Bayesian Vector Autoregression Analysis on EUR/USD Exchange Rate Forecast

Cosimo Schiavoni - 964563
DSE Unimi - Bayesian Statistics

April 2025

Abstract

This project explores the use of Bayesian statistics to a multivariate time series forecasting application, focusing on the joint dynamics of the EUR/USD exchange rate, Crude Oil price, and the S&P 500 Volatility Index (VIX). The analysis employs a Bayesian Vector Autoregression (BVAR) model with a Minnesota prior to capture temporal interdependencies among the variables using daily data from 2021 to 2025. To enhance the selection of the four prior hyperparameters ($\lambda_1 - \lambda_4$), a Bayesian Neural Network (BNN) is trained to learn them, based on the input values obtained by running a Grid Search algorithm over a set of possible parameters. The BNN architecture, implemented using the torch package in R, consists of a dense neural network with ReLU activation and is optimized via stochastic gradient descent using the Adam optimizer. The optimal prior hyperparameters predicted by the BNN are fed into the BVAR model using the BMR package for posterior simulation and out-of-sample forecasting. Forecast performance is evaluated over a short-time period covering the first quarter of 2025. Results show that the model provides accurate short- to medium-term forecasts of the EUR/USD exchange rate, with root mean square error (RMSE) and mean absolute error (MAE) metrics indicating strong performance over a four-month horizon. However, the predictive accuracy declines over longer horizons (e.g. one year), highlighting the known limitations of BVAR models in capturing long-term structural changes in financial series.

Keywords: Bayesian Statistics, Financial Markets, Forecast Accuracy, Time Series, Machine Learning, Econometrics

Introduction

Forecasting financial time series is a main topic in econometrics and financial analytics, with significant implications for policymakers and investors. Exchange rates, in particular, are among the most actively traded and analyzed economic indicators. The dynamic nature of financial markets, characterized by non-linear dependencies, volatility and structural breaks, poses considerable challenges to traditional modeling approaches. Multivariate time series models, such as Bayesian Vector Autoregressions (BVARs), are well-suited to capturing such joint dynamics, incorporating prior information that shrinks the parameter space and address the limitations of classical VAR models. This project proposes an integration of Bayesian neural networks (BNNs) and BVAR models to configure a data-driven prior hyperparameters selection. The final objective of this analysis is to obtain an accurate method to perform the forecast of the EUR/USD exchange rate.

Data Source and Variables

This analysis employs daily financial market data sourced from Yahoo Finance, covering the period from January 2021, to April 2025. The data are retrieved automatically using the "quantmod" package in R, ensuring both reproducibility and up-to-date data availability. This project focuses on three critical financial time series:

EUR/USD exchange rate: is among the most traded currency pairs, influenced by interest rate differentials, monetary policies (the ECB and the Federal Reserve), inflation expectations, capital flows, and relative economic performance between the Eurozone and the United States. Shocks from U.S. or Eurozone monetary and fiscal announcements often manifest rapidly in this indicator.

Crude Oil price: is a proxy for global energy markets, is shaped by global demand and supply, geopolitical tensions, OPEC+ decisions, market speculation and broader macroeconomic trends such as industrial activity. Oil price is also a significant component of inflation expectations, particularly in energy-importing economies (e.g. Eurozone).

VIX: reflects investor expectations of short-term volatility derived from S&P 500 option prices in U.S. equity markets and tends to spike in response to macroeconomic uncertainty, policy shifts, or financial instability.

Theory Foundations

Fluctuations in financial indicators such as exchange rates, commodity prices, and volatility indices exhibit important signals about macroeconomic trends and monetary policy. These asset prices contain a wealth of information, responding in real time to new data as market participants adjust expectations related to growth, inflation, and interest rates. However, interpreting these price movements is challenging because they are simultaneously influenced by a wide array of overlapping economic shocks (that rarely occur in isolation and can move financial variables in different directions), driven by both domestic and international dynamics, especially from dominant markets, such as the U.S. (L. Brandt, A. Saint Guilhem, M. Schröder, I. Van Robays - 2021). The variables analyzed in this research are deeply interconnected. For example, a spike in the VIX may drive investors into U.S. dollar assets, leading to appreciation of the USD and depreciation of the Euro. Nonetheless, S&P 500 spillover on EUR/USD exchange rate is not an easy to define. Referring to the correlation between these indexes: "Market observations give strong evidence that financial quantities are correlated in a strongly nonlinear, non-deterministic way." (L. Teng, M. Ehrhardt and M. Günther - 2016). The evidence of stochastic relationship basically means that the strength and even the direction of the correlation can change depending on market conditions. Regarding Oil price, it is plausible that the shock could produce impacts on the EUR/USD exchange rate by affecting Eurozone trade balances and inflation. "In recent years rising oil prices have often coincided with a strengthening of the US dollar – which has potentially intensified inflation dynamics in the euro area." (M. Ricci - ECB Economic Bulletin, Issue 7/2024). The analysis must consider the major macroeconomic shocks during 2021 to 2025:

2021 – Covid-19: recovery phase brought supply bottlenecks and inflation;

2022 – Russia/Ukraine war: triggered an energy crisis that significantly raised oil prices and heightened market risk. Central banks responded with aggressive interest rate hikes, influencing both volatility and exchange rates;

2025 – U.S. trade war: initiated under the Trump administration, including the imposition of tariffs and growing protectionist sentiment, is currently provoking side effects on global trade, exchange rate volatility, and investor confidence, contributing to elevated uncertainty in currency and commodity markets and pushing the VIX higher (2025).

Data Preprocessing

All data are collected at the daily frequency, and only closing prices are used to ensure comparability across variables and to minimize intraday noise. Once downloaded, the series are merged into a single multivariate time series object and cleaned to remove missing values via listwise deletion "na.omit()", thus preserving the temporal alignment of all observations. To prepare the data for training the Bayesian Neural Network (BNN), the time series are then standardized using z-score normalization, applying the "scale()" function across each variable. This preprocessing step ensures that all variables contribute equally to the loss function and prevents the optimization from being dominated by variables with larger numerical scales, such as oil prices. In the final data frame, the time series is then split into a training sample (up to end-2024) and a test sample (from 2025 onward), enabling out-of-sample forecast validation. This preprocessing pipeline ensures that the data input into both the neural and econometric components of the model is clean, standardized, and representative of real-world financial conditions, making the results robust and suitable for empirical inference.

Bayesian Vector Autoregression Model - BVAR($\boldsymbol{p}$)

A Bayesian Vector Autoregression (BVAR) model is a multivariate time series model estimated using Bayesian methods, where prior distributions are imposed on the model parameters. This is particularly useful in high-dimensional systems, such as those involving multiple macro-financial indicators, where traditional estimation (e.g. OLS) may suffer from overfitting or multicollinearity. In this research, a BVAR(10) model is analyzed on three daily financial variables from 2021 to 2024: the EUR/USD exchange rate (EURUSD_t), the VIX volatility index (VIX_t), and crude oil prices (OIL_t). The goal is to generate probabilistic forecasts for 2025 and evaluate out-of-sample forecast performance. Bayesian econometrics is based on **Bayes' Theorem**, which updates prior beliefs about parameters using observed data. Mathematically, for parameters $\boldsymbol{\theta}$ and data $\mathcal{D}$:

$$\underbrace{p(\boldsymbol{\theta} \mid \mathcal{D})}_{\text{Posterior}} = \frac{\underbrace{p(\mathcal{D} \mid \boldsymbol{\theta})}_{\text{Likelihood}} \cdot \underbrace{p(\boldsymbol{\theta})}_{\text{Prior}}}{\underbrace{p(\mathcal{D})}_{\text{Marginal likelihood}}} \tag{1}$$

