Master's Paper of the Department of Statistics, the University of Chicago

High-Dimensional Time Series: VAR Model Efficacy Across Lag Structures

Yiyuan Zhang

Advisor: Wei Biao Wu

Approved: Wei Biao Wu

Data: Apr 25, 2024

Abstract

The nuanced analysis of high-dimensional time series data holds paramount importance across various fields, including economics, finance, and engineering, demands sophisticated modeling techniques that can accurately capture dynamic interdependencies and forecast future observations. This paper centers on the critical role of high-dimensionality within the realm of Vector Autoregression (VAR) models, emphasizing the pivotal influence of lag order in sculpting the analytical structure of time series data. Acknowledging the complexity and variety of temporal relationships in high-dimensional datasets, this study innovates by analyzing the performance of various established methodologies, specifically, factor adjusted models (DFM and FAVAR), Lasso regression, and Information Criteria, across different lag order structures. Our study embarks on a detailed examination of these methodologies' ability to navigate the challenges posed by high-dimensional time series data, particularly emphasizing the differential impact of various lag order configurations on model performance. By systematically comparing these models across structures, we aim to unveil nuanced insights into each approach's forecasting accuracy and adaptability in capturing intricate data dependencies.

Keywords: High-dimensional time series, vector autoregression, lag order selection, structured sparsity, factor analysis

Contents

1 Introduction

As data-driven strategies become increasingly central to decision-making in areas like economics, finance, engineering, and environmental science, the analysis of high-dimensional time series stands as a critical task. The challenge is not just the volume of data but its complexity, which calls for sophisticated analytical methods. Vector Autoregression (VAR) models have become particularly prominent for their capacity to elucidate the dynamic interconnections within multivariate time series data, providing valuable insights for prediction and policy-making.
The omnipresence of high-dimensional data in real-world applications brings a wealth of information and analytical challenges. Whether it's the interconnected movements in financial markets or the multifaceted variables in environmental studies, the detailed interplay of numerous data points over time necessitates a refined approach to reveal significant patterns and trends. High-dimensionality's importance thus lies in its comprehensive perspective of phenomena, demanding equally advanced analytical techniques to decode its complexities.
The concept of lag order is essential in VAR model deployment, acting as a keystone that dictates the model's structure and effectiveness. By determining the extent to which past observations are used to forecast future events, lag order influences the precision of the model's predictions and the depth of insight into causal interactions. The careful selection of lag order is therefore crucial; it determines whether the model will provide a clear window into the data's dynamics or a distorted view clouded by noise. Hence, the choice of lag order transcends technicality, shaping the model's capacity to interpret and forecast the intricacies of high-dimensional time series data.

Building upon the foundational concept of lag order, these structures follows the framework of Nicholson, Wilms, Bien and Matteson (2020), proceeds to dissect and scrutinize the performance of prominent modeling frameworks under three distinct lag order structures: Componentwise, Elementwise, and Data-based. These structures are carefully chosen to reflect the varying degrees of interdependency and temporal resolution that may exist within a high-dimensional dataset.
In the Componentwise structure, the lag order is determined separately for each variable, allowing for tailored forecasting that can accentuate the unique characteristics and influences of individual data series. The Elementwise structure takes this customization a step further by permitting each variable within a component to have its own lag order, thus refining the model's resolution and potential predictive power. Fi-

nally, the Data-based structure employs empirical data to inform the selection of lag order, grounding the model's temporal framework in observed data patterns.

Within the context of these structured approaches, the performance of factor models—specifically Dynamic Factor Models (DFM) and Factor-Augmented Vector Autoregression (FAVAR)—is analyzed. DFM is evaluated for its adeptness in distilling a wealth of information into fewer factors, potentially unraveling the core dynamics of multivariate datasets. FAVAR, in parallel, is examined for its capacity to enhance traditional VAR models by incorporating factor analysis, aiming to capture the influence of unobserved common factors. These models are renowned for their ability to condense information from high-dimensional datasets into more manageable forms without significant loss of information.

The study contrasts these factor models with the Lasso regression approach, acclaimed for its regularization strength that enforces sparsity, thus mitigating the risk of overfitting in environments saturated with data. Lasso's ability to selectively shrink coefficients and retain only the most relevant predictors is scrutinized, offering insight into its practicality for high-dimensional forecasting.

Furthermore, the study evaluates the efficacy of Information Criteria, the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), which are established tools for model selection, particularly in determining the optimal number of lags. Alongside these methods, the traditional VAR(1) model—often employed for its simplicity and ease of interpretation—is also examined for its ability to perform in high-dimensional contexts.

This comprehensive comparative analysis is designed to do more than identify the superior methods for managing high-dimensional time series data. By identifying the strengths and weaknesses of each modeling approach within the various lag order frameworks, this research contributes to a deeper understanding of time series analysis and offers guidance for model selection in practical applications where predictive accuracy is paramount.

Paper Organization. In Section 2, we introduced three distinct lag order structures, and three approaches based on these, are investigated in Section 3. Subsequently in Section 4, the model's theoretical properties is introduced, followed by formulation of the function. Simulations in Section 5 and applications in Section 6 highlight model's advantages in forecasting and lag order selection.

2 Lag Order Structures

we delve into the background of the three distinct lag order structures that form the foundation of our comparative analysis: Componentwise, Elementwise, and Data-based structures. These structures are crucial as they represent different approaches to integrating lag order within high-dimensional VAR models, each with its implications for capturing the temporal dynamics in the data.

1. Componentwise Structure.
This structure adopts a segmented approach to lag selection, wherein each variable within the time series is allowed to have its own lag order. Within a single equation, all components adhere to the same lag order. This shared lag within each equation ensures a unified temporal depth for all variables, reflecting a balance between individual dynamics and a cohesive model structure. This is particularly useful in cases where the variables are believed to operate at different frequencies or possess unique temporal patterns that do not align synchronously with the rest of the system. The Componentwise structure facilitates the construction of models that can cater to the specific predictive needs of each variable, potentially leading to more accurate and tailored forecasting results. This approach respects the individuality of each series, ensuring that the model does not impose an artificial uniformity that could dilute the distinct signals each variable offers.

