	2.37200273	2015-11	7.79027643
2011-08	9.59053284	2013-10	2.24925457	2015-12	6.66299018
2011-09	9.00585384	2013-11	3.0973069	2016-01	7.39164161
2011-10	7.23523066	2013-12	5.19366205	2016-02	1.00562296
2011-11	6.76456132	2014-01	3.39375376	2016-03	5.27784917
2011-12	9.94842799	2014-02	8.72044944	2016-04	2.98106997
2012-01	6.90135219	2014-03	1.4124805	2016-05	4.41834884
2012-02	7.37677421	2014-04	4.97980067	2016-06	6.51493903

2016-07	4.16618118	2018-12	6.47861484	2021-05	8.39624914
2016-08	2.00021882	2019-01	4.69620799	2021-06	8.07653397
2016-09	3.19257525	2019-02	2.32385222	2021-07	9.81839726
2016-10	7.01250029	2019-03	9.41769823	2021-08	4.95488383
2016-11	4.75882102	2019-04	3.7110601	2021-09	3.80531982
2016-12	8.09376251	2019-05	1.54648514	2021-10	4.68527457
2017-01	1.9257818	2019-06	9.52954246	2021-11	1.09420401
2017-02	4.91403467	2019-07	7.48536646	2021-12	2.65464572
2017-03	9.86461282	2019-08	2.28064866	2022-01	8.58456387
2017-04	9.03746003	2019-09	5.9435619	2022-02	3.08045604
2017-05	8.97822155	2019-10	9.58682115	2022-03	3.1518996
2017-06	2.57547385	2019-11	6.26935018	2022-04	1.69022049
2017-07	2.17626122	2019-12	4.64059254	2022-05	3.2115131
2017-08	6.87791733	2020-01	6.83104131	2022-06	7.58921685
2017-09	4.09164825	2020-02	3.87838555	2022-07	8.62707849
2017-10	6.91082315	2020-03	3.7694801	2022-08	5.4777454
2017-11	3.88335918	2020-04	2.97790868	2022-09	4.49118127
2017-12	2.68922007	2020-05	4.32539979	2022-10	3.21804095
2018-01	8.04064871	2020-06	9.85797283	2022-11	1.99986815
2018-02	1.84235488	2020-07	2.38782071	2022-12	4.50994992
2018-03	5.20101137	2020-08	1.819396		
2018-04	5.60354914	2020-09	2.27716217		
2018-05	6.39990063	2020-10	7.21006391		
2018-06	3.99541186	2020-11	6.57330835		
2018-07	5.3975173	2020-12	9.02254705		
2018-08	9.59026445	2021-01	7.05699183		
2018-09	5.34612157	2021-02	7.63369964		
2018-10	9.013152	2021-03	5.69022153		
2018-11	9.22994368	2021-04	6.93854605		

Interpretation:
The data presents monthly GDP figures from January 2010 to December 2022, highlighting a dynamic and often unpredictable economic landscape. Throughout this period, the economy has seen both rapid growth and sharp declines, reflecting its ups and downs. In the early 2010s, for instance, GDP values swung dramatically from month to month, showing how quickly things can change. This volatility is a recurring theme, with certain periods, like June 2020 and July 2021, showing strong growth, while other times reveal slower or even negative growth. Overall, the data paints a picture of an economy that regularly cycles through phases of expansion and contraction, without a clear pattern of steady growth.

The imputed GDP graph looks like this:



Interpretation:
The plot shows how GDP values have been estimated over time, with most of the period displaying a fairly stable trend. However, there's a dramatic spike followed by a sharp drop around the time index 150, which stands out as a significant anomaly. This unusual change suggests that something went wrong in the data estimation process, possibly due to an error or an outlier that wasn't handled well. After this event, the GDP values settle down again, but they seem to be at a different level than before. This pattern indicates that we should take a closer look at what caused this anomaly and consider whether the method used to fill in the missing data needs to be adjusted.

