

Vector Autoregressive Model Selection: Gross Domestic Product and Europe Oil Prices Data Modelling

S. Zamani Mehreyan[†] and A. Sayyareh^{‡,*}

[†] Imam Khomeini International University

^{‡,*} K. N. Toosi University of Technology

Received: 2021/29/06 Approved: 2022/01/05

Abstract. We consider the problem of model selection in vector autoregressive model with Normal innovation. Tests such as Vuong's and Cox's tests are provided for order and model selection, i.e. for selecting the order and a suitable subset of regressors, in vector autoregressive model. We propose a test as a modified log-likelihood ratio test for selecting subsets of regressors. The Europe oil prices, Brent, and the real gross domestic product, GDP, data are considered as real data. Since the Brent data does Granger-cause the GDP data, so we suggest the vector autoregressive model and select optimal model based on the model selection test. The analysis provides analytic results show that the Vuong's, Cox's and proposed test are the appropriate test for order and model selection for vector autoregressive models with Normal innovation. In simulation study, the power of proposed test at least is as good as the power of Vuong's test.

Keywords. Cox's test, maximum likelihood estimation, mis-specified model, nested models, vector autoregressive model, Vuong's test.

* Corresponding author

1 Introduction

The vector autoregression, VAR, model is one of the most successful and flexible to use models for the analysis of multivariate time series. It is a natural extension of the univariate autoregressive model to dynamic multivariate time series. Vector autoregressive was introduced in to empirical economics by Sims (1980), who demonstrated that VAR's provide a flexible and tractable framework for analyzing economic time series. It is well known that the dynamic properties of responses may depend critically on the lag order of the VAR model fitted to the data. These differences can be large enough to affect the substantive interpretation of VAR response estimation, see Kilian (2001) and Hamilton and Herrera (2004). An important preliminary step in empirical studies is to select the order of the regression based on the same data used subsequently to construct the response estimates. The most common strategy in empirical studies is to select the lag-order by some information criterion.

Let \mathbf{Y} be a $(k \times 1)$ random vector and $\mathbf{X}_t = (\mathbf{Y}_t, \mathbf{Z}_t)$, $t = 1, \dots, T$, has common unknown true joint distribution $H(\cdot)$ on a complete probability space (\mathcal{X}, σ_X) , where \mathcal{X} is the Euclidean space \mathbb{R}^{k+1} and σ_X is the Borel σ -field on \mathcal{X} . Let (\mathcal{Y}, σ_Y) and (\mathcal{Z}, σ_Z) be the measurable spaces associated with \mathbf{Y}_t and \mathbf{Z}_t . We shall be interested in the true conditional distribution $H_{Y|Z}(\cdot|\cdot)$ of \mathbf{Y}_t given \mathbf{Z}_t . Let H_Z be the true marginal distribution of \mathbf{Z}_t , and ν_Y be a σ -finite measure on (\mathcal{Y}, σ_Y) . For H_Z -almost all z , $H_{Y|Z}(\cdot|z)$ has a Radon-Nikodym density $h(\cdot|z)$ relative to ν_Y , which is strictly positive for ν_Y -almost all y . We now consider two competing parametric families of distributions defined on $\sigma_Y \times \mathcal{Z}$ for \mathbf{Y}_t given \mathbf{Z}_t :

$$\mathcal{F}_\gamma = \{f^\gamma(y|z), \gamma \in \Gamma \subseteq \mathbb{R}^m\} \quad \text{and} \quad \mathcal{G}_\beta = \{g^\beta(y|z), \beta \in \mathcal{B} \subseteq \mathbb{R}^n\}.$$

Let the competing models satisfies the assumptions A1-A6, Vuong's (1989), which are stated below:

A1: The independent $(k \times 1)$ vector of random variables, ϵ_t $t = 1, 2, \dots, T$ have common joint distribution function H_{ϵ_t} with measurable Radon-Nikodim density $h(\cdot)$.

A2: Γ is a compact subsets of \mathbb{R}^m , and the joint density function $f^\gamma(Y_t|Z_t = z_t)$ is continuous in γ .

A3: $|\log f^\gamma(Y_t|Z_t = z_t)|$ is dominated by \mathcal{M} where \mathcal{M} is integrable with

respect to $H(\cdot)$ and independent of γ . Also the function $\int \log f^\gamma(Y_t|Z_t = z_t)H(dy)$ has a unique maximum on Γ in γ_* . Where $H(\cdot)$ is true conditional joint distribution and the value γ_* is called the pseudo-true value of γ for the conditional model \mathcal{F} .

A4: $\log f^\gamma(Y_t|Z_t = z_t)$ is twice continuously differentiable on Γ and $|\frac{\partial \log f^\gamma(Y_t|z_t)}{\partial \gamma} \cdot \frac{\partial \log f^\gamma(Y_t|z_t)}{\partial \gamma'}|$ and $|\frac{\partial^2 \log f^\gamma(Y_t|z_t)}{\partial \gamma \partial \gamma'}|$ are dominated by H-integrable function and independent of γ .

A5: γ_* is an interior point of Γ and γ_* is a regular point of $A_f(\gamma)$ where

$$A_f(\gamma) = \mathcal{E}_h \left\{ \frac{\partial^2 \log f^\gamma(\epsilon_t)}{\partial \gamma \partial \gamma'} \right\}$$

and \mathcal{E}_h denotes the expectation with respect to the true joint distribution.

A6: For H-almost all (y, z) the functions $|\log f^\gamma(Y_t|z_t)|^2$ and $|\log g^\beta(Y_t|z_t)|^2$ are dominated by H-integrable functions independent of γ and β .

The distance of $f^\gamma(y|z)$ from the true conditional density $h(y|z)$ measured by Kullback-Leibler divergence, $\mathcal{KL}\{h(\cdot|z), f^{\gamma_*}(\cdot|z)\}$, where γ_* is the pseudo-true value of γ , White (1982). The best model in \mathcal{F}_γ being the one for which Kullback-Leibler divergence, \mathcal{KL} , is the smallest or equally $E_h\{\log f^{\gamma_*}(Y|z)\}$ is the largest. The important part of the Kullback-Leibler divergence is $E_h\{\log f^{\gamma_*}(Y|z)\}$ which has an estimates as

$$\frac{1}{n} \sum_{t=1}^n \log f^{\hat{\gamma}_T}(y_t|z_t).$$

where the Quasi Maximum Likelihood Estimator, QMLE, $\hat{\gamma}_T$, is a consistent estimator of γ_* , White (1982).

Determination of the model order is an important step in vector autoregressive, VAR(p), modelling, where p is order of vector autoregressive model. The lag length for the VAR(p) model may be determined using model selection criteria. The general approach is to fit VAR(p) models with orders $p = 0, \dots, p_{max}$ and choose the model which have the lowest value of model selection criterion. Recall that p_{max} is the maximal order of vector autoregressive model. Model selection criteria for VAR(p) models have the form

$$IC(p) = \ln |\hat{\Omega}_T| + C_T \cdot \varphi(k, p)$$

where $\hat{\Omega}_T$ is the maximum likelihood estimate of the innovation covariance matrix Ω , C_T is a sequence indexed by the sample size T , and $\varphi(k, p)$ is

a penalty function. The three most common information criteria are the Akaike (1973), AIC, Schwarz (1978), SIC and Hannan and Quinn (1979), HQ, that defined as

$$AIC(p) = \ln |\hat{\Omega}_T| + \frac{2}{T}pk^2$$

$$SIC(p) = \ln |\hat{\Omega}_T| + \frac{\ln T}{T}pk^2$$

$$HQ(p) = \ln |\hat{\Omega}_T| + \frac{\ln(\ln T)}{T}pk^2.$$

The performance of an order selection criterion is optimal if the selected model is the most accurate model in the considered set of estimated competing models, see Quinn (1980), Paulsen and Tjostheim (1985) and Quinn (1988). The Akaike information criterion initially was proposed as an estimate of minus twice the expected log-likelihood by Akaike (1973). It is an asymptotically unbiased estimator of the Kullback-Leibler divergence and is known to suffer from overfit, selected order of model can be greater than the optimal model order, see Shibata(1984).

Cox (1961, 1962) and Vuong (1989) modified the classical hypothesis testing to test the non-nested hypotheses based on the generalization of the likelihood ratio test (LRT). Vuong (1989) has proposed a statistic for testing the null hypothesis that the competing models are equivalent related to true distribution against the alternative hypothesis that one model is closer to the true model. In this test one accept null hypothesis, it means that two competing models are equivalent, but it is less clear that they are close to the true model or far from it. To make inference after model selection, we use Cox's test and select two competing models as suitable or unsuitable equivalent models, see Sayyareh et al. (2011).

The rest of the paper is structured as follows: in section 2, we presented an example. It is the comparison of a autoregressive and vector autoregressive model of the Europe oil prices, Brent, and the real gross domestic product, GDP. In section 3, some of the model selection tests such as Vuong's test, Cox's test and the proposed test are improved for the vector autoregressive model with Normal innovation. In section 4, the obtained theoretical results are studied by simulation. We continue the motivating example and select optimal model between competing models based on the Vuong's, Cox's and proposed test in section 5.