2. Elementwise Structure.
Taking the principle of specificity a step further, the Elementwise structure permits each element—every individual lag of each variable—to have a unique order. This granular control allows for an even finer tuning of the model, giving it the flexibility to precisely capture the influence of past values on future outcomes. This structure is particularly beneficial when dealing with datasets where the inter-relationships between the series are complex and the influence of past values is not uniform across time. By allowing each lagged value to be considered on its merit, the Elementwise structure affords a nuanced modeling approach that can accommodate a high degree of heterogeneity within the data.

3. Data-based Structure.
Rather than relying on pre-set criteria or assumptions, the Data-based structure employs empirical evidence from the data itself to determine the lag order. It's neutral. This data-driven approach ensures that the lag structure is intimately connected with

the inherent dynamics present in the dataset. It relies on algorithmic or statistical methods to discern patterns and correlations within the data, using these insights to inform the lag selection process. The strength of this structure lies in its adaptive nature, which can be particularly powerful in rapidly changing environments or when the dataset exhibits non-stationary characteristics. The Data-based structure is designed to evolve as new data become available, potentially offering a dynamic and responsive modeling approach that remains attuned to the unfolding patterns in the data.

3 Approaches Review

3.1 Factor Models - DFM & FAVAR

Factor models are statistical methods used for dimensionality reduction, particularly useful in handling high-dimensional time series data. These models simplify the analysis by capturing the common variations across multiple time series through a few latent factors. The basic assumption is that while the observed variables are numerous, the underlying factors driving them are fewer in number.

The general form of a factor model can be expressed as:

$$y_t = Ax_t + \epsilon_t \tag{1}$$

where x_t is a r x 1 latent process with fixed but unknown dimension $r \leq p$, A is a p x r unknown constant matrix, and ϵ_t is a vector white noise process with mean 0 and covariance matrix $\sum_\epsilon$.

High dimensional data can lead to significant challenges related to overfitting and computational burdens primarily. Instead of dealing with all original variables, factor models identify a smaller number of latent factors that capture the majority of the variance in the dataset. Methods like Principal Component Analysis (PCA) or Singular Value Decomposition (SVD) are used to extract latent factors from the original high-dimensional dataset. This step dramatically reduces the number of variables in the model, thus simplifying the complexity of the VAR model. A VAR model is then applied to these factors.

FAVAR (Factor-Augmented VAR)

A factor-augmented vector autoregressive (FAVAR) model is defined by a VAR equation that captures lead-lag correlations amongst a set of observed variables X and latent factors F :

$$\begin{pmatrix} f_t \\ x_t \end{pmatrix} = A_1 \begin{pmatrix} f_{t-1} \\ x_{t-1} \end{pmatrix} + \cdots + A_p \begin{pmatrix} f_{t-p} \\ x_{t-p} \end{pmatrix} + \epsilon_t, \tag{2}$$

where f_t represents the latent factors at time t, extracted from a large dataset of time series, x_t are the observed variables of particular interest (e.g., GDP growth, interest rates), $A_1, \ldots, A_p$ are matrices of coefficients, ϵ_t is the vector of innovations or shocks at time t.

Coupled with a calibration equation that integrates an additional observed dataset Y with F and X. $Y_t \in \mathbb{R}^q$, which is influenced by both X_t and F_t; that is, the calibration equation is:

$$Y_t = \Lambda F_t + \Gamma X_t + \epsilon_t. \tag{3}$$

The calibration equation is pivotal for deducing the latent factors, which are then employed to calculate the parameters within the VAR setup.
Developed by Bernanke, Boivin, and Eliasz (2005), the FAVAR approach extends the traditional VAR model by incorporating both observed economic indicators and latent factors extracted from a large panel of data. This model is particularly useful in capturing the effects of monetary policy shocks and other macroeconomic dynamics across a wide array of variables.

The Factor-Augmented Vector Autoregression (FAVAR) model, despite its complexity, has found widespread application in empirical economic and finance research. Introduced by Bernanke et al. (2005), the FAVAR model initially included a set of core economic indicators like industrial production, consumer price index, and the federal funds rate. Caggiano et al. (2014) have applied the FAVAR approach under the constraints of a fixed dimensionality assumption. Stock and Watson (2016) utilized the FAVAR model to dissect the macroeconomic effects of oil supply shocks.

DFM (Dynamic Factor Model)

The DFM model aims to decompose a large number of time series into a few common latent factors and idiosyncratic components. The premise is that these common factors are the key drivers of the observed data, which themselves can exhibit temporal dynamics.

The dynamics of the latent factors f_t are typically modeled using a vector autoregressive (VAR) process:

$$x_t = \Phi_1 x_{t-1} + \Phi_2 x_{t-2} + \cdots + \Phi_p x_{t-p} + \epsilon_t \tag{4}$$

where $\Phi_1, \Phi_2, ..., \Phi_p$ are the matrices that capture the relationship between the factors at different lags, and ϵ_t is the error term.

They have been extensively used for forecasting purposes in economics (Stock and Watson, 2002), while their statistical properties have been studied in depth (Bai and Ng, 2008).

3.2 Lasso

VAR estimation is a natural high-dimensional problem, their applicability is curtailed because the quantity of model parameters increases quadratically with each added time series for every lag period considered. This complexity becomes a limitation, especially when the availability of sufficient time points for precise estimation is a challenge in many applications. Nevertheless, there is a recent body of technical work that leveraging structured sparsity and the corresponding regularized estimation framework has established results for consistent estimation of the VAR parameters under high dimensional scaling. Basu and Michailidis (2015) investigated Lasso-penalized Gaussian VAR models and established consistency results. They also provided valuable technical instruments for the analysis of sparse models that deal with temporally dependent data. Furthermore, Melnyk and Banerjee (2016) expanded these findings to include additional regularizers.

Lasso (Least Absolute Shrinkage and Selection Operator) is a regression analysis method that performs both variable selection and regularization in order to enhance the prediction accuracy and interpretability of the statistical model it produces. In the context of time series analysis and particularly when dealing with high-dimensional data, Lasso can be very useful for identifying the most relevant variables and avoiding overfitting.

