



Vector Autoregressive (VAR) Model for Economic Variables Forecast

Aliyu Abubakar Musa¹, Hamisu Abdulhamid² and Abubakar Musa¹

¹Department of Economics, Umar Suleiman College of Education, Gashua, Yobe State, Nigeria.

²Department of Agricultural Education, Umar Suleiman College of Education, Gashua, Yobe State, Nigeria

ABSTRACT

The aim of this study is to describe the procedure for estimation of Vector Autoregressive (VAR) Model parameters which is mostly used for forecasting and structural analysis. The procedures for the estimation of VAR and Vector Error Correction Model (VECM) have been explained. The study also described the procedures for Diagnostic test, Lag Order Selection (Information Criteria), Unit Roots test (Augmented Dickey-Fuller (ADF)), Structural Analysis (Granger Causality and Impulse Response Analysis), Random Walk Model, test for Cointegration and Forecasting.

Key words: VAR, VECM, Unit Roots Test, Cointegration, Forecast

1. Introduction

Autoregressive (AR) model is used to forecast future values of a variable where the future values respond to the present and past values of that particular variable, this is a univariate case. But most of the economic variables depend not only on their past values but past values of other related variables. For example, the future rate of unemployment may depend on its present and past values, present and a previous amount of GDP and even current and past level of other variables. When all the variables in the model are considered to be dependent variables and their past values as regressors for each and every dependent variable, that is the multivariate autoregressive and is considered as Vector Autoregressive (VAR) (Lutkepohl, 2005). AR model is expressed as

$$X_t = \mu + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \dots + \beta_p X_{t-p} + \epsilon_t \quad (1)$$

Where μ is a constant, $\beta_1 \dots \beta_p$ are parameters, $X_{t-1} \dots X_{t-p}$ are the lags of X_t and ϵ_t is the white noise. X_t process can also depend on the past values of the disturbance term, where it then becomes Moving Average (MA) process. At the same time also, it may simultaneously depend on its past values and past values of the white noise where it is called Autoregressive Moving Average (ARMA) process. It's commonly written as ARMA(p,q) where p is the number of lags in the AR process, and q is the number of lags in the MA process (Pfaff, 2008).

2. Vector Autoregressive

AR process is scalar because it depends on its lags only. But most of the times, the dynamic behaviour of economic variables are determined jointly, where the process (X_t) has now become vector ($N \times 1$) of the dependent variables (Box and Reinsel, 2008). There might be $X_{1t}, X_{2t}, \dots, X_{Nt}$ where N is the number of variables in the model. Each variable, X_{it} , depends on its lags and lags of every other variable in the model. In modelling exchange rate for example, if there are three variables (three currency combinations for instance), GBP/USD, GBP/AUD and USD/EUR, where the current value of each depends on its lags and lags of other variables, the VAR(1) (if it depends on only one lag) model can be represented as:

$$X_t = \mu_1 + \beta_1 X_{t-1} + \beta_2 Y_{t-1} + \beta_3 Z_{t-1} + \epsilon_{tx}$$

$$Y_t = \mu_2 + \alpha_1 X_{t-1} + \alpha_2 Y_{t-1} + \alpha_3 Z_{t-1} + \epsilon_{ty}$$

$$Z_t = \mu_3 + \gamma_1 X_{t-1} + \gamma_2 Y_{t-1} + \gamma_3 Z_{t-1} + \epsilon_{tz}$$

The model can more compactly be written as:

$$A_t = \mu + PA_{t-1} + \epsilon_t \quad (2)$$

$$A_t = \begin{bmatrix} X_t \\ Y_t \\ Z_t \end{bmatrix} P = \begin{bmatrix} \beta_1 & \beta_2 & \beta_3 \\ \alpha_1 & \alpha_2 & \alpha_3 \\ \gamma_1 & \gamma_2 & \gamma_3 \end{bmatrix} A_{t-1} = \begin{bmatrix} X_{t-1} \\ Y_{t-1} \\ Z_{t-1} \end{bmatrix} \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} \epsilon_t = \begin{bmatrix} \epsilon_{tx} \\ \epsilon_{ty} \\ \epsilon_{tz} \end{bmatrix}$$