2 Motivating Example

Our datasets consist of the Europe oil prices, Brent, and the real gross domestic product, GDP. We select optimal model between competing models based on the model selection test. These data can be found in

“<https://fred.stlouisfed.org/series/DCOILBRENTU>”

and

“<https://fred.stlouisfed.org/series/GDPC96>” respectively. The Brent dataset consists daily returns of the Europe oil prices with the sample extending from May 1987 to December 2020 for a total of $n = 8537$ observations. We denote $r_{t,i}$ as the i^{th} daily return for month t , then the monthly realized volatility is defined as

$$\sigma^2 =_{def} \left(\sum_{i=1}^m (r_{t,i} - \mu_t)^2 \right)^{1/2},$$

where m is the number of days and μ_t is monthly mean. The associated volatility of the Brent dataset, VB, was constructed by summing daily squared returns. The GDP dataset is quarterly returns of the real gross domestic product index with the sample extending from May 1987 to December 2020 for a total of $n = 136$ observations. The series QG is obtained by substituting the series GDP in function $Q(X)$,

$$Q(X) = \Delta(\log(X)),$$

where $\Delta(X)$ denotes the first order differences operator applied to a time series $\{X_t\}$, $\Delta(X) = X_t - X_{t-1}$. The dataset VB describe the information of oil volatility and the dataset QG contains information of economic growth. Due to the presence of missing data and select data with the same date, so the number of sample for both datasets is reduced to 108.

The curve of the VB and QG returns is given in Figure 1. The descriptive statistics of our datasets are given in Table 1 which shows that series VB has mean that is different from zero. The series QG has negative skewness and the series VB has positive skewness. Also both are characterized by heavy tails since they have positive the sample excess kurtosis. The hypothesis of normality is accepted for all series since $P - value > 0.05$.

The sample autocorrelation function, ACF, suggests that an autoregressive model might provide a reasonable model for given data. The ACF is shown in Figure 2.

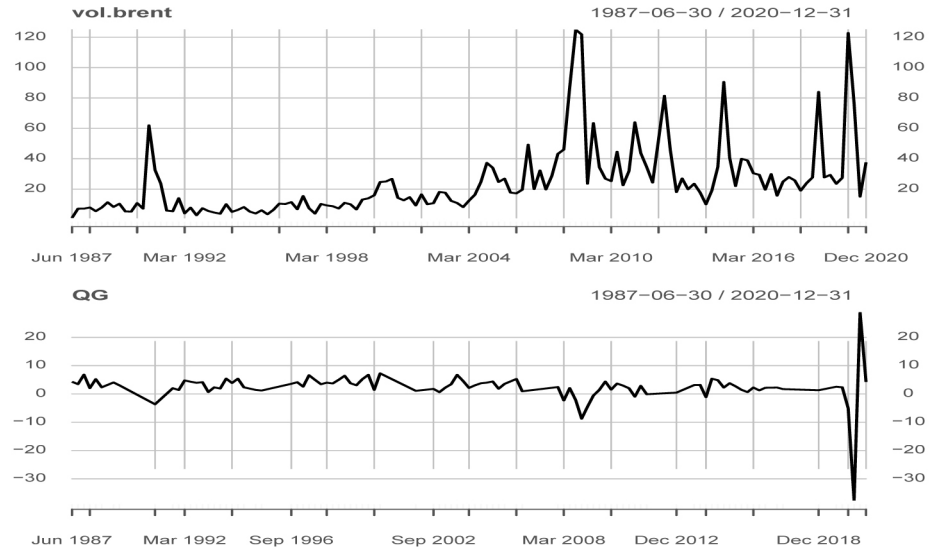


Figure 1. The time series plot of the VB and QG returns.

Table 1. Descriptive Statistics for Empirical Series.

series	n	\bar{x}	$\hat{\sigma}$	\mathcal{S}	\mathcal{K}	$\mathcal{P} - value$
QG	108	0.0005	0.0001	-2.9941	4.6446	0.2374
VB	108	11.9523	30.1373	1.1080	0.9974	0.1008

Notes:

1. \mathcal{S} denotes the sample skewness, \mathcal{K} denotes the sample excess kurtosis.
2. \mathcal{P} is the p-value of the Kolmogorov-Smirnov test for normality of the underlying series.

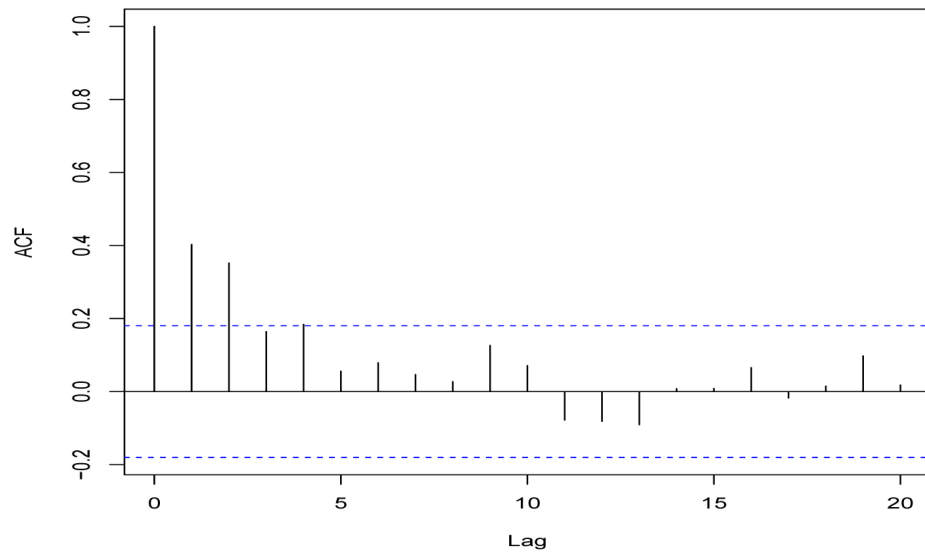


Figure 2. The sample autocorrelation function of the real gross domestic product data.

The rolling analysis of time series is widely used. See for example Aaltonen and Ostermark (1997) and Zivot and Wang (2006). The rolling correlation estimate at time t with window width R is the usual sample estimates using the most recent R observations. The monthly rolling correlation analysis with the width of a sub-sample or window, $R=80$, is performed for correlations between QG and VB. The results of this rolling correlation analysis are given in Figure 3. It shows that there is approximately a correlation -0.5 between VB and QG.

To test the null hypothesis that Z does not Granger-cause Y , one first finds the proper lagged values of Y to include in a univariate autoregression of Y

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \epsilon_t.$$

Next, the autoregression is augmented by including lagged values of Z

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varphi_1 Z_{t-1} + \varphi_2 Z_{t-2} + \cdots + \varphi_q Z_{t-q} + \epsilon_t$$

The variable Z is said to cause Y , provided some φ_j is non-zero. One retains in this regression all lagged values of Z that are individually significant according to their t -statistics, provided that collectively they add explanatory

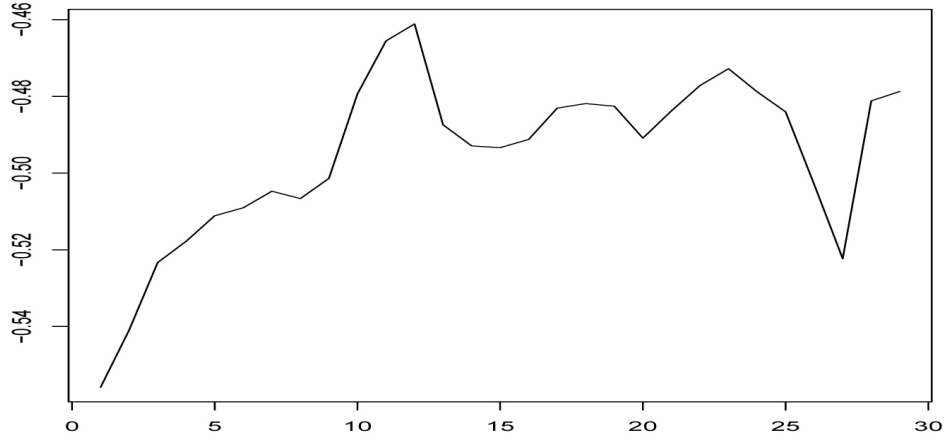


Figure 3. Thr rolling correlation between QG and VB.

power to the regression according to an F-test, whose null hypothesis is no explanatory power jointly added by the Z 's. The F-test is

$$F = \frac{(SSE_r - SSE_f)/q}{SSE_f/(T - p - q - 1)},$$

where SSE_r and SSE_f are the sum of squared errors of reduced model and full model respectively, T is total number of observation, p is the number of lags for the Y -variable and q is the number of lags for the Z -variable. Note that F has an asymptotic F distribution with q and $T - p - q - 1$ degrees of freedom. See Granger (1969, 2004) and Eichler (2012).

The monthly rolling test of causality with the width of window, $R=80$, for the first-order autoregressive model with Normal innovation of series is considered. The results are given in Table 2. In this Table, the significant causality at the 5% level is observed. In otherworld VB does Granger-cause QG but the variable QG does not Granger-cause VB.

The main question is whether vector autoregressive model is suitable model or autoregressive model is good model. Consequently, we study the model selection tests such as Vuong's test, Cox's test and a proposed test that they are based on the likelihood ratio test.

Table 2. The Rolling Causality Testing for QG and VB when AR(1) model is fitted.

	QG does not GC VB	No IC between QG	VB does not GC QG
2011-12-31	0.5165	0.0500	0.0003
2012-03-31	0.3513	0.0559	0.0005
2012-06-30	0.2398	0.0767	0.0004
2012-09-30	0.2847	0.0699	0.0004
2012-12-31	0.3588	0.0717	0.0004
2013-09-30	0.3669	0.0590	0.0002
2013-12-31	0.3370	0.0527	0.0003
2014-06-30	0.3238	0.0429	0.0003
2014-09-30	0.0820	0.1234	0.0002
2014-12-31	0.0539	0.1143	0.0002
2015-03-31	0.1701	0.1052	0.0009
2015-06-30	0.2935	0.0864	0.0010
2015-09-30	0.3160	0.0877	0.0013
2015-12-31	0.2650	0.0853	0.0012
2016-03-31	0.2615	0.0833	0.0016
2016-06-30	0.2639	0.0632	0.0007
2016-09-30	0.2964	0.0591	0.0007
2016-12-31	0.3086	0.0502	0.0004
2017-03-31	0.3299	0.0514	0.0009
2017-06-30	0.3292	0.0451	0.0006
2017-09-30	0.3933	0.0341	0.0003
2017-12-31	0.4167	0.0301	0.0002
2018-09-30	0.4765	0.0279	0.0002
2019-06-30	0.4836	0.0246	0.0001
2019-09-30	0.4758	0.0242	0.0001
2020-03-31	0.4789	0.0014	0.0007
2020-06-30	0.5178	0.0273	0.0013
2020-09-30	0.0297	0.0513	0.0000
2020-12-31	0.0233	0.0423	0.0000

3 Testing Order Selection

In this section the asymptotic distribution of statistics related to the Vuong's test and Cox's test are derived. The main purpose of this section is to compute the Cox's test statistic in vector autoregressive models with Normal innovation for stationary case.

