Vol. 25, No. 2, pp. 184-194 (2025)
Journal of Agricultural Physics
ISSN 0973-032X
https://doi.org/10.56093/JAP.v25i2.8

Research Article

Decoding Potato Price Dynamics using Bayesian VAR Models

HIMADRI SHEKHAR ROY†[1], CHIRANJIT MAZUMDER†[2], SUNIL KUMAR YADAV[1], MRINMOY RAY[2]* AND PRAKASH KUMAR[1]*

[1]*ICAR-Indian Agricultural Statistics Research Institute, New Delhi-110 012*
[2]*ICAR-Indian Agricultural Research Institute, New Delhi-110 012*

ABSTRACT

Vector autoregression (VAR) is a natural multivariate extension of the univariate autoregressive model and is widely used to model dynamic relationships among multiple time series. However, classical VAR models often overfit when higher lag lengths are used with limited sample sizes, leading to poor forecasting performance. Bayesian vector autoregression (BVAR) addresses this issue by treating model parameters as random variables and incorporating prior information to achieve shrinkage and a more stable estimation. In this study, a BVAR framework is applied to multivariate time-series data using alternative lag lengths and prior specifications. Model selection is carried out using information criteria and Bayesian measures such as AIC, BIC, marginal log-likelihood, and marginal log-posterior likelihood. While AIC and BIC tend to favour lower lag orders, Bayesian criteria tend to favour higher lag orders. The forecasting performance was evaluated using out-of-sample data and standard accuracy measures, including the RMSE, MAE, and MAPE. The empirical results show that increasing the lag length from one to 12 consistently improves the forecasting accuracy of the BVAR models. For example, in the case of Delhi, the RMSE decreases from 363.69 at lag 1 to 361.10 at lag 12, and the MAPE declines from 29.86% to 28.00%. Similar improvements were observed for Shimla and Trivandrum. In contrast, the classical VAR model exhibits declining performance with higher lags, with Delhi's MAPE increasing from 31.57% to 37.54%. Furthermore, a prior sensitivity analysis indicated that the Normal–Wishart prior outperformed the flat-flat prior, yielding lower RMSE (393.64 vs. 427.28) and MAE (306.17 vs. 364.67). Overall, this study demonstrates the superiority of BVAR over VAR for multivariate time-series forecasting.

Key words: Autoregressive (AR), Bayesian VAR, Overfitting, Prior, Vector autoregressive (VAR)

Introduction

Multivariate time series models play a central role in macroeconomic and financial analysis, and since the seminal contribution of Sims, Vector Autoregressive (VAR) models have become one of the most widely used frameworks for studying dynamic relationships among multiple variables (Sims, 1980; Sims *et al*., 1986). The main appeal of VAR models lies in their flexibility, as they treat all variables as endogenous and avoid relying on strong a priori theoretical restrictions, thereby allowing the data to reveal the underlying dynamic structure (Sims, 1980; Kadiyala and Karlsson, 1997; Karlsson, 2012). Recent research confirms that VARs remain a workhorse for practical forecasting in policy institutions and applied macroeconomics (Ahelegbey *et al*., 2014).

Despite their popularity, classical VAR models face important limitations in empirical applications.

*Corresponding author, Email: mrinmoy4848@gmail.com; prakash289111@gmail.com
†equally contributed

First, the assumption of time-invariant coefficients is often unrealistic in environments characterized by structural breaks, policy regime shifts, and evolving economic relationships, which may lead to parameter instability and degraded forecasting performance (Sims, 1980; Sims *et al.*, 1986). Second, VAR models are highly parameter-intensive: the number of coefficients increases rapidly with the number of variables and lag length, and this "curse of dimensionality" becomes particularly severe when modeling systems with several variables and higher-order lags or allowing for time-varying error covariance structures (Kadiyala and Karlsson, 1997; Karlsson, 2012). In finite samples, such over-parameterization typically results in imprecise estimates, overfitting, and poor out-of-sample forecasts (Litterman, 1980, 1986; Carriero, Clark, and Marcellino, 2011). These problems become even more pronounced in high-dimensional settings, motivating the use of structured shrinkage priors and regularization techniques.

To address these challenges, shrinkage and regularization techniques have been proposed as effective ways to reduce model complexity while retaining relevant dynamics. Shrinkage methods impose restrictions on parameters or pull them toward zero, thereby stabilizing estimation and improving forecast accuracy in high-dimensional settings (Litterman, 1980, 1986; Carriero, Clark, and Marcellino, 2011; Huber and Feldkircher, 2019). Bayesian methods provide a natural and coherent framework for implementing shrinkage via prior distributions on the VAR coefficients and covariance matrix, and their use in VAR modeling has increased substantially in recent years (Koop and Korobilis, 2009; Karlsson, 2012; Gruber, 2025). Recent contributions show that adaptive hierarchical and local–global shrinkage priors, such as normal–gamma and horseshoe variants, can yield substantial gains in both estimation and forecasting in large VARs (Huber and Feldkircher, 2019; Giannone *et al*., 2015; Medeiros *et al*., 2021).