Here, X_t can be GBP/USD, Y_t could be GBP/AUD and Z_t could be USD/EUR. A_t is $N \times 1$ matrix (N number of dependent variables), μ is $N \times 1$, P is $N \times N$, A_{t-1} is $N \times 1$ and ϵ_t is also $N \times 1$,

$$E[\epsilon_t] = 0 \quad \text{and} \quad E[\epsilon_t' \epsilon_s] = \begin{cases} \Sigma & \text{for } s = t \\ 0 & \text{otherwise} \end{cases}$$

Σ is $N \times N$ positive definite matrix. In this model, the number of parameters to estimate is $N^2 + N$. As the lags (p) or number of variables (N) increases, the estimation becomes more complicated. If the number of the lags increases by 1, the total number of parameters to estimate will increase by N which is greater than 1. For a VAR(p) model in equation (3), the number of parameters to estimate becomes $N + N^2(p)$, where p is the number of lags (Hamilton, 1994).

$$A_t = \mu + P_1 A_{t-1} + P_2 A_{t-2} + \dots + P_p A_{t-p} + \epsilon_t \tag{3}$$

Equation(2) may includes deterministic part (such as drift or trend or both) and exogenous variables;

$$A_t = P A_{t-1} + B D_t + G Y_t + \epsilon_t$$

Where D_t is $K \times 1$ deterministic variable, Y_t is $H \times 1$ vector of exogenous variables, and B and G are the parameters to be estimated (Zivot and Wang, 2006).

2.1 Stable VAR

In Lutkepohl (2005), stability is the necessary feature of VAR(p) process. In equation (2), considering the system that generates the process, A_t are determined by $A_0, \epsilon_1, \dots, \epsilon_t$ uniquely.

$$A_t = \mu + P A_{t-1} + \epsilon_t$$

$$A_t = (I_N + P + P^2 + \dots + P^{t-1})\mu + P^t A_0 + \sum_{i=0}^{t-1} P^i \epsilon_{t-i} \tag{4}$$

Equation (3) can also be written in terms of lag polynomial:

$$P(L)A_t = \mu + \epsilon_t$$

In practical studies, companion matrix is used to derive the stability of a VAR (p) process by computing the eigen values of the matrix. VAR(p) process can then equivalently be represented as VAR(1) process as:

$$\tilde{A}_t = \mu + P \tilde{A}_{t-1} + \epsilon_t$$

where:

$$A_t = \begin{bmatrix} A_t \\ A_{t-1} \\ \vdots \\ A_{t-1+p} \end{bmatrix}, P = \begin{bmatrix} P_1 & P_2 & \dots & P_{p-1} & P_p \\ I_N & 0 & \dots & 0 & 0 \\ 0 & I_N & & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I_N & 0 \end{bmatrix}, \mu = \begin{bmatrix} \mu \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \epsilon_t = \begin{bmatrix} \epsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

This process is stable if:

$$\det(I_{Np} - Pz) \neq 0, \text{ for } |z| \leq 1$$

with $E[A_t] = (I_{Np} - Pz)^{-1} \mu$. If the moduli of eigen values of P are less one, then A_t process is stable which implies that the means, variances and covariances are time-invariant. The parameters of the endogenous variables can be estimated using OLS from each equation and its equivalent to MLE if the ϵ_t is normally distributed (Pfaff, 2008).

If VAR (p) process is generated by a stable process, it fluctuates around its fixed mean and its variance remains constant as it fluctuates. But in practice most of the time series are not stable as they exhibit trend and non-constant variability (Lutkepohl, 2005).

2.1.1 Augmented Dickey Fuller Test of Unit Roots

One of the fundamental assumptions of VAR model is stationarity. The model is stationary if all of its roots lie outside the unit circle. Augmented Dickey-Fuller (ADF) test is usually conducted to know whether the model is stable or not. Before conducting the ADF test, the parameters of the model are estimated using OLS or MLE and the test is performed to know if the values are significantly different from 1 or not (Enders, 2015). In equation (1), the H_0 and H_1 should be tested as:

$H_0: P = 1$, and $H_1: P < 1$.

Accepting H_0 means the process is non-stationary and has unit root while rejecting H_0 implies stationarity (Martin and Harris, 2013).