3.1 Vuong's Test for VAR Model

Consider p - lag vector autoregressive, $VAR(p)$,

$$Y_t = C + \Phi_1 Y_{t-1} + \dots + \Phi_p Y_{t-p} + \epsilon_t, \quad t = 1, \dots, T, \quad T > 1 \quad (1)$$

as true model, where Φ_i 's are $(k \times k)$ coefficient matrices and ϵ_t is an $(k \times 1)$ unobservable zero mean white noise vector process with time invariant covariance matrix Ω . The covariance matrix Ω is positive definite matrix and

$$E(\epsilon_{it} \epsilon_{js}') = \begin{cases} \sigma_{ij} & t = s \\ 0 & t \neq s \end{cases}$$

In lag operator notation, the VAR(p) can be written as

$$\Phi(B)Y_t = C + \epsilon_t$$

where $\Phi(B) = I_k - \Phi_1 B - \dots - \Phi_p B^p$ and $B^j Y_t = Y_{t-j}$. For stationary, we require that no zeros of the determinant of the autoregressive matrix polynomial, $|\Phi(B)|$, line on or inside the unit circle, i.e., $|\Phi(B)| \neq 0$ for $|B| \leq 1$.

Consider model (1), where innovation terms are distributed as multivariate Normal distribution, $N(\mathbf{0}, \Omega)$, with conditional joint density function,

$$f_{Y_t|Z_t}^\gamma(y_t|z_t) = (2\pi)^{-\frac{k}{2}} |\Omega^{-1}|^{\frac{1}{2}} \exp \left(-\frac{1}{2} (y_t - \Pi' x_t)' \Omega^{-1} (y_t - \Pi' x_t) \right)$$

where

$$\gamma = \left(C', \text{vec}(\Phi_1)', \dots, \text{vec}(\Phi_p)', \text{vec}(\Omega)' \right)',$$

$$z_t = \begin{pmatrix} 1 \\ y_{t-1} \\ \vdots \\ y_{t-p} \end{pmatrix}_{(kp+1) \times 1}$$

and

$$\Pi' = \begin{pmatrix} C & \Phi_1 & \Phi_2 & \dots & \Phi_p \end{pmatrix}_{k \times (kp+1)}.$$

We define $\mu_{3f} = E\{Y_t^3|Z_t\}$, $\mu_{4f} = E\{Y_t^4|Z_t\}$ as third and fourth moments with respect to f and v_f as a column vector of the diagonal elements of Ω_f^{-2} . Also μ_{3g} , μ_{4g} and v_g are defined with respect to g . The log-likelihood function is

$$L(\gamma) = -\frac{kT}{2} \log(2\pi) + \frac{T}{2} \log |\Omega^{-1}| - \frac{1}{2} \sum_{t=1}^T (y_t - \Pi' z_t)' \Omega^{-1} (y_t - \Pi' z_t).$$

The maximum likelihood estimators of Π and Ω are obtained by solving the estimating equations. Using some algebraic calculations, the maximum likelihood estimator of Π is

$$\hat{\Pi}'_T = \left(\sum_{t=1}^T y_t z_t' \right) \left(\sum_{t=1}^T z_t z_t' \right)^{-1}$$

The j^{th} row of $\hat{\Pi}'_T$ is

$$u_j' \hat{\Pi}'_T = u_j' \left(\sum_{t=1}^T y_t z_t' \right) \left(\sum_{t=1}^T z_t z_t' \right)^{-1}.$$

where u_j is a vector of 0 and 1 that j^{th} element of this vector is 1. Also the maximum likelihood estimator of Ω is given by

$$\hat{\Omega}_T = \frac{1}{T} \sum_{t=1}^T \hat{\epsilon}_t \hat{\epsilon}_t'.$$

The maximum likelihood estimators $\hat{\Pi}'_T$ and $\hat{\Omega}_T$ will give consistent estimates of the population parameters. Properties of the obtained maximum

likelihood estimators have been discussed by Nicholls (1976, 1977). They have shown that $\hat{\gamma}_T$ is asymptotically unbiased and consistent and that $\sqrt{T}(\hat{\gamma}_T - \gamma)$ has an asymptotic multivariate Normal distribution,

$$\sqrt{T}(\hat{\gamma}_T - \gamma) \xrightarrow{d} N(\mathbf{0}, \Sigma_\gamma)$$

where Σ_γ is the inverse of the information matrix,

$$\Sigma_\gamma = \left[-E \left(\frac{\partial L(\gamma)}{\partial \gamma \partial \gamma'} \right) \right]^{-1}.$$

For more illustration see Wei (2006).

Definition 1. The conditional model \mathcal{G}_β is nested in \mathcal{F}_γ if and only if

$$\mathcal{G}_\beta \subset \mathcal{F}_\gamma.$$

It means that any conditional distribution in \mathcal{G}_β is equal to a conditional distribution in \mathcal{F}_γ . Vuong (1989) has proposed a LR-based test for selecting between two nested models. This test reduces to the classical Neyman-Pearson LR test when the largest model is correctly specified. Given Assumptions A1-A6, Vuong (1989), under null hypothesis, we have

$$2LR_n \xrightarrow{d} M_m(x, \hat{\lambda}_T)$$

where $\hat{\lambda}_T$ is the vector of eigenvalues of the sample analog W_T of W ,

$$W = B_f(\gamma_*) \left[\frac{\partial \varphi(\beta_*)}{\partial \beta'} A_g^{-1}(\beta_*) \frac{\partial \varphi'(\beta_*)}{\partial \beta} - A_f^{-1}(\gamma_*) \right],$$

$$A_f(\gamma) = \left[E_h \left\{ \frac{\partial^2 \log f^\gamma(Y_t|Z_t)}{\partial \gamma \partial \gamma'} \right\} \right]$$

$$B_f(\gamma) = \left[E_h \left\{ \frac{\partial \log f^\gamma(Y_t|Z_t)}{\partial \gamma} \cdot \frac{\partial \log f^\gamma(Y_t|Z_t)}{\partial \gamma'} \right\} \right],$$

$$LR_n = \sum_{t=1}^n \log \frac{f^{\hat{\gamma}_T}(Y_t|Z_t)}{g^{\hat{\gamma}_T}(Y_t|Z_t)},$$

$\tilde{\gamma}_T \equiv \phi(\hat{\beta})$ is the constrained maximum likelihood estimator of γ_* subject to the constraints that γ belongs to $\phi(\mathcal{B})$ and $M_p(\cdot, \lambda)$ denotes the weighted

sum of Chi-square distributions. If $A_f(\gamma_*) + B_f(\gamma_*) = 0$, then under the null hypothesis we have

$$2LR_n \xrightarrow{d} \chi_{m-n}^2.$$

3.2 Cox's Test for VAR Model

Sometimes Vuong's test selects two competing models as equivalent models. It is less clear that they are close to the unknown true model or far from it. Cox (1962) proposed a test of separated families. Cox's test as a modified log-likelihood ratio statistic involves centering the log-likelihood ratio statistic under the null hypothesis $H_0^f : h = f^\gamma$ against $H_1^g : h = g^\beta$. This test is based on the T_{fg} statistic

$$T_{fg} = \frac{\left\{ L_f(\hat{\gamma}_T) - L_g(\hat{\beta}_T) \right\} - \mathcal{E}_f \left\{ L_f(\hat{\gamma}_T) - L_g(\hat{\beta}_T) \right\}}{\hat{\sigma}^{fg}}$$

comparing the observed difference of log-likelihoods with an estimate of that to be expected under H_f , where L_f is log-likelihood function and σ^{fg} is the standard deviation of the numerator of T_{fg} . It is known that T_{fg} has asymptotically standard Normal distribution, see Cox (1962). This test has four rejection and acceptance regions as

- (i) Reject both H_0^f and H_1^g if $|T_{fg}| > C_\alpha$ and $|T_{gf}| > C_\alpha$,
- (ii) Reject neither H_0^f and H_1^g if $|T_{fg}| < C_\alpha$ and $|T_{gf}| < C_\alpha$,
- (iii) Reject H_0^f but not H_1^g if $|T_{fg}| > C_\alpha$ and $|T_{gf}| < C_\alpha$,
- (iv) Reject H_0^g but not H_1^f if $|T_{gf}| > C_\alpha$ and $|T_{fg}| < C_\alpha$,

where C_α is the critical value from the standard Normal distribution for some significance level α . We compute Cox's test statistics for null hypothesis contain p-order vector autoregressive model, VAR(p), against q-order vector autoregressive model, VAR(q), with Normal innovation, where $q \leq p$. The log-likelihood functions for these models are

$$L_f(\gamma) = -\frac{KT}{2} \log 2\pi - \frac{T}{2} \log |\Omega_f| - \frac{1}{2} \sum_{t=1}^T \epsilon_t \Omega_f^{-1} \epsilon_t'$$

and

$$L_g(\beta) = -\frac{KT}{2} \log 2\pi - \frac{T}{2} \log |\Omega_g| - \frac{1}{2} \sum_{t=1}^T \epsilon_t \Omega_g^{-1} \epsilon_t'.$$

