

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
 Maths 2021; 6(2): 06-20
 © 2021 Stats & Maths
www.mathsjournal.com
 Received: 04-01-2021
 Accepted: 06-02-2021

Ahmed Mohamed Mohamed
 Elsayed
 Al-Obour High Institute for
 Management & Informatics
 Department of Basic Science
 Kilo 21 Cairo-Belbies Road,
 P.O. Box 27 Obour City, Egypt

Forecasting EGX30 index time series using vector autoregressive models VARS

Ahmed Mohamed Mohamed Elsayed

DOI: <https://doi.org/10.22271/math.2021.v6.i2a.658>

Abstract

Time series analysis is considered one of the most important analysis processes at the present time, especially if it is a multivariate analysis. This analysis helps the decision maker in making his future decision based on the behavior of a phenomenon in the past. This is done for many economic, financial, engineering, medical, and other important fields. So we were keen in this article to address a multivariate time series using the vector autoregressive models analysis of the practical time series. This analysis is also used in the process of forecasting the future of multiple time series. Three packages from the R program, are used for numerical analysis these data, that are "vars", "MTS(VAR)", and "forecast" packages. For these data ARIMA (0,1,0) model is the best model for non-stationary time series, but the best model is ARIMA(0,0,0) with zero mean for stationary time series. VAR(1) model is better than VAR(2) model for analyzing the data. Also, VAR(1) model is better than VAR(2) model for identifying the Lags order. Comparing between VAR(p) and VARS(p), we found that the VAR(p) model order is better than the VARS(p) model order for identifying lag order. The significance predictor time series is Institutional investor, the other predictors are non-significance. No serial correlation in error terms. The Institutional investor and Trade values time series are useful for forecasting other time series. Only the Exchange rate time series has instantaneous causality between it and other time series. Finally, A five forecasts values of all stationary time series, using VARS(1) models, are predicted.

Keywords: VARS models, ARIMA models, autoregressive models, serial test, accuracy measures, ACF, PACF, time series, stationarity, multivariate time series

1. Introduction

Time series analysis is considered one of the most important analysis processes at the present time, especially if it is a multivariate analysis. This analysis helps the decision maker in making his future decision based on the behavior of a phenomenon in the past. In many economic, financial, engineering, medical, and other important fields, so we were keen in this article to address a multivariate time series using the vector autoregressive models analysis of time series data taken from the Egyptian Stock Exchange to measure the impact of the EGX30 index with many variables, which are all time series during specified period. This analysis is also used in the process of forecasting the future of multiple time series. Within this framework, we will use three packages from the R program, which are vars, MTS, and forecast, to analyze these time series, and focus on comparing the results between vars and MTS (var), especially in determining the maximum lags, this is the value order of dependent variable in the previous series.

There are many articles and applications for using the VAR models. Primiceri ^[1] used time varying structural vector autoregressions and monetary policy. VAR models are also increasingly used in health research for automatic analyses of diary data or sensor data ^[2]. Ankargren† and Jonéus ^[3] discussed the issue of estimating large-scale VAR models with stochastic volatility in real-time situations where data are sampled at different frequencies. Econometric models taking information in unequal frequencies into account, and thereby avoiding a loss of information stemming from aggregation to the lower frequency, have gained attention under the name of mixed-frequency methods. Multiple approaches are available in Foroni and Marcellino ^[4], for example the mixed-frequency vector autoregressive (VAR)

Corresponding Author:
 Ahmed Mohamed Mohamed
 Elsayed
 Al-Obour High Institute for
 Management & Informatics
 Department of Basic Science
 Kilo 21 Cairo-Belbies Road,
 P.O. Box 27 Obour City, Egypt

model. McCracken *et al.* [5] found that the mixed-frequency information is important for now-casting. Schorfheide and Song [6] found that the model improved forecasting as compared to a quarterly VAR model. Eraker *et al.* [7] formulated a Bayesian VAR based on this idea and Schorfheide and Song [6] proceeded with a Gibbs sampling approach based on simulation smoothing and forward-filtering, backward-smoothing along the lines of Carter and Kohn [8]. Ankargren *et al.* [9] developed a steady-state mixed-frequency VAR model for a real-time US dataset.

This paper is organized as following: Section 2 presents some algorithms and packages that are used in this paper. Section 3 presents the numerical analysis for the collected time series data. Section 4 presents the results discussion. Finally, Section 5 presents the conclusions of this article.

2. Algorithms and Packages

This section presents three types of topics, the first is a theoretical and mathematical definition of the concept of vector autoregressive models (VARs), the cross correlation matrix (CCM), and the other is R program packages (vars, MTS and forecast).

2.1 Vector Autoregressive Models (VARs)

The VAR models generalize the univariate autoregressive model by allowing for multivariate time series. VAR models are used in economics and the natural sciences. Each variable has an equation modelling its evolution over time. This equation includes the variable's lagged values, the lagged values of the other variables in the model. VAR models require only a prior knowledge about the list of variables which can be hypothesized to affect each other over time. VAR model describes the evolution of a set of k variables, called endogenous variables, over time. Each period of time is numbered t . The k -variables modeled as a linear function of only their past values. The variables are collected in a vector, y_t , length k . The vector's components are referred to as $y_{i,t}$, meaning the observation at time t of the i -th variable. VAR models are characterized by their order, which refers to the number of earlier time periods the model will use. For example a 5-th order VAR model is a linear combination of the last five years. A lag is the value of a variable in a previous time period. So in general a p -th order VAR refers to a VAR model which includes lags for the last p time periods. In this case VAR model is denoted "VAR(p)". A p -th order VAR model is written as:

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \dots + A_p y_{t-p} + e_t$$

The variables of the form y_{t-i} are called the " i -th lag" of y_t . The variable c is a k -vector of constants serving as the intercept of the model. A_i is a time-invariant ($k \times k$)-matrix and e_t is a k -vector of error terms. The error terms must satisfy the conditions: $E(e_t) = 0$, $E(e_t e_t') = \Omega$, and $E(e_t e_{t-k}') = 0$. This means that: Every error term has a mean zero. The contemporaneous covariance matrix of error terms is $k \times k$ positive-semidefinite matrix denoted Ω . For non-zero k , there is no correlation across time. There is no serial correlation in error terms. For a general example of a VAR(p) with k variables. A VAR(1) in two variables is:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} e_{1,t} \\ e_{2,t} \end{bmatrix}$$

Single A matrix appears because this example has a maximum lag = 1, or system of two equations:

$$y_{1,t} = c_1 + a_{1,1} y_{1,t-1} + a_{1,2} y_{2,t-1} + e_{1,t}$$

$$y_{2,t} = c_2 + a_{2,1} y_{1,t-1} + a_{2,2} y_{2,t-1} + e_{2,t}$$

Each variable in the model has one equation. The VAR (2) model can be written as:

$$y_t = C + A_1 y_{t-1} + A_2 y_{t-2} + e_t,$$

And can be as the VAR (1) model:

$$\begin{bmatrix} y_t \\ y_{t-1} \end{bmatrix} = \begin{bmatrix} C \\ 0 \end{bmatrix} + \begin{bmatrix} A_1 & A_2 \\ I & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \end{bmatrix} + \begin{bmatrix} e_t \\ 0 \end{bmatrix},$$

where I is the identity matrix.

Choosing the maximum lag p in the VAR model requires special attention because inference is dependent on correctness of the selected lag order [10-11].