Within this Bayesian framework, several prior specifications have been developed to mitigate overparameterization and improve forecasting performance. The Minnesota prior of Litterman (1980, 1986) shrinks own-lag coefficients toward a random-walk process and other-lag coefficients toward zero, while diffuse and conjugate priors such as the Normal–Wishart and Normal–diffuse priors offer alternative trade-offs between flexibility and shrinkage (Zellner, 1971; Litterman, 1986; Karlsson, 2012). More recent work extends Minnesota-type priors to large systems via adaptive hierarchical formulations that selectively shrink small coefficients while preserving important signals (Huber and Feldkircher, 2019; Chan, 2021; Gruber, 2025). Empirical evidence indicates that Bayesian Vector Autoregressive (BVAR) models equipped with such priors can deliver more accurate and robust forecasts than classical VARs, particularly in small samples and high-dimensional applications (Koop and Korobilis, 2009; Carriero, Clark, and Marcellino, 2011; McCracken *et al*., 2021; Gruber, 2025). Kuschnig and Vashold (2021) introducd the BVAR framework for estimating Bayesian Vector Autoregression models in R, emphasizing hierarchical prior selection to address over-parameterization in high-dimensional VARs. Uddin (2023) studied Bayesian Vector Autoregression (Bayesian VAR) framework to explore how shocks to renewable energy interact with key macroeconomic and financial variables, including green finance, private energy investment, GDP growth, and foreign direct investment. Hartwig (2024) examined how Bayesian vector autoregression (BVAR) models can be adjusted to better handle the unprecedented economic data during the COVID-19 pandemic. van der Drift *et al*. (2024) investigates how forecasting accuracy for house prices can be improved by incorporating credit conditions and Bayesian econometric techniques. It compares the out-of-sample predictive performance of three Bayesian models—a Bayesian vector autoregression in levels (BVAR-l), in differences (BVAR-d), and a Bayesian vector error correction model (BVECM)—against their non-Bayesian counterparts.

Motivated by these considerations, the present study investigates the forecasting performance of BVAR models relative to classical VAR models using multivariate monthly potato price series for three major markets—Delhi, Shimla, and Trivandrum—from January 2002 to December 2013 (National Horticulture Board, 2013). Alternative lag lengths

and prior specifications, including the Normal–Wishart and flat-flat priors, are considered. Models are evaluated using standard information criteria and Bayesian measures, as well as out-of-sample forecast accuracy indicators such as RMSE, MAE, and MAPE (Koop and Korobilis, 2009; Karlsson, 2012; McCracken *et al*., 2021). By comparing the forecasting performance of BVAR and classical VAR models across different lag structures and prior choices, this study aims to provide practical guidance on the use of Bayesian shrinkage in multivariate time series forecasting for agricultural price analysis, complementing recent applications of BVARs to commodity and agricultural markets (Bondarenko, 2023).

Materials and Methods

Bayesian Inference

The main objective of Bayesian forecasting is based on the (posterior) predictive distribution, as described by Karlson (2012). Let the distribution of future data points, $y_{T+1:T+H}=(y'_{T+1},...,y'_{T+H})'$ based on the currently observed data $Y_T=\{y_t\}_{t=1}^{T}$ is given as $P(y_{T+1:T+H}|Y_T)$. Predictive distribution catches all the relevant information about the unknown future events by itself. So it depend upon the researcher of the forecast which features of the predictive distribution are relevant for the situation. So for example mean, median and mode of the predictive distribution together with a probability interval indicating the range of likely outcomes.

So, it is a decision problem which requires the specification of problem dependent loss function $L(a, y_{T+1:T+H})$, where a is the action taken, the vector of real numbers to report as the forecast, and $y_{T+1:T+H}$ represents the unknown future state of nature. The Bayesian decision is to choose action (forecast) that minimizes the expected loss conditional on the available information Y_T,

$$E[L(a, y_{T+1:T+H} \mid Y_T)] = \int L(a, y_{T+1:T+H}) p(y_{T+1:T+H} \mid Y_T) dy_{T+1:T+H} \quad (1)$$

For a given Loss function and predictive distribution, $L(a, y_{T+1:T+H} \mid Y_T)$ the solution to be the minimization problem in the function of the data, $a(Y_T)$. Now it need to specify the form of the predictive distribution, this requires the specification of three different distributions that completes the description of problem, the distribution of future observation conditional on unknown parameter values, β, and the observation the distribution of the observed data, $p(y_{T+1:T+H} \mid Y_T, \beta)$, the distribution of the observed data, that is, the model or likelihood - conditional on the parameters, $L(Y_T \mid \beta)$, and the prior distribution, $\pi(\beta)$, representing prior notions about likely or "reasonable" values of the unknown parameters, β. In a time series and forecasting context, the likelihood usually takes the form

$$L(Y_T \mid \beta) = \prod_{t=1}^{T} f(y_t \mid Y_{t-1}, \beta) \quad (2)$$

with the history in $Y_1, Y_2,...$ suitably extended to include initial observations that the likelihood is conditional on. The distribution of future observations is of the same form,

$$f(y_{T+1:T+H} \mid Y_T, \beta) = \prod_{t=T+1}^{T+H} f(y_t \mid Y_{t-1}, \beta) \quad (3)$$

With these in hand straightforward application of Bayes Rule yields the predictive distribution as

$$p(y_{T+1:T+H} \mid Y_T) = \frac{p(y_{T+1:T+H}, Y_T)}{m(Y_T)} = \frac{\int f(y_{T+1:T+H} \mid Y_T, \beta) L(Y_T \mid \beta) \pi(\beta) d\beta}{\int L(Y_T \mid \beta) \pi(\beta) d\beta} \quad (4)$$