Griffiths and Lim (2008), the normal procedure for conducting ADF test, is taking the first difference of the process as:

$$A_t - A_{t-1} = \mu + PA_{t-1} - A_{t-1} + \epsilon_t$$

$$\Delta A_t = \mu + \rho A_{t-1} + \epsilon_t$$

where $\rho = (P-1)$. In this case, the procedure involves testing whether ρ is significantly different from 0 or not with H_0 and H_1 modified as:

$H_0: \rho = 0$, and $H_1: \rho < 0$,

where the hypotheses can be stated more compactly as:

$H_0: \rho = 0 \Rightarrow H_0: P = 1$ and $H_1: \rho < 0 \Rightarrow H_1: P < 1$

According to Vinod (2008), to be sure that the errors are not serially correlated, more than one lag is used in the test as in the following equation:

$$\Delta A_t = \mu + \rho A_{t-1} + \sum_{c=1}^m b_c \Delta A_{t-c} + \epsilon_t$$

• Critical Values of ADF: After the estimation of the model, hypothesis that $\rho = 0$ is tested. But one challenge is that t-statistic does not follow t-distribution, hence τ -statistic is used for the test. A value of τ -statistic is compared with the value of τ critical (τ_c), if $\tau \leq \tau_c$, the null hypothesis is rejected, otherwise it is accepted (Griffiths and Lim, 2008).

2.2 Lag Order Selection

To select the appropriate order of VAR, the most common approaches are likelihood ratio test and information criteria

(IC) selection. The widely used information criteria are Akaike information criteria (AIC), Hannan information criteria (HIC) and Schwarz information criteria (SIC), where each IC may suggest different value of p. p that minimizes the information criteria is the most appropriate one (Tsay, 2014).

Martin and Harris (2013), gave the following values of information criteria:

$$AIC = -2 \ln L_T(\hat{\theta}) + \frac{2k}{T-s}$$

$$HIC = -2 \ln L_T(\hat{\theta}) + \frac{2k \ln(\ln(T-s))}{T-s}$$

$$SIC = -2 \ln L_T(\hat{\theta}) + \frac{k \ln(T-s)}{T-s} \quad (5)$$

where k represents the total amounts of parameters in the model, s is the maximum number of lags that the model can have (q_{max} and p_{max}), and

$$\ln L_T(\hat{\theta}) = -\frac{N}{2} (1 + \ln 2\pi) - \frac{1}{2} \ln |\hat{V}| \quad (6)$$

where $|V|$ is the estimate of covariance matrix. The first part of equation (6) is neglected since it is constant and doesn't influence the value of any of the ICs, hence equation (5) can be reformed to equation (7).

$$AIC = \ln |\hat{V}| + \frac{2k}{T-s}$$

$$HIC = \ln |\hat{V}| + \frac{2k \ln(\ln(T-s))}{T-s}$$

$$SIC = \ln |\hat{V}| + \frac{k \ln(T-s)}{T-s} \quad (7)$$

2.3 VAR Estimation

Maximum likelihood, least squares (both ordinary least squares and generalized least squares) and bayesian estimators are used in the estimation of VAR model parameters. When using least squares for a VAR(p) model, equation (3) can be considered where Σ is the covariance matrix of ϵ . For simplicity, the VAR(p) model as in equation (3) can be written as:

$$x_t' = y_t' \beta' + \epsilon_t'$$

where $\beta' = (P_0, P_1, \dots, P_p)$ is $k \times (kp + 1)$ matrix of parameters to be estimated, $y_t = (1, x_{t-1}, \dots, x_{t-p})'$. Putting

equation (3) in form of (8) is more convenient for the estimation the this transformation.

$$X = Y\beta + M \tag{8}$$

Applying the attributes of Kronecker product and vectorization of equation (8), one gets:

$$vec(X) = (I_k \otimes Y)vec(\beta) + vec(M) \text{ where the covariance matrix of the } vec(M) \text{ is } \epsilon \otimes I_{T-p} \Sigma \text{ (Tsay, 2014).}$$

With the same transformation (from equation 8), GLS estimates of β can be used to minimize:

$$S(\beta) = [vec(M)]'(\Sigma_\epsilon \otimes I_{T-p})^{-1}vec(A)$$

$$S(\beta) = [vec(X - Y\beta)]'(\Sigma_\epsilon^{-1} \otimes I_{T-p})vec(X - Y\beta)$$

$$S(\beta) = tr[(X - Y\beta)\Sigma_\epsilon^{-1}(X - Y\beta)']$$

Σ_ϵ being symmetric matrix and using the fact that $tr(QRS) = vec(S)'(R' \otimes I)vec(Q)$, one gets:

$$\begin{aligned} S(\beta) &= [vec(X) - (I_k \otimes Y)vec(\beta)]'(\Sigma_\epsilon^{-1} \otimes I_{T-p})[vec(X) - (I_k \otimes Y)vec(\beta)] \\ S(\beta) &= vec(X)'(\Sigma_\epsilon^{-1} \otimes I_{T-p})vec(X) - 2vec(\beta)'(\Sigma_\epsilon^{-1} \otimes Y')vec(X) \\ &\quad + vec(\beta)'(\Sigma_\epsilon^{-1} \otimes Y'Y)vec(\beta) \end{aligned} \tag{9}$$