The log likelihood ratio is

$$L_f(\hat{\gamma}) - L_g(\hat{\beta}) = -\frac{T}{2} \log |\hat{\Omega}_f| + \frac{T}{2} \log |\hat{\Omega}_g| - \frac{1}{2} \sum_{t=1}^T \epsilon_t \hat{\Omega}_f^{-1} \epsilon_t' + \frac{1}{2} \sum_{t=1}^T \epsilon_t \hat{\Omega}_g^{-1} \epsilon_t',$$

where $\hat{\gamma}$ and $\hat{\beta}$ are the maximum likelihood estimators of γ and β . We have calculated expectation and variance of $\log f^\gamma - \log g^\beta$ under the f^γ as,

$$\begin{aligned} E_f \left(\log f^\gamma(Y_t|Z_t) - \log g^\beta(Y_t|Z_t) \right) &= \frac{1}{2} \log |\Omega_g| - \frac{1}{2} \log |\Omega_f| \\ &\quad - \frac{1}{2} E_f \left(\text{tr} \left[\epsilon_t \Omega_f^{-1} \epsilon_t' \right] \right) + \frac{1}{2} E_f \left(\text{tr} \left[\epsilon_t \Omega_g^{-1} \epsilon_t' \right] \right) \\ &= \frac{1}{2} \log |\Omega_g| |\Omega_f^{-1}| \\ &\quad - \frac{1}{2} \text{tr} \left[\Omega_f^{-1} E_f \left(\epsilon_t \epsilon_t' \right) \right] + \frac{1}{2} \text{tr} \left[\Omega_g^{-1} E_f \left(\epsilon_t \epsilon_t' \right) \right] \\ &= \frac{1}{2} \log |\Omega_g| |\Omega_f^{-1}| - \frac{K}{2} + \frac{1}{2} \text{tr} \left[\Omega_g^{-1} \Omega_f \right] \end{aligned}$$

and

$$\begin{aligned} V_f \left(\log f^\gamma(Y_t|Z_t) - \log g^\beta(Y_t|Z_t) \right) &= \frac{1}{4} V_f \left(\epsilon_t \Omega_g^{-1} \epsilon_t' - \epsilon_t \Omega_f^{-1} \epsilon_t' \right) \\ &= \frac{1}{4} V_f \left(\epsilon_t \Omega_f^{-1} \epsilon_t' \right) + \frac{1}{4} V_f \left(\epsilon_t \Omega_g^{-1} \epsilon_t' \right) \\ &\quad - \frac{1}{2} \text{Cov}_f \left(\epsilon_t \Omega_f^{-1} \epsilon_t', \epsilon_t \Omega_g^{-1} \epsilon_t' \right) \end{aligned}$$

Since $ABA' = \text{tr}(ABA')$ when ABA' is (1×1) , so we have

$$\begin{aligned} V_f \left(\log f^\gamma(Y_t|Z_t) - \log g^\beta(Y_t|Z_t) \right) &= \frac{1}{4} V_f \left(\text{tr} \left(\epsilon_t \Omega_f^{-1} \epsilon_t' \right) \right) + \frac{1}{4} V_f \left(\text{tr} \left(\epsilon_t \Omega_g^{-1} \epsilon_t' \right) \right) \\ &\quad - \frac{1}{2} \text{Cov}_f \left(\text{tr} \left(\epsilon_t \Omega_f^{-1} \epsilon_t' \right), \text{tr} \left(\epsilon_t \Omega_g^{-1} \epsilon_t' \right) \right) \\ &= \frac{1}{4} V_f \left(\text{tr} \left(\Omega_f^{-1} \epsilon_t' \epsilon_t \right) \right) + \frac{1}{4} V_f \left(\text{tr} \left(\Omega_g^{-1} \epsilon_t' \epsilon_t \right) \right) \\ &\quad - \frac{1}{2} \text{Cov}_f \left(\text{tr} \left(\Omega_f^{-1} \epsilon_t' \epsilon_t \right), \text{tr} \left(\epsilon_t' \epsilon_t \Omega_g^{-1} \right) \right). \end{aligned}$$

Note that $\text{tr}(BA'A) = \text{vec}(B)\text{vec}(A'A)$, then we can write

$$\begin{aligned} V_f \left(\log f^\gamma(Y_t|Z_t) - \log g^\beta(Y_t|Z_t) \right) &= \frac{1}{4} V_f \left(\text{vec}(\Omega_f^{-1}) \text{vec}(\epsilon_t' \epsilon_t) \right) + \frac{1}{4} V_f \left(\text{vec}(\Omega_g^{-1}) \text{vec}(\epsilon_t' \epsilon_t) \right) \\ &\quad - \frac{1}{2} \text{Cov}_f \left(\text{vec}(\Omega_f^{-1}) \text{vec}(\epsilon_t' \epsilon_t), \text{vec}(\epsilon_t' \epsilon_t) \text{vec}(\Omega_g^{-1}) \right) \\ &= \frac{1}{4} \text{vec}(\Omega_f^{-1}) V_f \left(\text{vec}(\epsilon_t' \epsilon_t) \right) \text{vec}(\Omega_f^{-1}) \\ &\quad + \frac{1}{4} \text{vec}(\Omega_g^{-1}) V_f \left(\text{vec}(\epsilon_t' \epsilon_t) \right) \text{vec}(\Omega_g^{-1}) \\ &\quad - \frac{1}{2} \text{vec}(\Omega_f^{-1}) \text{Cov}_f \left(\text{vec}(\epsilon_t' \epsilon_t), \text{vec}(\epsilon_t' \epsilon_t) \right) \text{vec}(\Omega_g^{-1}), \end{aligned}$$

where $\epsilon_t' \epsilon_t$ has the Wishart distribution, $W_k(\Omega_f, T)$. Thus

$$T_{fg} = \frac{T \left(\text{tr} [\Omega_g^{-1} \Omega_f] - K \right) + \sum_{t=1}^T \epsilon_t \Omega_g^{-1} \epsilon_t' - \sum_{t=1}^T \epsilon_t \Omega_f^{-1} \epsilon_t'}{2\sqrt{V_f}(\log f^\gamma - \log g^\beta)}.$$

similarly

$$T_{gf} = \frac{T \left(\text{tr} [\Omega_f^{-1} \Omega_g] - K \right) + \sum_{t=1}^T \epsilon_t \Omega_f^{-1} \epsilon_t' - \sum_{t=1}^T \epsilon_t \Omega_g^{-1} \epsilon_t'}{2\sqrt{V_g}(\log g^\beta - \log f^\gamma)}.$$

3.3 Proposed Test for VAR Model

Consider a vector autoregressive model as

$$Y_t^* = \Phi Y_{t-1}^* + \epsilon_t \quad (2)$$

where $Y_t^* = Y_t - \mu_y$ and $\mu_y = E\{Y_t\}$. We can write model (2) as

$$Y_t^* = \Theta Y_{t-1}^* + \Psi Y_{t-1}^* + \epsilon_t \quad (3)$$

where

$$\Theta = \begin{bmatrix} \phi_{11}^1 & 0 & \dots & 0 \\ 0 & \phi_{22}^1 & \dots & 0 \\ \vdots & & & \\ 0 & 0 & \dots & \phi_{kk}^1 \end{bmatrix} \quad \text{and} \quad \Psi = \begin{bmatrix} 0 & \phi_{12}^1 & \dots & \phi_{1k}^1 \\ \phi_{21}^1 & 0 & \dots & \phi_{2k}^1 \\ \vdots & & & \\ \phi_{k1}^1 & \phi_{k2}^1 & \dots & 0 \end{bmatrix}.$$

In this subsection the asymptotic distribution of LR statistic for hypothesis test, $H_0 : \Psi = \mathbf{0}$, is derived. Let $\hat{\gamma}_T$ be the unrestricted MLE that obtained in the previous section under the hypothesis $H_1 : \Psi \neq \mathbf{0}$ and $\hat{\gamma}_0$ be the MLE when the parameter space is restricted by null hypothesis. Also, let $L_T = -2L(\hat{\gamma}_T)$ and $L_0 = -2L(\hat{\gamma}_0)$ be minus twice the log-likelihood evaluated at the unrestricted and restricted maximum likelihood estimate scheme, respectively and let $L = L_T - L_0$ be a statistic for testing H_0 against H_1 .

Theorem 1. (*Asymptotic Distribution*) Suppose $H_0 : \Psi = \mathbf{0}$ holds, then as $T \rightarrow \infty$

$$L = L_T - L_0 = \text{tr} \left[\left(\Psi - \hat{\Psi}_T \right) \sum_{t=1}^T y_{t-1}^* y_{t-1}^{*'} \left(\Psi - \hat{\Psi}_T \right)' \hat{\Omega}_0^{-1} \right] \xrightarrow{D} \chi_{K(K-1)}^2$$

One chooses a critical value c_α from the chi-square distribution for some significance level α . If the value of statistic $|L| = |L_T - L_0|$ is smaller than c_α then one does not reject the null hypothesis.