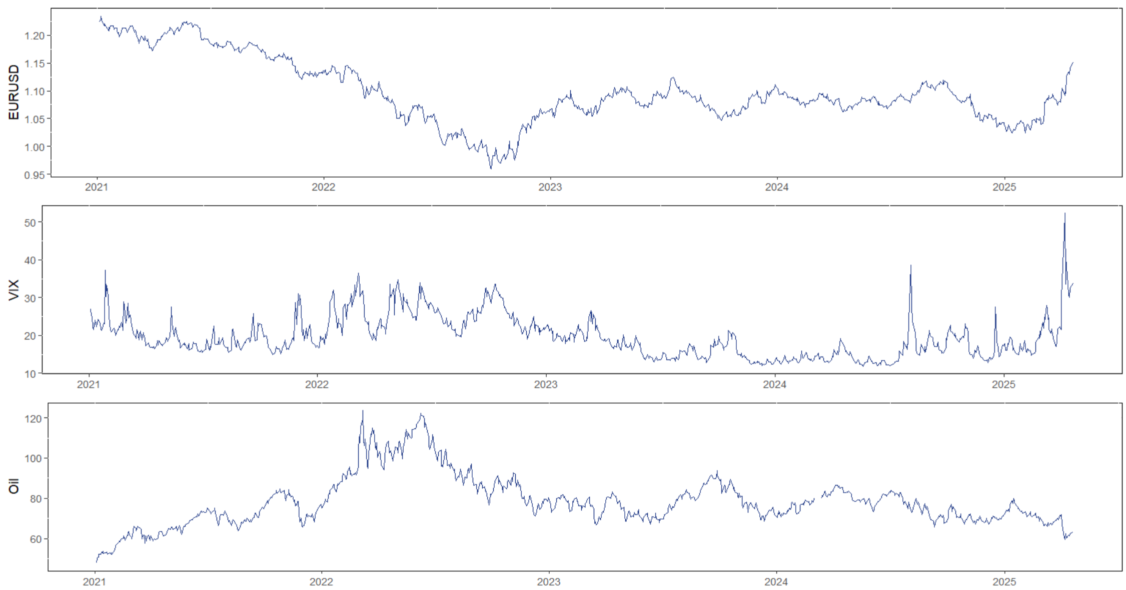

Figure 1: This figure illustrates the Multivariate Time Series for EUR/USD Exchange Rate, Crude oil Prices, VIX (2021 - 2025).

The **posterior** represents updated beliefs about $\boldsymbol{\theta}$ after observing $\mathcal{D}$. In that regard, Bayesian approach in BVAR model estimation foresees the following steps: specify a prior: $p(\boldsymbol{\theta})$, derive the likelihood from the data: $p(\mathcal{D} \mid \boldsymbol{\theta})$, compute the posterior using Bayes' Theorem.[1] In such a model, the parameter vector is defined as:

$$\boldsymbol{\theta} = \{\Phi_1, \ldots, \Phi_p, \Sigma\} \tag{2}$$

where $\boldsymbol{\theta}$ includes the VAR coefficients and the error covariance matrix, hence the parameters to estimate in a BVAR model are $\Phi = \{\Phi_1, \ldots, \Phi_{10}\}$ and Σ. As a result, the Bayesian inference combines the likelihood of the data with a prior distribution on Φ and Σ to form a posterior distribution:

$$p(\Phi, \Sigma \mid \mathbf{Y}) \propto p(\mathbf{Y} \mid \Phi, \Sigma) \cdot p(\Phi, \Sigma) \tag{3}$$

The Bayesian approach in BVAR usually shows several advantages with respect to traditional frequentist models: priors prevent overfitting in high-dimensional systems (Shrinkage), priors encode economic beliefs (Interpretability), posterior predictive distributions incorporate parameter uncertainty (Forecasting Power), bayesian approach accommodates structural restrictions and alternative prior beliefs (Flexibility).

In the following paragraphs, the BVAR model is structured based on the scope of this reasearch. Let $\mathbf{y}_t = (y_{1t}, y_{2t}, y_{3t})'$ denote a $n \times 1$ vector of observed time series at time t, where $N = 3$. The BVAR(p) model is defined as:

$$\mathbf{y}_t = \Phi_1 \mathbf{y}_{t-1} + \Phi_2 \mathbf{y}_{t-2} + \cdots + \Phi_p \mathbf{y}_{t-p} + \boldsymbol{u}_t, \quad \boldsymbol{u}_t \sim \mathcal{N}(\mathbf{0}, \Sigma) \tag{4}$$

The model with $n = 3$ variables (EURUSD$_t$, VIX$_t$, and OIL$_t$) and $p = 10$, can be represented as follows:

[1] Since the posterior often lacks a closed-form solution, it is possible to use Markov Chain Monte Carlo (MCMC) methods like the Gibbs sampler.

$$\mathbf{y}_t = \begin{bmatrix} \mathrm{EURUSD}_t \\ \mathrm{VIX}_t \\ \mathrm{OIL}_t \end{bmatrix}, \quad \Phi_l = \begin{bmatrix} \phi_{11}^{(l)} & \phi_{12}^{(l)} & \phi_{13}^{(l)} \\ \phi_{21}^{(l)} & \phi_{22}^{(l)} & \phi_{23}^{(l)} \\ \phi_{31}^{(l)} & \phi_{32}^{(l)} & \phi_{33}^{(l)} \end{bmatrix} \quad \text{for } l = 1, \ldots, 10 \tag{5}$$

Each equation includes its own lags and the lags of the other two variables up to 10 periods, resulting in a total of $3 \times 10 = 30$ regressors per equation [2]. In the Bayesian approach, prior distributions are imposed on the parameters $\phi_{i,j}^{(k)}$ and Σ, often to shrink the estimates toward a simple benchmark (such as a unit root or white noise process), which helps avoid overfitting and improves out-of-sample forecast performance. The lag order p determines how many past observations of each variable are used to model the dynamics of the system. Here follows and example of EURUSD_t equation:

$$\begin{aligned} \mathrm{EURUSD}_t = {} & \phi_{1,1}^{(1)}\mathrm{EURUSD}_{t-1} + \phi_{1,2}^{(1)}\mathrm{VIX}_{t-1} + \phi_{1,3}^{(1)}\mathrm{OIL}_{t-1} \\ & + \phi_{1,1}^{(2)}\mathrm{EURUSD}_{t-2} + \phi_{1,2}^{(2)}\mathrm{VIX}_{t-2} + \phi_{1,3}^{(2)}\mathrm{OIL}_{t-2} \\ & \vdots \\ & + \phi_{1,1}^{(10)}\mathrm{EURUSD}_{t-10} + \phi_{1,2}^{(10)}\mathrm{VIX}_{t-10} + \phi_{1,3}^{(10)}\mathrm{OIL}_{t-10} + u_{1,t} \end{aligned}$$

where $\mathbf{y}_t = \begin{bmatrix} \mathrm{EURUSD}_t \\ \mathrm{VIX}_t \\ \mathrm{OIL}_t \end{bmatrix}$ is a 3×1 vector of endogenous variables, Φ_i are 3×3 coefficient matrices for lags $i = 1, \ldots, 10$, $\mathbf{u}_t$ is an 3×1 vector of error terms, Σ is a 3×3 covariance matrix of the residuals. Similarly, the equations for VIX_t and OIL_t would follow the same structure (each equation in the system includes 10 lags of all three variables):

$$\begin{aligned} \mathrm{EURUSD}_t &= \sum_{l=1}^{10} \left(\phi_{1,1}^{(l)}\mathrm{EURUSD}_{t-l} + \phi_{1,2}^{(l)}\mathrm{VIX}_{t-l} + \phi_{1,3}^{(l)}\mathrm{OIL}_{t-l} \right) + u_{1,t} \\ \mathrm{VIX}_t &= \sum_{l=1}^{10} \left(\phi_{2,1}^{(l)}\mathrm{EURUSD}_{t-l} + \phi_{2,2}^{(l)}\mathrm{VIX}_{t-l} + \phi_{2,3}^{(l)}\mathrm{OIL}_{t-l} \right) + u_{2,t} \\ \mathrm{OIL}_t &= \sum_{l=1}^{10} \left(\phi_{3,1}^{(l)}\mathrm{EURUSD}_{t-l} + \phi_{3,2}^{(l)}\mathrm{VIX}_{t-l} + \phi_{3,3}^{(l)}\mathrm{OIL}_{t-l} \right) + u_{3,t} \end{aligned}$$

In this research, the **Minnesota prior** (Canova - 2007), has been introduced as a prior distribution to encode beliefs about the parameters of the model before observing the data (to incorporate prior macroeconomic theory or empirical regularities). It is a widely used informative prior that imposes shrinkage on the autoregressive coefficients to improve forecasting performance, particularly when the number of parameters is large relative to the available observations. This prior distribution is particurlarly effective for shrinking the model toward a random walk or white noise process (for each variable), addressing overparameterization, considering a decaying influence across lags and cross-variable. Moreover, it is relatively simple to implement and computationally efficient. Under a

[2]Details are provided in "Annex - Posterior Densities for BVAR(10) Coefficients".

