International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

IJIEMR Transactions, online available on 7th July 2024. Link

https://ijiemr.org/downloads.php?vol=Volume-13& issue=issue07

DOI: 10.48047/IJIEMR/V13/ISSUE 07/17

Title A Theoretical Approach to Parameter Estimation in Multivariate Time Series VAR Models

Volume 13, ISSUE 07, Pages: 156 – 162

Paper Authors
K. Murali, K. Sreenivasulu, M.Vasudeva Reddy, S. Hariprasad, N. Ramachandra

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

A Theoretical Approach to Parameter Estimation in Multivariate Time Series VAR Models

K. Murali[1], K. Sreenivasulu[2], M.Vasudeva Reddy[3], S. Hariprasad[4], N. Ramachandra[5]**
[1,5]Department of Statistics, S V University, Tirupati.
[2]Associate Professor, D.K. Govt. College for Women(A), Nellore.
[3]Lecturer in Statistics, S.G.S Arts College,TTD, Tirupati.
[4]Lecturer in Statistics, P.V.K.N Govt. College (A), Chittoor.
[**]Corresponding Author: **dr.murali4stat@gmail.com**

Abstract
Building a multivariate time series model involves five key steps: identification, specification, estimation and hypothesis testing, diagnostic checking, and forecasting. Estimating the parameters of a multivariate Vector Autoregressive (VAR) model is more complex than for univariate autoregressive models. When the error terms follow a normal distribution, parameter estimation can be performed using Maximum Likelihood Estimation (MLE), and the Likelihood Ratio (LR) test can be applied for hypothesis testing in VAR models.

This research article focuses on estimating the parameters of a multivariate VAR model using the MLE method based on Ordinary Least Squares (OLS) regression. The dispersion matrix of errors is estimated using Internally Studentized residuals. Furthermore, a test procedure has been developed for determining the optimal number of lags in a multivariate VAR model using the Likelihood Ratio test. The proposed methodology enhances the accuracy and reliability of multivariate VAR model estimation, making it a valuable approach for time series analysis.

Keywords: Vector Autoregressive (VAR), Maximum Likelihood Estimation (MLE), Likelihood Ratio (LR), Ordinary Least Squares (OLS) regression

Introduction
Multivariate time series modelling plays a crucial role in understanding the dynamic relationships among multiple interdependent variables over time. One of the most widely used models for analysing multivariate time series data is the Vector Autoregressive (VAR) model, which extends the univariate autoregressive model to capture dependencies across multiple variables. The VAR model is extensively applied in economics, finance, and various other domains to analyze and forecast complex time-dependent data.

Building a multivariate time series model involves five fundamental steps: identification, specification, estimation and hypothesis testing, diagnostic checking, and forecasting. Among these, parameter estimation is a critical and complex task, particularly in the context of VAR models,

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

due to the presence of multiple interrelated variables. Unlike univariate autoregressive models, estimating the parameters of a multivariate VAR model requires sophisticated statistical techniques to ensure accuracy and reliability.

Under the assumption of normally distributed errors, Maximum Likelihood Estimation (MLE) is a widely used approach for parameter estimation in VAR models. Additionally, the Likelihood Ratio (LR) test is an essential statistical tool for hypothesis testing, particularly in determining the appropriate lag structure of the VAR model. The optimal number of lags is crucial for capturing the true underlying relationships among the variables while avoiding overfitting or underfitting.

In this study, we estimate the parameters of the multivariate VAR model using the MLE approach based on Ordinary Least Squares (OLS) regression. The dispersion matrix of errors is estimated using Internally Studentized residuals, providing a robust framework for evaluating the model's accuracy. Furthermore, we propose a test procedure to determine the optimal number of lags in a multivariate VAR model using the Likelihood Ratio test. The findings of this study contribute to improving the efficiency and reliability of multivariate time series modeling, making it a valuable tool for forecasting and decision-making in various fields.

Vector Autoregression (VAR) Models

Introduction

Vector Autoregression (VAR) models have emerged as a fundamental statistical tool for analyzing multivariate time series data. Originally introduced by Sims (1980), the VAR model extends the concept of univariate autoregressive models by incorporating multiple interdependent variables. This model is particularly advantageous in econometrics, finance, and various applied sciences, where multiple time series evolve together over time. Unlike structural equation models, which impose strong theoretical restrictions, the VAR model treats all variables symmetrically as endogenous, thereby allowing the data to determine their dynamic interrelationships.

Mathematical Formulation of the VAR Model

A VAR model of order **p**, denoted as **VAR(p)**, represents a system of equations where each variable is regressed on its own past values as well as the past values of all other variables in the system.