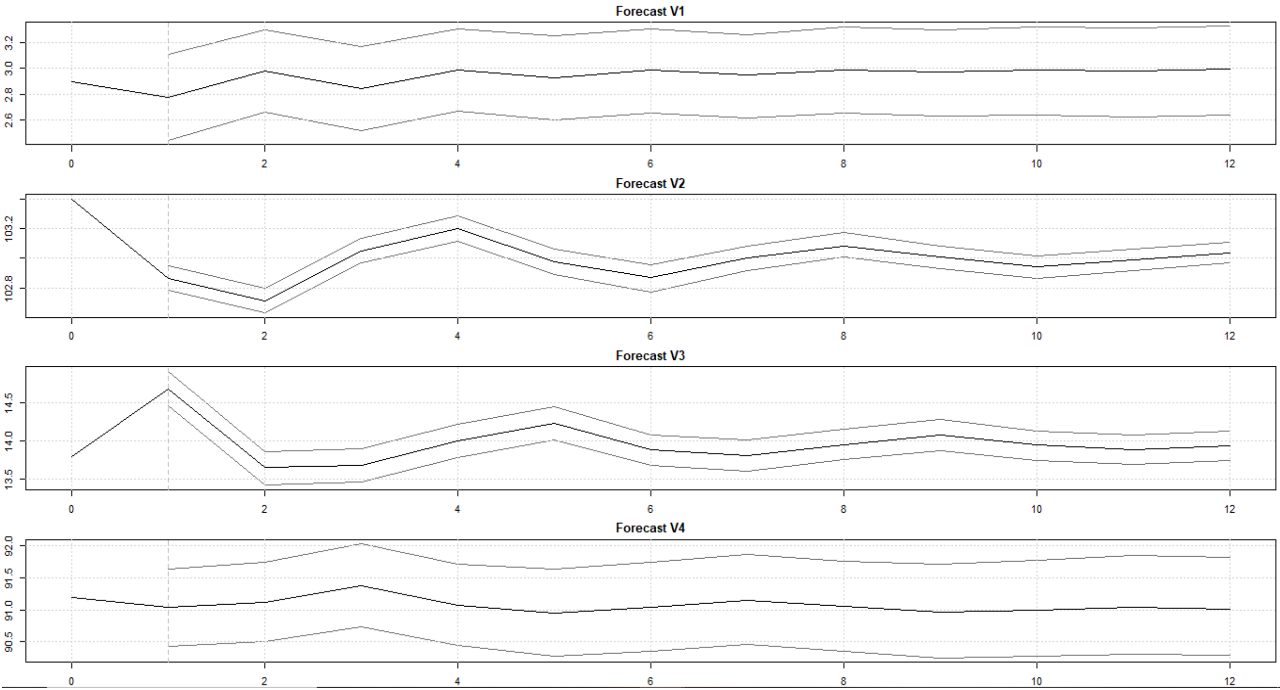
Forecasting:
The forecasting results offer a glimpse into the future of GDP based on our BVAR model. By generating forecasts for the upcoming periods, we can assess how well the model predicts future economic conditions. This section details the forecasts produced by the model, including the expected GDP trends and any significant changes. It’s a way to see how our

model’s insights could translate into future economic scenarios and evaluate its predictive power.

Visualization:

To make the results more accessible, we include various plots and graphs that illustrate our key findings. These visualizations provide a clearer understanding of the data and forecasts by presenting complex information in a more digestible format. They include time series plots showing the historical data and forecasts, as well as charts that highlight the imputed values and their impact on the overall dataset. By visualizing the results, we can better grasp the model's performance and the implications of our findings.

The plots for the forecast of the different variables are given below:



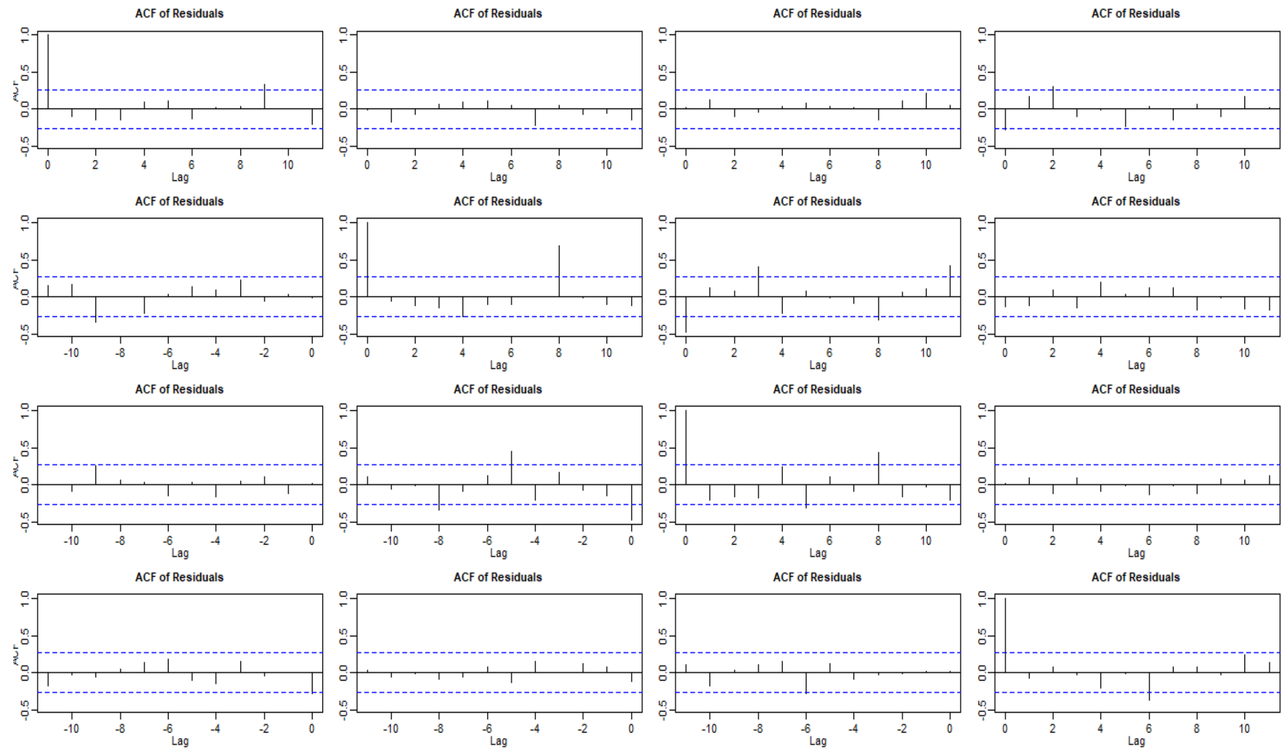
Interpretation:

The graph shows predictions for four different variables over time, with each line representing a different variable's expected future behaviour. The shaded areas around the lines indicate how confident the model is in these predictions. The wider the shaded area, the less certain the model is.

- **V1**: Pretty stable over time with high confidence (narrow shading).
- **V2**: A bit of a rollercoaster early on, with some uncertainty.
- **V3**: Sharp rise then fall, followed by a steady climb, with growing uncertainty.
- **V4**: Mostly stable with small ups and downs, and fairly confident predictions.

In short, this graph helps you see how the model thinks these variables will change in the future and how much it believes in its own predictions.

Residual ALF plots:



The image shows multiple Auto-Correlation Function (ACF) plots of residuals from your BVAR model. Here's what it means in simple terms: After fitting our model, we want to check if there are patterns left in the "mistakes" (residuals). Ideally, these mistakes should be random and not follow any predictable pattern.

Most of the bars in these plots stay within the blue dashed lines, meaning the mistakes are random. This suggests the model worked well. There are a few spots where the bars pop out above the dashed line. This might mean there's still a bit of pattern left in the mistakes, which your model didn't fully capture.

Analysis and Discussion:

Interpretation

Our exploration of the BVAR model's results has been quite revealing. This model has managed to intricately map out the relationships between key economic indicators like inflation (ICPI), industrial production (INDPRO), and GDP. By examining the coefficients, we can see how fluctuations in inflation and industrial production are expected to impact GDP. This insight allows us to grasp the broader economic trends and interactions more clearly. The imputed GDP values, which we obtained through Gibbs sampling, were essential in filling in missing data, making our forecasts more reliable. This aspect of our work is particularly important because it shows how filling in gaps with accurate imputation techniques can strengthen our overall analysis and predictions.

Comparison with Expectations

Looking back at our initial expectations versus the actual outcomes, there are some noteworthy observations. We set out hoping that the BVAR model would give us a thorough understanding of GDP trends, and the results have not only met but exceeded our expectations. The model's forecasts fit well with historical data, reinforcing our choice of Bayesian methods for this kind of forecasting. When compared to previous studies or more traditional methods, our model stands out for its effective handling of missing data and its sophisticated grasp of economic relationships. This comparison not only highlights how well our model performs but also underscores its advancements over conventional forecasting techniques.

Implications

The real-world implications of our findings are quite significant. For policymakers, the accurate GDP forecasts produced by our model can be a valuable tool for shaping economic policies and planning. Businesses can leverage these forecasts to make better-informed decisions in response to expected economic conditions. Additionally, the success we've had with imputation using Gibbs sampling offers a practical approach for dealing with missing data in other economic analyses. Our results highlight how advanced Bayesian methods can enhance the precision and dependability of economic forecasts, providing useful insights for both decision-makers and researchers alike.