When constructing a time series model, especially in a high-dimensional setting where the number of variables can exceed the number of observations, it's crucial to select a

subset of variables that are truly predictive. This is where Lasso comes in, by imposing a penalty on the size of the coefficients:

$$\hat{\beta}^{lasso} = \underset{\beta}{\operatorname{argmin}} \left\{ \frac{1}{2T} \sum_{t=1}^{T} (y_t - X_t\beta)^2 + \lambda \|\beta\|_1 \right\}, \tag{5}$$

where: $\hat{\beta}^{lasso}$ is the vector of estimated coefficients after applying the Lasso. y_t is the observed time series data at time t. X_t is the matrix of potential predictive variables at time t. λ is the tuning parameter that controls the strength of the penalty imposed on the coefficients. $\|\beta\|_1$ represents the L1 norm of β, which is the sum of the absolute values of the coefficients.

The key aspect of Lasso is the selection of the tuning parameter λ. A larger λ will shrink more coefficients to zero, resulting in a sparser model. The choice of λ can be determined through cross-validation, typically aiming to minimize forecast error or another loss function relevant to the specific context.

The Lasso methodology provides a powerful tool for modeling in high-dimensional spaces, where traditional regression techniques would falter due to overfitting and the curse of dimensionality. By incorporating Lasso into the time series analysis, one can construct more parsimonious models that retain only the most informative predictors, leading to better interpretability and predictive power.

3.3 Information Criteria

In time series analysis, the Vector Autoregression (VAR) model is a fundamental tool for modeling the dynamic relationships between multiple interrelated time series. However, a critical issue in VAR modeling is the determination of the appropriate lag order p. Selecting too few lags can result in omitted variable bias, whereas too many lags can lead to overfitting and unnecessary complexity. Information criteria address this problem by providing a formal approach to selecting a model that appropriately balances model fit and complexity.

The most commonly used information criteria in the context of VAR models are the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), with each having different properties and implications for model selection.

4 Methodology

Let $\{y_t \in \mathbb{R}^k\}_{t=1}^T$ denote a k-dimensional vector time series of length T. A pth order vector autoregression $VAR_k(p)$ may be expressed as a multivariate regression

$$y_t = A^{(1)}y_{t-1} + \cdots + A^{(p)}y_{t-p} + u_t, \quad \text{for } t = 1, \ldots, T, \tag{6}$$

conditional on initial values $\{y_{-(p-1)}, \ldots, y_0\}$, where $\nu \in \mathbb{R}^k$ denotes an intercept vector, $\{A^{(\ell)} \in \mathbb{R}^{k\times k}\}_{\ell=1}^p$ are lag-ℓ coefficient matrices, and $\{u_t \in \mathbb{R}^k\}_{t=1}^T$ is a mean zero white noise vector time series with unspecified $k \times k$ nonsingular contemporaneous covariance matrix Σ_u.

Estimation Procedure.
Based on the k-dimensional time series data in VAR model of order p in Equation(4), we seek to estimate the parameters that minimize the residuals of the model. We express the VAR model in matrix notation as follows:

$$\mathbf{Y} = \mathbf{1}\boldsymbol{\nu}^\top + \mathbf{\Phi}\mathbf{Z} + \mathbf{U}, \tag{7}$$

where:

- $\mathbf{Y} = [y_1 \ldots y_T]^\top$ is the matrix of observed time series,
- $\mathbf{1}$ is a $T \times 1$ column vector of ones,
- $\boldsymbol{\nu}$ is the $k \times 1$ vector of intercepts,
- $\mathbf{\Phi} = [\mathbf{\Phi}^{(1)} \ldots \mathbf{\Phi}^{(p)}]$ is the $k \times kp$ matrix of autoregressive coefficients,
- $\mathbf{Z} = [\mathbf{Z}_1 \ldots \mathbf{Z}_T]^\top$ with $\mathbf{Z}_t = [y_{t-1}^\top \ldots y_{t-p}^\top]^\top$ is the matrix of lagged values,
- $\mathbf{U} = [u_1 \ldots u_T]^\top$ is the matrix of error terms.

The objective function for the least squares estimation can be expressed as:

$$\min_{\boldsymbol{\nu}, \mathbf{\Phi}} \sum_{t=1}^{T} \|\mathbf{y}_t - \boldsymbol{\nu} - \mathbf{\Phi}\mathbf{z}_t\|_2^2, \tag{8}$$

where $\|\cdot\|_2$ denotes the Euclidean norm.

To solve for the autoregressive coefficients $\mathbf{\Phi}$, we vectorize the system and employ the ordinary least squares procedure, leveraging the Kronecker product properties:

$$\text{vec}(\mathbf{Y}) = (\mathbf{I}_k \otimes \mathbf{1})\text{vec}(\boldsymbol{\nu}) + (\mathbf{I}_k \otimes \mathbf{Z})\text{vec}(\mathbf{\Phi}) + \text{vec}(\mathbf{U}), \tag{9}$$

where $\otimes$ denotes the Kronecker product, and $\mathbf{I}_k$ is the $k \times k$ identity matrix. The formulation facilitates the use of linear regression techniques to estimate the VAR model parameters in a high-dimensional setting.

4.1 Factor Models

The integration of factor models into the Vector Autoregression (VAR) framework allows for a significant reduction in the dimensionality of high-dimensional time series data. In this methodology, we extract common factors from a large dataset of time series, which are then used as inputs into the VAR model. This procedure efficiently captures the co-movements among the time series, reducing the problem's complexity.

4.1.1 Factor Extraction

The first step in our methodology is the extraction of common factors. Given a dataset of k time series $\{X_t\}$ of length T, we assume that each observation can be represented as:

$$X_t = \Lambda F_t + \epsilon_t, \tag{10}$$

where X_t is the vector of observed variables, Λ is a matrix of factor loadings, F_t represents the vector of common factors, and ϵ_t denotes idiosyncratic noise.

4.1.2 VAR Modeling with Factors

After extracting the factors, we model their dynamics in a state-space form to incorporate both the observed time series and the latent factors, see calibration equation (2).

The representation allows for the modeling of time series dynamics and latent factor evolution simultaneously.

$$Y_t = \Lambda F_t + \Gamma X_t + \epsilon_t. \tag{11}$$

where the parameters Λ and Γ are identified by imposing restrictions, such as setting the upper $p_1 \times p_1$ block of Λ to the identity matrix and constraining the first p_1 rows of Γ to zero. These restrictions are critical to ensuring the unique identification of the model, enabling the latent factors F_t to be consistently estimated and used in forecasting and economic analysis. Moreover, identification strategies like these are essential in high-dimensional settings, where the dimensionality of X_t and F_t can complicate the estimation process. Lin and Michailidis(2020).