Minnesota prior, the distribution of the vectorized coefficient matrix is given by [3]:

$$\text{vec}(\Phi) \sim \mathcal{N}(\mu_\Phi, \Omega_\Phi), \tag{6}$$

where the prior mean vector μ_Φ is specified as:

$$\mu_\Phi = \begin{cases} 1 & \text{if } \Phi_{jj}^{(1)} \text{ (first own lag)} \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

The Minnesota prior imposes the following structure on individual coefficients:

$$\begin{aligned} &\phi_{j,j}^{(1)} \sim \mathcal{N}(1, \sigma^2_{\phi_{j,j}^{(1)}}), \\ &\phi_{j,k}^{(l)} \sim \mathcal{N}(0, \sigma^2_{\phi_{j,k}^{(l)}}), \quad \text{for } j \neq k \text{ or } l > 1, \end{aligned} \tag{8}$$

where $\phi_{j,k}^{(l)}$ denotes the coefficient on variable k in the j-th equation at lag l. The prior variance for each coefficient $\Phi_{ij}^{(\ell)}$ depends on whether it corresponds to an own lag, a cross-variable lag, or an exogenous variable (e.g. constant):

$$\text{Var}(\Phi_{ij}^{(\ell)}) = \begin{cases} \dfrac{\lambda_1}{d(\ell)} & \text{if } i = j \quad \text{(own lag)} \\ \dfrac{\lambda_1 \cdot \lambda_2 \cdot \sigma_j^2}{d(\ell) \cdot \sigma_i^2} & \text{if } i \neq j \quad \text{(cross lag)} \\ \lambda_1 \cdot \lambda_3 & \text{if exogenous variable} \end{cases} \tag{9}$$

where σ_i^2 and σ_j^2 are the residual variances from univariate AR models for equations i and j, respectively, and $d(\ell)$ is a lag decay function [4]. The prior variances depend on the hyperparameters $(\lambda_1, \lambda_2, \lambda_3)$, which are pre-selected combining Grid Search and Bayesian Neural Network algorithms in this research.

In a BVAR(p) model with n variables and p lags, there are n^2p autoregressive coefficients to estimate. As p or n increases, this could quickly leads to overparameterization, resulting in overfitting and poor out-of-sample performance. The Minnesota prior addresses this issue by introducing structured shrinkage: own-lag coefficients are expected to be more important than lags of other variables, more recent lags are assumed to carry more information than distant ones and cross-variable effects are typically weaker and therefore shrunk more aggressively toward zero.

The Minnesota prior assumes that each coefficient is centered around a prior mean (typically zero) with a variance that decays with lag order and adjusts for cross-variable relationships. Importantly, the prior mean for the first own lag (e.g. the variable's own first lag coefficient) is often set to one, reflecting the belief that macro-financial variables such as exchange rates follow a near-random-walk process.

The prior variance Ω_Φ is specified using four hyperparameters as follows:

[3] These prior means and variances are plugged into the BVAR likelihood to derive the posterior (analytically in conjugate cases or via Gibbs sampling otherwise). In practice, the posterior draws of the VAR coefficients Φ and the covariance matrix Σ reflect a combination of the observed data and the prior beliefs.

[4] In this research:

$$d(\ell) = \begin{cases} \ell^{\lambda_4} \\ \lambda_4^{-\ell+1} \end{cases}$$

Overall Shrinkage (λ_1): Controls global tightness of the prior, large values induce strong regularization, shrinking coefficients toward zero. On the contrary, low values make the model highly flexible but prone to overfitting.

Cross-variable Shrinkage (λ_2): Determines how strongly the lagged values of other variables are penalized. Large values limit spillover effects, so that each variable acts as an independent univariate process. On the other hand low values encourage capturing dynamic interdependencies, such as VIX–EURUSD co-movements.

Exogenous Shrinkage (λ_3): Sets the tightness of the prior on the intercept term. High values downplay trend components while low values may overfit the sample mean.

Lag Decay (λ_4): Implements exponential decay across lags, reducing the influence of distant lags. Large values apply strong penalty on higher-order lags (model effectively ignores distant history). Conversely, low values retain long memory (useful when dynamics persist over time).

Writing the BVAR(p) model from Equation (4) in regression form (stacked over time):

$$y = X\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, I_T \otimes \Sigma) \tag{10}$$

where $y \in \mathbb{R}^{Tm\times 1}$ is the stacked response vector, $X \in \mathbb{R}^{Tm\times km}$ is the regressor matrix (block-diagonal), $\beta \in \mathbb{R}^{km\times 1}$ are the vectorized coefficients, and $\Sigma \in \mathbb{R}^{m\times m}$ is the residual covariance matrix (assumed fixed). Under the Minnesota prior, the coefficients β [5] follow a multivariate normal distribution:

$$\beta \sim \mathcal{N}(\beta_0, V_\beta) \tag{11}$$

where β_0 is the prior mean and V_β is the prior variance-covariance matrix, typically diagonal. $\Sigma = \bar{\Sigma}$ is assumed to be a fixed diagonal matrix (not estimated).[6]
The likelihood of the data given β is:

$$p(y \mid \beta) \propto \exp\left(-\frac{1}{2}(y - X\beta)'(I_T \otimes \Sigma^{-1})(y - X\beta)\right) \tag{12}$$

The posterior distribution is proportional to the product of the prior and the likelihood:

$$\begin{aligned} p(\beta \mid y) \propto \exp&\left(-\frac{1}{2}(\beta - \beta_0)'V_\beta^{-1}(\beta - \beta_0)\right) \\ \times \exp&\left(-\frac{1}{2}(y - X\beta)'(I_T \otimes \Sigma^{-1})(y - X\beta)\right) \end{aligned} \tag{13}$$

Expanding the prior quadratic term [7]:

$$(\beta - \beta_0)'V_\beta^{-1}(\beta - \beta_0) = \underbrace{\beta' V_\beta^{-1}\beta}_{\text{quadratic in } \beta} - \underbrace{2\beta' V_\beta^{-1}\beta_0}_{\text{linear in } \beta} + \underbrace{\beta_0' V_\beta^{-1}\beta_0}_{\text{constant}}$$

[5] Note that β corresponds to vec(Φ) in standard form Equation (5)

[6] V_β is the prior variance-covariance matrix of the VAR coefficients β under the Minnesota prior. It encodes beliefs about the uncertainty and structure of the coefficients before seeing the data. Do not confuse this with Σ, the covariance matrix of the residuals in the BVAR, which reflects the volatility of the shocks and is fixed via hyperparameters.

[7] A quadratic form is an expression of the type $x'Ax$, where x is a vector and A is a symmetric matrix, resulting in a second-degree polynomial in x. A useful identity for expanding expressions like $(a - Bx)'M(a - Bx)$ is $(a - Bx)'M(a - Bx) = a'Ma - 2x'B'Ma + x'B'MBx$, which decomposes the expression into constant, linear, and quadratic terms in x, simplifying posterior derivations.

Expanding the likelihood quadratic term and summing the prior and likelihood terms:

$$(y - X\beta)'(I_T \otimes \Sigma^{-1})(y - X\beta) = \underbrace{y'(I_T \otimes \Sigma^{-1})y}_{\text{constant}} - \underbrace{2\beta' X'(I_T \otimes \Sigma^{-1})y}_{\text{linear in } \beta} + \underbrace{\beta' X'(I_T \otimes \Sigma^{-1})X\beta}_{\text{quadratic in } \beta}$$

$$-2 \log p(\beta \mid y) =$$

$$\underbrace{\beta' V_\beta^{-1}\beta + \beta' X'(I_T \otimes \Sigma^{-1})X\beta}_{\text{quadratic in } \beta} - \underbrace{2\beta' \left(V_\beta^{-1}\beta_0 + X'(I_T \otimes \Sigma^{-1})y\right)}_{\text{linear in } \beta} + \underbrace{\beta_0' V_\beta^{-1}\beta_0 + y'(I_T \otimes \Sigma^{-1})y}_{\text{constant}}$$

Defining $K_\beta = V_\beta^{-1} + X'(I_T \otimes \Sigma^{-1})X$ and $\eta = V_\beta^{-1}\beta_0 + X'(I_T \otimes \Sigma^{-1})y$, then:

$$-2 \log p(\beta \mid y) = \underbrace{\beta' K_\beta \beta}_{\text{quadratic in } \beta} - \underbrace{2\beta'\eta}_{\text{linear in } \beta} + \text{const}$$