A structural VAR(p), is noted as SVAR, can be written as:

$$B_0 y_t = c_0 + B_1 y_{t-1} + B_2 y_{t-2} + \dots + B_p y_{t-p} + \varepsilon_t,$$

where c_0 is a $k \times 1$ vector of constants, B_i is a $k \times k$ matrix (for every $i = 0, \dots, p$) and ε_t is a $k \times 1$ vector of error terms. The main diagonal terms of the B_0 matrix are scaled to 1. The error terms ε_t (structural shocks) satisfy that all the elements in the off diagonal of the covariance matrix $E(\varepsilon_t \varepsilon_t') = \Sigma$ are zero. The structural shocks are uncorrelated. Two variable structural VAR (1) is:

$$\begin{bmatrix} 1 & B_{0;1,2} \\ B_{0;2,1} & 1 \end{bmatrix} \begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} C_{0;1} \\ C_{0;2} \end{bmatrix} + \begin{bmatrix} B_{1;1,1} & B_{1;1,2} \\ B_{1;2,1} & B_{1;2,2} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix},$$

where

$$E(\varepsilon_i \varepsilon_i') = \Sigma = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \text{ the variances of the structural shocks, are denoted as } var(\varepsilon_i) = \sigma_i^2 \text{ (} i = 1, 2), \text{ and the covariance is}$$

$cov(\varepsilon_1, \varepsilon_2) = 0$. Writing the first equation explicitly and passing $y_{2,t}$ to the right hand side one obtain:

$$y_{1,t} = C_{0;1} + B_{0;1,2} y_{2,t} + B_{1;1,1} y_{1,t-1} + B_{1;1,2} y_{2,t-1} + \varepsilon_{1,t}$$

Note that $y_{2,t}$ can have a contemporaneous effect on $y_{1,t}$ if $B_{0;1,2}$ is non-zero. This is different from the case when B_0 is the identity matrix, when $y_{2,t}$ can impact directly $y_{1,t+1}$ and subsequent future values, but not $y_{1,t}$. Because of the parameter identification problem, ordinary least squares estimation of the structural VAR would yield inconsistent parameter estimates. This problem can be overcome by rewriting the VAR in reduced form:

By multiplying the structural VAR with B_0^{-1} , we have:

$$y_t = B_0^{-1} c_0 + B_0^{-1} B_1 y_{t-1} + B_0^{-1} B_2 y_{t-2} + \dots + B_0^{-1} B_p y_{t-p} + B_0^{-1} \varepsilon_t$$

$$\text{Putting } c = B_0^{-1} c_0, \quad B_0^{-1} B_i = A_i \quad (i = 1, \dots, p), \quad B_0^{-1} \varepsilon_t = e_t$$

The p -th order reduced VAR form

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \dots + A_p y_{t-p} + e_t.$$

However, the error terms in the reduced VAR are composites of the structural shocks $e_t = B_0^{-1} \varepsilon_t$.

The occurrence of one structural shock $\varepsilon_{i,t}$ can potentially lead to the occurrence of shocks in all error terms $e_{i,t}$.

The covariance matrix of the reduced VAR is:

$$\Omega = E(e_t e_t') = E(B_0^{-1} \varepsilon_t \varepsilon_t' (B_0^{-1})') = B_0^{-1} \Sigma (B_0^{-1})'$$

have non-zero off-diagonal elements, and allowing non-zero correlation between error terms.

Starting from the concise matrix notation:

$$Y = BZ + U$$

The MLS approach for estimating B yields:

$$\hat{B} = YZ'(ZZ')^{-1}$$

This can be written as:

$$\text{Vec}(\hat{B}) = ((ZZ')^{-1} Z \otimes I_k + U) \text{Vec}(Y)$$

where denotes the Kronecker product, and Vec the vectorization of the indicated matrix. This estimator is consistent and asymptotically efficient. It is furthermore equal to the conditional maximum likelihood estimator ^[12]. The multivariate least squares estimator is equivalent to the ordinary least squares estimator applied to each equation separately ^[13]. As in the standard case, the MLE of the covariance matrix differs from the ordinary least squares (OLS) estimator.

The MLE estimator is:

$$\hat{\Sigma} = \frac{1}{T} \sum_{t=1}^{T-\Sigma} \hat{\varepsilon}_t \hat{\varepsilon}_t'$$

The OLS estimator: for a model with a constant, k variables and p lags.

$$\hat{\Sigma} = \frac{1}{T-kp-1} \sum_{t=1}^{T-\Sigma} \hat{\varepsilon}_t \hat{\varepsilon}_t'$$

In a matrix notation, this gives:

$$\hat{\Sigma} = \frac{1}{T-kp-1} (Y - \hat{B}Z)(Y - \hat{B}Z)'$$

Vector autoregressive models often involve the estimation of many parameters. For example, with 7 variables and 4 lags, each matrix of coefficients for a given lag length is 7 by 7, and the vector of constants has 7 elements, so a total of $49 \times 7 = 343$ parameters are estimated. The degrees of freedom of the regression is (the number of data points minus the number of parameters to be estimated). This can hurt the accuracy of the parameter estimates and hence of the forecasts given by the model.

Properties of the VAR model are usually summarized using structural analysis using Granger causality, impulse responses, and forecast error variance decompositions.

Consider the first-order case, equation of evolution is:

$$y_t = Ay_{t-1} + e_t$$

Write above equation of evolution one period lagged:

$$y_{t-1} = Ay_{t-2} + e_{t-1}$$

Substitute, we have

$$y_t = A^2y_{t-2} + Ae_{t-1} + e_t$$

Repeat substitution, we obtain

$$y_t = A^3y_{t-3} + A^2y_{t-2} + Ae_{t-1} + e_t$$

It can be seen from this induction process that any shock will have an effect on the elements of y infinitely far forward in time, although the effect will become smaller and smaller over time assuming that the AR process is stable — that is, that all the eigenvalues of the matrix A are less than 1 in absolute value.

An estimated VAR model can be used for forecasting, and the quality of the forecasts can be judged, in ways that are completely analogous to the methods used in univariate autoregressive modelling.

2.2 Cross Correlation Matrix (CCM)

A cross correlation is a measure of similarity of two series as a function of the displacement of one relative to the other. That is one signal is delayed relative to the other and the correlation between the two is calculated, kind of similar to the autocorrelation but with another signal rather than itself.

Cross correlation is only one measure - which is referring to the correlation of one signal with another. Time series can also be auto correlated, i.e. a signal for a particular time period can be correlated with the one previous. So, correlation is not necessarily time independent. When you say one signal is delayed relative to the other, you are referring to the time period for which one signal is expected to follow another. e.g. during a thunderstorm, you will first see light and then hear thunder X seconds later. However, cross-correlation of a signal does not necessarily imply Granger Causality, i.e. that one causes the other, so this would still have to be investigated separately: Cross Correlation Does Not Imply Granger Causation.

Let $D = \text{diag}(\sigma_1, \dots, \sigma_k)$ with $\sigma_i = \sqrt{\gamma_{ii}(0)}$, the diagonal matrix of the standard deviations of y_i 's. Then, $\rho_p = D^{-1}\Gamma_p D^{-1}$, with $\rho_p = [\rho_{ij}(p)]$, is the cross correlation matrix of order p . Where,

$$\rho_{ij}(p) = \frac{\gamma_{ij}(p)}{\sqrt{\gamma_{ii}(0)}\sqrt{\gamma_{jj}(0)}} = \frac{\text{cov}(y_{it}, y_{jt-p})}{\sigma_i \sigma_j}$$

where $\rho_{ij}(p)$ is the correlation coefficient between y_{it} and y_{jt-p} [14].

2.3 Used Packages of R program

There are three packages will be used in this paper. These packages are: vars, MTS and forecast packages.

2.3.1 MTS Package

Multivariate Time Series (MTS) is a general package for analyzing multivariate linear time series and estimating multivariate volatility models. For the multivariate linear time series analysis, the package performs model specification, estimation, model checking, and prediction for many widely used models, including vector AR models, vector MA models, vector ARMA models,

seasonal vector ARMA models, VAR models with exogenous variables, multivariate regression models with time series errors, augmented VAR models, and Error correction VAR models for co-integrated time series. For model specification, the package performs structural specification to overcome the difficulties of identifiability of VARMA models. The methods used for structural specification include Kronecker indices and Scalar Component Models. The package also considers multiple tests for conditional heteroscedasticity, including rank-based statistics. Finally, the MTS package also performs forecasting using diffusion index, transfer function analysis, Bayesian estimation of VAR models, and multivariate time series analysis with missing values [15-16].

2.3.2 VARS Package

Estimation, lag selection, diagnostic testing, forecasting, causality analysis, forecast error variance decomposition and impulse response functions of VAR models and estimation of SVAR and SVEC models. VAR Model with Selected Lags is a modified version of VAR command by allowing the users to specify which AR lags to be included in the model [17, 18, 19].

