MJFAS MALAYSIAN JOURNAL OF FUNDAMENTAL AND APPLIED SCIENCES
PRINT ISSN: 2289-5981 | ONLINE ISSN: 2289-599X

RESEARCH ARTICLE

A robust vector autoregressive model for forecasting economic growth in Malaysia

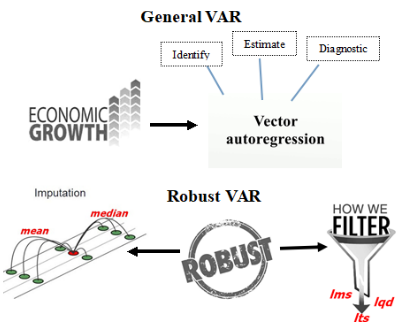
Azme Khamis*, Nur Azreen Abdul Razak, Mohd Asrul Affendi Abdullah

Faculty of Applied Science and Technology, University of Tun Hussein Onn Malaysia, 86400 Parit Raja, Batu Pahat, Johor, Malaysia.

* *Corresponding author: azme@uthm.edu.my*

Article history
Submitted 20 February 2018
Revised 15 March 2018
Accepted 16 April 2018
Published Online 3 September 2018

Graphical abstract



Abstract

Economic indicator measures how solid or strong an economy of a country is. Basically, economic growth can be measured by using the economic indicators as they give an account of the quality or shortcoming of an economy. Vector Auto-regressive (VAR) method is commonly useful in forecasting the economic growth involving a bounteous of economic indicators. However, problems arise when its parameters are estimated using least square method which is very sensitive to the outliers existence. Thus, the aim of this study is to propose the best method in dealing with the outliers data so that the forecasting result is not biased. Data used in this study are the economic indicators monthly basis starting from January 1998 to January 2016. Two methods are considered, which are filtering technique via least median square (LMS), least trimmed square (LTS), least quartile difference (LQD) and imputation technique via mean and median. Using the mean absolute percentage error (MAPE) as the forecasting performance measure, this study concludes that Robust VAR with LQD filtering is a more appropriate model compare to others model.

***Keywords*: Forecasting, economic growth, robust, imputation, filtering**

INTRODUCTION

Vector autoregression (VAR) is a stochastic process model used to capture the linear interdependencies among multiple time series data. The VAR models generalize the univariate autoregressive model (AR model) by allowing for more than one evolving variable, where it described at least two factors where the dependent factors are found as a lag on the correct hand side of the equation. Also, VAR is a tools that take control of both the changing and interdependent relationships of the numerous variables (Onwukwe, 2014).

VAR model is a very popular tool in multiple time series analysis. Its parameters are usually estimated by the least squares procedure which is very sensitive to the presence of errors in data known as outliers. If outliers were present, the estimation results would become unreliable (Jonas, 2009). Outliers in forecasting are data points that are not considered to be part of the overall pattern of demand and it can affects the forecasting accuracy. Since outliers can affect and skew forecast accuracy, it can be useful to exclude them from overall forecasting calculations to improve forecast accuracy (Anscombe (1960), Chen (1993) and Yu (2014)).

In order to overcome this problem, an alternative method have been developed that are not so easily affected by outliers that is filtering and imputation technique. Filtering technique is done by employing robust regression methods (Shumacker *et. al*., 2002). Robust regression is an essential method for analysing data that are contaminated with outliers. It can be employed to detect outliers and to provide resistant results in the presence of outliers (Almonem, 2015). The robust methods are Least Median of Squares (LMS), Least Trimmed Squares (LTS) and Least Quartile Difference (LQD) (Berk (1990), Birkes & Dodge (1993). Meanwhile, imputation is a common approach for dealing with missing values in statistical databases. The imputer fills in missing values with draws from predictive models estimated from the observed data such as mean and median, resulting in completed versions of the database (Fan & Jerome, 2017).

An economic indicator such as exchange rate and interest rate acts as the important indicator in the analysis of economic performance of a country and the predictions of future performance as stated by Agalega & Antwi (2013). Economic growth is defined as the increase in the capacity of the economy, where it can be measured by using an economic indicators as they move markets and measure how vigorous an economy of a country is. Besides that, they also can quantify particular divisions of an economy, for example, the lodging or retail division, or they provide quantification or estimations of an economy in general, such as unemployment or Gross Domestic Product (GDP) (Gaspar *et. al*., 2017).

Thus, this study aims to propose and investigate the performance of the robust VAR models as an alternative tool in forecasting the economic growth in Malaysia in a specific divisions such as currency in circulation, exchange rate, external reserve and reserve money.