Completing the square for:

$$\beta' K_\beta \beta - 2\beta'\eta = \underbrace{(\beta - K_\beta^{-1}\eta)' K_\beta (\beta - K_\beta^{-1}\eta)}_{\text{quadratic form centered at } K_\beta^{-1}\eta} - \underbrace{\eta' K_\beta^{-1}\eta}_{\text{constant}}$$

Therefore, the posterior distribution of β is:

$$p(\beta \mid y) \propto \exp\left(-\frac{1}{2}(\beta - K_\beta^{-1}\eta)' K_\beta (\beta - K_\beta^{-1}\eta)\right) \tag{14}$$

$$\beta \mid y \sim \mathcal{N}(K_\beta^{-1}(V_\beta^{-1}\beta_0 + X'(I_T \otimes \Sigma^{-1})y),\ K_\beta^{-1}) \tag{15}$$

To define the optimal parameters of the Minnesota prior a **Bayesian Neural Network (BNN)** has been trained, using the output of a **Grid Search** analysis performed on a list of possible hyparameters as input of the neural network. This learning approach allows for a data-driven adjustment of the prior structure based on evolving market dynamics, combining the forecasting strength of structured Bayesian econometric models with the flexibility of neural networks in extracting empirical regularities. By learning how optimal priors relate to economic conditions, the model not only achieves superior predictive accuracy but also opens a path toward adaptive prior specification in future forecasting. This approach is particularly valuable in volatile markets, where the relationships among financial variables are complex and evolving. Formally, the process can be described in the following steps:

At first, an extensive Grid Search over the BVAR prior hyperparameters and lag lengths has been conducted. For each combination of $\mathcal{HP} = \{\lambda_1, \lambda_2, \lambda_3, \lambda_4, p\}$, a BVAR model has been estimated and its forecasting accuracy has been evaluated using the Root Mean Squared Error (RMSE) on the EUR/USD exchange rate.

$$\text{RMSE} = \sqrt{\frac{1}{h}\sum_{t=T+1}^{T+h}(\hat{y}_t - y_t)^2}, \tag{16}$$

where y_t denotes the actual EUR/USD exchange rate and $\hat{y}_t$ the forecast. As a result, a full Cartesian product of all hyperparameter combinations and lag lengths defined the search space:

$$\mathcal{HP} = \{(\lambda_1, \lambda_2, \lambda_3, \lambda_4, p)\}_{i=1}^{N}, \tag{17}$$

where N is the total number of configurations evaluated. For each tuple in $\mathcal{HP}$, a BVAR model has been estimated using the training dataset. The prior has been defined as shown in Equations (7) and (9). After estimating the model using 10.000 Gibbs sampling iterations, h-step-ahead forecasts over the test and relative RMSE were computed (as by Equation (16)). This yielded a training dataset of the form:

$$\mathcal{D}_{\text{train}} = \left\{(x^{(i)}, y^{(i)})\right\}_{i=1}^{N} \tag{18}$$

where $x^{(i)} \in \mathbb{R}^d$ are the input features computed from the training period (e.g. averages, standard deviations, and correlations of EURUSD, VIX, and Oil), $y^{(i)} \in \mathbb{R}^4$ are the corresponding optimal hyperparameters $\{\lambda_1, \lambda_2, \lambda_3, \lambda_4\}$ that yielded the lowest RMSE.

Secondly, the BNN used a fully connected feedforward architecture implemented via the torch package in R. It took as input standardized features of the recent historical data and outputs four hyperparameters $\{\lambda_1, \lambda_2, \lambda_3, \lambda_4\}$ governing the tightness, decay, and overall scale of the Minnesota prior.
Let $\mathbf{X} \in \mathbb{R}^{T \times K}$ denote the matrix of standardized input features, where T is the number of daily observations and $K = 3$ corresponds to the three variables: EUR/USD exchange rate, VIX, and crude Oil price. The BNN is defined as a two-layer feedforward neural network with rectified linear unit (ReLU) activation:

$$\begin{aligned} \mathbf{H}_1 &= \mathrm{ReLU}(\mathbf{X}\mathbf{W}_1 + \mathbf{b}_1), \quad \mathbf{W}_1 \in \mathbb{R}^{K \times H},\ \mathbf{b}_1 \in \mathbb{R}^{H}, \\ \mathbf{Y}_{\text{pred}} &= \mathbf{H}_1\mathbf{W}_2 + \mathbf{b}_2, \quad \mathbf{W}_2 \in \mathbb{R}^{H \times 4},\ \mathbf{b}_2 \in \mathbb{R}^{4}, \end{aligned} \tag{19}$$

where H is the number of hidden units (set to 5.000 in this research), and $\mathbf{Y}_{\text{pred}} \in \mathbb{R}^{T \times 4}$ represents the predicted prior hyperparameters for each input row: $\lambda_1, \lambda_2, \lambda_3, \lambda_4$.

Thirdly, the network has been trained using stochastic gradient descent with the Adam optimizer for 6.000 epochs to minimize the MSE between the predicted hyperparameters and a predefined target vector $\mathbf{Y}_{\text{true}}$ derived from grid search results:

$$\mathcal{L}(\theta) = \frac{1}{T} \sum_{t=1}^{T} \left\| \mathbf{Y}_{\text{pred}}^{(t)} - \mathbf{Y}_{\text{true}} \right\|_2^2, \tag{20}$$

where $\theta = \{\mathbf{W}_1, \mathbf{W}_2, \mathbf{b}_1, \mathbf{b}_2\}$ represents all trainable parameters of the network. The network has been trained for 6.000 epochs, and the final predicted hyperparameters were extracted from the most recent observation (the last row of $\mathbf{X}$).
The Bayesian Neural Network was trained on the dataset $\mathcal{D}_{\text{train}}$ to approximate the function:

$$f_{\text{BNN}} : x \mapsto y = \{\lambda_1, \lambda_2, \lambda_3, \lambda_4\} \tag{21}$$

Unlike standard neural networks, BNNs introduce uncertainty in weights through prior distributions, allowing for posterior inference over functions. Mathematically, let $\mathcal{W}$ represent the weights of the BNN with prior $p(\mathcal{W})$ and observed data $\mathcal{D}_{\text{train}}$. The posterior predictive distribution for a new input x^* is:

$$p(y^* \mid x^*, \mathcal{D}_{\text{train}}) = \int p(y^* \mid x^*, \mathcal{W})\, p(\mathcal{W} \mid \mathcal{D}_{\text{train}})\, d\mathcal{W} \tag{22}$$

In practice, this integral is approximated using Monte Carlo dropout, variational inference, or Hamiltonian Monte Carlo. The BNN learns to map input features (economic signal) to hyperparameters, capturing uncertainty along the way.
After training, the BNN predicted a new set of optimal prior hyperparameters $\hat{\lambda} = f_{\text{BNN}}(x_{\text{new}})$ for unseen macroeconomic conditions. These predicted values were then used to estimate a final BVAR model:

$$\Phi \sim \mathcal{N}(\mu_\Phi, \Omega_\Phi(\hat{\lambda})), \quad \Sigma \sim \mathcal{IW}(S_0, \nu_0) \tag{23}$$

To ensure numerical stability and model convergence, the hyperparameters were transformed post-estimation to ensure positivity:

$$\lambda = (|\hat{\lambda}_1|, |\hat{\lambda}_2|, |\hat{\lambda}_3|, |\hat{\lambda}_4|) \tag{24}$$

where $\hat{\lambda}_i$ denotes the raw output from the network for each hyperparameter foreseen by Minnesota prior. The resulting vector $\boldsymbol{\lambda}$ was then supplied as input to the BVAR prior specification. This approach allows the BVAR framework to incorporate a flexible, data-driven prior that adapts over time to changes in financial conditions, reducing the reliance on static or arbitrarily selected hyperparameter values, which is expected to generalize better than any fixed-prior counterpart.

Once the Minnesota prior hyperparameters have been defined, the estimation of the posterior distribution in the BVAR model proceeded via **Gibbs Sampling** (a Markov Chain Monte Carlo technique - MCMC). When employing the Minnesota prior, a specific prior structure that imposes shrinkage on the BVAR coefficients is assumed, while the covariance matrix of residuals is treated as fixed or known (rather than sampling from its posterior as in fully conjugate priors).