Mathematically, a **k-variable VAR(p)** model expressed as follows:

$$Y_t = c + A_1 Y_{t-1} + A_2 Y_{t-2} + \cdots + A_p Y_{t-p} + \varepsilon_t$$

Where:

- $Y_t = (Y_{1t}, Y_{2t}, \ldots, Y_{kt})'$ is a $K \times 1$ vector of endogenous variable at time t.
- $c\ is\ a\ k \times 1$ vector of interception terms.
- $A_i\ is\ a\ k \times 1$ coefficient matrices capturing the relationships among variables.
- $\varepsilon_t\ is\ a\ k \times 1$ vector of white noise error terms with zero mean and covariance matrix $\sum$. i.e.,

$$E(\varepsilon_t) = 0,\quad E(\varepsilon_t \varepsilon_t') = \sum,\quad E(\varepsilon_t \varepsilon_s' = 0\ \forall\, t \neq s.$$

The lag order p is a crucial parameter that determines how much past information influences the present values. Selecting an optimal lag length is critical for balancing model complexity and forecasting accuracy.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Assumptions of the VAR Model
For valid estimation and inference, the VAR model relies on the following key assumptions:

- Stationarity: The time series should be weakly stationary, meaning that their mean and variance are constant over time. If non-stationary, differencing or transformation techniques may be required.
- No Perfect Multicollinearity: The explanatory variables should not be perfectly correlated to ensure the stability and interpretability of parameter estimates.
- White Noise Errors: The residuals are assumed to be serially uncorrelated and homoscedastic. If autocorrelation is present, modifications such as higher-order VAR models or alternative estimation techniques may be necessary.

Estimation Techniques for VAR Models
Estimating the parameters of a VAR model involves solving a system of equations, typically using one of the following methods:

1. Ordinary Least Squares (OLS): Since each equation in a VAR system contains the same set of regressors, OLS can be applied separately to each equation, yielding consistent and efficient estimates under the assumption of normally distributed residuals.
2. Maximum Likelihood Estimation (MLE): When the errors are assumed to follow a multivariate normal distribution, MLE provides asymptotically efficient parameter estimates.
3. Bayesian VAR (BVAR): To address the problem of overparameterization in large VAR models, Bayesian methods incorporate prior information to shrink estimates and improve forecasting performance.

Selection of Lag Order
The choice of lag length ppp significantly impacts the model's performance. Commonly used criteria for selecting the optimal lag order include:

- Akaike Information Criterion (AIC)
- Bayesian Information Criterion (BIC)
- Hannan-Quinn Criterion (HQC)
- Likelihood Ratio (LR) Test

A lower value of these criteria typically indicates a better-fitting model with an optimal balance between complexity and explanatory power.

Advantages and Limitations of VAR Models
Advantages

- Captures Dynamic Relationships: VAR models effectively capture interdependencies between multiple time series variables.
- Flexible Model Specification: Unlike structural models, VAR models impose minimal theoretical restrictions, allowing data-driven analysis.
- Forecasting Utility: VAR models provide reliable multi-step-ahead forecasts, making them widely used in economic and financial forecasting.
- Facilitates Granger Causality Testing: VAR models can help identify causal relationships between time series variables.

Limitations

- High Dimensionality: As the number of variables and lag order increase, the number of parameters grows exponentially, leading to overfitting and estimation inefficiencies.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

- Stationarity Requirement: Non-stationary series require transformation (e.g., differencing or cointegration techniques).
- Difficulty in Interpretation: The presence of multiple lagged variables complicates the direct interpretation of individual coefficients.

Applications of VAR Models

VAR models are extensively used in various domains, including:

- Macroeconomic Analysis: Studying the impact of monetary policy shocks on inflation, interest rates, and GDP growth.
- Financial Markets: Analyzing interactions between stock returns, interest rates, and exchange rates.
- Energy Economics: Modeling the relationship between oil prices, energy demand, and macroeconomic indicators.
- Epidemiology: Examining the spread of infectious diseases using time series data on infection rates and control measures.

Maximum Likelihood Estimation of Parameters of VAR Model Using OLS Regression

Vector Autoregressive (VAR) models are widely used in multivariate time series analysis to capture dynamic relationships among multiple interdependent variables. A crucial step in building a VAR model is the estimation of its parameters, which directly impacts the model's predictive accuracy and inferential reliability. Among various estimation techniques, Maximum Likelihood Estimation (MLE) is one of the most efficient methods, particularly under the assumption of normally distributed residuals. However, due to the structural form of VAR models, Ordinary Least Squares (OLS) regression provides consistent and computationally efficient estimates, which coincide with the MLE under Gaussian error assumptions.