EXPERIMENTAL

Materials

Data that were used in this study are the economic indicators consist of Currency in Circulation (CIC), Exchange Rate (EXC), External Reserve (EXT) and Reserve Money (RM) that were monthly basis starting from January 1998 to January 2016, which a sample of 217 observations is available. The first 171 values of actual data that is January 1998 to December 2011 is used in the estimation period to help select the model and to estimate its parameters. Then, the last 46 values, January 2012 to January 2016 are hold out for validation and 12 forecasts for the future are generated. The data were gained from Department of Statistics Malaysia. All statistical analyses are carried out by using R language.

Method

This study consider with two techniques which are filtering via least median of squares (LMS), (Hampel, 1975 and Rousseeuw, 1984), least trimmed squares (LTS), (Rousseeuw, 1983) and least quartile differences (LQD), (Croux and Hossjer, 1994). LMS, LTS and LQD filter are reasonable for removing low or high frequency parts from a time series which might be sullied with outliers and can contain shifts level. Meanwhile, the imputation methods are employed using mean and median values. Various studies had been explored regarding on the VAR model procedure (Onwukwe & Nwafor, 2014; Khodaparasti & Moslehi, 2014; Lanne & Nyberg, 2015; Petrovska & Naumovski, 2016; Gupta & Wohar, 2017 and Nyberg, 2017).

Least median of square (LMS)

Median filter has been widely used, refer (Arce 1998, Fried 2006, Fried 2007). The LMS filter offers the highest robustness against outliers and is able to track level shifts and trend changes well (Reform, 2006). LMS is a less delicate, or more robust fitting methodology than least squares. It pictured that an estimator used the median would be less touchy to outrageous values than OLS. LMS estimator is given by:

$$\min_{\hat{\theta}} (med r_i^2) \tag{1}$$

where, $i = 1,2,3,...$ and the residuals r_i is minimized. Least median of squares estimation is robust to outliers due to its high breakdown value of 50%.

Least trimmed squares (LTS)

Least trimmed squares (LTS) regression can be noticed as an alteration of the LMS (Koivunen, 1995). The LTS has preferred asymptotic conduct over the LMS, particularly a non-zero Gaussian productivity, however it performs correspondingly to the LMS and significantly more costly in limited examples. The least trimmed squares (LTS) estimator is defined as:

$$\min_{\hat{\theta}} \sum_{i=1}^{h} r^2(\beta : n) \tag{2}$$

where $h=[n+2]+1$, $i = 1,2,3,...$, β is the minimum LTS cost and n is an element point. Moreover, for $h=[n/2]+[(p+1)/2]$ the LTS reaches the maximal possible value for the breakdown point. When n is very small, it is possible to generate all subsets of size h to determine which one minimizes the LTS criterion.

Least quartile difference (LQD)

The interquartile range (IQR) is often utilized to find outliers in data. Outliers here are defined as observations that fall below Q_1 -1.5 IQR or above Q_3 + 1.5 IQR. The LQD estimator, introduced by Croux *et. al.* (1994) presents a much better performance at Gaussian distributed errors than other most extreme breakdown strategy such as LMS and LTS. The LQD estimator has a breakdown point of $h= [(n+k+1)/2]$ if the data fulfill certain requirements. This means that up to 50% of the data can be contaminated without ruining the fit. Also, 50% represents an upper bound for the breakdown point in the class of regression-equivariant estimator. The LQD estimator is defined as, consider a line $L: y = \beta x + \alpha$ with slope β and intercept α, and let

$$r_i(L) = y_i - \beta x - \alpha \tag{3}$$

Denote the residual of the point $p_i = (x_i, y_i)$ with respect to the line L and $i = 1,2,3,....$

Mean and median imputation

The imputation techniques by using mean and median replacement is highlighted in this study. Mean imputation is the replacement of a missing value with the mean of the non-missing observations for that variable while median substitution replaces all missing data in a variable by the median value for that variable (Cankaya, *et. al*., 2015). Imputation approach is defined as:

$$\hat{y}_j = x_i \beta \tag{4}$$

where y is a missing values, x is an observed values, β is an unknown parameters, $i=1,2,3,...$ and $j=1,2,3,....$ The principle purpose behind leading the imputation is to lessen non-response bias, which occurs as the conveyance of the missing values. Instead of erasing, the sample size is kept up by applying the imputation, bringing about a perhaps higher flexibility than case deletion as discussed by Burke & Wycombe (2012).