1. *Set priors:* under the Minnesota prior, the coefficients on the first own lag of each variable are given a prior mean close to one (e.g. random walk), while cross-variable and higher-order lags are shrunk toward zero. The residual covariance matrix Σ is assumed to be diagonal and known (It has to be algebraically calculated and inverted only once). The following priors are assumed [8]:

$$\text{vec}(\Phi) \sim \mathcal{N}(\mu_\Phi, \Omega_\Phi), \quad \Sigma = \text{diag}(\hat{\sigma}_1^2, \hat{\sigma}_2^2, \ldots, \hat{\sigma}_n^2) \tag{25}$$

   where $\Phi = [\Phi_1, \Phi_2, \ldots, \Phi_p]$ and $\text{vec}(\cdot)$ stacks the matrices column-wise. The variance terms $\hat{\sigma}_i^2$ are fixed and estimated outside the Gibbs sampling loop.

2. *Initialize:* Start with initial values for Φ, often based on OLS estimates. The covariance matrix Σ was fixed throughout.

3. *Iterate:* The posterior predictive distribution for future observations was constructed by sampling $\Phi \mid \Sigma, \mathbf{Y}$ from its conditional posterior $\mathcal{N}(\bar{\Phi}, \bar{V})$ and simulating the BVAR forward, using the fixed covariance matrix Σ. Since Σ is fixed, only the regression coefficients are updated [9]. The sample of $\boldsymbol{\Phi} \sim \mathcal{N}(\bar{\boldsymbol{\Phi}}, \bar{\mathbf{V}})$ from its conditional posterior (as shown in Equation (15)) has been performed for 10.000 iterations to ensure convergence.

[8] In the standard form priors are defined as by Equations (7) and (9)

[9] Unlike the Normal-Inverse-Wishart prior, Σ is not sampled and remains constant throughout the Gibbs sampling procedure.

4. *Discard burn-in:* it removes initial samples (e.g. first 5.000) to allow convergence.
5. *Posterior inference:* used the retained posterior draws to compute posterior means, credible intervals, impulse response functions, and predictive densities.

Finally, the forecasts for 2025 were generated using the **posterior predictive distribution** computed through the Gibbs Sampler. The accuracy is evaluated using RMSE, MAE[10], and forecast error bands. Forecasting involves drawing from the posterior predictive distribution:

$$p(\mathbf{y}_{T+h} \mid \mathcal{D}) = \int p(\mathbf{y}_{T+h} \mid \boldsymbol{\theta}) \cdot p(\boldsymbol{\theta} \mid \mathcal{D}) \, d\boldsymbol{\theta} \tag{26}$$

This integrates over uncertainty in $\boldsymbol{\theta}$, offering a way to quantify forecast uncertainty.

Analysis Results

The Grid Search resulted in the following hyperparameters pre-selection: λ_1 (0,1), λ_2 (0,1), λ_3 (1) and λ_4 (0,1), with lag equal to 10. Starting from these inputs, the BNN computed the hyperparameters values actually used in the Minnesota prior: λ_1 (0,1696379), λ_2 (0,02871177), λ_3 (0,9307866) and λ_4 (0,136559).
Table 1 shows the out-of-sample forecast accuracy of BVAR models with lag lengths of 10 and 68, evaluated over two periods: the short-term period (January 2025 - April 2025) and the long-term period (January 2024 - April 2025). Forecast performance was assessed using RMSE and MAE across three key variables. The BVAR(10) model results demonstrate relatively strong performance for forecasts restricted to 2025, particularly in predicting the EUR/USD and oil prices. For example, RMSE and MAE are respectively 0,41 and 0,34 (for EUR/USD) and 0,36 and 0,274 (for Crude Oil Price), which points out a better result with respect to the BVAR(68) counterparts. These results suggest that a more parsimonious model may be better suited for short-term forecasting, while higher-order lag models may produce overfitting and degrade the general performance. In contrast, when the forecast horizon extends to a long-term period (2024-2025), BVAR(68) shows improvements in predicting VIX and Crude Oil Price, with a noticeable reduction in both RMSE and MAE (respectively reducing the scores for EUR/USD from 0,740 to 0,524 and from 0,643 to 0,394). This could support the notion that, over longer periods, richer lag structures can better capture the dynamics of financial variables,

Despite the general improvement of BVAR(68) for longer-term forecasts, its performance is still mixed. For instance, in 2025, its MAE for VIX equal to 1,104 is substantially higher than that of BVAR(10) equal to 0,925, while RMSE scores are respectively 1,758 and 1,561, indicating that higher model complexity does not always translate into superior accuracy, particularly for volatile variables like VIX. Overall, these findings underscore the importance of adapting BVAR model configurations to the forecast horizon and the characteristics of the target variables. Simpler specifications may outperform

[10] MAE is defined as:

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^{n} |y_t - \hat{y}_t|$$

where:

y_t	is the actual value at time t
$\hat{y}_t$	is the forecast or predicted value at time t
n	is the total number of observations

Table 1: Forecast Accuracy by BVAR Model, Period, and Variable

BVAR Model	**Period**	**Variable**	**RMSE**	**MAE**
BVAR(10)	2025	EURUSD	0,405	0,334
		VIX	1,561	0,925
		Oil	0,360	0,274
BVAR(10)	2024–2025	EURUSD	0,740	0,643
		VIX	0,932	0,674
		Oil	0,577	0,503
BVAR(68)	2025	EURUSD	0,518	0,453
		VIX	1,758	1,104
		Oil	0,692	0,599
BVAR(68)	2024–2025	EURUSD	0,524	0,394
		VIX	1,541	1,025
		Oil	0,423	0,338

in short-term settings, whereas more complex models may be justified for multi-period forecasting, taking care to avoid potential overfitting.

Figure 7 illustrates both the out-of-sample short-term period (January 2025 - April 2025) and long-term period (January 2024 - April 2025) trend line analysis. In the upper panels, corresponding to BVAR(10), the forecasts for the short-term period 2025 maintain a relatively good alignment with actual values, especially for EURUSD and Oil. The forecasted projections capture the general direction and volatility in these series. However, for the 2024–2025 time period, the forecast quality noticeably deteriorates, particularly in the case of the VIX, where the model underestimates the abrupt spikes and increasing volatility. Once again, this suggests that BVAR(10), while reasonably effective over shorter horizons, may lack sufficient historical information to generalize accurately across multi-year periods.

In contrast, the BVAR(68) model displayed in the lower panels demonstrates a general robust predictive structure. For the 2025 window, its forecasts for EURUSD and Oil maintain similar performance to the shorter-lag model, but the VIX forecast slightly improves in capturing the sharp variance near the end of the period. More importantly, for the extended 2024–2025 horizon, BVAR(68) exhibits an evident enhancement in forecast accuracy, especially for the EURUSD and Oil series. The VIX forecast remains somewhat conservative, yet it shows marginally better adaptability to volatility.

Overall, the findings suggest that for stable and moderately volatile periods, a shorter-lag BVAR may be sufficient. However, during periods of economic uncertainty or when forecasting across regime shifts, a long-lag structure provides superior adaptability. Moreover, the posterior distribution's evolution indicates that while priors play a role, their influence diminishes as the data richness increases, especially in models with deeper lag structures. This underscores the importance of calibrating prior tightness and model memory according to the forecast horizon and volatility characteristics of the macro-financial environment.