In practice an intermediate step through the posterior distribution of the parameters,

$$p(\beta \mid Y_T) = \frac{L(Y_T \mid \beta)\pi(\beta)}{\int L(Y_T \mid \beta)\pi(\beta) d\beta} \propto L(Y_T \mid \beta)\pi(\beta) \quad (5)$$

is used with the predictive distribution given by

$$p(y_{T+1:T+H} \mid Y_T) = \int f(y_{T+1:T+H} \mid Y_T, \beta) p(\beta \mid Y_T) d\beta \quad (6)$$

So, by the latter form of the predictive distribution makes it clear how Bayesian forecasts accounts for both the inherent uncertainty about the future embodied by $f(y_{T+1:T+H} \mid Y_T, \beta)$ and the uncertainty about the true parameter values described by the posterior distribution $p(\beta \mid Y_T)$.

While the posterior distribution of the parameters may be available in closed form in special cases when

conjugate prior distributions are used closed form expressions for the predictive distribution are generally unavailable when lead times greater than 1 are considered. This makes the equation no. 6 of the predictive distribution especially attractive. Marginalizing out the parameters of the joint distribution of $y_{T+1:T+H}$ and β analytically may be difficult or impossible, on the other hand it also suggests a straightforward simulation scheme for the marginalization. Supposing that random numbers can be generated from the posterior $p(\beta | Y_T)$, for each draw of β generate a sequence of draws of $y_{T+1}...y_{T+H}$ by repeatedly drawing from $f(y_t|Y_{t-1}, \beta)$ and adding the draw of y_t to the conditioning set for the distribution of y_{t+1}. This gives a draw from the joint distribution of $(\beta, y_{T+1}... y_{T+H})$ conditional on Y_T and marginalization is achieved by simply discarding the draw of β. Repeating this R times gives a sample from the predictive distribution that can be used to estimate $E(y_{T+1:T+H}|Y_T)$ or any other function or feature a (Y_T) of the predictive distribution of interest.

The denominator in 4 and 5,

$$m(Y_T) = \int L(Y_T|\beta)\pi(\beta)d\beta \quad (7)$$

is known as marginal likelihood or or prior predictive distribution and plays a crucial role in Bayesian hypothesis testing and model selection. Consider two alternative models, M_1 and M_2, with corresponding likelihoods $L(Y_T|\beta_1,M_1)$, $L(Y_T|\beta_2, M_2)$ and priors $\pi(\beta_1|M_1)$, $\pi(\beta_2|M_2)$ Supposing that one of M_1 and M_2 is the true model but it is not certain which of the competing hypothesis or theories embodied in the models is the correct which can be assigned prior probabilities, $\pi(M_1)$ and $\pi(M_2) = 1-\pi(M_1)$, that each of the models is the correct one. With these in hand Bayes Rule yields the posterior probabilities that the models are correct as

$$p(M_i) = \frac{m(Y_T | M_i)\pi(M_i)}{m(Y_T | M_1)\pi(M_1) + m(Y_T | M_2)\pi(M_2)} \quad (8)$$

with $m(Y_T | M_i) = \int L(Y_T | \beta_i, M_i)\pi(\beta_i | M_i)d\beta_i$. The posterior odds for model (4) against model (5) is given by

$$\frac{p(M_1)}{p(M_2)} = \frac{m(Y_T | M_1)\pi(M_1)}{m(Y_T | M_2)\pi(M_2)} = \frac{m(Y_T | M_1)}{m(Y_T | M_2)} \times \frac{\pi(M_1)}{\pi(M_2)} \quad (9)$$

the Bayes factor $BF_{1,2} = m(Y_T|M_1)=m(Y_T|M_2)$ comparing M_1 to M_2 times the prior odds. Model choice can be based on the posterior odds but it is also common to use the Bayes factors directly, implying equal prior probabilities. The Bayes factor captures the data evidence and can be interpreted as measuring how much opinion about the models have changed after observing the data. The choice of model should of course take account of the losses associated with making the wrong choice. Alternatively, conditioning can be avoided on one single model being the correct one by averaging over the models with the posterior model probabilities as weight. That is, instead of basing forecasts on the predictive distribution $p(Y_{T+1:T+H}|Y_T, M1)$ and conditioning on being the correct model Bayesian Model Averaging (BMA) is conducted to obtain the marginalized (with respect to the models) predictive distribution

$$p(y_{T+1:T+H} | y_T) = p(y_{T+1:T+H} | Y_T, M_1)\pi(M_1) + p(y_{T+1:T+H}|Y_T, M_2)\pi(M_2) \quad (10)$$

which accounts for both model and parameter uncertainty.

The calculations involved in equation no. 8 are non-trivial and the integral $\int L(Y_T | \beta_i, M_i)\pi(\beta_i | M_i)d\beta_i$ is only well defined if the prior is proper. That is, if $\int \pi(\beta_i | M_i)d\beta_i = 1$. For improper priors, such as a uniform prior on the whole real line, the integral is not convergent and the scale is arbitrary. For the uniform prior it can be written as $\pi(\beta_i | M_i) = k_i$ and it follows that $m(Y_T) \propto k_i$ and the Bayes factors and posterior probabilities are arbitrary. There is, however, one circumstance where improper prior can be used. This is when there are parameters that are common to all models, for example an error variance. Then partitioning $\theta = (\theta_i, \sigma^2)$ and using proper priors for θ_i and an improper prior, such as $\pi(\sigma^2) \propto 1/\sigma^2$ for the variance since the common scale factor cancels in the calculation of posterior model probabilities and Bayes factors.