2.3.3 Forecast Package

Methods and tools for displaying and analyzing univariate time series forecasts including exponential smoothing via state space models and automatic ARIMA modelling. The "mforecast" is a class of objects for forecasting from multivariate time series or multivariate time series models. The function invokes particular methods which depend on the class of the first argument. The function summary is used to obtain and print a summary of the results, while the function plot produces a plot of the multivariate forecasts and prediction intervals [20-21].

3. Numerical Analysis

In this study, we obtained monthly data of the net transactions of both the foreign investor, the institutional investor, the capital, and the trade values of the EGX30 Index. In the period from Jan 2015 to Feb 2019. Thus, we obtained 50 monthly time series related to these values. Therefore, we study these data in light of dividing the data into 6 variables, the dependent variable being the time series of the EGX30 index time series. The independent variables are both the series of net transactions of the (foreign investor, the institutional investor, the capital, and the trade values on the values of the EGX30 Index. The aim of the study is to find out whether there is a significant impact from the previous independent variables on the dependent variable, EGX30 index. The plots of all time series can be displayed simultaneously as shown in Figure 1:

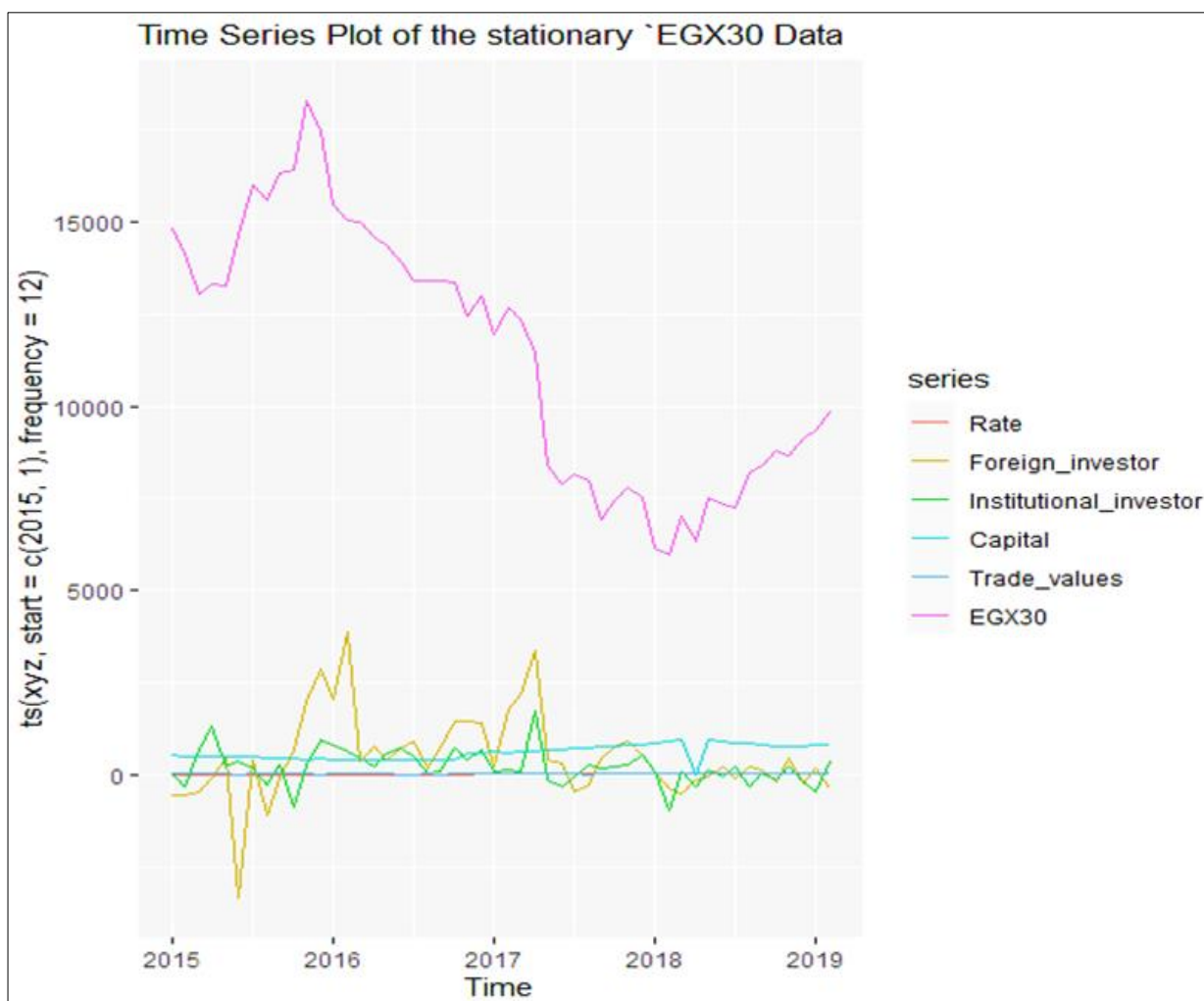


Fig 1: Plots of time series EGX30 data

To be clear, let us to indicate the shape of each time series plots as shown in Figure 2

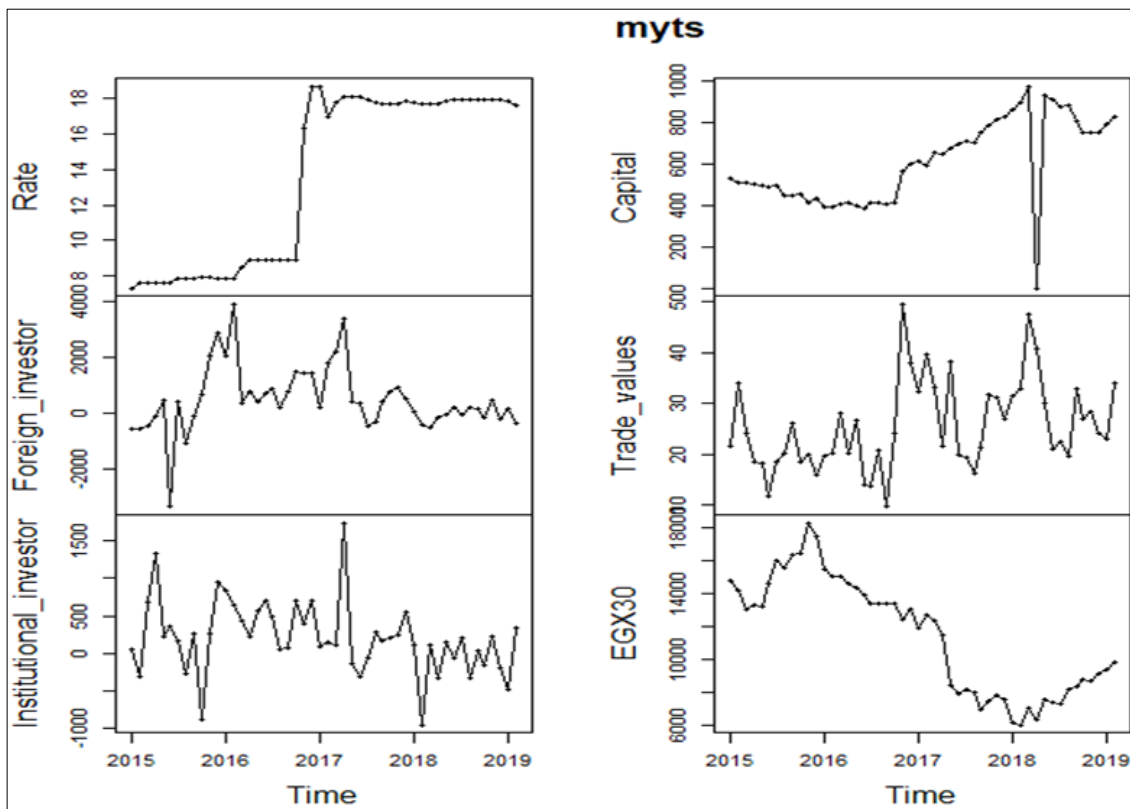


Fig 2: Time series data plots for each variable

Figure 3 explains the seasonal time series data for EGX30 index dependent variable:

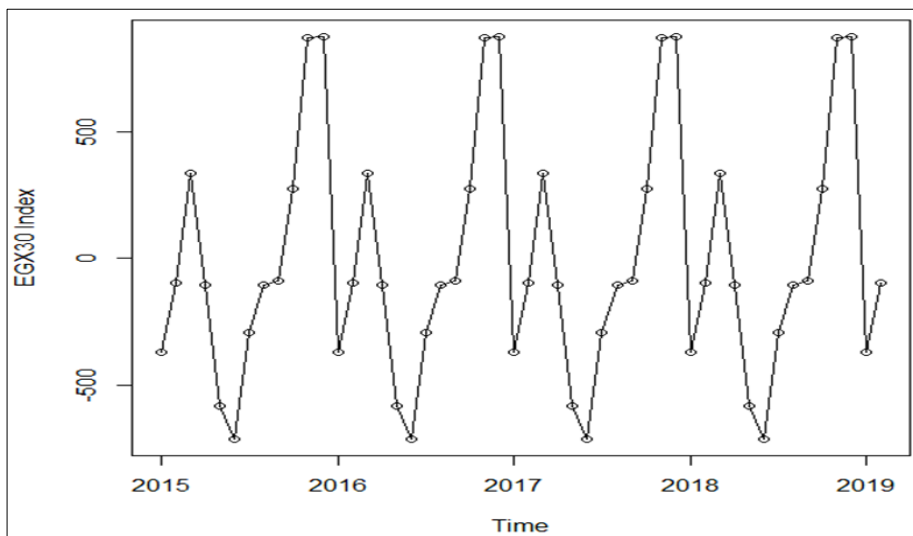


Fig 3: Seasonal time series EGX30 Index

3.1 Stationarity Process of Time Series Data

Before analyzing the time series data, first we check the stationary of all time series variables using the Augmented Dickey-Fuller Test. We found that the time series Exchange rate (Rate) and EGX30 index are non-stationary time series. Reset time series are stationary. Therefore, we make some modifications for all time series, the differencing process, to get the full stationary time series to make the process analysis. Table 1 explains the results of the ADF test before and after differencing process.

Table 1: The ADF test for stationarity

Time series	Before differencing process			After differencing process		
	ADF	P-value	Case	ADF	P-value	Case
Exchange rate (Rate)	-1.115	0.6399	Non-stationary	-5.3394	0.01	Stationary
Foreign investor	-4.2332	0.01	Stationary	-11.6007	0.01	Stationary
Institutional investor	-5.8245	0.01	Stationary	-10.3177	0.01	Stationary
Capital	-3.7553	0.01	Stationary	-12.1951	0.01	Stationary
Trade values	-3.9804	0.01	Stationary	-8.9325	0.01	Stationary
EGX30 index	-1.0647	0.6581	Non-stationary	-6.1537	0.01	Stationary

Let us construct the ARIMA models for EGX30 index as a dependent variable before differencing process as shown in Table 2:

Table 2: ARIMA models (Before Differencing process)

ARIMA (0,1,0)	Error = 803.9104
ARIMA (0,1,0)	805.4105
ARIMA (0,1,1)	805.5287
ARIMA (0,1,2)	807.7817
ARIMA (0,1,3)	809.7861
ARIMA (0,1,4)	811.0361
ARIMA (0,1,4)	813.1424
ARIMA (0,1,5)	812.8776
ARIMA (0,1,5)	815.2216
ARIMA (1,1,0)	805.5183
ARIMA (1,1,0)	807.2375
ARIMA (1,1,1)	809.6104
ARIMA (1,1,2)	810.0465
ARIMA (1,1,3)	811.9858
ARIMA (1,1,3)	813.9621
ARIMA (1,1,4)	813.3469
ARIMA (1,1,4)	815.6458
ARIMA (2,1,0)	807.7906
ARIMA (2,1,0)	809.607
ARIMA (2,1,1)	810.0408
ARIMA (2,1,1)	811.9896
ARIMA (3,1,0)	810.1073
ARIMA (3,1,0)	811.9853
ARIMA (3,1,1)	812.2708
ARIMA (3,1,1)	814.2451
ARIMA (4,1,0)	810.8853
ARIMA (4,1,0)	813.0626
ARIMA (4,1,1)	813.4187
ARIMA (4,1,1)	815.7542
ARIMA (5,1,0)	813.3303
ARIMA (5,1,0)	815.6933

Best model: ARIMA (0,1,0)

The measures of accuracy for ARIMA (0,1,0) model, before differencing process, can be presented as

ME	RMSE	MAE	MPE	MAPE	MASE	ACF(1)
-98.92132	856.4993	620.7519	-1.170764	5.995428	0.1929128	0.09332822

These measures indicated accurate models for non-stationary time series.

Figure 4 displays the ACF and PACF measures of fitting the EGX30 index time series before differencing.

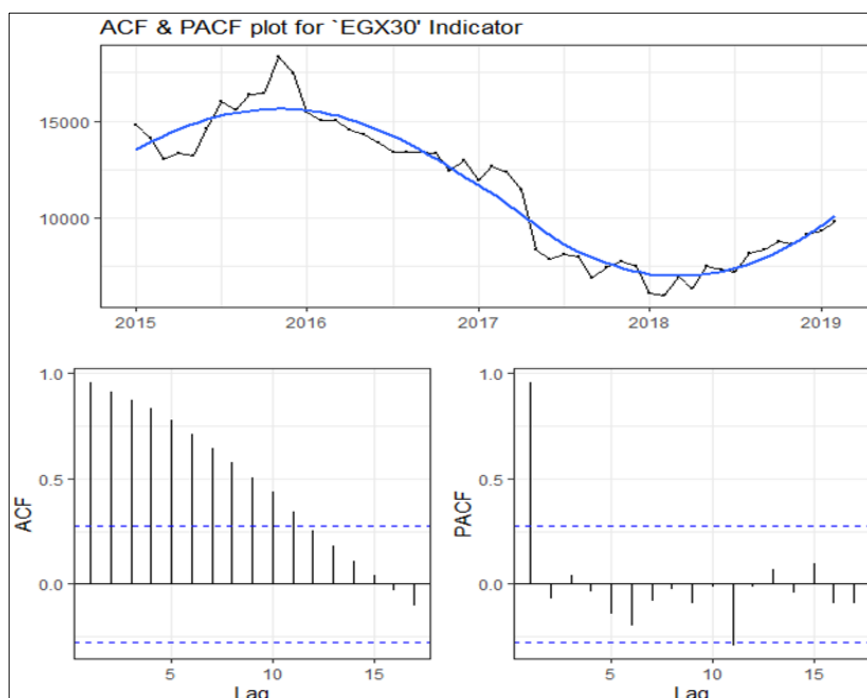


Fig 4: AFC and PAFC measures of EGX30 time series before differencing

From Figure 4, we see the sharp shape of ACF and PACF before differencing process. Let us construct the ARIMA models for EGX30 index after differencing process as shown in Table 3:

Table 3: ARIMA models (After Differencing process)

ARIMA (0,0,0) with zero mean Error	: 803.9104
ARIMA (0,0,0) with non-zero mean	: 805.4105
ARIMA (0,0,1) with zero mean	: 805.5287
ARIMA (0,0,1) with non-zero mean	: 807.2368
ARIMA (0,0,2) with zero mean	: 807.7817
ARIMA (0,0,2) with non-zero mean	: 809.6077
ARIMA (0,0,3) with zero mean	: 809.7861
ARIMA (0,0,3) with non-zero mean	: 811.6387
ARIMA (0,0,4) with zero mean	: 811.0361
ARIMA (0,0,4) with non-zero mean	: 813.1424
ARIMA (0,0,5) with zero mean	: 812.8775
ARIMA (0,0,5) with non-zero mean	: 815.2216
ARIMA (1,0,0) with zero mean	: 805.5183
ARIMA (1,0,0) with non-zero mean	: 807.2375
ARIMA (1,0,1) with non-zero mean	: 809.6104
ARIMA (1,0,2) with zero mean	: 810.0465
ARIMA (1,0,3) with zero mean	: 811.9857
ARIMA (1,0,3) with non-zero mean	: 813.9621
ARIMA (1,0,4) with zero mean	: 813.3469
ARIMA (1,0,4) with non-zero mean	: 815.6458
ARIMA (2,0,0) with zero mean	: 807.7906
ARIMA (2,0,0) with non-zero mean	: 809.607
ARIMA (2,0,1) with zero mean	: 810.0408
ARIMA (2,0,1) with non-zero mean	: 811.9896
ARIMA (3,0,0) with zero mean	: 810.1073
ARIMA (3,0,0) with non-zero mean	: 811.9853
ARIMA (3,0,1) with zero mean	: 812.2708
ARIMA (3,0,1) with non-zero mean	: 814.2451
ARIMA (4,0,0) with zero mean	: 810.8853
ARIMA (4,0,0) with non-zero mean	: 813.0626
ARIMA (4,0,1) with zero mean	: 813.4187
ARIMA (4,0,1) with non-zero mean	: 815.7542
ARIMA (5,0,0) with zero mean	: 813.3303
ARIMA (5,0,0) with non-zero mean	: 815.6933