Estimation of VAR Parameters Using Maximum Likelihood

The VAR Model Representation

Consider a k-dimensional multivariate time series $Y_t = (Y_{1t}, Y_{2t}, \ldots, Y_{kt})'$, which follows a **p-th order VAR model:**

$$Y_t = c + A_1 Y_{t-1} + A_2 Y_{t-2} + \cdots + A_p Y_{t-p} + \varepsilon_t$$

Where:

- Y_t *is a* $k \times 1$ vector of endogenous variables.
- c *is a* $k \times 1$ vector of intercept terms.
- $A_i (for\ i = 1,2, \ldots, p)$ *are* $k \times k$ coefficient matrices to be estimated.
- ε_t *is a* $k \times 1$ vector of error terms, assumed to be normally distributed as

$$\varepsilon_t \sim N(0, \textstyle\sum)$$

Where $\sum$ is the $k \times k$ covariance matrix of residuals.

Likelihood Function for the VAR Model

Under the assumption that ε_t follows a multivariate normal distribution, the likelihood function for the sample $\{Y_1, Y_2, \ldots, Y_T\}$ is given by:

$$L(\Theta) = (2\pi)^{\frac{kT}{2}} |\textstyle\sum|^{-\frac{T}{2}} \exp\left(-\frac{1}{2}\sum_{t=1}^{T} \varepsilon_t' \Sigma^{-1} \varepsilon_t\right)$$

Where $\Theta = \{c, A_1, A_2, \ldots, A_p, \sum\}$ represents the set of parameters to be estimated. Taking the natural logarithms, the log-likelihood function simplifies to:

$$\ln L(\Theta) = -\frac{kT}{2}\ln(2\pi) - \frac{T}{2}\ln|\Sigma| - \frac{1}{2}\sum_{t=1}^{T} \varepsilon_t' \Sigma^{-1} \varepsilon_t$$

To maximize $\ln L(\Theta)$, we need to estimate the parameters in a way that minimizes the residual sum of squares.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Estimation Using Ordinary Least Squares (OLS) Regression

OLS-Based Parameter Estimation

Since each equation in a VAR model consists of the same set of lagged variables as regressors, OLS Estimation can be applied to each equation separately. Rewriting the VAR model in matrix form:

$$Y = XB + E$$

Where:

- Y is a $T \times k$ matrix of endogenous variables.
- X is a $T \times (kp + 1)$ matrix of lagged endogenous variables and intercepts.
- B is a $(kp + 1) \times k$ matrix of coefficient to be estimated.
- E is a $T \times k$ matrix of residuals.

The OLS estimator for B is given by:

$$\hat{B} = (X'X)^{-1}X'Y$$

Which provides unbiased and consistent estimates of the VAR parameters under the assumption that X is exogenous and full-rank.

Estimation of Residual Covariance Matrix

The covariance matrix of the residuals is estimated using:

$$\hat{\sum} = \frac{1}{T}E'E = \frac{1}{T}\left(Y - X\hat{B}\right)'\left(Y - X\hat{B}\right)$$

Which represents the dispersion of errors across different time series. The **Internally Studentized residuals** are computed as:

$$r_{it} = \frac{\varepsilon_{it}}{\sqrt{\hat{\sigma}_{ii}}}$$

Where $\hat{\sigma}_{ii}$ is the estimated variance of the residuals for the i-th equation, ensuring robustness in variance estimation.

Likelihood Ratio Test for Lag Selection

The selection of the optimal lag length in a VAR model is crucial for capturing the underlying dynamics of the time series. The **Likelihood Ratio (LR) test** is commonly used method for testing whether a higher lag order significantly improves model performance. The test statistic is given by:

$$LR = T\left[\ln|\hat{\sum}_r| - \ln|\hat{\sum}_r|\right]$$

Where:

- $\hat{\sum}_r$ is the restricted (lower lag order) covariance matrix.
- $\hat{\sum}_u$ is the unrstricted (higher lag order) covariance matrix.

Under the null hypothesis that the additional lags do not contribute significantly, the LR statistic follows a **chi-square** distribution with degrees of freedom equal to the number of additional parameters.

Conclusion

In this study, we explored the estimation of parameters for Vector Autoregressive (VAR) models using Maximum Likelihood Estimation (MLE) based on Ordinary Least Squares (OLS) regression. The research highlights the fundamental steps involved in VAR model building, including identification, specification, estimation, diagnostic checking, and forecasting.