VAR Models

Consider a typical VAR model

$$Y_t = B_1 Y_{t-1} + B_2 Y_{t-2} + ... + B_p Y_{t-p} + Dz_t + \varepsilon_t \quad t=1,...,T \quad (11)$$

where Y_t is a nx 1 vector of endogenous variables and ε_t is a nx 1 vector of error terms independently, identically and normally distributed with variance-covariance matrix $\sum, \epsilon_t \sim iidN(0, \Sigma)$, $B_t(\iota = 1, ...p)$ and D are nxn *and* nxd matrices of parameters respectively, and z_t is a d x 1 vector of exogenous variables.

Now defining $x_t = \left(y'_{t-1} ... y'_{t-p} z'_t \right)'$ as a vector containing p lags of y_t, the VAR model can be written as:

$$y_t = \beta x_t + \varepsilon_t \qquad (12)$$

Classical estimation of models like equation no. 11 may yield imprecisely estimated relations that fit the data well only because of the large number of variables included, a problem known as 'overfitting' in the literature. In fact, the number of parameters to be estimated, $n\ (np + d)$, grows geometrically with the number of variables (n) and proportionally with the number of lags (p). When the number of variables on the right hand side of (1.1) is relatively high and the sample information is relatively loose, it is likely that the estimates are influenced by noise. When this is the case, it is recommendable to estimate models like equation no. 11 by imposing some restrictions to reduce the dimension of the parameter space. Therefore, the problem is to find restrictions that are as credible (Sims, 1980) as possible.

A Bayesian approach of VAR was originally advocated by Litterman (1980) as a solution to the 'overfitting' problem. The solution he proposed is to avoid overfitting without necessarily imposing exact zero restrictions on the coefficients. The researcher cannot be sure that some coefficients are zero and should not ignore their possible range of variation. A Bayesian perspective fits precisely this view. Without putting too much weight on certain values, one may think of this uncertainty over the exact value of the model's parameters as a probability distribution for the parameter vector. The degree of uncertainty represented by this distribution can then be altered by the information contained in the data if the two sources of information are different. More specifically, Litterman (1986) specified a new technique which can be expressed formally by introducing a vector of parameters (called hyperparameters), say $\prod$ say (π_1,..., π_H). For instance, π_1 may control the value of the mean of the first own lag coefficient, π_2 controls the variance of the lags of one variable, π_3 controls the variance of the lags of another variable, π_4 controls the speed of decreasing of the variance as the number of lags increase, π_5 controls the variance of the deterministic/exogenous part, and π_6 controls the overall degree of prior uncertainty. Therefore, the original problem of estimating $n\ (np + d)$ parameters can be converted in a problem of estimating just six "hyper-parameters". Additional data characteristics can be accommodated by using different hyperparameter to control the other relevant information.

Forecasting and Estimation by Using Bayesian VAR

To describe the Bayesian estimation principle, the equation (3.1) can be written as

$$Y_t = X_t \beta + \varepsilon_t \qquad t = 1, ..., n \qquad (13)$$

Where $X_t = \left(I_n \otimes W_{t-1} \right)$ is $n \times nk$, $W_{t-1} = \left(Y'_{t-1}, ..., Y_{t-p}, z'_t \right)$ is $k \times 1$ and $\beta = vec(B_1, B_2, ..., B_p, D)$ is $nk \times 1$. The unknown parameters of the model are β and Σ.

The Bayesian estimation of the equation (4.1) can be worked out as follows.

The likelihood function of (β, Σ) based on Y is then

$$= L(\beta, \Sigma) = \frac{1}{|\Sigma|^{\frac{T}{2}} \exp\left\{ -\frac{1}{2} \Sigma_{t=1}^{T} \ (Y_t - X_t\beta)'(Y_t - X_t\beta) \right\}} \qquad (14)$$

$$= \frac{1}{|\Sigma|^{\frac{T}{2}} \exp\left\{ -\frac{1}{2} \text{trace}(Y_t - X_t\beta)'(Y_t - X_t\beta) \right\}}$$

The maximum Likelihood estimators (MLE) of β and Σ are

$$\hat{\beta}_M = (X'_t X_{\partial t})^{-1} X'_t Y_t \text{ and } \hat{\Sigma}_M = S(\beta_M) / T \qquad (15)$$

respectively, when

$$S(\beta_M) = (Y_t - X_t\beta)'(Y_t - X_t\beta)$$

A joint distribution on the parameters β and Σ are p(β, Σ), the joint posterior distribution of the parameters conditional on the data is obtained through Bayes rule

$$p(\beta, \Sigma \mid Y_t) = \frac{p(\beta, \Sigma) L(Y_t \mid \beta, \Sigma)}{p(Y)} \propto p(\beta, \Sigma) L(Y_t \mid \beta, \Sigma) \qquad (16)$$

By definition of conditional probability, the joint distribution of data and parameters p(β, Σ, Y_t) can be written as

$$p(\beta,\Sigma,Y_t) = L(Y_t \mid \beta,\Sigma)\, p(\beta,\Sigma) \\ = p(\beta,\Sigma \mid Y_t)\, p(Y_t) \quad (17)$$

where ∝ denotes "proportional to". Given p(β, Σ| Y_t), the marginal posterior distribution conditional on data, $p(\Sigma \mid Y_t)$ and $p(\beta \mid Y_t)$ can be obtained by integrating out β and Σ from $p(\beta, \Sigma \mid Y_t)$, respectively. Finally the location and dispersion of $p(\Sigma \mid Y_t)$ and $p(\beta \mid Y_t)$ can be analyzed to yield point estimates of the parameter of interest and measure of precision, comparable to those obtained by using a classical approach to estimation.