4.1.3 Consistency & Asymptotic Normality

We have established that common factors can be consistently estimated and are asymptotically normal. We have derived corollaries concerning the consistency and asymptotic distribution of the common factors, based on the estimation methods developed by Lam et al. (2011), Lam and Yao (2012) and Chan er al.(2016).

Corollary 1.1. For consistency, the factor model estimates have the following convergence rates:
(1) The estimated factor loadings matrix $\hat{\Lambda}$ is consistent for Λ, with convergence rate given by:

$$\|\hat{\Lambda} - \Lambda\| = O_k\left(k^{-1/2}\right).$$

(2) The estimated common factors $\hat{f}_t$ are consistent for the true factors f_t, with convergence rate given by:

$$\|\hat{f}_t - f_t\| = O_k\left(k^{-1/2}\right).$$

Remark 1.1. Note that although T and p go to infinity together, the convergence rate is independent of p. This phenomenon is called blessing of dimensionality.

Corollary 1.2. For asymptotic normality, as T, p goes to infinity, then for each t: For the normalized factor $\hat{f}_t^0$, we have:

$$\sqrt{k}(\hat{f}_t^0 - f_t^0) \xrightarrow{d} N(0, Avar(\hat{f}_t^0)),$$

where $\text{Avar}(\hat{f}_t^0) = H\Gamma H^T$ is the asymptotic variance matrix of $\hat{f}_t^0$.

4.1.4 Estimation

The estimation process involves two main steps: the extraction of latent factors and the estimation of the VAR model incorporating these factors.

The parameters of the factor models are estimated by maximizing the likelihood function, which is constructed based on the assumption that ϵ_t and u_t are normally distributed. The estimation can be performed using the Expectation-Maximization (EM) algorithm or Bayesian estimation methods, depending on the specificities of the dataset and the model.

Expectation-Maximization (EM) Algorithm.
The EM algorithm is particularly well-suited for scenarios where the model parameters cannot be observed directly or are difficult to estimate using standard techniques. It alternates between estimating the missing data given the parameters (E-step) and optimizing the parameters given the estimated data (M-step), thus iterating to convergence.

Bayesian Estimation.
Bayesian estimation methods are employed when incorporating prior beliefs about the parameters or when the model is complex relative to the amount of data. We specify prior distributions for the parameters and then update these priors with the observed data to obtain the posterior distributions, which are used for inference.

In summary, this methodology enables the handling of high-dimensional time series data by integrating factor analysis and VAR modeling, which reduces dimensionality and captures the dynamics of common movements within the data. The theoretical properties ensure that the model is identifiable, the estimators are consistent, and the chosen model is the most suitable given the data.

4.2 Lasso

Lasso regression is employed to impose sparsity by penalizing the absolute size of the coefficients.

4.2.1 Formulation of Lasso-VAR

The Lasso-VAR model incorporates the Lasso penalty into the traditional VAR framework, see Equation (5). For a VAR model of order p, the Lasso-VAR estimation problem can be formulated as:

$$\hat{B}^{lasso} = \underset{B}{\operatorname{argmin}} \left\{ \frac{1}{2T} \sum_{t=1}^{T} (Y_t - \sum_{i=1}^{p} B_i Y_{t-i})^2 + \lambda \sum_{i=1}^{p} \|B_i\|_1 \right\}, \tag{12}$$

where Y_t is the vector of time series data at time t, B_i is the matrix of autoregressive coefficients at lag i, λ is the regularization parameter, and $\|B_i\|_1$ denotes the L1 norm of the matrix B_i.

Implementation and Optimization.
The practical implementation of the Lasso-VAR involves the use of efficient optimization algorithms such as coordinate descent. The objective function, consisting of a least squares loss and the L1 penalty term, is iteratively minimized to obtain sparse coefficient estimates.

$$\hat{\beta}^{lasso} =_{\beta \in \mathbb{R}^q} \left\{ -2\beta' \hat{\gamma} + \beta' \hat{\Gamma} \beta + \lambda \|\beta\|_1 \right\}, \tag{13}$$

where $\hat{\gamma} = (W \otimes Z')Y/N$ and $\hat{\Gamma} = (W \otimes Z')Z/N$ with W being a weight matrix, typically set to the identity matrix or the inverse of the error covariance matrix for weighted Lasso.

4.2.2 Consistency

The Lasso estimator's theoretical properties, particularly its consistency and selection ability, rely on specific conditions. The choice of λ is crucial, and various methods exist to select an appropriate λ, such as information criteria or cross-validation.

Underpinning the practical implementation are key theoretical properties, which guarantee the consistency and efficiency of the Lasso-VAR estimator. These properties are derived under a general penalized M-estimation framework and include:

1. *Consistency:* For consistent estimation in high-dimensional settings, we require that the design matrix satisfies certain regularity conditions, such as the restricted eigenvalue condition. This ensures that the Lasso estimator does not suffer from multicollinearity among a large number of predictors.

2. *Asymptotic normality:* Assuming that the error terms are independently and identically distributed, the Lasso estimator is asymptotically normal. This allows for valid statistical inference about the effects of the selected predictors.

These conditions ensure that the Lasso can recover the correct sparsity pattern with high probability as the sample size grows. Basu and Michailidis (2015) results show that consistent estimation of β^* is possible, as long as the predictor and noise processes are stable.

4.2.3 Tuning Parameter Selection

The tuning parameter λ controls the trade-off between the model's sparsity and its fit to the data. We use cross-validation to select λ, minimizing a loss function such as the Mean Squared Prediction Error (MSPE):

$$\lambda_{opt} = \underset{\lambda}{\operatorname{argmin}} \left\{ \text{MSPE}(\lambda) \right\}. \tag{14}$$

Cross-validation, information criteriam, stability selection method and gradient descent are among the methods used to select an optimal λ, which is critical for the estimator to enjoy good theoretical properties.