Best model: ARIMA (0,0,0) with zero mean

After differencing process, we get the accuracy measures for the ARIMA (0,0,0) model as below

ME	RMSE	MAE	MPE	MAPE	MASE	ACF (1)
-101.2422	865.1924	633.1182	100	100	0.675666	0.09517107

Figure 5 displays the ACF and PACF measures of fitting the EGX30 index time series after differencing.

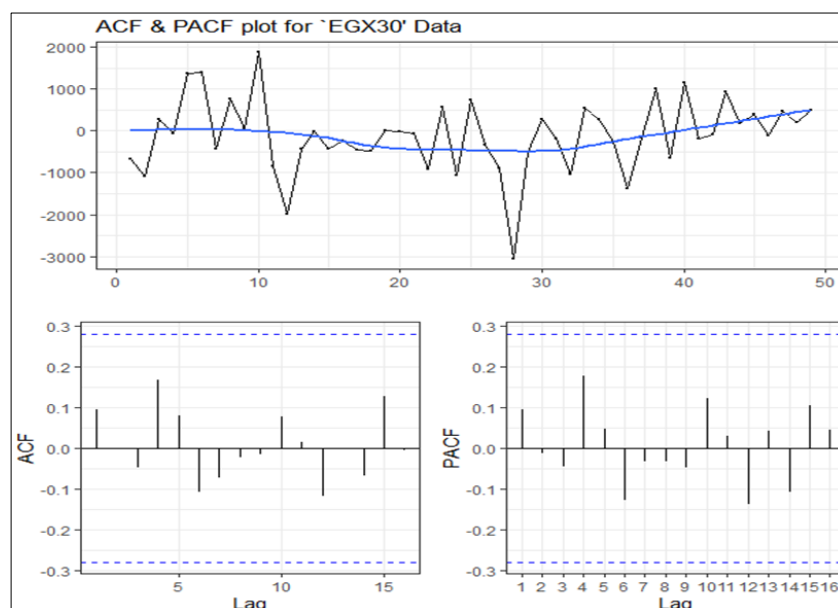


Fig 5: AFC and PAFC measures of EGX30 time series after differencing

From Figure 5, we see the moderate shape of ACF and PACF after differencing process.

3.2 Time series With MTS package (VAR () Function)

Figure 6 displays the EGX30 data stationary time series with VAR model

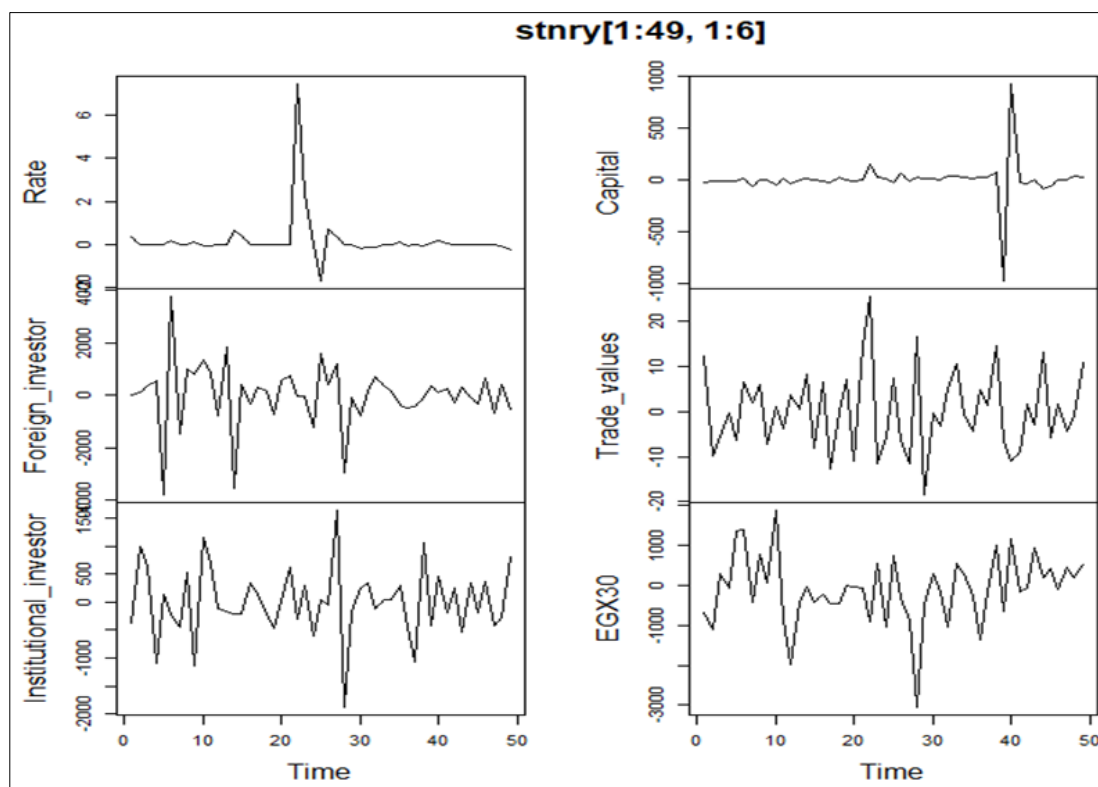


Fig 6: EGX30 data stationary time series separately with VAR model

Figure 7 display the stationary time series EGX30 data together with MTS(VAR model):

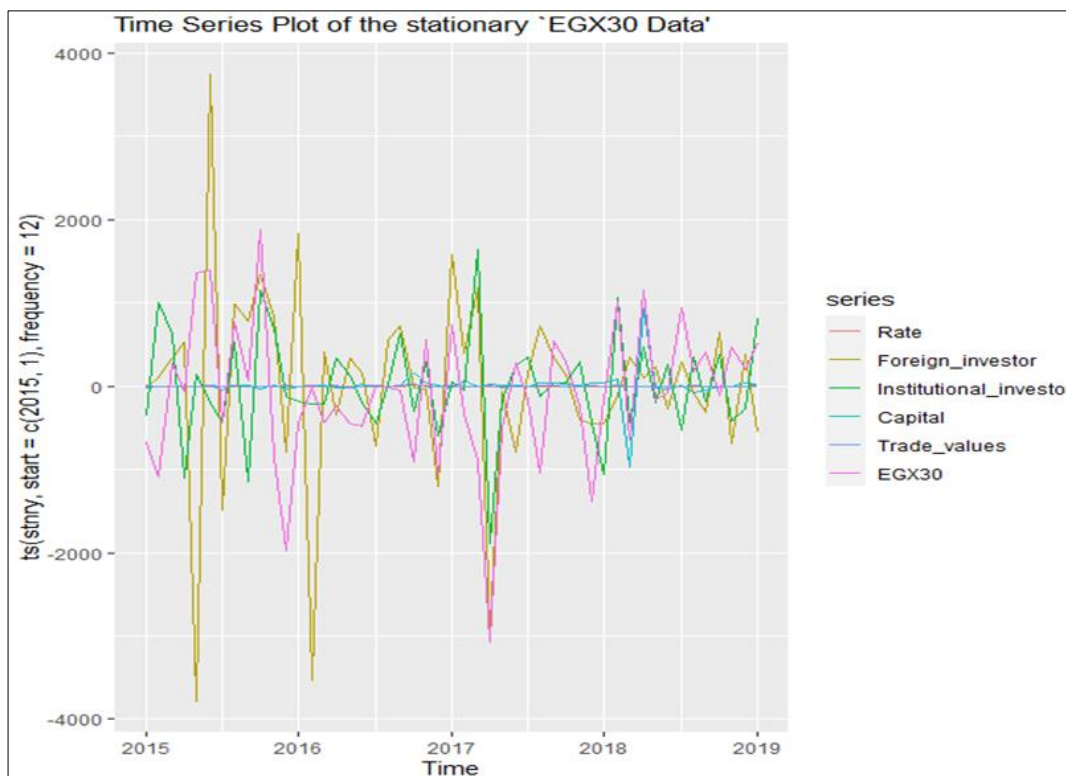


Fig 7: EGX30 data stationary time series with VAR model

In this subsection, we are fitting with two cases for VAR model: the first one when the autoregressive model is with lag. Max=1, the second when lag. Max=2. In each case we construct the estimators of each time series and standard error for each estimators. Then, we construct the identification of lags of the VAR model to get the maximum lags with minimum P-value for CCM.