4 Simulation Analysis

In this section the obtained theoretical results are studied by simulation study. In this study, all calculations are performed using R software. It is of interest to estimation, select an optimal model based on the maximum likelihood, ML, approach. We have used ML approach to obtain estimates for the structural parameters of proposed models and use the model selection test which define in section 3 to select the optimal model. A set of data x_1, \dots, x_n is generated under first-order vector autoregressive model, $x_t =$

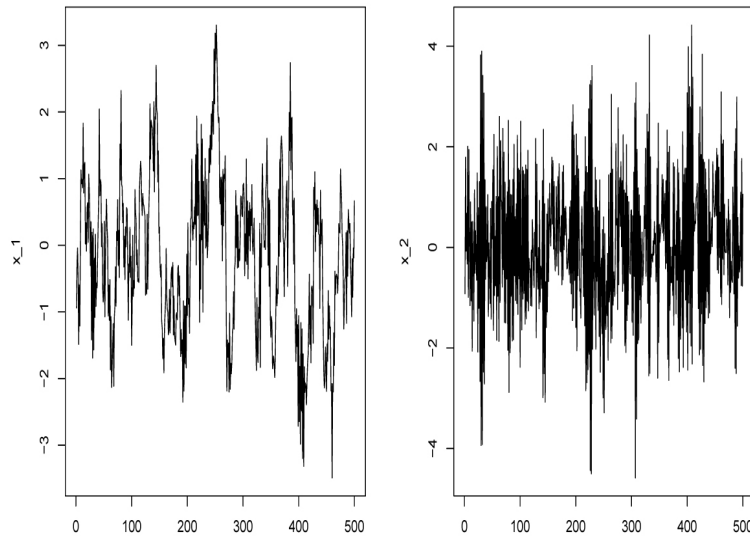


Figure 4. The time series plot of x_1 and x_2 .

$\phi x_{t-1} + \epsilon_t$, where $\phi = \begin{pmatrix} 0.8 & -0.3 \\ -0.7 & -0.7 \end{pmatrix}$, ϵ_t 's are i.i.d, $N_2(\mu, \Sigma)$, $\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and $\Sigma = \begin{pmatrix} 0.4 & 0.3 \\ 0.3 & 0.5 \end{pmatrix}$. The curve of the x_1 and x_2 is given in Figure 4. It

shows that the dataset follows the stationary assumption. We consider the first-order vector autoregressive, VAR(1), second-order vector autoregressive, VAR(2) and third-order vector autoregressive, VAR(3) models with Normal innovation as competing models. We will ignore the true model and estimate competing models using MLE and available data and do 10^4 replications. The results for all estimation procedures are given for different sample sizes, $n=50, 150, 250, 500$, which have summarized in Table 3. In this Table the average, across replications, estimates of the parameters are presented. It shows that, as the sample size increase the value of estimators are closed to the true parameters.

Using generated data and estimated parameters, the value of Vuong's test statistic for each pair of competing models is computed. The results are given in Table 4. The corresponding generated data and competing model curves are given in Figure 5. It shows that the estimated VAR(1) is appropriate

Table 3. The value of estimated parameters in mis-specification case

model	n	μ	Σ	ϕ
VAR(1)	50	$\begin{pmatrix} -5.0671e-05 \\ 8.7081e-04 \end{pmatrix}$	$\begin{pmatrix} 0.3703 & 0.2783 \\ 0.2783 & 0.4674 \end{pmatrix}$	$\begin{pmatrix} 0.7024 & -0.2595 \\ -0.7564 & -0.6658 \end{pmatrix}$
	150	$\begin{pmatrix} -4.2445e-05 \\ 5.9353e-04 \end{pmatrix}$	$\begin{pmatrix} 0.3916 & 0.2938 \\ 0.2938 & 0.4900 \end{pmatrix}$	$\begin{pmatrix} 0.7090 & -0.2871 \\ -0.7170 & -0.6885 \end{pmatrix}$
	250	$\begin{pmatrix} -1.4359e-05 \\ 3.8985e-04 \end{pmatrix}$	$\begin{pmatrix} 0.3952 & 0.2963 \\ 0.2963 & 0.4942 \end{pmatrix}$	$\begin{pmatrix} 0.7834 & -0.2924 \\ -0.7091 & -0.6927 \end{pmatrix}$
	500	$\begin{pmatrix} 2.4512e-05 \\ 1.0329e-04 \end{pmatrix}$	$\begin{pmatrix} 0.3975 & 0.2980 \\ 0.2980 & 0.4969 \end{pmatrix}$	$\begin{pmatrix} 0.7920 & -0.2963 \\ -0.7043 & -0.6963 \end{pmatrix}$
VAR(2)	50	$\begin{pmatrix} 0.0004 \\ 0.0012 \end{pmatrix}$	$\begin{pmatrix} 0.3533 & 0.2656 \\ 0.2656 & 0.4461 \end{pmatrix}$	$\begin{pmatrix} 0.7534 & -0.2948 & -0.0753 & -0.0046 \\ -0.7160 & -0.7190 & -0.0746 & -0.0318 \end{pmatrix}$
	150	$\begin{pmatrix} 6.4697e-05 \\ 7.1787e-04 \end{pmatrix}$	$\begin{pmatrix} 0.3861 & 0.2897 \\ 0.2897 & 0.4832 \end{pmatrix}$	$\begin{pmatrix} 0.7883 & -0.2988 & -0.0241 & -0.0020 \\ -0.7028 & -0.7074 & -0.0259 & -0.0113 \end{pmatrix}$
	250	$\begin{pmatrix} 3.3283e-05 \\ 4.3998e-04 \end{pmatrix}$	$\begin{pmatrix} 0.3920 & 0.2939 \\ 0.2939 & 0.4902 \end{pmatrix}$	$\begin{pmatrix} 0.7927 & -0.2988 & -0.0127 & -0.0009 \\ -0.7016 & -0.7035 & -0.0142 & -0.0068 \end{pmatrix}$
	500	$\begin{pmatrix} 2.8029e-05 \\ 3.1587e-04 \end{pmatrix}$	$\begin{pmatrix} 0.3959 & 0.2968 \\ 0.2968 & 0.4949 \end{pmatrix}$	$\begin{pmatrix} 0.7959 & -0.2989 & -0.0054 & -0.0005 \\ -0.7007 & -0.7015 & -0.0068 & -0.0032 \end{pmatrix}$
VAR(3)	50	$\begin{pmatrix} 0.0003 \\ 0.0012 \end{pmatrix}$	$\begin{pmatrix} 0.3361 & 0.2527 \\ 0.2527 & 0.4240 \end{pmatrix}$	$\begin{pmatrix} 0.7497 & -0.2930 & -0.0511 & -0.0044 & -0.0216 & 0.0142 \\ -0.7186 & -0.7202 & -0.0113 & -0.0892 & -0.0995 & -0.0207 \end{pmatrix}$
	150	$\begin{pmatrix} -9.1812e-05 \\ -9.8953e-04 \end{pmatrix}$	$\begin{pmatrix} 0.3807 & 0.2856 \\ 0.2856 & 0.4763 \end{pmatrix}$	$\begin{pmatrix} 0.7878 & -0.2986 & -0.0163 & -0.0014 & -0.0064 & 0.0048 \\ -0.7032 & -0.7074 & -0.0051 & -0.0286 & -0.0307 & -0.0057 \end{pmatrix}$
	250	$\begin{pmatrix} 3.8602e-05 \\ 4.7270e-04 \end{pmatrix}$	$\begin{pmatrix} 0.3887 & 0.2914 \\ 0.2914 & 0.4861 \end{pmatrix}$	$\begin{pmatrix} 0.7926 & -0.2983 & -0.0085 & -0.0001 & -0.0032 & 0.0030 \\ -0.7019 & -0.7036 & -0.0018 & -0.0170 & -0.0180 & -0.0035 \end{pmatrix}$
	500	$\begin{pmatrix} 3.3795e-05 \\ 3.7285e-04 \end{pmatrix}$	$\begin{pmatrix} 0.3942 & 0.2955 \\ 0.2955 & 0.4929 \end{pmatrix}$	$\begin{pmatrix} 0.7959 & -0.2988 & -0.0032 & -0.0001 & -0.0018 & 0.0012 \\ -0.7007 & -0.7015 & -0.0005 & -0.0084 & -0.0092 & -0.0017 \end{pmatrix}$

Table 4. The value of Vuong's statistic

	n	Value of statistic	Result
VAR(1)-VAR(2)	50	- 117.7982	VAR(1) is better than VAR(2)
VAR(1)-VAR(3)		-554.1708	VAR(1) is better than VAR(3)
VAR(2)-VAR(3)		-436.3726	VAR(2) is better than VAR(3)
VAR(1)-VAR(2)	150	-628.5196	VAR(1) is better than VAR(2)
VAR(1)-VAR(3)		-2528.114	VAR(1) is better than VAR(3)
VAR(2)-VAR(3)		-3156.633	VAR(2) is better than VAR(3)
VAR(1)-VAR(2)	250	-506.4068	VAR(1) is better than VAR(2)
VAR(1)-VAR(3)		-3540.47	VAR(1) is better than VAR(3)
VAR(2)-VAR(3)		-3034.063	VAR(2) is better than VAR(3)
VAR(1)-VAR(2)	500	-360.1438	VAR(1) is better than VAR(2)
VAR(1)-VAR(3)		-828.507	VAR(1) is better than VAR(3)
VAR(2)-VAR(3)		-468.3631	VAR(2) is better than VAR(3)

model to fit the generated data. Also the relative frequency of proposed model selection test for each of rejection-acceptance regions are computed and the results are summarized in Table 5. For example, if $n=500$, when we test $H_0^f : VAR(1)$ against $H_1^g : VAR(2)$, the value of the relative frequency of Vuong's test for rejection-acceptance region is 0.9036, see column 3. It shows that Vuong's test select estimated VAR(1) as optimal model. Also we see that the power of proposed test at least is as good as the power of Vuong's test.