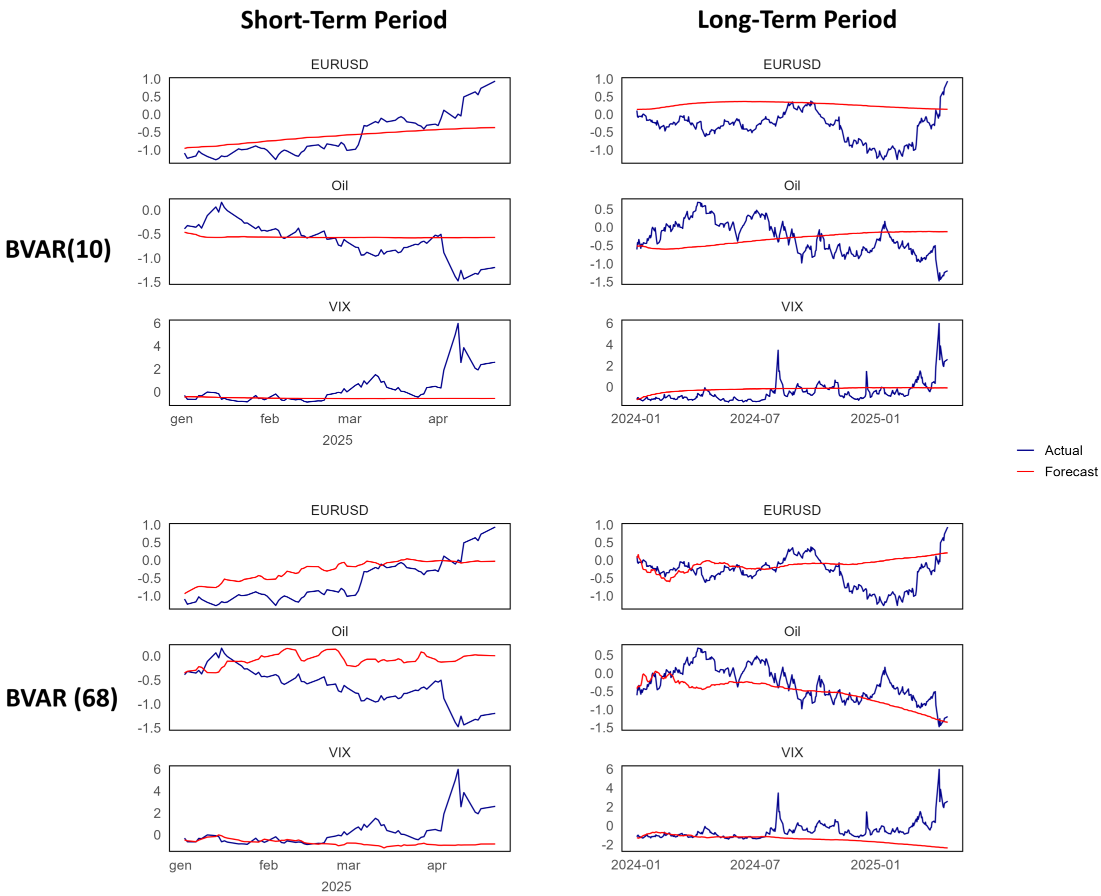

Figure 2: This figure illustrates the out-of-sample forecast analysis for EUR/USD Exchange Rate, Crude Oil Prices, and VIX (on the right side), for both lag lengths 10 (top) and 68 (bottom). The graphs show the actual (blue) and forecasted (red) time series for two distinct out-of-sample time periods (short-term and log-term).

The Impulse Response Function (IRF) analysis[11] shown in Figure 3 from the BVAR(10) model for the short-term period 2025 provides valuable insights into the dynamic interdependencies among the 3 variables, highlighting the responses to structural shocks (accompanied by 68% confidence intervals). Here follows a brief analysis:

At first, the responses of EUR/USD to the EUR/USD Shock itself show a positive reaction, which gradually dissipates over time, indicating a transitory impact with no long-run effect. The response to the Oil Shock is almost insignificant and around zero throughout, while the response to the VIX Shock is negatively durable over time.

Secondly, the response of the Crude Oil price to the Oil Shock is strongly negative and immediate, followed by a gradual reversion toward zero.The response to the EUR/USD Shock is positive but quite insignificant over time (around zero throughout the period),

[11]The IRF captures how a variable $y_{i,t+h}$ responds to a shock $\varepsilon_{j,t}$ over time, given by $IRF_{i,j,h} = \frac{\partial y_{i,t+h}}{\partial \varepsilon_{j,t}}$. In the R code, the function `bvar_obj$IRF(horizon)` simulates multiple draws of the IRF for each response-shock pair at each forecast horizon. These draws are reshaped into a 4D array, where the dimensions correspond to response variables, shocks, horizons, and simulation draws. The mean response and 68% credible intervals (16th and 84th percentiles) are computed across the simulation draws.

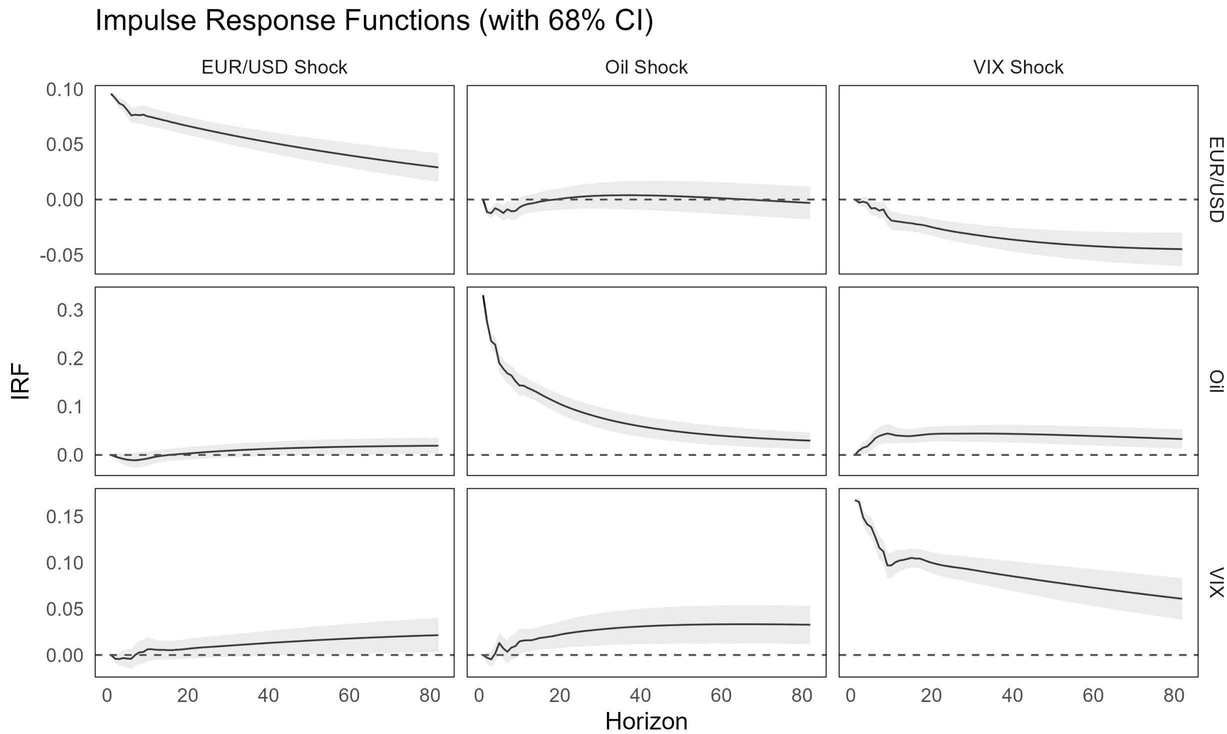

Figure 3: This figure shows the Impulse Response Functions (IRFs) of a BVAR model to one-standard-deviation shocks in EUR/USD (left column), Crude Oil Prices (center column), and the VIX Index (right column). The responses of EUR/USD, Oil, and VIX are plotted over a 70-day horizon, with 68% posterior Confidence Intervals shown in gray. The figure illustrates the dynamic interactions among financial variables in response to exogenous shocks.

while the response to the VIX Shock, although initially noisy, settles into a slightly positive and persistent pattern, potentially reflecting a delayed economic effect.

Finally, the response of VIX to VIX shock is a persistent negative trend, while the responses to the EUR/USD and the Oil shocks show a sligthly increasing effect over time.

Conclusions

The comparative analysis of BVAR(10) and BVAR(68) models yields encouraging results about the main scope of this research, which is the forecasting of the EUR/USD exchange rate. Empirical findings suggest that the BVAR(10) model shows a better predictive accuracy over short to medium-term period (1–3 months). However, over longer greater periods (exceeding one year), the BVAR(68) model outperforms its shorter-lag counterpart, highlighting the value of considering extended historical information when modeling persistent macro-financial relationships. This underscores the crucial trade-off between model complexity and forecast accuracy, as higher lag specifications (dealing with the risk of overfitting). It is thus important for the data scientist to calibrate the lag structure, balancing informational richness with statistical robustness. Overall, the results highlight the efficacy of Bayesian methods in financial forecasting applications, due to it's flexibility in integrating prior beliefs and managing model uncertainty.