For VAR (1) model we have

T. Series	Rate	Foreign investor	Institutional investor	Capital	Trade value	EGX30
Estimates	0.177	70.9143	13.520	0.8809	0.1721	-60.683
Std. Error	0.164	161.155	88.318	24.820	1.319	104.416

For VAR (2) model we have

T. Series	Rate	Foreign investor	Institutional investor	Capital	Trade value	EGX30
Estimates	0.192	137.61	22.296	-4.504	0.112	-54.369
Std. Error	0.184	176.31	75.3	24.321	1.4603	106.91

We found that the standard errors of most of estimators for VAR (1) is lower than VAR (2) model

The rule indicates that the VAR (p) identification Lags with the lower measure is better

Measure	VAR (1)	VAR (2)
SSE	1.211271e+23	3.416435e+22
AIC	54.62052	54.82425
BIC	56.01042	57.60406
HQ	55.14784	55.8789

From the obtained results, the VAR (1) model is better than VAR (2) model for identifying the Lags. Where the previous measures are lower in VAR (1) model. Also, Figures 8, 9 displays the significance plot of Cross-Correlation Matrices (CCM) at each lag to identify Maximum Lag with minimum P. values for VAR (1) and VAR (2) models as shown below:

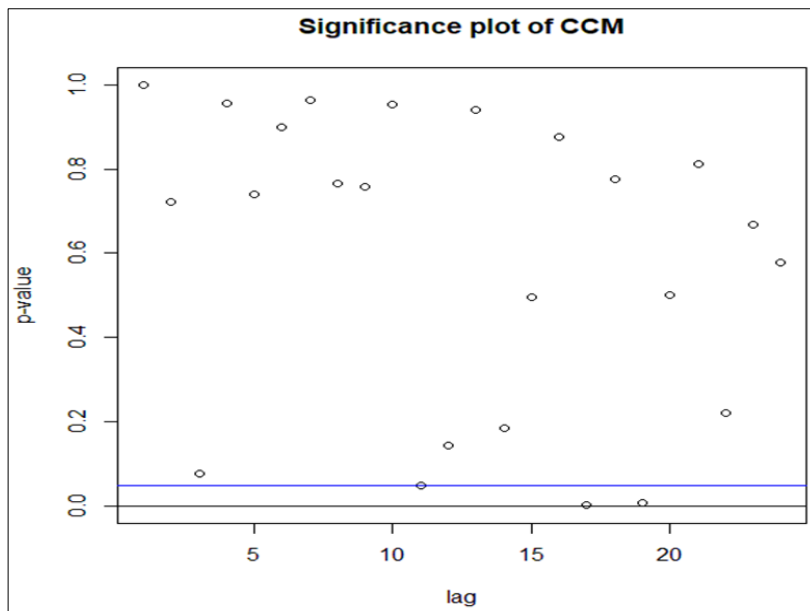


Fig 8: VAR (1) lag. Max identification

Figure 8 explains that when VAR(1) the minimum P.value is achieved when lag.max=11.

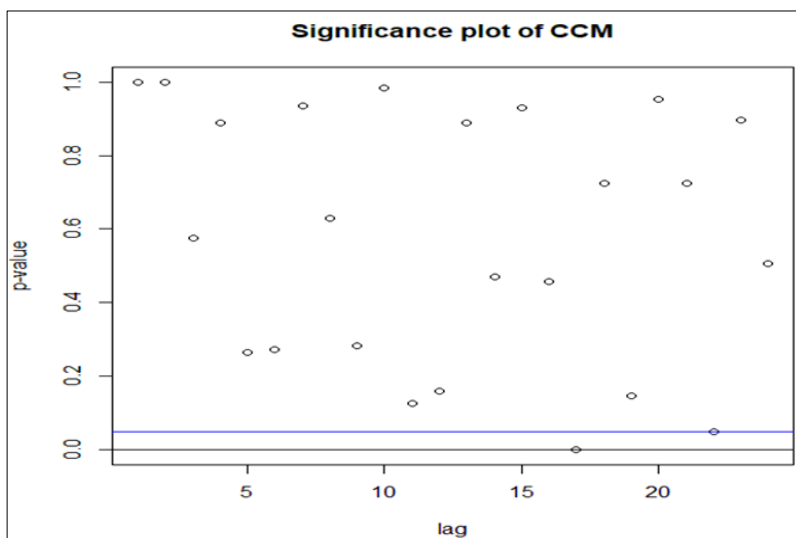


Fig 9: VAR (2) lag. Max identification

Figure 9 explains that when VAR (2) the minimum P. value is achieved when Maximum lag=22.

A lag is the value of a variable in a previous time period. So in general a p -th order VAR refers to a VAR model which includes lags for the last p time periods. Then, the VAR (1) form is more convenient for analytical derivations and allows more compact statements. This is explained also, graphically, Figure (8, 9), since the maximum lag for VAR (1) is 11. This means that 11-th order VAR refers to a VAR model which includes lags for the last 11 time periods.

3.3 Time series with vars Package (vars: VAR Function)

Also, the VARS (p) identification Lags with the lower measure is better

Measure	VARS (1)	VARS (2)
SSE	1.150954e+23	3.146287e+22
AIC	54.56944	54.74187
BIC	55.95934	57.52169
HQ	55.09677	55.79653

Comparing between VAR(p) and VARS(p) identification lag order, we found that: For the order of two models VAR(p) or VARS(p), ($p = 1,2$), we found that the VAR(p) model order is better than the VARS(p) model order for identifying lag order, where all measures are lower values for VAR(p) model. We can obtain the VARS (1) model fitting for the time series EGX30 index as a dependent variable, and the other time series as independent variables. Sample size: 48, Log Likelihood: -1684.281.

Roots of the characteristic polynomial

Rate	Foreign investor	Institutional investor	Capital	Trade values	EGX30
0.6687	0.5066	0.271	0.271	0.1818	0.1818

The VARS model can be defined as:

EGX30 = Intercept + Rate + Foreign investor + Institutional investor + Capital + Trade values.

Table 4 presents the VARS (1) model predictors and their test:

Table 4: VARS (1) model

VARS Model	Estimate	Std. Error	t value	P. value
Intercept	0.40072	0.12940	3.097	0.00348 **
Rate	43.36084	91.74175	0.473	0.63892
Foreign investor	-0.09015	0.08864	-1.017	0.31499
Institutional investor	-0.89162	0.18200	-4.899	1.48e-05 ***
Capital	-0.72615	0.52759	-1.376	0.17600
Trade values	-5.29609	12.08011	-0.438	0.66333

Residual standard error: 698.3, R-Squared: 0.4346, F-statistic: 5.381, P-value: 0.0003384

Serial test

H_0 : There is no serial correlation in error terms.

H_1 : There is serial correlation in error terms.

Chi-squared = 423.61, p-value = 0.9999

No serial correlation in individual error terms.

Granger test for causality of EGX30 index

H_0 : EGX30 does not Granger-cause.

H_1 : EGX30 does Granger-cause.