Tsay (2014), obtaining the normal equations, partial derivatives of equation (9) is taken with respect to $vec(\beta)$ and set it equal to 0.

$$\frac{\partial vec(\beta)}{\partial vec(\beta)} = \dots \tag{10}$$

lastly, the estimate of VAR(p) from GLS method is given as:

$$vec(\beta) = (\Sigma_\epsilon^{-1} \otimes Y'Y)^{-1}(\Sigma_\epsilon^{-1} \otimes Y')vec(X)$$

$$vec(\beta) = [\Sigma_\epsilon^{-1} \otimes (Y'Y)^{-1}](\Sigma_\epsilon^{-1} \otimes Y')vec(X)$$

$$vec(\beta) = [I_k \otimes (Y'Y)^{-1}Y']vec(X) \text{ } vec(\beta) = vec[(Y^0Y)^{-1}(Y'X)]$$

hence,

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

This shows that GLS estimate of this model is the same as that of OLS in equation 8.

3. Residual Analysis

After the estimation of a VAR (p) model, the next most important task is to check whether the errors are consistent with the assumptions of the model. The following test should be conducted for the checks (Pfaff, 2008). Value of the estimated parameters from a model that has right specification should be so close to that of the real parameter (adequacy of the fitted model), and the residuals of the model should also behave normally with mean of zero and constant variance (white noise behaviour). The presence of a serial correlation is a sign of dynamic misspecification of the model. Plots of the model residuals should wander very close to the value of zero (Cryer and Chan, 2008).

1. Test for Autocorrelation

Testing for zero correlation in the data involves testing the null hypothesis that all the $\rho_i = 0$ against alternative that $\rho_i \neq 0$ (ρ is the univariate cross-correlation). For the multivariate data is best achieved using the Ljung-Box test which is known as the multivariate portmanteau test which is defined as:

$$Q_k(m) = T^2 \sum_{l=1}^m \frac{1}{T-l} tr(\hat{\Gamma}_l' \hat{\Gamma}_0^{-1} \hat{\Gamma}_l \hat{\Gamma}_0^{-1}) \tag{11}$$

where $\hat{\Gamma}_l = T^{-1} \sum_{t=l}^T \hat{\epsilon}_t \hat{\epsilon}_{t-l}' =$ Multivariate cross correlation matrix.

$$H_0 : \Gamma_1 = \dots = \Gamma_m = 0 \text{ against } H_1 : \Gamma_j \neq 0$$

for $j: 1 \leq j \leq m$ and $Q_k(m)$ asymptotically $\sim \chi_{(m-p)k}^2$ (Tsay, 2014).

H_0 : There is no serial correlation in the model residuals, while

H_1 : There is a serial correlation in the model residuals.

2. Normality Test

Zivot and Wang (2006) explained that normality test involves testing whether the data of the model are normally distributed, is best carried out by using Jarque-Bera statistic and follows χ^2 (Pfaff, 2008). In JB (Jarque-Bera) test, a random variable that is normally distributed should have skewness of 0 and kurtosis of 3, and the test statistic is given as;

$$JB = \frac{T}{6} \left(\hat{S}^2 + \frac{(\hat{K}-3)^2}{4} \right)$$

Where \hat{S} is the sample skewness and \hat{K} is the sample kurtosis, the hypothesis:

H_0 : Data is normally distributed

H_1 : Data is not normally distributed.

4. Structural Analysis

VAR has a lot of parameters which makes analysing the dynamic interrelationships of its variables much difficult. This problem is dealt with by VAR structural analysis that uses three types of analysis as; Granger causality, Impulse response analysis and variance decomposition.