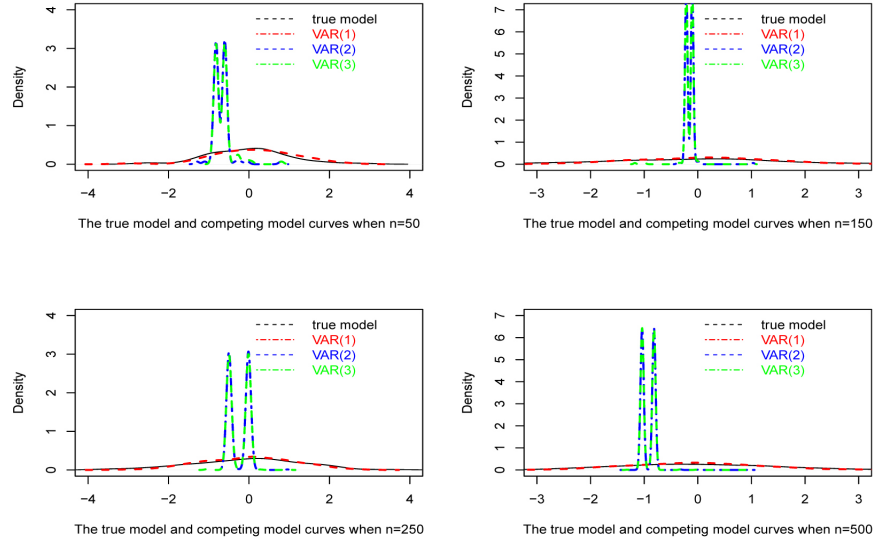


Figure 5. The true model and competing model curves.

Table 5. The relative frequency of Vuong's and proposed model selection tests

n	Conclusion	$H_0^f : VAR(1)$	$H_0^f : VAR(1)$	$H_0^f : VAR(2)$	$H_0 : \Psi = \mathbf{0}$
		$H_1^g : VAR(2)$	$H_1^g : VAR(3)$	$H_1^g : VAR(3)$	$H_1 : \Psi \neq \mathbf{0}$
50	f is better	0.7086	0.8116	0.7227	0.0120
	g is better	0.1650	0.0743	0.1640	0.9797
	f & g equi	0.1264	0.1141	0.1133	0.0252
150	f is better	0.8235	0.8860	0.7942	0.0091
	g is better	0.0033	0.0200	0.1129	0.9877
	f & g equi	0.1732	0.0940	0.0929	0.0083
250	f is better	0.8892	0.9056	0.8870	0.0066
	g is better	0.0259	0.0108	0.0321	0.9908
	f & g equi	0.0849	0.0836	0.0809	0.0059
500	f is better	0.9036	0.9476	0.8912	0.0071
	g is better	0.0237	0.0056	0.0380	0.9911
	f & g equi	0.0727	0.0468	0.0708	0.0035

In the following, we will show that, when Vuong's test select the two competing models as equivalent models, Cox's test select two models as suitable or unsuitable equivalent models. Consider $y_t = \phi y_{t-1} + \epsilon_t$ where $\phi = \begin{pmatrix} 0.6 & -0.01 \\ -0.02 & 0.1 \end{pmatrix}$, ϵ_t 's are i.i.d random variables with $N(\mu, \Sigma)$, $\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and $\Sigma = \begin{pmatrix} 0.8 & 0.01 \\ 0.01 & 0.6 \end{pmatrix}$ as true model and VAR(1), VAR(2), VAR(3) as competing models. The relative frequency of Vuong's and Cox's results are given in Table 6. It shows that for different n, Vuong's test consider the competing models as equivalent models and Cox's test emphasize that the models are suitable equivalent models. Since competing models, VAR(1), VAR(2) and VAR(3), are suitable equivalent models, so we select estimated VAR(1) model as optimal model.

Also, in a similar simulation, the first-order autoregressive model $x_t = \phi x_{t-1} + \epsilon_t$ is considered, where $\phi = \begin{pmatrix} 0.8 & -0.3 \\ -0.7 & -0.7 \end{pmatrix}$, ϵ_t 's are i.i.d, $N_2(\mu, \Sigma)$, $\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and $\Sigma = \begin{pmatrix} 5 & 3 \\ 3 & 7 \end{pmatrix}$. Due to the similarity, the obtained results are not presented here.

5 Continue Motivating Example

Based on the sample autocorrelation function and causality test, we consider VAR(1)-VAR(3) models with Normal innovation as competing models. The parameters of competing models are estimated by using MLE and available data. The value of maximum likelihood estimators are given in Table 7. Using data and estimated parameters, the value of the proposed, Vuong's and Cox's tests for paired of competing models are computed. The results are given in Table 8. It shows if $\alpha = 0.05$, the Vuong's test decides VAR(1) is better than VAR(2) and VAR(3) and selects second-order and third-order estimated vector autoregressive models with Normal innovations as equivalent models. Also the Vuong's test find that VAR(1) model as optimal model when $\alpha = 0.1$. When $\alpha = 0.05$ (and 0.1), the Cox's test decides VAR(1) is better than VAR(2) and VAR(3) and selects VAR(2) and VAR(3) models with Normal innovation as equivalent models. For both α , the proposed test

Table 6. The relative frequency of Vuong's and Cox's tests

	n	Conclusion	$H_0^f : VAR(1)$	$H_0^f : VAR(1)$	$H_0^f : VAR(2)$
			$H_1^g : VAR(2)$	$H_1^g : VAR(3)$	$H_1^g : VAR(3)$
Vuong's test	50	f is better	0.142	0.236	0.162
		g is better	0.037	0.038	0.043
		f & g are equ	0.821	0.726	0.795
	150	f is better	0.048	0.006	0.040
		g is better	0.057	0.095	0.080
		f & g are equ	0.895	0.899	0.880
	250	f is better	0.002	0.020	0.054
		g is better	0.029	0.031	0.048
		f & g are equ	0.969	0.949	0.898
	500	f is better	0.011	0.049	0.001
		g is better	0.007	0.002	0.020
		f & g are equ	0.982	0.949	0.979
Cox's test	50	f is better	0.001	0.298	0.294
		g is better	0.000	0.000	0.006
		reject both	0.000	0.000	0.000
		accept both	0.999	0.702	0.700
	150	f is better	0.000	0.008	0.005
		g is better	0.000	0.000	0.000
		reject both	0.000	0.000	0.000
		accept both	1.000	0.922	0.995
	250	f is better	0.000	0.000	0.000
		g is better	0.000	0.000	0.000
		reject both	0.000	0.000	0.000
		accept both	1.000	1.000	1.000
	500	f is better	0.000	0.000	0.000
		g is better	0.000	0.000	0.000
		reject both	0.000	0.000	0.000
		accept both	1.000	1.000	1.000

Table 7. The value of estimated parameters of competing models

model	μ	Σ	ϕ
VAR(1)	$\begin{pmatrix} 3.308e^{-18} \\ -6.799e^{-15} \end{pmatrix}$	$\begin{pmatrix} 0.0001 & -0.0077 \\ -0.0077 & 15.5977 \end{pmatrix}$	$\begin{pmatrix} 0.0210 & 2.5370 \\ -0.3510 & -0.1105 \\ -0.0011 & 0.7492 \end{pmatrix}$
VAR(2)	$\begin{pmatrix} 3.363e^{-18} \\ -2.173e^{-14} \end{pmatrix}$	$\begin{pmatrix} 0.0001 & -0.0091 \\ -0.0091 & 15.3042 \end{pmatrix}$	$\begin{pmatrix} 0.0206 & 1.3535 \\ -0.3690 & 0.1138 \\ -0.0015 & 0.6433 \\ -0.1292 & 0.2830 \\ 0.0004 & 0.1880 \end{pmatrix}$
VAR(3)	$\begin{pmatrix} -2.157e^{-17} \\ 5.432e^{-14} \end{pmatrix}$	$\begin{pmatrix} 0.0001 & -0.0079 \\ -0.0079 & 14.6203 \end{pmatrix}$	$\begin{pmatrix} 0.0205 & 0.6645 \\ -0.3522 & 0.1004 \\ -0.0014 & 0.59691 \\ -0.1986 & 0.6175 \\ 0.0006 & 0.0775 \\ 0.1573 & -0.3408 \\ -0.0002 & 0.2259 \end{pmatrix}$

does reject null hypothesis, $\Psi = 0$ and select first-order estimated vector autoregressive model with Normal innovation as optimal model. Also this test decides VAR(1) is better than VAR(2) and VAR(3).

Now we consider two important performance measures, Mean Absolute Error, MAE, and Mean Squared Error, MSE, which are frequently used by researchers. The mean absolute error is defined

$$MAE = \frac{1}{T} \sum_{t=1}^T |e_t|,$$

where $e_t = y_t - \hat{y}_t$, y_t is the actual value and \hat{y}_t is the forecasted value. It measures the average absolute deviation of forecasted values from original

Table 8. The results of the proposed, Vuong's and Cox's tests.

Vuong's test		The results when $\alpha = 0.05$		$\chi^2_{0.975}$	$\chi^2_{0.95}$
$H_0^f : VAR(1)$ $H_1^g : VAR(2)$	9.2507	VAR(1) is better than VAR(2)		7.3777	5.9914
$H_0^f : VAR(3)$ $H_1^g : VAR(2)$	-4.2992	VAR(2) and VAR(3) are equivalent		7.3777	5.9914
$H_0^f : VAR(1)$ $H_1^g : VAR(3)$	12.5500	VAR(1) is better than VAR(3)		11.1432	9.4877

Cox's test	T_{fg}	T_{gf}	The results when $\alpha = 0.05$	$C_{0.05}$	$C_{0.1}$
$H_0^f : VAR(1)$ $H_1^g : VAR(2)$	0.5160	-1.9666	VAR(1) is better than VAR(2)	1.9599	1.6448
$H_0^f : VAR(3)$ $H_1^g : VAR(2)$	0.6337	-1.1807	VAR(2) and VAR(3) are equivalent	1.9599	1.6448
$H_0^f : VAR(1)$ $H_1^g : VAR(3)$	0.3616	-1.9606	VAR(1) is better than VAR(2)	1.9599	1.6448

Proposed test	The results when $\alpha = 0.05$		$\chi^2_{0.05}$	$\chi^2_{0.025}$
$H_0^f : VAR(1)$ $H_1^g : VAR(2)$	5.6547	VAR(1) is better than VAR(2)	9.4877	11.1432
$H_0^f : VAR(1)$ $H_1^g : VAR(3)$	14.6824	VAR(1) is better than VAR(3)	15.5073	17.5345
$H_0^f : VAR(2)$ $H_1^g : VAR(3)$	9.1343	VAR(2) is better than VAR(3)	9.4877	11.1432

Table 9. The values of MSE, MAE.

model	VAR(1)	VAR(2)	VAR(3)
MSE	1.7099	1.7493	1.7482
MAE	1.3218	1.3342	1.3334

ones. On the other hand the mean squared error is

$$MSE = \frac{1}{T} \sum_{t=1}^T e_t^2.$$

This criteria emphasizes on the fact that the total forecast error is in fact much affected by large individual errors, i.e. large errors are much expensive than small errors. The main purposes is fitting an vector autoregressive model that has the smallest MSE and MAE.