F-Test = 0.81678, P-value = 0.5386.

The EGX30 index does not a Granger-cause.

Instant test of EGX30 index

H_0 : No instantaneous causality between EGX30 and other time series.

H_1 : There is instantaneous causality between EGX30 and other time series.

Chi-squared = 3.0484, P-value = 0.6925.

No instantaneous causality between: EGX30 and other time series.

Granger test for causality of Exchange rate

H_0 : Exchange rate does not Granger-cause.

H_1 : Exchange rate does Granger-cause.

F-Test = 0.69355, P-value = 0.6288.

The Exchange rate does not Granger-cause.

Instant test of Exchange rate

H_0 : No instantaneous causality between Exchange rate and other time series.

H_1 : There is instantaneous causality between Exchange rate and other time series.

Chi-squared = 11.506, P-value = 0.04222

There is instantaneous causality between Exchange rate and other time series.

Granger test for causality of Foreign investor

H₀: Foreign investor does not Granger-cause.

H₁: Foreign investor does Granger-cause.

F-Test = 0.63365, P-value = 0.6742

Foreign investor does not Granger-cause.

Instant test of Foreign investor

H₀: No instantaneous causality between Foreign investor and other time series.

H₁: There is instantaneous causality between Foreign investor and other time series.

Chi-squared = 7.9036, P-value = 0.1616

No instantaneous causality between Foreign investor and other time series.

Granger test for causality of Institutional investor

H₀: Institutional investor does not Granger-cause.

H₁: Institutional investor does Granger-cause.

F-Test = 6.295, P-value = 1.59e-05.

Institutional investor does Granger-cause.

Instant test of Institutional investor

H₀: No instantaneous causality between Institutional investor and other time series.

H₁: No instantaneous causality between Institutional investor and other time series.

Chi-squared = 7.1944, P-value = 0.2066.

No instantaneous causality between Institutional investor and other time series.

Granger test for causality of Capital

H₀: Capital does not Granger-cause.

H₁: Capital does Granger-cause.

F-Test = 0.43485, P-value = 0.824

Capital does not Granger-cause.

Instant test of Capital

H₀: No instantaneous causality between Capital and other time series.

H₁: There is instantaneous causality between Capital and other time series.

Chi-squared = 6.1428, df = 5, p-value = 0.2926

No instantaneous causality between Capital and other time series.

Granger test for causality of Trade values

H₀: Trade values does not Granger-cause.

H₁: Trade values does Granger-cause.

F-Test = 3.7191, P-value = 0.002888.

Trade values does Granger-cause.

Instant test of Trade values

H₀: No instantaneous causality between Trade values and other time series.

H₁: There is instantaneous causality between Trade values and other time series.

Chi-squared 9.1624, df = 5, p-value = 0.1028

No instantaneous causality between Trade values and other time series.

We can conduct that: Only the Institutional investor and Trade values time series, are useful for forecasting other time series, and only the Exchange rate time series has instantaneous causality between it and other time series with 5% significance level.

Figure 9 displays the forecasts of VARS (1) model for all time series.

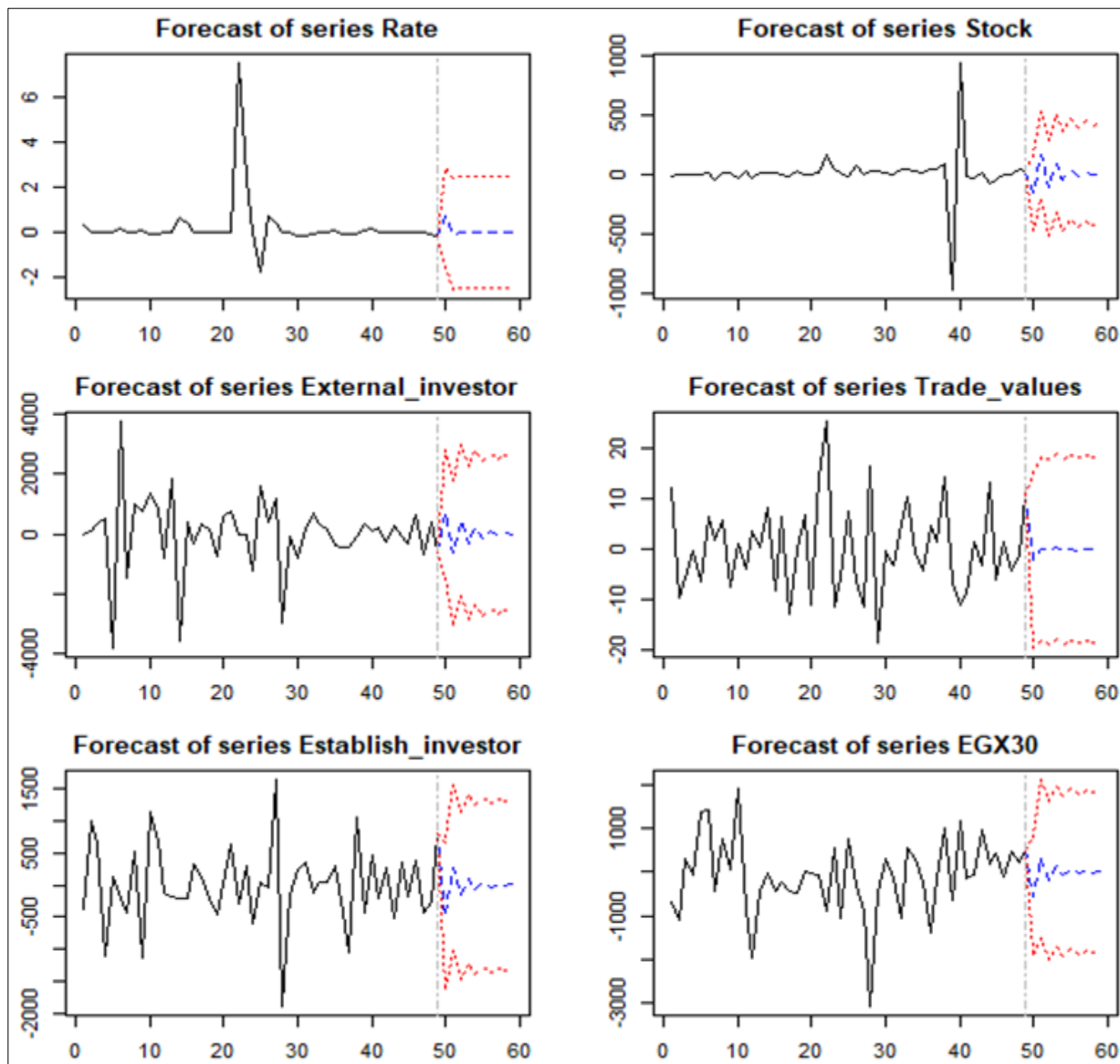


Fig 9: VARS forecasts

A five forecasts values of all stationary time series, using VARS (1) models for each time series with confidence interval estimation, can be presented in the following Tables (5:10):

Table 5: Forecasts values for "Exchange rate"

No.	Forecasts	Lower	Upper	C.I.
1	0.69475821	1.468953	-2.858469	2.163711
2	-0.0594831	-2.509962	2.390995	2.450479
3	0.00894993	-2.444626	2.462526	2.453576
4	-0.0131895	-2.467881	2.441502	2.454691
5	0.01127459	-2.443564	2.466113	2.454838

Table 6: Forecasts values for "Foreign investor"

No.	Forecasts	Lower	Upper	C.I.
1	671.63554	-1436.916	2780.187	2108.551
2	-643.04071	-3061.363	1775.281	2418.322
3	460.62971	-2048.356	2969.616	2508.986
4	-315.38669	-2861.070	2230.297	2545.684
5	207.37946	-2354.365	2769.124	2561.744

Table 7: Forecasts values for "Institutional investor"

No.	Forecasts	Lower	Upper	C.I.
1	-479.68631	-1632.853	673.4807	1153.167
2	279.282871	-1006.683	1565.2490	1285.966
3	-160.62885	-1467.334	1146.0763	1306.705
4	99.424170	-1214.485	1413.3334	1313.909
5	-67.681391	-1384.385	1249.0221	1316.703