In summary, the Lasso methodology within a high-dimensional VAR framework addresses the challenges of overfitting and enables the construction of parsimonious models that retain only the most significant predictors. The integration of Lasso with VAR models allows for efficient estimation and improved forecast accuracy in the presence of a large number of time series variables.

4.3 Information Criteria

To implement information criteria for VAR model selection, we first estimate the VAR model for a range of lag orders and then calculate the information criterion for each

model. The lag order that minimizes the chosen information criterion is selected as the optimal lag length for the VAR model.

4.3.1 Akaike Information Criterion (AIC)

The AIC is derived from information theory and is defined for a VAR model as:

$$\text{AIC}(p) = \ln |\hat{\Sigma}(p)| + 2 \times \frac{kp^2}{T}, \tag{15}$$

where $\hat{\Sigma}(p)$ is the estimated covariance matrix of the residuals from the VAR model with lag order p, k is the number of variables in the VAR model, and T is the sample size. The term $2 \times kp^2/T$ penalizes the complexity of the model.

4.3.2 Bayesian Information Criterion (BIC)

The BIC is similar to the AIC but imposes a heavier penalty for the number of parameters, and is defined as:

$$\text{BIC}(p) = \ln |\hat{\Sigma}(p)| + \ln(T) \times \frac{kp^2}{T}. \tag{16}$$

The BIC is particularly useful in high-dimensional settings as it tends to select models with fewer parameters, which can be advantageous when dealing with limited observations.

4.3.3 Implementation Procedure

The procedure for implementing information criteria in the selection of VAR models includes the following steps:

1. Estimate the VAR model for a pre-specified range of lag orders, say $p = 1, 2, \ldots, P_{\max}$.
2. For each lag order, calculate the AIC and BIC values using the formulae provided.

3. Identify the model with the lowest AIC and BIC values. These represent the optimal models according to each criterion.

4. Perform diagnostic checks to ensure the adequacy of the selected models, such as testing for serial correlation in the residuals.

In summary, information criteria play a critical role in the empirical application of VAR models, guiding the selection of an appropriate model structure. By penalizing excessive complexity, they help to prevent overfitting and ensure that the selected model is well-suited for both forecasting and causal inference in multivariate time series analysis.

5 Simulation Study

The simulation study is designed to assess the forecasting capabilities of a suite of time series models, including the Dynamic Factor Model (DFM), Factor Augmented VAR (FAVAR), Lasso-VAR, AIC-VAR, and BIC-VAR. Our evaluation focuses on different lag structures within the VAR framework to determine the efficacy of these models in various simulated environments.

5.1 Simulation Scenarios

We have developed three distinct simulation scenarios to test our proposed methods, each with its own specific lag structure and data-generating process (DGP).

- In Scenarios 1 and 2, we simulate data with $k = 45$ components over a series length of $T = 100$. The scenarios are distinguished by their lag structures: the first scenario features a componentwise lag structure, while the second exhibits an elementwise pattern. These scenarios are particularly insightful for situations where the number of parameters is substantial relative to the sample size.

- Scenario 3 employs a robust DGP with $k = 40$ and $T = 100$, incorporating a comprehensive approach to ensure the reliability of the simulations. This scenario employs a "data-based Monte Carlo method" for DGP, avoiding any particular lag structure preference and assigning $p = 4$ non-zero lags across all variables.

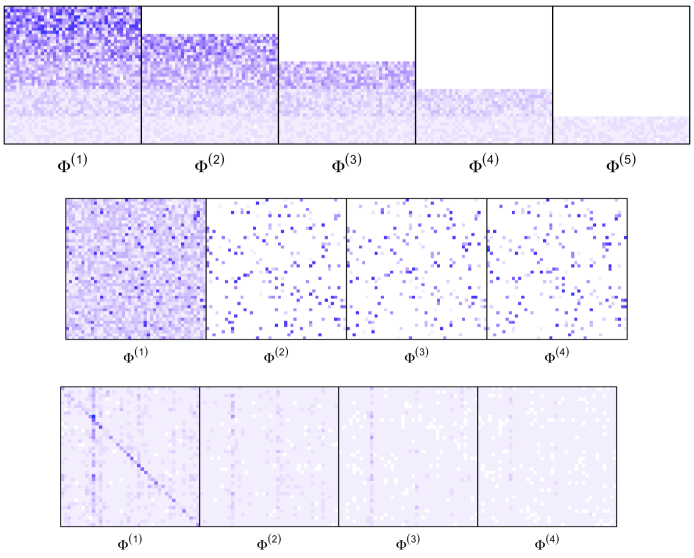

Figure 1: Sparsity patterns and magnitudes of the simulation scenarios. Darker shading indicates coefficients that are larger in magnitude.

Each simulated dataset is generated from stationary coefficient matrices. Detailed steps are taken to ensure the stationarity of each model's structure. The error covariance for each scenario is standardized, with $\Sigma_u = 0.01 \cdot I_k$, where I_k is the identity matrix of dimension k. We run 250 simulations in each scenarios.

Sparsity Patterns and Magnitudes

Sparsity patterns refer to which coefficients are non-zero and their relative magnitudes,reflecting the degree of influence of different time lag values on the current value. These patterns are a visual representation of the model's structure, where typically a heatmap or matrix is used to illustrate the presence and strength of the relationships between different variables across various lags.Therefore, it is particularly important to select the most informative lag term by applying a certain degree of regularization (such as Lasso method). Our model is also designed to discover and utilize sparsity, to keep the model as concise as possible and only include the most influential lagged terms. The coefficient matrices used in these scenarios are depicted in Figure 1.

	Method	Scenarios 1: Componentwise	Scenarios 2: Elementwise	Scenarios 3: Data-based
Factor	FAVAR	0.05387	0.03901	0.5634
	DFM	0.05433	0.04091	0.6308
VAR	Lasso-VAR	0.03448	0.01781	0.4663
	AIC	0.2552	0.2651	1.1630
	BIC	0.2552	0.2651	0.6212

Table 1: Out-of-sample mean squared forecast error for VARs in Scenario 1 to 3

5.2 Performance Measurement

Our main interest lies in point forecast accuracy. For each generated time series, we fit our proposed VAR model, and possibly alternative models, to the data up to time T. We then forecast the value at time $T+1$ and compare the forecast $\hat{y}_{i,t+1}$ against the true value $y_{i,t+1}$. This process is repeated for each time point within the evaluation period, and the one-step-ahead mean squared forecast error is computed as follows:

$$MSFE(T_1, T_2) = \frac{1}{k(T_2 - T_1)} \sum_{i=1}^{k} \sum_{t=T_1}^{T_2-1} (\hat{y}_{i,t+1} - y_{i,t+1})^2, \tag{17}$$

where $\hat{y}_{i,t+1}$ represents the forecast for time $t+1$ and component i, and k is the number of time series components.