For variable QG, consider p-order vector autoregressive as

$$\mathbf{y}_t = \Phi_1 \mathbf{y}_{t-1} + \dots + \Phi_p \mathbf{y}_{t-p} + \epsilon_t,$$

where \mathbf{y}_t is the observed process and ϵ_t is its innovation which is distributed as Normal, $N(\mu, \sigma^2)$. In bivariate autoregressive, $y_t = (y_{t,QG}, y_{t,VB})$, Φ_i 's are $(2 \times p)$ coefficient matrices, ϵ_t is an (2×1) vector of Normal variables. The 80 rolling forecasts of proposed models are computed and the MSE and MAE are given in Table 9. The presented results of 80 rolling prediction show that for QG, the first-order vector autoregressive model has the smallest MSE and MAE.

Figure 6 shows the histogram of observations and estimated competing models, VAR(1), VAR(2) and VAR(3). It shows that the VAR(1) is optimal model for observations.

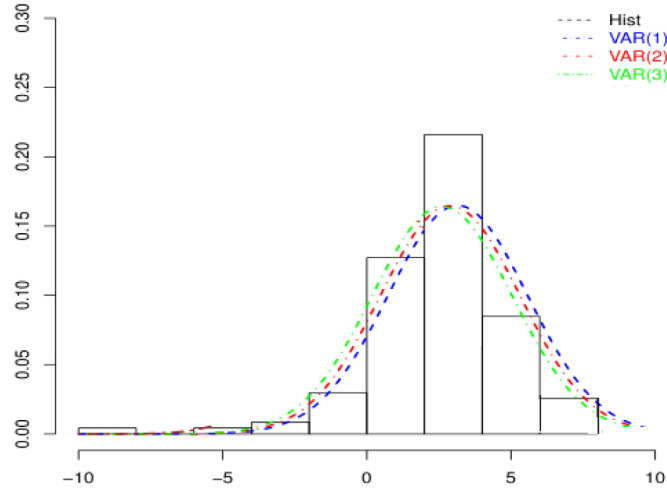


Figure 6. Histogram of observation and VAR(1), VAR(2) and VAR(3) model curve.

6 Conclusion

In this paper, we consider the Europe oil prices, Brent, and the real gross domestic product, GDP. These data describe the information of growth in oil price and economic growth. The results of the rolling test of correlation and Granger causality show that the Europe oil prices is related to the real gross domestic product and the Europe oil prices data does Granger-cause the real gross domestic product data. So, we have suggested the vector autoregressive model as competing models and have selected the order and a suitable subset of regressors in vector autoregressive model using model selection test such as Vuong's test, Cox's test and the proposed test. The results show that, the model selection tests confirm the causality test and results of predictive method and select first-order vector autoregressive model as optimal model for GDP data.

It is important to note that in this paper, linear models are considered as competing models. If other time series models such as GARCH or State Space models are considered as competing models, the data may be better fitted in other models.

References

- Aaltonen, J. and Ostermark, R. (1997). A rolling test of Granger causality between the Finnish and Japanese security. *Omega, Int. J. Mgmt. Sci.*, **25**, 635-642.
- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. Presented at the Second International Symposium on Information Theory, eds. by B.N. Petrov and F. Csake, Akademiai Kiado, Hungary, 267-281.
- Cox, D.R. (1961). Tests of non-nested families of hypotheses. Proceeding of the Fourth Berkeley Symposium on *Mathematical statistics and Probability*, **1**, 105-123.
- Cox, D.R. (1962). Further results on tests of separate families of hypotheses. *J. Royal Statistical Society*, **24**, 406-424.
- Eichler, M. (2012). *Causal inference in time series analysis*. In Berzuini, Carlo. *Causality: statistical perspectives and applications*. Hoboken, N.J.: Wiley, 327-352.
- Granger, C.W.J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, **37**, 424-438.
- Granger, C.W.J. (2004). Time series analysis, cointegration and applications. *American Economic Review*, **94**, 421-425.
- Hamilton, J.D., and Herrera, A.M. (2004). Oil shocks and aggregate macroeconomic behavior: the role of monetary policy. *J. Money. Credit. and Banking*, **36**, 265-286.
- Hannan, E.J., and Quinn, B.G. (1979). The determination of the order of an autoregression. *J. Royal Statistical Society*, **41**, 190-195.
- Kilian, L. (2001). Impulse response analysis in vector autoregressions with unknown lag order. *J. Forecasting*, **20**, 161-179.
- Nicholls, D, F. (1976). The efficient estimation of vector linear time series models. *J. Biometrika*, **63**, 381-390.
- Nicholls, D.F. (1977). A comparison of estimation methods for vector linear time series models. *J. Biometrika*, **64**, 85-90.
- Paulsen, J., and Tjostheim, D. (1985). On the estimation of residual variance and order in autoregressive time series. *J. Royal Statistical Society*, **47**, 216-228.
- Quinn, B.G. (1980). Order determination for a multivariate autoregression. *J. Royal Statistical Society*, **42**, 182-185.
- Quinn, B.G. (1988). A note on AIC order determination for multivariate autoregressions. *J. Time Series Analysis*, **9**, 241-245.

Sayyareh, A., Obeidi, R., and Bar-Hen, A. (2011). Empirical comparison between some model selection criteria. *J. Communications in Statistics Simulation and Computation*, **40**, 72-86.

Schwarz, G. (1978). Estimating the dimension of a model. *J. Annals of Statistics*, **6**, 461-464.

Shibata, R. (1984). Approximate efficiency of a selection procedure for the number of regression variables. *J. Biometrika*, **71**, 43-49.

Sims, C.A. (1980). Macroeconomics and reality. *J. Econometrica*, **48**, 1-48.

Vuong, Q.H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *J. Econometrica*, **57**, 307-333.

White, H. (1982). Maximum likelihood estimation of misspecified models, *J. Econometrica*, **50**, 1-25.

Wei, W.S. (2006). *Time series analysis: univariate and multivariate methods*. 2nd Edition.

Zivot, E. and Wang, J. (2006). *Modeling financial time series with S-Plus*. Springer, ISBN: 978-0-387-27965-7.

Appendix

Proof of Theorem 1: Assume $H_0 : \Psi = \mathbf{0}$. Then $Y_t^* = \Theta_1 Y_{t-1}^* + \epsilon_t$. In fact the vector autoregressive model reduce to autoregressive model. So,

$$Y_{1t}^* = \phi_{11}^1 Y_{1t-1}^* + \epsilon_{1t},$$

where ϵ_{1t} 's are independent and identically distributed as Normal distribution with zero mean and finite variance, σ_{11}^2 . From the fact that

$$L = 2L(\hat{\gamma}_0) - 2L(\hat{\gamma}_T)$$

We have

$$\begin{aligned} L &= -KT \log 2\pi - T \log |\hat{\Omega}_0| - \sum_{t=1}^T \hat{\epsilon}_t' \hat{\Omega}_0^{-1} \hat{\epsilon}_t \\ &\quad + KT \log 2\pi + T \log |\hat{\Omega}_T| + \sum_{t=1}^T \hat{\epsilon}_t' \hat{\Omega}_T^{-1} \hat{\epsilon}_t \\ &= T \left(\log |\hat{\Omega}_T| - \log |\hat{\Omega}_0| \right) \\ &= T \left(\text{tr} \left(\log \hat{\Omega}_T \right) - \text{tr} \left(\log \hat{\Omega}_0 \right) \right) \\ &= T \left(\text{tr} \left(\log \hat{\Omega}_T - \log \hat{\Omega}_0 \right) \right) \\ &= T \left(\text{tr} \left(\log \hat{\Omega}_T \hat{\Omega}_0^{-1} \right) \right) \\ &= T \left(\text{tr} \left(\hat{\Omega}_T \hat{\Omega}_0^{-1} - \mathbf{I} \right) \right) + o_p(1). \end{aligned} \tag{4}$$

Based on the first-order Taylor series expansion, the last equation is true where \mathbf{I} is a $K \times K$ identity matrix,

$$\begin{aligned}
\hat{\Omega}_T &= \frac{1}{T} \sum_{t=1}^T \hat{\epsilon}_t \hat{\epsilon}_t' = \frac{1}{T} \sum_{t=1}^T \left(y_t^* - \hat{\Theta}_T y_{t-1}^* - \hat{\Psi}_T y_{t-1}^* \right) \left(y_t^* - \hat{\Theta}_T y_{t-1}^* - \hat{\Psi}_T y_{t-1}^* \right)' \\
&= \frac{1}{T} \sum_{t=1}^T \left(\Theta y_{t-1}^* + \Psi y_{t-1}^* + \epsilon_t - \hat{\Theta}_T y_{t-1}^* - \hat{\Psi}_T y_{t-1}^* \right) \left(\Theta y_{t-1}^* + \Psi y_{t-1}^* + \epsilon_t - \hat{\Theta}_T y_{t-1}^* - \hat{\Psi}_T y_{t-1}^* \right)' \\
&= \frac{1}{T} \sum_{t=1}^T \left((\Theta - \hat{\Theta}_T) y_{t-1}^* + (\Psi - \hat{\Psi}_T) y_{t-1}^* + \epsilon_t \right) \left((\Theta - \hat{\Theta}_T) y_{t-1}^* + (\Psi - \hat{\Psi}_T) y_{t-1}^* + \epsilon_t \right)' \\
&= \frac{1}{T} \sum_{t=1}^T (\Theta - \hat{\Theta}_T) y_{t-1}^* y_{t-1}^{*'} (\Theta - \hat{\Theta}_T)' + \frac{1}{T} \sum_{t=1}^T (\Psi - \hat{\Psi}_T) y_{t-1}^* y_{t-1}^{*'} (\Psi - \hat{\Psi}_T)' + \frac{1}{T} \sum_{t=1}^T \epsilon_t \epsilon_t' \\
&\quad + \frac{2}{T} \sum_{t=1}^T (\Theta - \hat{\Theta}_T) y_{t-1}^* y_{t-1}^{*'} (\Psi - \hat{\Psi}_T)' + \frac{2}{T} \sum_{t=1}^T (\Theta - \hat{\Theta}_T) y_{t-1}^* \epsilon_t' + \frac{2}{T} \sum_{t=1}^T (\Psi - \hat{\Psi}_T) y_{t-1}^* \epsilon_t'.
\end{aligned} \tag{5}$$