1. Granger Causality

Here, if there are two variables, say Z_t and X_t , if Z_t Granger causes X_t . Then it means that the information contains in Z_t helps in improving the forecasting of X_t . It involves testing if a variable in the model is determined by lags of any other variable in the model apart from its lags which encompasses testing if the combined parameters of the variable are all zeroes. For example if the P matrix in equation (1) is diagonal, it means none of the variables Granger cause any. That is best achieved using the Wald statistic (Tsay, 2014).

2. Impulse Response Analysis

Impulse Response Function examines the response of one variable as a result of change in another variable holding all other variable constant. This is achieved by converting the VAR(p) model to its VMA representation and analysing its properties through the parameters of VMA. For example,

$$A_t = \epsilon_t + \Psi_1 \epsilon_{t-1} + \Psi_2 \epsilon_{t-2} + \dots$$

and

$$A_{t+1} = \epsilon_{t+1} + \Psi_1 \epsilon_t + \Psi_2 \epsilon_{t-1} + \Psi_3 \epsilon_{t-2} + \dots$$

this shows that ϵ_t affects A_{t+1} via Ψ_1 and hence ϵ_t affects A_{t+h} via Ψ_h (Martin and Harris, 2013). If one does not react as a result of change in other variable, the impulse responses are zero (Lutkepohl, 2005).

5. Random Walk Model

In Finance and Economics, most of the series are non-stationary, series such as exchange rate, interest rate and assets returns are empirically non-stationary. These kinds of series are unit root, and are known as Random Walk Model because their first differences are usually stationary (Tsay, 2014).

6. Forecasting

The major aims of VAR are forecasting and structural analysis. Forecasting is mostly associated with a cost, the data generation process might be VAR(p), in which the past and present information which is used for the forecast can be denoted as F_t . The best forecast is the one with minimum cost. MSE forecast is virtually employed in VAR settings (Lutkepohl, 2005). According to Zivot and Wang (2006), when the parameters of VAR are known, A_{t+1} which is the 1-step prediction based on the available information up to time T (F_t) is given as:

$$A_{t+1}|T = \mu + P_1 A_t + P_2 A_{t-1} + \dots + P_p A_{t-p+1} \text{ (1-step forecast)}$$

$$A_{t+h}|T = \mu + P_1 A_t + P_2 A_{t-1} + \dots + P_p A_{t-p+h} \text{ (h-step forecast)}$$

and

$$A_{t+h} - A_{t+h}|T = \sum_{s=0}^{h-1} \Psi_s \epsilon_{T+s-h} \text{ (is the forecast error)}$$

where:

$$\Psi_s = \sum_{j=1}^{p-1} \Psi_s \cdot P_j$$

MSE matrix of the forecast is given as:

$$\Sigma(h) = MSE(A_{t+h} - A_{t+h}|T) = \sum_{s=0}^{h-1} \Psi_s \Sigma \Psi_s'$$

But when the parameters are not known and hence are estimated, MSE matrix of the forecast error is given as:

$$\hat{\Sigma}(h) = \sum_{s=0}^{h-1} \hat{\Psi}_s \hat{\Sigma} \hat{\Psi}_s'$$

6.1 Cointegrated VAR Forecast

If series are cointegrated, Vector Error Correction Model can be estimated. After the estimation, the fitted model can be employed to generate forecasts of the differenced series, and can be applied to make the prediction of the original series. In VECM method, cointegration relationship is used in generating the forecasts (Tsay, 2014).

7. Cointegration and VEC Model

Using regression in non-stationary series may possibly lead to spurious regression. Where the test statistic is highly significant with high value of R^2 but with a small value of DW statistic (not close to 2) which means that the relationship between the variables is spurious. Although, if the first differences of all the series are stationary and if the errors of the linear combination of the series is stationary, this may infer an important issue about the series which is probably long term relationship between the variables (Griffiths and Lim, 2008).

One of the fundamental assumptions of VAR is stability, where the process fluctuates around a constant mean and a fixed variance. Macroeconomic variables in some cases have a long term relationship or equilibrium and are modelled by a process integrated of order 1 (I(1)); the process then becomes stationary if differenced once (Zivot and Wang, 2006).

Economic variables are normally empirically nonstationary and sometimes exhibit a trend. If two time series processes X_t and Y_t are non-stationaries, regressing Y_t on X_t , where the model can be specified as $Y_t = \alpha X_t + \epsilon_t$, when both of the series are I(1) and if $\epsilon_t = Y_t - \alpha X_t$ is stationary process [I(0)], then Y_t and X_t are said to be cointegrated where $(1, -\alpha)$ is the cointegrating vector. Cointegration means long term relationship or equilibrium between series (Greene, 2013).