Here N-IW prior is used:

$\beta \mid \Sigma \sim N(\beta_0, \Sigma \otimes \Omega_0)$ where Σ~IW (s_0, v_0)

As N-IW prior is conjugate, the conditional posterior distribution of the model is also N-IW (Zellner, 1973).

$$\beta \mid \Sigma, Y_t \sim N(\beta, \Sigma \otimes \bar{\Omega}), \Sigma \mid Y_t \sim IW(\bar{S}, \bar{v}) \quad (18)$$

Defining β and ε as the OLS estimates and $\bar{\beta} = (\Omega_0^{-1} + X_t'X_t)^{-1} (\Omega_0^{-1}\beta_0 + X_t'Y_t)$, $\bar{\Omega} = (\Omega_0^{-1} + X'X)^{-1}$, $\bar{v} = v_0 + T$ and $\bar{S} = \hat{\beta}'X'X\hat{\beta} + \beta_0'\Omega_0^{-1}\beta_0 + \beta_0 + \hat{\varepsilon}'\hat{\varepsilon} - \hat{\beta}\Omega^{-1}\hat{\beta}$

The one-step ahead forecasts are obtained by using the posterior mean β.

$$\hat{Y}_{t+1} = \bar{A} + \bar{B}_1 Y_t + \bar{B}_2 Y_{t-1} + ... + \bar{B}_p Y_{t-p+1} + \varepsilon_t \qquad t=1,...,n$$

The h-step ahead forecasts are obtained by iteration:

$$\hat{Y}_{t+h} = \bar{B}_1 \hat{Y}_{t+h-1} + \bar{B}_2 \hat{Y}_{t+h-2} + ... + \bar{B}_p \hat{Y}_{t+h-p} + \bar{D}\hat{z}_{t+h} + \varepsilon_t \qquad t=1,...,T$$

where for h ≤ p, alternatively by using the notation (3), one can write

$$\hat{y}_{t+h} = \bar{\beta}^h x_t$$

Here the prior expectation and variance of the coefficient matrices may be expressed as

$$E(\beta_k^{(ij)}) = \begin{cases} \beta^* & \text{if i=j, k=1} \\ 0 & \text{otherwise} \end{cases}, \quad Var[\beta_k^{(ij)}] = \frac{\theta}{k^2}\frac{\sigma_i^2}{\sigma_j^2}, k = 1,..., p, \quad (19)$$

where $\beta_k^{(ij)}$ denotes the element in position (i, j) in the matrix β_k, and where the covariances among the coefficients in are zero. The prior mean is typically set to 1 in the traditional Minnesota prior to account for the persistence of the data, but if the VAR is estimated in the first difference β should be set to zero. The shrinkage parameter θ measures the tightness of the prior: when θ↔0 the prior is imposed exactly and the data do not influence the estimates, while as θ↔∞ the prior becomes loose and the prior information does not influence the estimates (Carriero and Clark, 2011) and it will approach the standard OLS estimates.

Priors

A crucial issue in Bayesian VAR estimation is the choice of prior distributions for the model parameters. Since VAR models involve a large number of parameters, appropriate priors are required to control over-parameterization and improve estimation efficiency. In a fully Bayesian framework, prior distributions must also be specified for the hyperparameters, but integrating them out is often computationally difficult. Hence, practical implementations typically rely on either Empirical Bayesian estimation, where hyperparameters are fixed using preliminary estimates, or Hierarchical Bayesian methods, where hyperparameters are integrated out numerically. This study focuses on commonly used priors that balance tractability and forecasting performance.

Minnesota Prior

The Minnesota prior (Litterman, 1986) is one of the most widely used priors in the BVAR literature. It assumes prior independence across equations and a fixed, diagonal residual variance–covariance matrix. Coefficients are shrunk toward a random walk for own lags and toward zero for other variables and higher-order lags. Under this prior, each equation can be estimated separately, and in the case of a diffuse prior variance, the posterior mean coincides with the OLS estimator.

Diffuse Prior

The diffuse (non-informative) prior relaxes the restrictive assumptions of the Minnesota prior by allowing for a non-diagonal residual variance–covariance matrix. This prior leads to a joint posterior distribution in which the coefficient matrix follows

a multivariate Student-t distribution after integrating out the error variance. Although flexible, diffuse priors may suffer from weaker shrinkage, particularly in small samples.