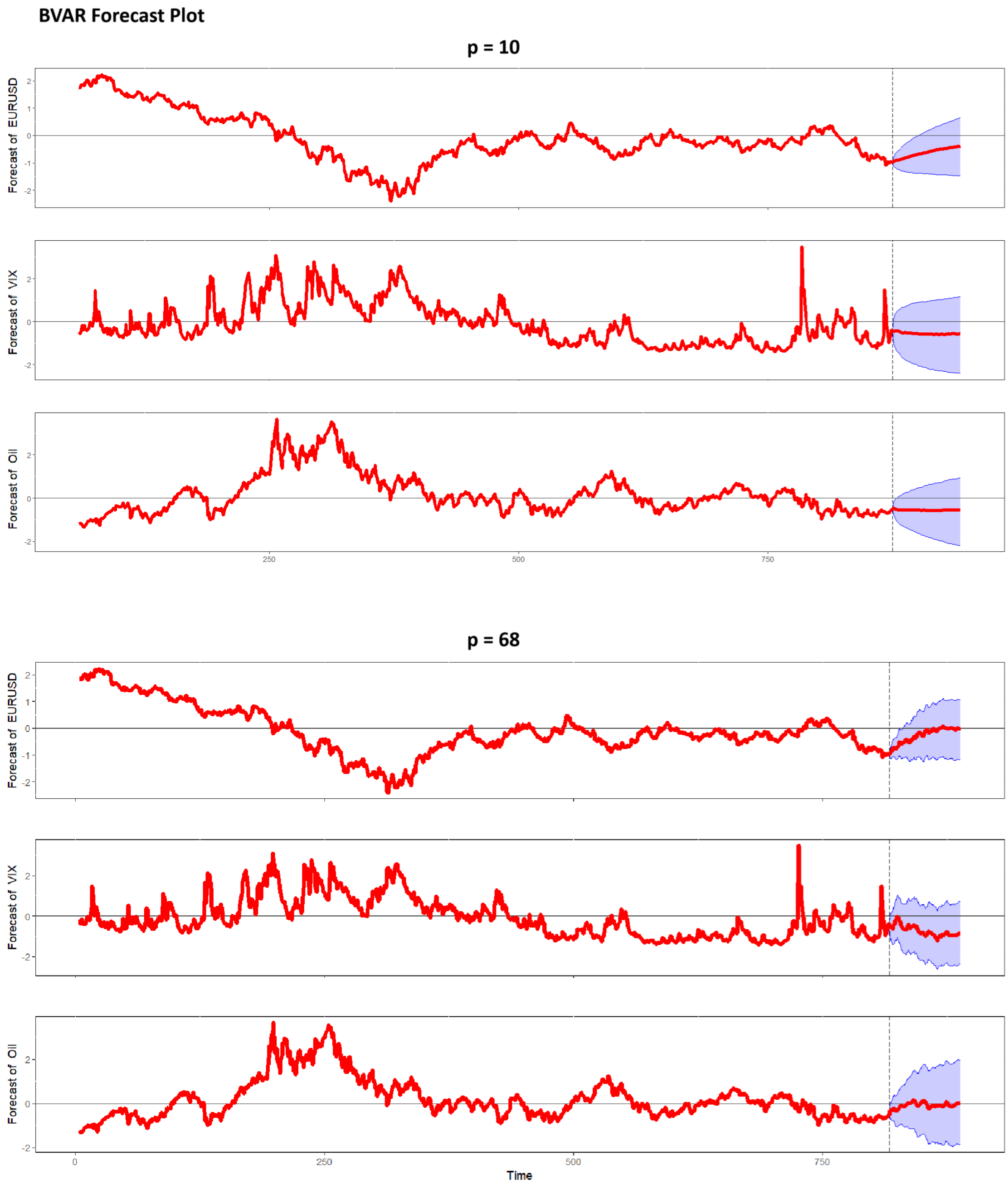

Figure 4: This figure displays the forecasts for 70 days from a Bayesian Vector Autoregression (BVAR) model with 10 and 68 lags for EUR/USD (top), VIX (middle), and Crude Oil Prices (bottom). The red lines represent historical (in-sample) observations, while the blue shaded areas indicate the forecast distribution with associated credible intervals. The vertical dashed line marks the beginning of the forecast period.

References

Keith O'Hara (2015) - "Bayesian Macroeconometrics in R" Version: 0.5.0 - Department of Economics, New York University

Keith O'Hara (2015) - "Package 'BMR'"

Kevin Kotzé - "Bayesian Vector Autoregression Models" - https://kevinkotze.github.io/ts-9-bvar/

Kevin Kotzé - "Tutorial: Bayesian Vector Autoregression Models" - https://kevinkotze.github.io/ts-9-tut/

L. Brandt, A. Saint Guilhem, M. Schröder, I. Van Robays (2021) - "What drives euro area financial market developments? The role of US spillovers and global risk" - ECB Working Paper Series No 2560

M. Ricci (2024) - "The link between oil prices and the US dollar: evidence and economic implications" - ECB Economic Bulletin, Issue 7/2024

L. Teng, M. Ehrhardt and M. Günther (2016) -"Modelling stochastic correlation" - Journal of Mathematics in Industry

Annex - Posterior Densities for BVAR(10) Coefficients

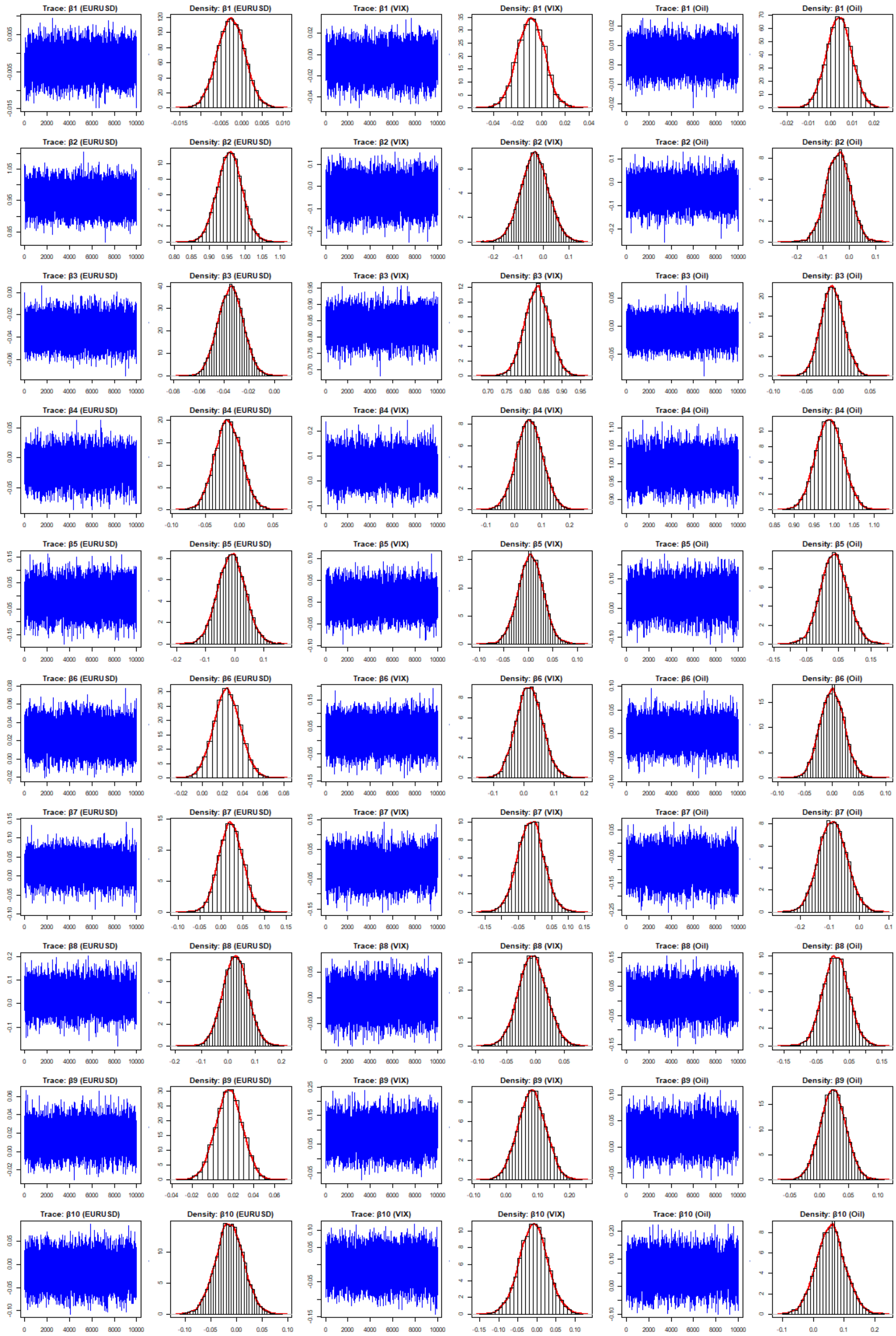

Figure 5: This figure displays trace plots and posterior densities for selected BVAR coefficients (β_1 to β_{10}) across the EUR/USD, VIX, and Crude Oil equations. The visualizations assess convergence and distributional properties of the Gibbs sampler draws used for posterior inference.