The usual way of knowing whether series are cointegrated or not is to conduct stationarity test on the residuals of the model ($\epsilon_t = Y_t - \alpha X_t$) using Augmented Dickey-Fuller test. When there is enough evidence that the residuals are

stationary, then the series are cointegrated. Otherwise the series not cointegrated and if there is a significant relationship between the series, this relationship is spurious. In testing this hypothesis, both null and alternative hypothesis are expressed respectively as; H_0 : the series are stationary and H_1 : the series are non-stationary. The relationship among the variables are of two form; long term and short term. Long term relationship is the relationship among the series that are integrated of order 1 [I(1)]. While a short-term relationship is the relationship among stationary series that are I(0) (Griffiths and Lim, 2008).

When series are cointegrated, a model known as Vector Error Correction Model (VECM) is used by enclosing their cointegrating relationships to determine both long run and short run relationships between the variables (Koop, 2013). Supposing bivariate model X_t and Y_t , which are both I(1) with stationary residuals and can be regressed as: $X_t = \alpha + \beta Y_t + \mu$ which is the long term relationship between the series. In the short run, the series remain at equilibrium by making adjustments in X_t , Y_t or both. The series can be transformed as:

$$\begin{aligned} X_t - X_{t-1} &= \gamma_1 \mu_{t-1} + v_{x,t} = \gamma_1 (X_{t-1} - \alpha - \beta Y_{t-1}) + v_{x,t} \\ &= \gamma_1 X_{t-1} - \gamma_1 \alpha - \gamma_1 \beta Y_{t-1} + v_{x,t} \quad (12) \end{aligned}$$

$$\begin{aligned} Y_t - Y_{t-1} &= \gamma_2 \mu_{t-1} + v_{y,t} = \gamma_2 (X_{t-1} - \alpha - \beta Y_{t-1}) + v_{y,t} \\ &= \gamma_2 X_{t-1} - \gamma_2 \alpha - \gamma_2 \beta Y_{t-1} + v_{y,t} \quad (13) \end{aligned}$$

Equation (12) and (13) are known as the Vector Error Correction Model (VECM) because the two processes adjust each other in the subsequent time through the errors (Juselius, 2008). Martin and Harris (2013) asserts that γ_1 and γ_2 are known as the Error Correction Parameters, the two equations can more compactly be written as:

$$\begin{bmatrix} \Delta X_t \\ \Delta Y_t \end{bmatrix} = \begin{bmatrix} -\gamma_1 \alpha \\ -\gamma_2 \alpha \end{bmatrix} + \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} \begin{bmatrix} 1 & -\beta \end{bmatrix} \begin{bmatrix} X_{t-1} \\ Y_{t-1} \end{bmatrix} + \begin{bmatrix} v_{xt} \\ v_{yt} \end{bmatrix}$$

This can also be written in form of VAR(1):

$$\begin{bmatrix} X_t \\ Y_t \end{bmatrix} = \begin{bmatrix} -\gamma_1 \alpha \\ -\gamma_2 \alpha \end{bmatrix} + \begin{bmatrix} 1 + \gamma_1 & -\gamma_1 \beta \\ \gamma_2 & 1 - \gamma_2 \beta \end{bmatrix} \begin{bmatrix} X_{t-1} \\ Y_{t-1} \end{bmatrix} + \begin{bmatrix} v_{xt} \\ v_{yt} \end{bmatrix}$$

$$Bt = \lambda + \Phi Bt-1 + Vt$$

where:

$$B_t = \begin{bmatrix} X_t \\ Y_t \end{bmatrix} \Phi = \begin{bmatrix} 1 + \gamma_1 & -\gamma_1\beta \\ \gamma_2 & 1 - \gamma_2\beta \end{bmatrix} B_{t-1} = \begin{bmatrix} X_{t-1} \\ Y_{t-1} \end{bmatrix} V_t = \begin{bmatrix} v_{xt} \\ v_{yt} \end{bmatrix} \lambda = \begin{bmatrix} -\gamma_1\alpha \\ -\gamma_2\alpha \end{bmatrix}$$

This kind of VAR can encompasses p lags which has:

$$\Phi(L)B_t = \lambda + V_t$$

where: $\Phi(L) = I_N - \Phi_1 L - \dots - \Phi_p L^p$. Hence, the resultant VECM is written as:

$$\Delta B_t = \lambda + \Phi(1)B_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta B_{t-j} + V_t \quad (14)$$

with p-1 lags, where: $\Phi(1) = I_N - \Phi_1 L - \dots - \Phi_p$.