Conjugate Priors

To achieve analytical convenience and flexibility, conjugate priors are commonly used. The Normal–Wishart prior allows joint modeling of coefficients and the residual variance–covariance matrix, resulting in closed-form posterior distributions. It provides effective shrinkage and improved forecasting performance. Alternatively, the Normal-Diffuse prior assumes independence between coefficients and the variance–covariance matrix, offering greater flexibility but weaker regularization. In practice, the Normal–Wishart prior is preferred due to its stability and superior forecasting ability.

Results and Discussion

The data set which is used is the monthly price of fresh potato of three different cities (Delhi, Shimla and Trivandrum) from January, 2002 to December, 2013 (Data source: National Horticulture Board-http://www.nhb.gov.in). Here, Delhi, Shimla and Trivandrum are taken as variables of the forecasting exercise. Seasonality and trend detection test is done and here it is seen that seasonality is present. Hence, seasonality decomposition is done before the forecasting exercise. The monthly price data of 3 different cities (Delhi, Shimla, Trivandrum) are plotted in Fig. 1.

The 3 different variables are used with different lags. So to choose the appropriate model, it is needed to look on the model selection criteria. Also the coefficient matrix of VAR(1) and BVAR(1) is given in Table 1. The list of model selection criteria for BVAR for different lag with their corresponding values are given in Table 2.

Forecasting Performance of BVAR(p)

Table 3 presents the out-of-sample forecasting performance of the Bayesian Vector Autoregression (BVAR) model for different lag orders (p = 1 to 12)

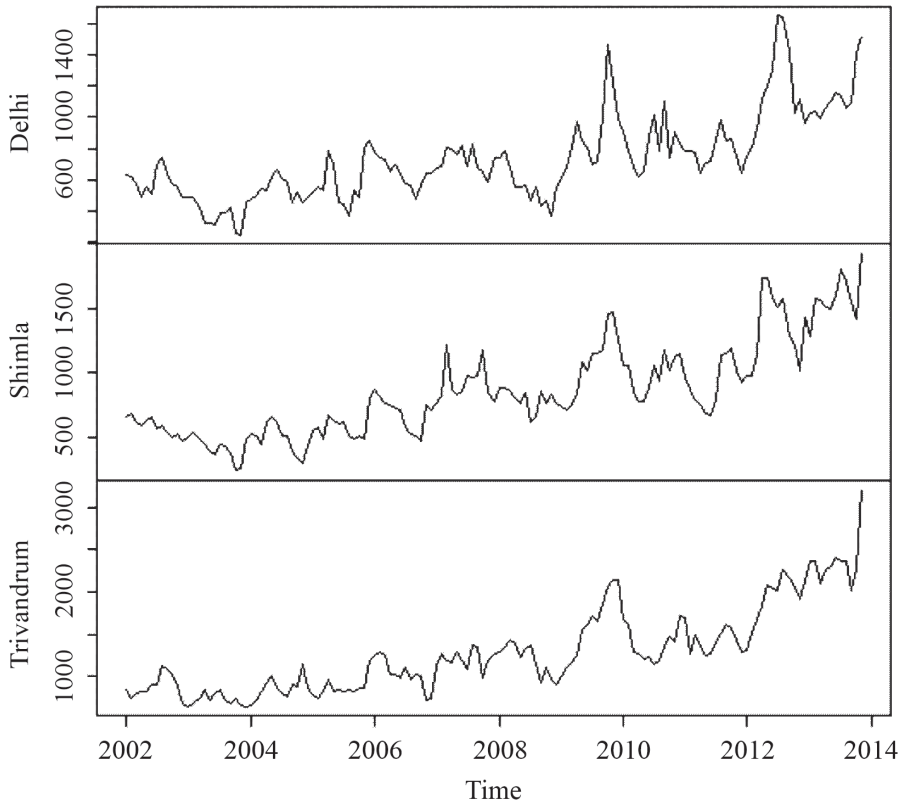

Fig. 1. Monthly price data of three different cities

Table 1. Matrix of coefficients of VAR(1) and BVAR(1)

VAR Model				BVAR Model			
Constant	Lag-1 Coefficient Matrix			Constant	Lag-1 Coefficient Matrix		
66.855	0.641	0.140	0.062	12.693	0.917	0.019	0.024
9.241	0.158	0.652	0.137	7.958	0.027	0.919	0.034
56.040	0.322	0.234	0.616	6.484	0.094	0.061	0.905

Table 2. Lag-wise Evaluation of BVAR Model Using AIC, BIC, and Marginal Likelihoods

Lag	AIC	BIC	Marginal log likelihood	Marginal log posterior likelihood
1	28.96003	29.23878	414.64	197.5831
2	28.95784	29.44565	408.9658	195.0372
3	28.93959	29.63646	400.99	191.8184
4	29.01062	29.91656	396.5855	188.5569
5	29.04705	30.16204	369.3965	185.3816
6	29.125	30.44906	358.1024	182.1419
7	29.13811	30.67123	355.6727	179.0123
8	29.15182	30.894	347.1312	175.8475
9	29.16188	31.11312	343.8141	172.7028
10	29.26265	31.42296	332.4804	169.5635
11	29.30068	31.67005	326.4582	166.2839
12	29.31373	31.89216	321.6718	163.2952