Table 1 gives the forecast performance of the methods in Scenarios 1-3. Concerning the VAR-based methods, we report the results for known (p = 5 in Scenario 1 and p = 4 in Scenario 2-3) maximal lag order.

Scenario 1: Componentwise HLag.
In the componentwise scenario, the Lasso-VAR method yields the lowest mean-squared forecast error (MSFE), indicating its superior performance in scenarios where the true model is likely to have a sparse structure. The FAVAR and DFM methods show similar performance, with FAVAR slightly outperforming DFM. The AIC and BIC methods display considerably higher MSFEs, which suggests that these criteria may be selecting models that are either too complex (AIC) or too simplified (BIC), resulting in poorer out-of-sample forecast performance.

Scenario 2: Elementwise.
The elementwise structure seems to benefit significantly from the Lasso-VAR method, which again has the lowest MSFE. Its performance is even more pronounced than in Scenario 1, highlighting Lasso's effectiveness in situations with more complex inter-variable relationships and its ability to enforce sparsity where appropriate. FAVAR and DFM also perform well, with slightly higher MSFEs than Lasso-VAR. Their consistent performance across the first two scenarios indicates that factor models are generally robust and capable of handling various types of lag structures effectively. As in Scenario 1, AIC and BIC have higher errors, and their similarity in values may suggest that both information criteria lead to similar model choices, which not be optimal for elementwise structured scenarios.

Scenario 3: Data-based.
The data-based scenario presents a more challenging environment, likely due to the lack of a specific lag structure, and we observe an overall increase in MSFE for all methods. However, the increase is most dramatic for AIC-VAR, which suggests that AIC may be overfitting the model in this more complex scenario. BIC-VAR shows improvement over AIC-VAR, aligning with BIC's tendency to choose simpler models, which might be preventing overfitting to some extent. Lasso-VAR, while also experiencing a higher MSFE compared to the previous scenarios, maintains its position as the method with the lowest error, reinforcing the strength of Lasso in high-dimensional settings. FAVAR and DFM show an increase in MSFE compared to the previous scenarios. Though all the true parameters are non-zero, many of the coefficients are small and contribute less to the dynamics of the system, as in Figure 1, panel (3).

The essence of the model interaction with the DGP lies in the model's ability to discern and appropriately respond to the underlying structure of the time series. Factor models like FAVAR and DFM excel in situations where there are common movements shared across multiple series, which they can capture and use to make informed forecasts. But these approaches may not be optimal when the series' interactions are not factor-driven but are component-specific instead. Their performance, while not poor, isn't as competitive compared to models that are designed or tuned to accommodate component-specific behavior—something that Lasso-VAR, with its penalty-driven sparsity, seems to do more adeptly.

On the other hand, methods that enforce sparsity through penalization, such as Lasso-

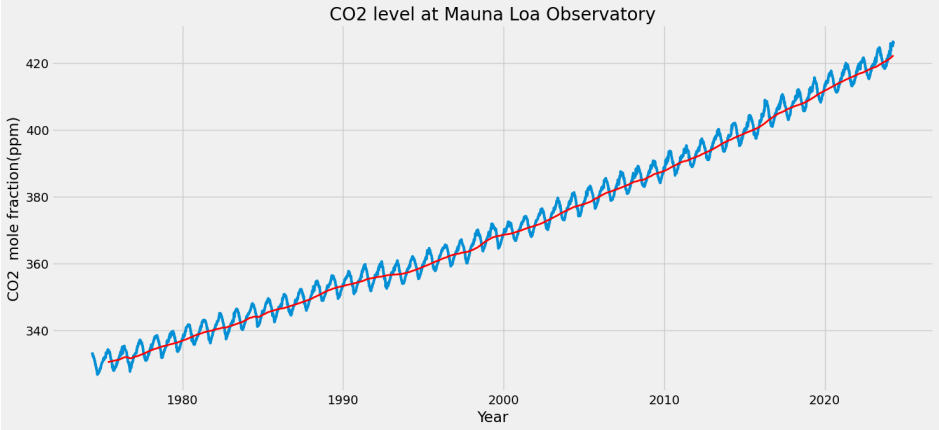

Figure 2: Atmospheric CO2 observed at Mauna Loa.

VAR, demonstrate flexibility and strength in both sparse and non-sparse environments. They can adaptively select the most significant lags and components that drive the system, avoiding unnecessary complexity that could degrade the forecast accuracy.

6 Real Data Analysis

6.1 Data Structure

This study utilizes the longest continuous record of atmospheric carbon dioxide (CO2) measurements, which began in March 1958 by C. David Keeling of the Scripps Institution of Oceanography. The National Oceanic and Atmospheric Administration (NOAA) has maintained this dataset since May 1974, providing an unparalleled resource for monitoring the global trend of atmospheric CO2 levels [Keeling, 1976; Thoning, 1989]. The data represents the dry air mole fraction of CO2 in parts per million (ppm), which is the ratio of CO2 molecules to all other molecules in air, excluding water vapor.

Due to the careful selection process intended to represent "background" conditions, the daily mean values are filtered to remove influences from local emissions and volcanic activity. Data from the last year remain preliminary and are subject to pending recalibrations and quality checks. Notably, measurements were suspended during the recent eruption of the Mauna Loa Volcano from November 29, 2022, to July 4, 2023.