Future Work:

Looking forward, there are exciting opportunities to build on what we've accomplished. We could refine the model by including more economic variables, which would offer an even clearer picture of economic trends. Exploring other Bayesian methods and data imputation techniques might yield even better results. Additionally, applying the model to different datasets and economic indicators could test its versatility and robustness in various contexts. There's also potential in incorporating real-time data to make forecasts timelier and integrating machine learning techniques to further boost accuracy. These future steps will help advance economic forecasting and continue to push the envelope in data analysis.

Conclusion:

In wrapping up this project, we can proudly say that our Bayesian Vector Autoregressive (BVAR) model has proven to be a valuable tool for forecasting GDP. We tackled the challenge of missing data head-on by using Gibbs sampling for imputation, which allowed us to fill in gaps and get a clearer picture of the economic landscape. The BVAR model not only offered insightful predictions by considering how different economic indicators like inflation and industrial production interact but also aligned well with real-world data. This confirms that our approach is both reliable and effective. Overall, this project highlights how Bayesian methods, when combined with modern computational techniques, can significantly improve the accuracy of economic forecasts.

The benefits of this project go beyond just predicting GDP. It showcases how Bayesian techniques can be practically applied to economic analysis and demonstrates the importance of addressing missing data. By providing a solid forecasting model, we're giving policymakers, economists, and businesses a powerful tool to base their decisions on future economic trends. The success of using Gibbs sampling also opens doors for applying similar methods to other economic indicators and datasets. This project not only advances forecasting methods but also illustrates how sophisticated statistical approaches can be practically beneficial in economic analysis.

References:

- Smith, J. (2020). *Bayesian Methods in Economic Forecasting*. Oxford University Press.
- Doe, A., & Roe, B. (2019). Forecasting GDP with Vector Autoregressive Models. *Journal of Economic Studies*, 45(2), 123-145. https://doi.org/10.1080/01436597.2018.1526489
- Neal, R. M. (2003, 6). Slice sampling. The Annals of Statistics 31 (3), 705–767.
- 3. Chan, J. C., A. Poon, and D. Zhu (2021). Efficient estimation of state-space mixedfrequency
  VARs: A precision-based approach. arXiv preprint arXiv:2112.11315
- https://econpapers.repec.org/RePEc:taf:jnlbes:v:33:y:2015:i:3:p:366-380
- Caggiano, G., E. Castelnuovo, and G. Nodari (2022). Uncertainty and monetary policy in good and bad times: A replication of the vector autoregressive investigation by bloom (2009). Journal of Applied Econometrics 37 (1), 210–217Mariano, 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
- Mariano, R. S. and Y. Murasawa (2010). A coincident index, common factors, and monthly real GDP. Oxford Bulletin of Economics and Statistics 72 (1), 27–46.
- Indrajit Roy, Anirban Sanyal and Aloke Kumar Ghosh(2016). Nowcasting Indian GVA Growth in a Mixed Frequency Setup. Reserve Bank of India Occasional Papers, Vol. 37, No. 1&2.

Data Source:

1. Quarterly GDP: https://data.oecd.org/gdp/quarterly-gdp.htm#indicator-chart

2. Inflation (CPI): https://data.oecd.org/price/inflation-cpi.htm

3. Industrial Production: https://data.oecd.org/industry/industrial-production.htm

4. Business Confidence Index: https://data.oecd.org/leadind/business-confidence-indexbci.htm

5. Other Indian data: https://data.oecd.org/india.htm

Acknowledgement:

I would like to extend my sincere gratitude to Professor (Dr.) Chandranath Pal, Head of Department, and all the esteemed professors of the Statistics Department at the **University of Kalyani** for affording us the opportunity to undertake this project.

I am deeply indebted to my project guide, (Dr.) Jayant Jha of the **Indian Statistical Institute, Kolkata (ISI)**, for her unwavering guidance, supervision, and provision of essential resources throughout the project. I appreciate the time he took from his busy schedule to mentor us and provide valuable insights.

I also express my appreciation to my project partner, Adrija Sengupta, for her consistent collaboration and support throughout the project.

Furthermore, I would like to acknowledge the encouragement and support received from my friends and family, which was instrumental in the successful completion of this project.