across three locations: Delhi, Shimla, and Trivandrum, evaluated using RMSE, MAE, and MAPE. For Delhi, a consistent decline in forecasting errors is observed as the lag length increases. RMSE decreases steadily from 363.69 at lag 1 to 361.10 at lag 12, while MAE and MAPE also show monotonic reductions. This indicates that incorporating longer lag structures allows the BVAR model to better capture temporal dependencies and dynamic interactions among variables. The lowest MAPE (28.00%) at lag 12 suggests superior relative forecasting accuracy at higher lag orders. A similar pattern is evident for Shimla, where RMSE decreases from 396.33 (lag 1) to approximately 395.00 (lag 12). The reduction in MAE and MAPE is more pronounced, with MAPE declining from about 22.00% to 19.57%. This highlights the ability of the BVAR framework to efficiently borrow strength across equations and control over-parameterization, even when lag length increases. For Trivandrum, BVAR performance improves markedly with lag order. RMSE reduces from 418.71 to 414.28, and MAPE drops from 10.39% to 8.78% as p increases from 1 to 12. Compared to Delhi and Shimla, Trivandrum consistently exhibits lower MAPE values, indicating relatively more stable and predictable dynamics.

Overall, the BVAR model demonstrates systematic improvement in forecasting accuracy with higher lag orders, confirming the advantage of Bayesian shrinkage in handling large parameter spaces.

Forecasting Performance of VAR(p)

Table 4 reports the forecasting accuracy of the classical VAR(p) model. In contrast to BVAR, the VAR model shows deteriorating performance as lag length increases. For Delhi, RMSE rises sharply from 374.82 at lag 1 to over 401.99 at lag 12. MAE and MAPE also increase substantially, with MAPE exceeding 37% at higher lags. This indicates severe over-parameterization and loss of predictive efficiency in the classical VAR framework. In the case of Shimla, VAR forecasting accuracy remains relatively stable only at lower lags but worsens

Table 3. Forecasting performance of BVAR(p)

City-wise Accuracy		Lag											
		1	2	3	4	5	6	7	8	9	10	11	12
Delhi	RMSE	363.69	363.33	363.27	363.26	363.26	363.23	363.04	362.87	362.27	362.00	361.85	361.10
	MAE	319.03	318.77	318.70	318.69	318.69	318.66	318.66	318.15	317.79	317.52	317.46	317
	MAPE	29.86	29.86	29.85	29.84	29.84	29.83	29.45	29.05	28.78	28.74	28.07	28.00
Shimla	RMSE	396.33	396.91	396.9	396.74	396.58	396.41	396.16	396	395.82	395.73	395..44	395
	MAE	332.71	333..68	333.43	333.41	333.23	333	332.92	332.65	331.82	331.45	329.63	329.14
	MAPE	22	22.05	22.05	22	22.02	21.97	21.12	20.61	20.57	20.06	19.82	19.57
Trivandrum	RMSE	418.71	418.45	418.41	418.42	418.25	418	417.96	417.10	416.50	415.86	414.88	414.28
	MAE	265.09	265.35	265.32	265.30	265.19	264.58	263	263.08	263.89	262.11	262.37	262.45
	MAPE	10.39	10.40	10.40	10.39	10.35	10.20	10.18	10.01	9.51	9.07	8.93	8.78

Table 4. Forecasting performance of VAR(p)

City-wise Accuracy		Lag											
		1	2	3	4	5	6	7	8	9	10	11	12
Delhi	RMSE	374.82	376.12	377.39	378.95	379.01	381.41	382	383.44	396.93	398.10	401.61	401.99
	MAE	330.67	331.88	333.15	335.08	336.21	337.73	337.07	338.23	339.39	340.34	342.78	342.93
	MAPE	31.57	31.76	32.91	33.16	33.59	33.87	33.98	34.02	35.20	36.28	36.87	37.54
Shimla	RMSE	394.92	395.11	395.93	396.17	396.28	396.5	396	396.28	383.11	395.73	395..44	395
	MAE	338.82	333..68	333.43	333.41	333.23	333	351.64	332.65	334.82	331.45	329.63	329.14
	MAPE	22.16	22.89	23.15	23.42	23.52	23.78	23.95	24	24.19	24.07	24.16	24.28
Trivandrum	RMSE	420.30	419.45	418.41	418.42	418.25	418	417	415.71	416.37	417.18	419.87	421.28
	MAE	288.40	290.67	292.36	293.59	294	295.47	297	298.17	319.21	320.01	321.48	322.71
	MAPE	10.63	10.67	10.71	10.91	10.95	10.97	11.28	11.86	12.84	12.95	13.04	13.78

Table 5. Comparison of Normal-Wishart prior and Flat-flat prior under BVAR(12)

	Normal-Wishart prior	Flat-flat prior
RMSE	393.64	427.28
MAE	306.17	364.67

beyond lag 6, as reflected by increasing RMSE and MAPE values. Unlike BVAR, VAR fails to maintain stable performance when additional lags are introduced. For Trivandrum, the VAR model exhibits increasing MAE and MAPE beyond lag 7, with MAPE rising from 10.63% at lag 1 to 13.78% at lag 12. This again confirms the limited ability of classical VAR to handle higher-dimensional parameter spaces without regularization. These results clearly indicate that VAR models suffer from overfitting at higher lag orders, leading to inferior out-of-sample forecasting performance.