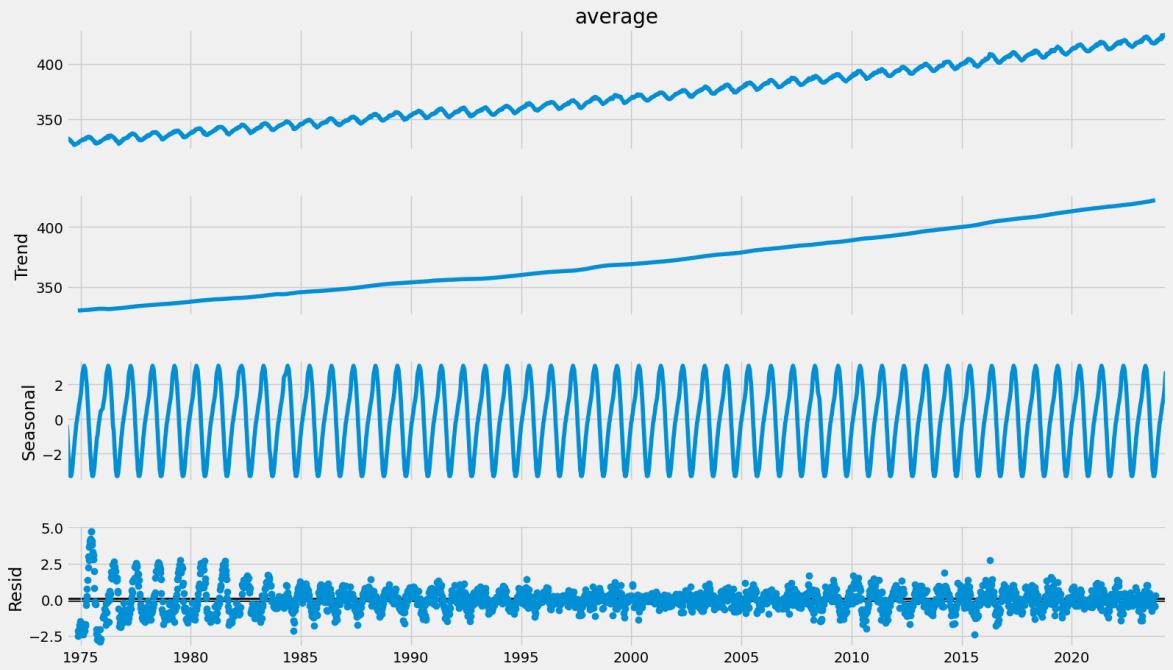

Figure 3: Decomposition

In light of this, a temporary observatory provided continuity for the observations.

The full record of combined Scripps data and NOAA data is shown in the Figure 2. The weekly mean (blue line) is simply the average of all days in the week (Sunday to Saturday) for which a background value could be defined. The red line extracted using a rolling mean with a window of 52 weeks, indicates a persistent increase in atmospheric CO2 levels. Commencing in 1974 with a mole fraction of 333.37 parts per million (ppm), the levels have escalated to 426.34 ppm as of mid-April 2024, encapsulating an alarming uptick in this marker of human impact on the planet's climate system.

The decomposition of the Mauna Loa CO2 data provides a detailed look into the underlying components of the time series, see Figure 3. The graph represents a broader period from 1974 onwards.

- *Trend.* It shows a clear upward trend, consistent with the overall increase in atmospheric CO2 levels discussed previously. The trend line slopes upwards, reflecting the ongoing accumulation of CO2 in the atmosphere. This trend aligns

with the increasing emissions from fossil fuels and deforestation.

- *Seasonality.* The seasonal component captures the cyclical pattern observed within each year. This oscillation is primarily due to the annual cycle of plant growth and decay, with photosynthesis drawing down CO2 levels during the spring and summer growing seasons and respiration from decomposition releasing CO2 during the fall and winter. The regularity and amplitude of the seasonal fluctuations are consistent over time, with no significant changes in seasonality observed between the graph. This periodicity underlines the Earth's biological processes' role in modulating CO2 levels.

- *Residuals.* The residuals, which represent the random and irregular fluctuations after accounting for the trend and seasonality, appear relatively stable and centered around zero. There are no obvious patterns or trends in the residuals, suggesting that the decomposition has successfully captured the main systematic components of the time series.

The decomposed components align with the larger trends discussed in previous sections. The upward trend in the data corresponds with the long-term increase in CO2 levels due to anthropogenic activities. The seasonality matches the known biological and ecological processes that affect CO2 levels on a regular basis. The residuals do not show any additional patterns, which indicates that the model has effectively captured the behavior of the CO2 time series.

6.2 Forecast Comparisons

The forecast performance of the various models—DFM, FAVAR, Lasso-VAR, AIC-VAR, and BIC-VAR—will be evaluated using the mean squared forecast error (MSFE). Each model will be fit to the data up to the last available observation, and the final data point will be used to compute one-step-ahead forecast errors. This methodology offers a direct comparison of each model's predictive accuracy in the context of long-term environmental data characterized by a clear upward trend and periodic fluctuations.

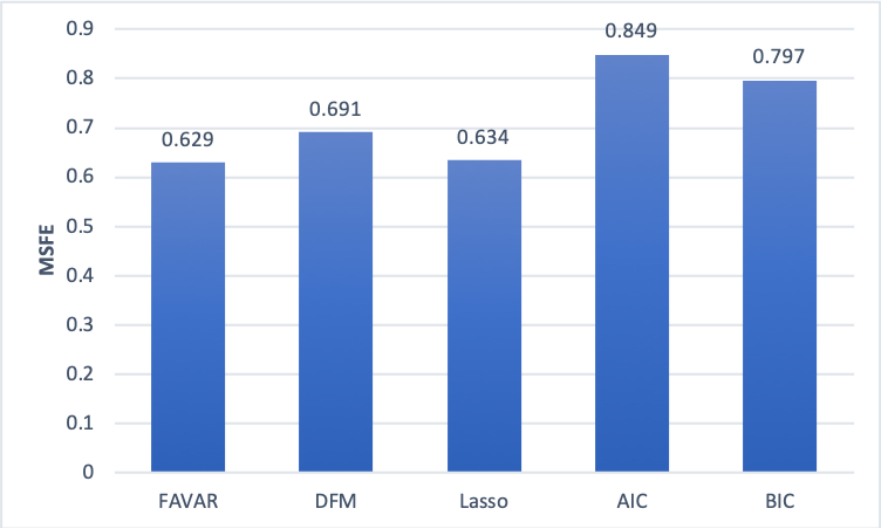

Figure 4: Rolling out-of-sample one-step ahead mean squared forecast error of CO_2 *levels*

For VAR models, and most time series models, it's important to work with stationary data. The ADF (Augmented Dickey-Fuller) test results show that the p-value is 0.00, suggesting that the series is stationary, as we expected.