The study demonstrates that while MLE provides an efficient estimation framework under normality assumptions, OLS regression serves as a computationally simple and asymptotically efficient method for parameter estimation in VAR models. Since each equation in the VAR system contains identical regressors, OLS can be applied independently to each equation, yielding consistent and unbiased estimates of the coefficient matrices. Additionally, the covariance matrix of residuals was estimated using Internally Studentized residuals, ensuring robustness in variance estimation.

To optimize model performance, an essential aspect of VAR estimation is the selection of the appropriate lag order. The study incorporated the Likelihood Ratio (LR) test, which aids in determining the optimal number of lags by comparing restricted and unrestricted models. This ensures that the VAR model effectively captures the dynamic interdependencies among time series variables while maintaining parsimony.

Despite the advantages of the VAR model, some limitations persist. Over parameterization in large-scale models can lead to inefficiencies, requiring alternative estimation techniques such as Bayesian VAR (BVAR) or penalized regression methods. Additionally, the assumption of stationarity is critical for valid inference, necessitating preliminary unit root testing and transformation techniques for non-stationary data.

Overall, this study reaffirms the importance of VAR models in multivariate time series analysis, particularly in fields such as econometrics, finance, and macroeconomic forecasting. The combination of MLE with OLS estimation provides a statistically sound and computationally feasible approach for parameter estimation. Future research can extend this work by incorporating structural VAR (SVAR) models, cointegration techniques, and machine learning-based forecasting methods to enhance model interpretability and predictive performance.
The next step in this research is model validation and diagnostic checking, ensuring that the estimated VAR model adheres to the assumptions of stationarity, stability, and residual independence. These steps are crucial for making reliable statistical inferences and improving the forecasting accuracy of the VAR framework.

References:-

1. Lin Botao, Guo Jiancheng. Discussion On Current Application Of Artificial Intelligence In Petroleum Industry. Petroleum Science Bulletin, 2019, 4(4): 403–413.
2. Jia Ailin, Guo Jianlin. Key Technologies And Understandings On The Construction Of Smart Fields. Petroleum Exploration And Development, 2012, 39(1): 118–122.
3. Li J, Castagna J, Li D, Et Al. Reservoir Prediction Via Svm Pattern Recognition: Seg Technical Program Expanded Abstracts 2004. Tulsa: Society Of Exploration Geophysicists, 2004: 425–428.
4. Xing Minghai, Chen Xiangguang, Wang Yu. Oilfield Output Forecast Using Fuzzy Neural Network. Computer Simulation, 2005, 22(2): 187–190.
5. Weber D, Edgar T F, Lake L W, Et Al. Improvements In Capacitance-Resistive Modeling And Optimization Of Largescale Reservoirs. Spe 121299, 2009.
6. Anifowose F A. Artificial Intelligence Application In Reservoir Characterization And Modeling: Whitening The Black Box. Spe 155413, 2011.
7. [Gherabati S A, Hughes R G, Zhang H, Et Al. A Large Scale Network Model To Obtain Interwell Formation Characteristics. Spe 153386, 2012.
8. Salazar M, Gonzalez H, Matringe S, Et Al. Combining Decline-Curve Analysis And Capacitance-Resistance Models To Understand And Predict The Behavior Of A Mature Naturally Fractured Carbonate Reservoir Under Gas Injection. Spe 153252, 2012.
9. Holdaway K R. Data Mining Methodologies Enhance Probabilistic Well Forecasting. Spe 167428-Ms, 2013.
10. Zhang Fangzhou, Yan Huyong, Yang Liquan, Et Al. Application Of Improved Grey Neural Network Model To Oil Yields. Computer Technology And Development, 2013, 23(6): 241–244.
11. Gupta S, Fuehrer F, Jeyachandra B C. Production Forecasting In Unconventional Resources Using Data

Mining And Time Series Analysis. Spe 171588-Ms, 2014.

12. Zhou Q, Dilmore R, Kleit A, Et Al. Evaluating Gas Production Performances In Marcellus Using Data Mining Zhang Rui Et Al. / Petroleum Exploration And Development, 2021, 48(1): 201–211 – 211 – Technologies. Journal Of Natural Gas Science And Engineering, 2014, 20(1): 109–120.
13. Li J, Lei Z, Li S, Et Al. Optimizing Water Flood Performance to Improve Injector Efficiency In Fractured Low-Permeability Reservoirs Using Streamline Simulation. Spe 182779, 2016.
14. Cao Q, Banerjee R, Gupta S, Et Al. Data Driven Production Forecasting Using Machine Learning. Spe 180984, 2016.
15. Zhao Hui, Kang Zhijiang, Sun Haitao, Et Al. An Interwell Connectivity Inversion Model fsor Waterflooded Multilayer Reservoirs. Petroleum Exploration and Development, 2016, 43(1): 99–106.