Effect of Prior Specification in BVAR

Table 5 compares the forecasting accuracy of BVAR(12) under two different prior structures: Normal–Wishart prior and Flat–flat prior. The results show that the Normal–Wishart prior substantially outperforms the Flat–flat prior, with RMSE reduced from 427.28 to 393.64 and MAE reduced from 364.67 to 306.17. This improvement highlights the importance of informative priors in Bayesian time-series modeling. The Normal–Wishart prior effectively shrinks coefficients toward plausible values, thereby reducing estimation variance and improving forecast precision.

Conclusion

This study used several accuracy measures to assess the forecasting performance of traditional Vector Autoregression (VAR) and Bayesian Vector Autoregression (BVAR) models across different lag orders. The results clearly show that the BVAR framework outperforms the conventional VAR technique for multivariate time series forecasting. Across all three locations—Delhi, Shimla, and Trivandrum—the BVAR model consistently improved forecasting accuracy as the lag order grew. In contrast, the VAR model's performance deteriorated significantly at greater delays, indicating a vulnerability to over-parameterization and estimating instability. BVAR successfully managed model complexity by using Bayesian shrinkage, which resulted in lower RMSE, MAE, and MAPE values across all lag structures. The durability of Bayesian estimation in high-dimensional situations was further supported by the comparative study, which showed that the performance gap between BVAR and VAR grew at increasing lag orders. Delhi and Shimla profited greatly from the regularization effects of the Bayesian framework, while Trivandrum regularly recorded smaller forecast errors, indicating considerably more stable underlying dynamics. Furthermore, the prior sensitivity analysis verified that BVAR performance is significantly impacted by previous selection. Under BVAR(12), the Normal–Wishart prior performed noticeably better than the Flat–flat prior, highlighting the significance of informative prior architectures in enhancing forecast accuracy and lowering estimation variance.

Overall, the results provide compelling evidence for using BVAR models with suitable prior specification for multivariate forecasting issues, especially when greater lag orders are needed. The findings imply that Bayesian VAR offers a dependable and effective substitute for traditional VAR models, with increased model stability and forecast accuracy. In order to improve forecasting performance, future research may expand this framework by investigating different prior formulations, time-varying parameter BVAR models, or hybrid Bayesian–machine learning techniques.

References

Ahelegbey, D., Billio, M. and Casarin, R. 2014. Sparse graphical vector autoregression: A Bayesian approach. *Annals of Economics and Statistics,* **123/124**: 333–361.

Bondarenko, O. 2023. *Agricultural commodity price dynamics: Evidence from BVAR models* (NBU Working Paper No. 03/2023). National Bank of Ukraine.

Carriero, A., Clark, T.E. and Marcellino, M. 2011. *Bayesian VARs: Specification choices and forecast accuracy*. Federal Reserve Bank of Cleveland.

Chan, J.C.C. 2021. Minnesota-type adaptive hierarchical priors for large Bayesian VARs.

International Journal of Forecasting **37**(4): 1212–1226.

Chan, J.C., Pettenuzzo, D., Poon, A. and Zhu, D. 2023. *Conditional forecasts in large bayesian VARs with multiple soft and hard constraints*. Available at SSRN, 4358152.

Hartwig, B. 2024. Bayesian VARs and prior calibration in times of COVID-19. *Studies in Nonlinear Dynamics & Econometrics* **28**(1): 1-24.

Huber, F. and Feldkircher, M. 2019. Adaptive shrinkage in Bayesian vector autoregressive models. *Journal of Business and Economic Statistics,* **37**(1): 27–39.

Kadiyala, K.R. and Karlsson, S. 1997. Numerical methods for estimation and inference in Bayesian VAR models. *Journal of Applied Econometrics,* **12**(2): 99–132.

Karlsson, S. 2012. *Forecasting with Bayesian vector autoregressions*. Örebro University Business School.

Koop, G. and Korobilis, D. 2009. Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and Trends in Econometrics* **3**(4): 267–358.

Kuschnig, N. and Vashold, L. 2021. BVAR: Bayesian vector autoregressions with hierarchical prior selection in R. *Journal of Statistical Software*, **100**: 1-27.

Litterman, R. 1980. *Techniques for forecasting with vector autoregressions* (Doctoral dissertation). University of Minnesota, Minneapolis.

Litterman, R. 1986. Forecasting with Bayesian vector autoregressions: Five years of experience. *Journal of Business and Economic Statistics* **4**(1): 25–38.

McCracken, M.W., Owyang, M.T. and Sekhposyan, T. 2021. Real-time forecasting and scenario analysis using a large mixed-frequency Bayesian VAR. *International Journal of Central Banking* **17**(2): 285–334.

National Horticulture Board. (n.d.). *Monthly price and arrival report*. Government of India. https://nhb.gov.in.

Sims, C.A. 1980. Macroeconomics and reality. *Econometrica* **48**(1): 1–48.

Sims, C.A., Stock, J.H. and Watson, M.W. 1990. Inference in linear time series models with some unit roots. *Econometrica* **58**(1): 113–144.

Uddin, M.J. 2023. Investigating the impulse responses of renewable energy in the context of China: A Bayesian VAR Approach. *Renewable Energy*, **219**: 119485.

van der Drift, R., de Haan, J. and Boelhouwer, P. 2024. Forecasting house prices through credit conditions: a Bayesian approach. *Computational Economics* **64**(6): 3381-3405.

Zellner, A. 1971. *An introduction to Bayesian inference in econometrics*. Wiley.

Received: 3 September 2025; Accepted: 10 December 2025