Dynamic Factor Model (DFM). Given the clear trend and seasonality in the CO2 data, DFM could perform well by capturing these common factors across multiple related time series. However, if there are non-linearities or complex dynamics that a basic DFM does not capture, this could affect its performance.

Factor-Augmented Vector Autoregression (FAVAR). This model might perform better than DFM if additional economic or environmental indicators (observable factors) provide useful information that is not captured solely by the past values of CO2 levels.

Lasso-VAR (Lasso-Vector Autoregression). If the underlying data has many predictors, Lasso-VAR can be beneficial due to its variable selection capabilities, which might result in a model that avoids overfitting and captures the most relevant dependencies.

AIC-VAR (Akaike Information Criterion VAR). AIC tends to choose models with more parameters, which might capture more of the complexity in the data. However, it also has a higher risk of overfitting, especially with a limited amount of data.

BIC-VAR (Bayesian Information Criterion VAR). BIC selects models with fewer parameters and can prevent overfitting better than AIC. However, it may underfit if the

true model is complex.

6.3 Discussion

We can see slightly difference in our simulation results and real data results.

In the simulations, it seems that the performance of the models depended heavily on whether the time series data exhibited common movements that could be captured by factors (as in FAVAR and DFM), or whether the time series' interactions were more component-specific, in which case methods enforcing sparsity like Lasso-VAR performed better.

When looking at the real CO2 data from Mauna Loa:
The CO2 data has a very clear upward trend and a strong seasonal component. Any model that does not account for these characteristics would likely perform poorly.
DFM and FAVAR models might capture the trend and seasonality effectively once they can model the underlying factors driving the CO2 levels. However, if the common factors are not the primary drivers behind the changes in CO2 levels, or if the factor models do not include variables that account for anthropogenic emissions or other specific environmental factors, their performance may be limited.
Given Lasso-VAR's ability to impose sparsity and select relevant lags, Lasso-VAR might outperform DFM and FAVAR in the real data scenario. Since CO2 emissions are influenced by a combination of highly specific factors such as economic activity, policy changes, and environmental regulations, the capacity to pinpoint these influences can be critical for forecasting. Nevertheless, we observed no unexpected changes throughout the data timeframe.
These criteria-based models choose lags based on information theoretic criteria rather than capturing underlying factors. They may either overfit (AIC) or underfit (BIC), especially with complex, non-linear time series data like CO2 levels. The traditional VAR models using AIC and BIC may not strike the right balance between capturing all necessary components and avoiding overfitting, particularly given the clear trend and seasonality in the CO2 data.

We might consider certain factors as influencers to determine model performance from real data.

- Data Complexity: Real-world data often contains more noise and complexity than simulated data, which can include non-linearities and abrupt changes not captured in simulations.
- Model Assumptions: Simulated data typically conforms more closely to model assumptions, as the Data Generating Process (DGP) is known and controlled. In contrast, real-world data may violate these assumptions.
- External Factors: Real data for CO2 levels is influenced by external factors not present in simulations, such as policy changes, economic growth, technological advancement, and natural events.

The real data analysis aims to contribute to a deeper understanding of the temporal dynamics of CO2 levels and provide a rigorous examination of the selected time series models' forecasting abilities. The results will not only shed light on the effectiveness of each method but also on their practical implications in the context of environmental monitoring and trend analysis.

Reference

Adamek, R., Smeekes, S. and Wilms, I. (2023). Lasso inference for high-dimensional time series. Journal of Econometrics, 235(2), pp.1114–1143.

Bai, J. and Ng, S. (2002). Determining the Number of Factors in Approximate Factor Models. Econometrica, 70(1), pp.191–221.

Barigozzi, M., Cho, H. and Owens, D. (2023). FNETS: Factor-Adjusted Network Estimation and Forecasting for High-Dimensional Time Series. Journal of Business & Economic Statistics, pp.1–13.

Basu, S. and Michailidis, G. (2015). Regularized estimation in sparse high-dimensional time series models. The Annals of Statistics, 43(4).

Chan, N.H., Lu, Y. and Yau, C.Y. (2016). Factor Modelling for High-Dimensional Time Series: Inference and Model Selection. Journal of Time Series Analysis, 38(2), pp.285–307.

Chen, X., Xu, M. and Wu, W.B. (2013). Covariance and precision matrix estimation for high-dimensional time series. The Annals of Statistics, 41(6).

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2000). The Generalized Dynamic-Factor Model: Identification and Estimation. Review of Economics and Statistics, 82(4), pp.540–554.

Gao, Z. and Tsay, R.S. (2021). Modeling High-Dimensional Time Series: A Factor Model With Dynamically Dependent Factors and Diverging Eigenvalues. Journal of the American Statistical Association, 117(539), pp.1398–1414.

Han, Z., Zhao, J., Leung, H., Ma, K.F. and Wang, W. (2021). A Review of Deep Learning Models for Time Series Prediction. IEEE Sensors Journal, 21(6), pp.7833–7848.

Lam, C. and Yao, Q. (2012). Factor modeling for high-dimensional time series: Inference for the number of factors. The Annals of Statistics, 40(2).

Lin, J. and Michailidis, G. (2018). Regularized Estimation of High-dimensional Factor-Augmented Vector Autoregressive (FAVAR) Models. SSRN Electronic Journal.

Nicholson, W., Bien, J., Matteson, D. and Nicholson, M. (2020). High Dimensional Forecasting via Interpretable Vector Autoregression. Journal of Machine Learning Research, 21, pp.1–52.

Rorni, M. and Reichlin, L. (1996). Dynamic common factors in large cross-sections. Empirical Economics, 21(1), pp.27–42.

Stock, J., Watson, M., Boivin, J., Eliasz, P., Evans, C., Morley, J., Nason, J., Ng, S., Rudebusch, G., Shapiro, M., Sheng, X. and Sims, C. (2005). IMPLICATIONS OF DYNAMIC FACTOR MODELS FOR VAR ANALYSIS.

Stock, J.H. and Watson, M.W. (2012). Dynamic Factor Models. The Oxford Handbook of Economic Forecasting, pp.35–60.