Using the Beveridge-Nelson decomposition; $\Phi(L) = \Phi(1)L + \Gamma(L)(1-L)$

in which: $\Gamma(L) = I_N - \sum_{j=1}^{p-1} \Gamma_j L^j$ and $\Gamma_j = -\sum_{i=j+1}^p \Phi_i$.

$$\text{Hence} \quad \Phi(L)B_t = \lambda + V_t.$$

$$\text{Substituting } \Phi(L), \quad [\Phi(1)L + \Gamma(L)(1-L)]B_t = \Phi(1)B_{t-1} + \Delta B_t - \sum_{j=1}^{p-1} \Gamma_j \Delta B_{t-j} = \lambda + V_t.$$

7.1 Johansen Test for Cointegration

To perform a test of cointegration, restriction is imposed on matrix Φ in equation 14 which contains two matrices; cointegrating parameters β and speed adjustments parameters γ (Enders, 2015). Two types of test are conducted in this case: trace and eigenvalue which are all based on the rank of the matrix Φ , if $\text{Rank}(\Phi) = 0$, it means that $\Phi = 0$ and conclusion is that there is no cointegration among the variables, and the analysis should be based on the differenced data which is VAR(p-1), on the other side, if the $\text{Rank}(\Phi) = m > 0$: A_t has m cointegrating vectors, the analysis should be based on the VECM and α denotes outcomes of the series (Tsay, 2014). Likelihood ratio test is conducted with the view of finding out if the rank is significantly different from 0. The hypothesis is:

$$H_0: r = 0 \text{ and } H_1: r = N \text{ where } N > 0.$$

The Likelihood ratio statistic is given as:

$$LR = -2(T-p)(\ln L_T(\theta_0) - \ln L_T(\theta_1))$$

where θ_0 relates to the model with reduced rank and θ_1 relates to the model with full rank. The trace statistic is given

as:

$$LR = -(T-p) \sum_{i=r+1}^N \ln(1 - \hat{\lambda}_i)$$

where $\hat{\lambda}_i$ is the eigenvalues. if the value of LR > critical value, the null hypothesis is rejected in favour of alternative hypothesis (Martin and Harris, 2013).

8. Conclusion

This research work described the procedure for the estimation of VAR model along with the VECM. It also described the procedures for VAR diagnostic test, lag order selection, unit roots test and structural analysis. Cointegration and forecasting were all explained.

For the empirical analysis, two or more related time series variables are required.

References

- Box, G. E., J. G. M. and Reinsel, G. C. (2008). *Time Series Analysis: Forecasting and Control*. Wiley, New Jersey.
- Cryer, J. D. and Chan, K.-S. (2008). *Time Series Analysis With Applications in R*. Springer, New York.
- Enders, W. (2015). *Applied Econometric Time Series*. Wiley, Hoboken, USA.
- Greene, W. H. (2013). *Econometric Analysis*. Pearson, New York.
- Griffiths, W. E., H. C. R. and Lim, G. C. (2008). *Principles of Econometrics*. Wiley, Hoboken, USA.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton, New Jersey.
- Juselius, K. (2008). *The Cointegrated VAR model; Methodology and Applications*. Oxford University Press, New York.

Koop, G. (2013). *Analysis of economic Data*. Wiley, UK.

Lutkepohl, H. (2005). *New Introduction To Multiple Time Series Analysis*. Springer, New York.

Martin, V., H. S. and Harris, D. (2013). *Econometric Modelling With Time Series: Specification, Estimation and Testing*. Cambridge University Press, New York.

Pfaff, B. (2008). *Analysis of Integrated and Cointegrated Time Series With R*. Springer, New York.

Tsay, R. R. (2014). *Multivariate Time Series Analysis With R and Financial Applications*. Wiley, Hoboken, USA.

Vinod, H. D. (2008). *Hands-on Intermediate Econometrics Using R: Templates for Extending Dozens of Practical Examples*. World Scientific, New York.

Zivot, E. and Wang, J. (2006). *Modelling Financial Time Series With S-Plus*. Springer, New York.