Table 8: Forecasts values for "Capital"

No.	Forecasts	Lower	Upper	C.I.
1	-156.979884	-480.9723	167.0125	323.9924
2	160.062297	-210.5202	530.6448	370.5825
3	-115.353245	-515.1681	284.4616	399.8148
4	81.081831	-330.6437	492.8074	411.7255
5	-56.483966	-473.6413	360.6733	417.1573

Table 9: Forecasts values for "Trade values"

No.	Forecasts	Lower	Upper	C.I.
1	-2.24964638	-19.46963	14.97033	17.21998
2	0.22600730	-17.96169	18.41370	18.18769
3	-0.46088683	-18.79314	17.87136	18.33225
4	0.44648179	-17.91542	18.80838	18.36190
5	-0.31895405	-18.68936	18.05145	18.37040

Table 10: Forecasts values for "EGX30 index"

No.	Forecasts	Lower	Upper	C.I.
1	-562.55364	-1931.127	806.0193	1368.573
2	297.75539	-1501.866	2097.3765	1799.621
3	-191.73497	-2009.621	1626.1510	1817.886
4	111.45639	-1715.387	1938.2996	1826.843
5	-77.36887	-1906.623	1751.8857	1829.255

4. Discussion Results

From the previous results, we found that

1. For stationarity process of EGX30 index dependent variable time series

ARIMA (0,1,0) model is the best model, the plots of time series show sharp shapes of ACF and PACF before differencing process. Where the best model is ARIMA (0,0,0) with zero mean, the plots of time series show moderate shapes of ACF and PACF after differencing process.

2. For fitting the time series with MTS package-VAR function

We found that the standard errors of most of estimators for VAR (1) is lower than VAR (2) model. Also, the VAR (1) model is better than VAR (2) model for identifying the Lags order. The measures are lower in VAR (1) model. The significance plots of Cross-Correlation Matrices (CCMs) reflect that the VAR (1) is more convenient for analytical derivations and allows more compact statements. Where the maximum lag in VAR (1) is 11 but it is 22 in VAR (2).

3. For fitting the time series with vars package-vars: VAR function

Comparing between VAR(p) and VARS(p) identification lag order, we found that: For the order of two models VAR(p) or VARS(p), (p = 1,2), we found that the VAR(p) model order is better than the VARS(p) model order for identifying lag order, where all measures are lower values for VAR(p) model.

The VARS (1) model can be defined as:

EGX30 index = Intercept + Rate + Foreign investor + Institutional investor + Capital + Trade values

The significance predictor time series is Institutional investor, the other predictors are non-significance.

(a). For Serial test

No serial correlation in error terms.

(b). Granger test for causality

1. The EGX30 index does not a Granger-cause.
2. The Exchange rate does not Granger-cause.
3. Foreign investor does not Granger-cause.
4. Institutional investor does Granger-cause.
5. Capital does not Granger-cause.
6. Trade values does Granger-cause.

(c). Instant test

1. No instantaneous causality between: EGX30 and other time series.
2. There is instantaneous causality between Exchange rate and other time series.
3. No instantaneous causality between Foreign investor and other time series.
4. No instantaneous causality between Institutional investor and other time series.
5. No instantaneous causality between Capital and other time series.
6. No instantaneous causality between Trade values and other time series.

We can conduct that: Only the Institutional investor and Trade values time series, are useful for forecasting other time series, and only the Exchange rate time series has instantaneous causality between it and other time series with 5% significance level. Finally, five forecasts values of all stationary time series, using VARS (1) models, are predicted.

5. Conclusions

In this paper, we were keen in to address a multivariate time series using the vector autoregressive models (VARS) for analyzing the time series data are taken from the Egyptian Stock Exchange to measure the impact of some time series as independent variables on the EGX30 index as dependent variable, during specified period. This analysis is also used in the process of forecasting the future of multiple time series, using three packages from the R program (vars, MTS, and forecast). A comparison between the results of vars and MTS(VAR) packages are explained, especially in determining the maximum lags.

ARIMA (0,1,0) model is the best model for non-stationary time series, but the best model is ARIMA (0,0,0) with zero mean for stationary time series. The VAR (1) model is better than VAR (2) model for analyzing these data. Also, the VAR (1) model is better than VAR (2) model for identifying the Lags order. Where the maximum lag in VAR (1) is 11 but it is 22 in VAR (2). Comparing between VAR (p) and VARS (p), we found that the VAR (p) model order is better than the VARS (p) model order for identifying lag order, where all accuracy measures are lower values for VAR (p) model. The significance predictor time series is Institutional investor, the other predictors are non-significance. No serial correlation in error terms. The Institutional investor and Trade values time series are useful for forecasting other time series. Only the Exchange rate time series has instantaneous causality between it and other time series with 5% significance level. A five forecasts values of all stationary time series, using VARS (1) models, are predicted.

6. Acknowledgement

To Al-Obour High Institute for Management & Informatics. Department of Basic Science.

7. References

1. Primiceri E. Time Varying Structural Vector Autoregressions and Monetary Policy. *The Review of Economic Studies* 2005;72(3):821-852.
2. Krieke *et al.* Temporal Dynamics of Health and Well-Being: A Crowdsourcing Approach to Momentary Assessments and Automated Generation of Personalized Feedback (2016). *Psychosomatic Medicine* 2016.
3. Ankargren S, Jonéus P. Estimating Large Mixed-Frequency Bayesian VAR Models. Department of Statistics, Uppsala University. arXiv:1912.02231v1[econ. EM] 2019.
4. Forni C, Marcellino M. A Survey of Econometric Methods for Mixed-Frequency Data. Working Paper No. 6, Norges Bank 2013.
5. McCracken W, Ng S. FRED-MD: A Monthly Database for Macroeconomic Research. *Journal of Business & Economic Statistics* 2016;34(4):574-589.
6. Schorfheide F, Song D. Real-Time Forecasting with a Mixed-Frequency VAR. *Journal of Business & Economic Statistics* 2015;33(3):366-380.
7. Eraker B, Chiu CW, Foerster AT, Kim TB, Seoane HD. Bayesian Mixed Frequency VARs. *Journal of Financial Econometrics* 2015;13(3):698-721.
8. Carter K, Kohn R. On Gibbs Sampling for State Space Models. *Biometrika* 1994;81(3):541-553.
9. Ankargren S *et al.* A Flexible Mixed-Frequency Vector Autoregression with a Steady-State Prior. Preprint available on arXiv:1911.09151 2019.
10. Hacker S, Ans Hatemi A. Optimal lag-length choice in stable and unstable VAR models under situations of homoscedasticity and ARCH. *Journal of Applied Statistics* 2008;35(6):601-615.
11. Hatemi A, Hacker S. Can the LR test be helpful in choosing the optimal lag order in the VAR model when information criteria suggest different lag orders? *Applied Economics* 2009;41(9):1489-1500.
12. Hamilton, James D. *Time Series Analysis*. Princeton University Press 1994, 293.
13. Zellner, Arnold. An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias. *Journal of the American Statistical Association* 1962;57(298):348-368.
14. Sims C. Macroeconomics and Reality. *Econometrica* 1980;48(1):1-48.
15. Tsay R. *Multivariate Time Series Analysis with R and Financial Applications (Chapter 2)*. John Wiley. Hoboken, NJ 2014.
16. Tsay R, Wood D. MTS Package: All-Purpose Toolkit for Analyzing Multivariate Time Series (MTS) and Estimating Multivariate Volatility Models. R packages, Version 1.0 2018.
17. Holtz-Eakin D, Newey W, Rosen HS. Estimating Vector Autoregressions with Panel Data. *Econometrica* 1988;56(6):1371-1395.
18. Hatemi A. Multivariate tests for autocorrelation in the stable and unstable VAR models. *Economic Modelling* 2004;21(4):661-683.
19. Bernhard B, Stigler M. Vars Package: VAR Modelling. R package, Version 2018;1:5-3.
20. Hyndman J, Athanasopoulos G. *Vector Autoregressions. Forecasting: Principals and Practice*. O Texts 2018, 333-335.
21. Hyndman R *et al.* Package forecast: Forecasting functions for time series and linear models. R package, version 8.13 2020.