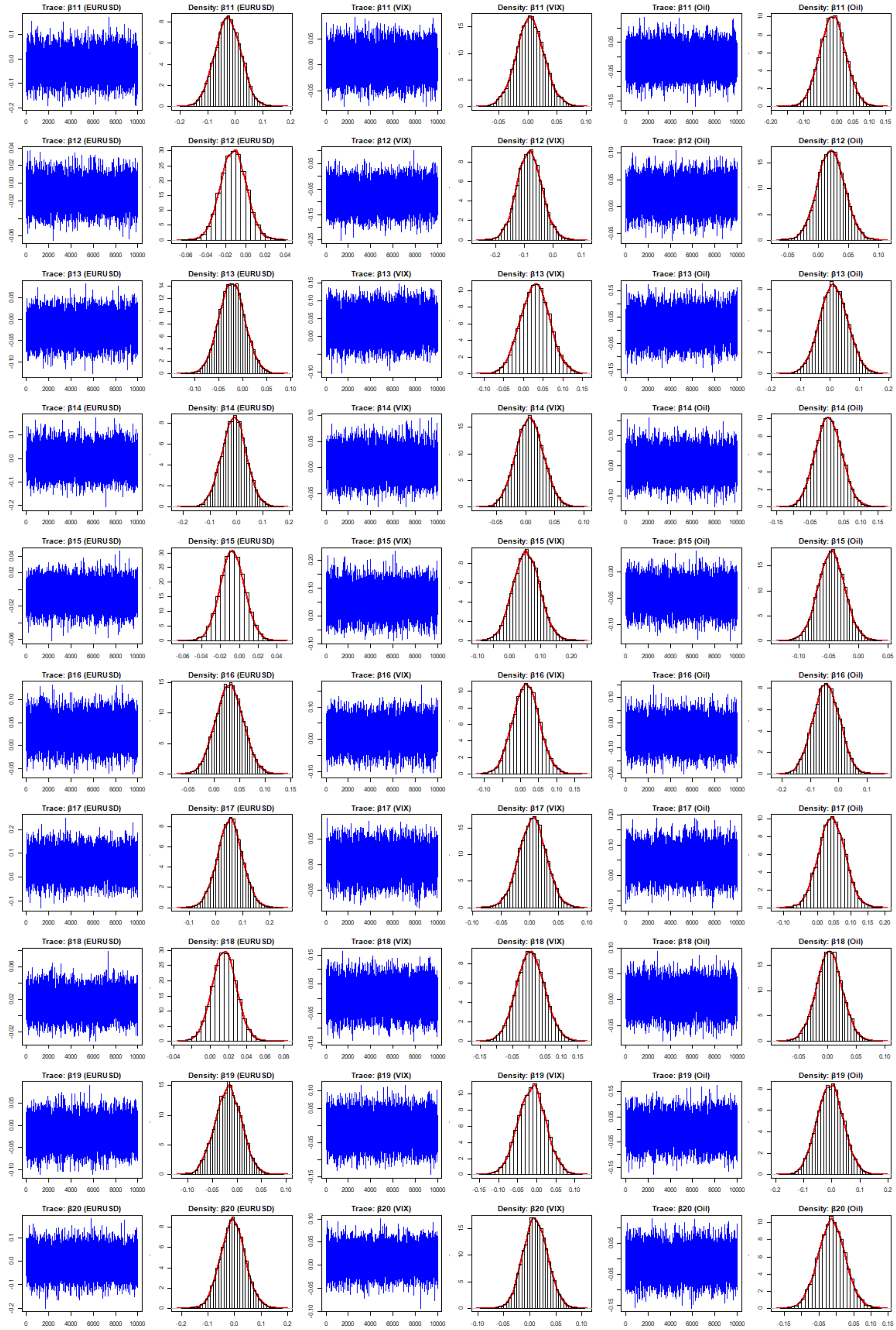

Figure 6: This figure displays trace plots and posterior densities for selected BVAR coefficients (β_{11} to β_{20}) across the EUR/USD, VIX, and Oil equations. The visualizations assess convergence and distributional properties of the Gibbs sampler draws used for posterior inference.

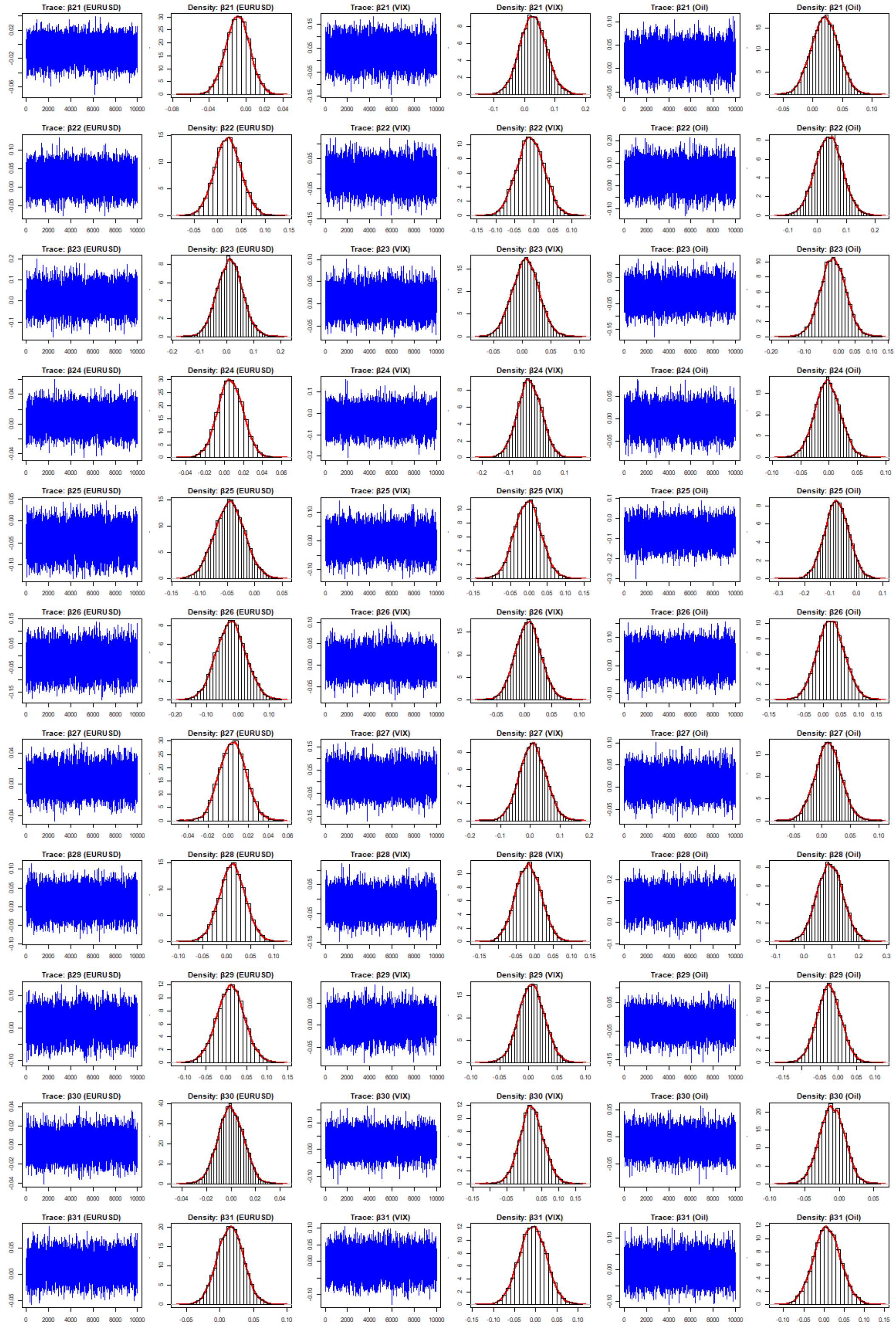

Figure 7: This figure displays trace plots and posterior densities for selected BVAR coefficients (β_{21} to β_{31}) across the EUR/USD, VIX, and Oil equations. The visualizations assess convergence and distributional properties of the Gibbs sampler draws used for posterior inference.