The straight calculations give

$$\frac{2}{T} \sum_{t=1}^T (\Theta - \hat{\Theta}_T) y_{t-1}^* y_{t-1}^{*'} (\Psi - \hat{\Psi}_T)' = 0$$

and the fifth and sixth term in (5) are $o_p(1)$, since

$$\frac{1}{T} \sum_{t=1}^T y_{t-1}^* \epsilon_t' \xrightarrow{P} E \left\{ Y_{t-1}^* \epsilon_t' \right\} = \mathbf{0}.$$

Similarly

$$\begin{aligned}
\hat{\Omega}_0 &= \frac{1}{T} \sum_{t=1}^T \hat{\epsilon}_t \hat{\epsilon}_t' = \frac{1}{T} \sum_{t=1}^T \left(y_t^* - \hat{\Theta}_T y_{t-1}^* \right) \left(y_t^* - \hat{\Theta}_T y_{t-1}^* \right)' \\
&= \frac{1}{T} \sum_{t=1}^T \left(\Theta y_{t-1}^* + \epsilon_t - \hat{\Theta}_T y_{t-1}^* \right) \left(\Theta y_{t-1}^* + \epsilon_t - \hat{\Theta}_T y_{t-1}^* \right)' \\
&= \frac{1}{T} \sum_{t=1}^T \left((\Theta - \hat{\Theta}_T) y_{t-1}^* + \epsilon_t \right) \left((\Theta - \hat{\Theta}_T) y_{t-1}^* + \epsilon_t \right)' \\
&= \frac{1}{T} \sum_{t=1}^T (\Theta - \hat{\Theta}_T) y_{t-1}^* y_{t-1}^{*'} (\Theta - \hat{\Theta}_T)' + \frac{1}{T} \sum_{t=1}^T \epsilon_t \epsilon_t' \\
&\quad + \frac{2}{T} \sum_{t=1}^T (\Theta - \hat{\Theta}_T) y_{t-1}^* \epsilon_t'.
\end{aligned} \tag{6}$$

The third term in (6) are $o_p(1)$, since

$$\frac{1}{T} \sum_{t=1}^T y_{t-1}^* \epsilon_t' \xrightarrow{P} E \left\{ Y_{t-1}^* \epsilon_t' \right\} = \mathbf{0}.$$

Substituted (5) and (6) in (3), so

$$\begin{aligned} L &= T \left(\text{tr} \left(\hat{\Omega}_T \hat{\Omega}_0^{-1} - \mathbf{I} \right) \right) + o_p(1) = T \left(\text{tr} \left[\left(\hat{\Omega}_T - \hat{\Omega}_0 \right) \hat{\Omega}_0^{-1} \right] \right) + o_p(1) \\ &= T \left(\text{tr} \left[\left(\frac{1}{T} \sum_{t=1}^T \left(\Psi - \hat{\Psi}_T \right) y_{t-1}^* y_{t-1}^{*'} \left(\Psi - \hat{\Psi}_T \right)' \right) \hat{\Omega}_0^{-1} \right] \right) + o_p(1) \\ &= \text{tr} \left[\left(\Psi - \hat{\Psi}_T \right) \sum_{t=1}^T y_{t-1}^* y_{t-1}^{*'} \left(\Psi - \hat{\Psi}_T \right)' \hat{\Omega}_0^{-1} \right] + o_p(1) \end{aligned} \quad (7)$$

where $\hat{\Phi}_T = \hat{\Phi}_0$, $\hat{\Omega}_0 \xrightarrow{P} \Omega_0$ and

$$\Omega_0 = \begin{bmatrix} \sigma_{11}^2 & 0 & \dots & 0 \\ 0 & \sigma_{22}^2 & \dots & 0 \\ \vdots & & & \\ 0 & 0 & \dots & \sigma_{kk}^2 \end{bmatrix}.$$

Note that the first component of main diagonal, C_1 of L , is

$$\begin{aligned} C_1 &= \hat{\sigma}_{11}^{-2} (\phi_{12}^1 - \hat{\phi}_{12}^1) \sum_{t=1}^T y_{2t-1}^* y_{2t-1}^* (\phi_{12}^1 - \hat{\phi}_{12}^1) + \hat{\sigma}_{11}^{-2} (\phi_{13}^1 - \hat{\phi}_{13}^1) \sum_{t=1}^T y_{3t-1}^* y_{3t-1}^* (\phi_{12}^1 - \hat{\phi}_{12}^1) + \dots \\ &\quad + \hat{\sigma}_{11}^{-2} (\phi_{1k}^1 - \hat{\phi}_{1k}^1) \sum_{t=1}^T y_{kt-1}^* y_{kt-1}^* (\phi_{12}^1 - \hat{\phi}_{12}^1) + \dots + \hat{\sigma}_{11}^{-2} (\phi_{12}^1 - \hat{\phi}_{12}^1) \sum_{t=1}^T y_{2t-1}^* y_{kt-1}^* (\phi_{1k}^1 - \hat{\phi}_{1k}^1) \\ &\quad + \hat{\sigma}_{11}^{-2} (\phi_{13}^1 - \hat{\phi}_{13}^1) \sum_{t=1}^T y_{3t-1}^* y_{kt-1}^* (\phi_{1k}^1 - \hat{\phi}_{1k}^1) + \dots + \hat{\sigma}_{11}^{-2} (\phi_{1k}^1 - \hat{\phi}_{1k}^1) \sum_{t=1}^T y_{kt-1}^* y_{kt-1}^* (\phi_{1k}^1 - \hat{\phi}_{1k}^1) \\ &= \hat{\sigma}_{11}^{-2} (\phi_{12}^1 - \hat{\phi}_{12}^1) \sum_{t=1}^T y_{2t-1}^* y_{2t-1}^* (\phi_{12}^1 - \hat{\phi}_{12}^1) + \hat{\sigma}_{11}^{-2} (\phi_{13}^1 - \hat{\phi}_{13}^1) \sum_{t=1}^T y_{3t-1}^* y_{3t-1}^* (\phi_{13}^1 - \hat{\phi}_{13}^1) \\ &\quad + \dots + \hat{\sigma}_{11}^{-2} (\phi_{1k}^1 - \hat{\phi}_{1k}^1) \sum_{t=1}^T y_{kt-1}^* y_{kt-1}^* (\phi_{1k}^1 - \hat{\phi}_{1k}^1) + o_P(1). \end{aligned} \quad (8)$$

Since under null hypothesis, when $i \neq j$, we have

$$\frac{1}{T} \sum_{t=1}^T y_{it-1}^* y_{jt-1}^* \xrightarrow{P} 0.$$

then the last term in (8) is $o_p(1)$. The first term in (8) is asymptotically distributed as χ_1^2 , since

$$\begin{aligned} & \hat{\sigma}_{11}^{-2} (\phi_{12}^1 - \hat{\phi}_{12}^1) \sum_{t=1}^T y_{2t-1}^* y_{2t-1}^* (\phi_{12}^1 - \hat{\phi}_{12}^1) \\ &= \hat{\sigma}_{11}^{-2} \frac{\sum_{t=1}^T y_{2t-1}^* e_{1t}}{\sum_{t=1}^T y_{2t-1}^* y_{2t-1}^*} \sum_{t=1}^T y_{2t-1}^* y_{2t-1}^* \frac{\sum_{t=1}^T y_{2t-1}^* e_{1t}}{\sum_{t=1}^T y_{2t-1}^* y_{2t-1}^*} \\ &= \hat{\sigma}_{11}^{-2} \frac{\left(\sum_{t=1}^T y_{2t-1}^* e_{1t} \right)^2}{\sum_{t=1}^T y_{2t-1}^* y_{2t-1}^*} \xrightarrow{D} \chi_1^2 \end{aligned}$$

Similarity the other terms in 8 have asymptotically chi-square, χ_1^2 , distribution, so

$$C_1 \xrightarrow{D} \chi_{k-1}^2.$$

The asymptotic distributions of other components of main diagonal are computed similarity, thus

$$L = tr \left[\left(\Psi - \hat{\Psi}_T \right) \sum_{t=1}^T y_{t-1}^* y_{t-1}^{*'} \left(\Psi - \hat{\Psi}_T \right)' \hat{\Omega}_0^{-1} \right] \xrightarrow{D} \chi_{k(k-1)}^2.$$

S. Zamani Mehreyan

Department of Statistics,
Imam Khomeini International University,
Qazvin, Iran.
email: zamani@sci.ikiu.ac.ir

A. Sayyareh

Department of Computer Science and Statistics,
K.N. Toosi University of Technology,
Tehran, Iran.
email: asayyareh@kntu.ac.ir