Forecast accuracy measurement

Forecast accuracy employed in this study is Mean Absolute Percentage Error (MAPE) that commonly used in the literature. It provides a measure of the distance of the true from the forecast value, see (Bruce, *et. al*, 2005). The forecast sample is $j = T+1, T+2,..., T+h$, and y_t denote the actual and forecast value in period t as $\hat{y}_t$, respectively. The forecast evaluation measures are defined as:

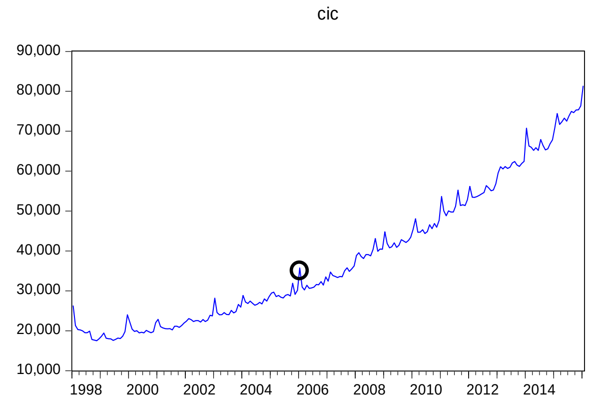
$$MAPE = 100 \times \sum_{t=T+1}^{T+h} \left| \frac{y_t - \hat{y}_t}{y_t} \right| / h \tag{5}$$

RESULTS AND DISCUSSION

The detection of outlier in time series data via robust approach such as imputation and filtering technique is discussed. Scatter plot is employed to identify the existence of the outlier. The outlier is then removed and replaced by using filtering and imputation technique. Next, the new or modified data is analyzed by VAR model. Finally, all models are compared and model with the highest accuracy measurement is selected as the best model for forecasting economic growth in Malaysia.

Data series plot

An observed time series can be decomposed into three components that are the trend, the seasonal and the irregular. Figure 1 shows the plot of the actual time series for CIC, EXC, EXT and RM.



a) Currency in Circulation

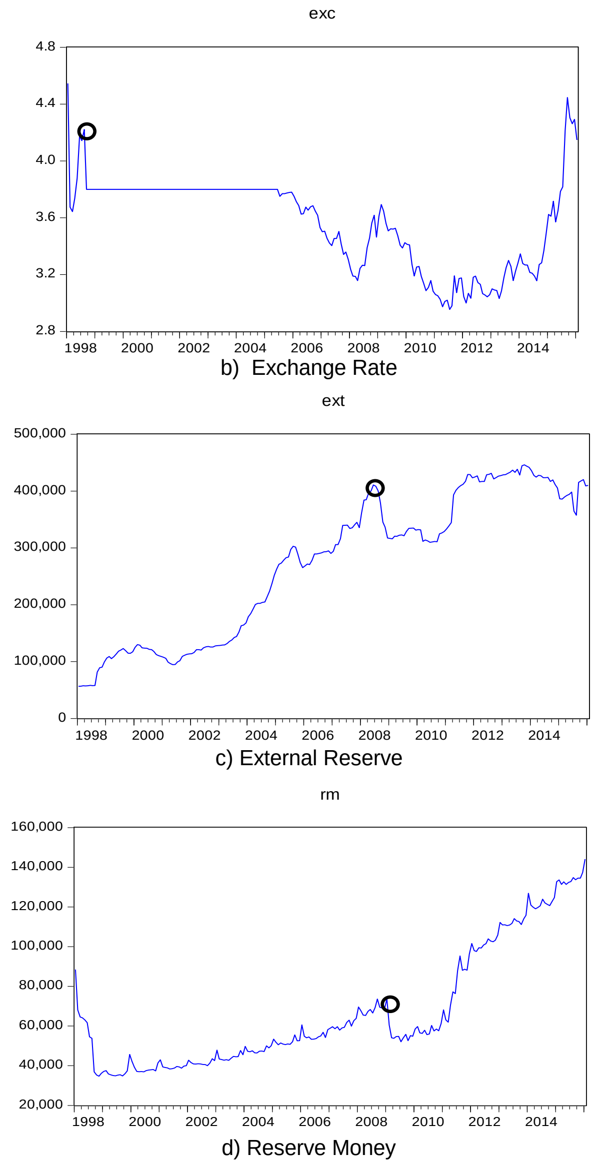


Fig. 1 Time series plot.

Figure 1 shows that the CIC, EXC, EXT and RM have an irregular components with random movements. It results from short term fluctuations (black dots) in the series which are neither systematic nor predictable. In a highly irregular series, these fluctuations can dominate movements, which will mask the trend and seasonality.

Unit root test

Unit root test is conducted to establish the integration order using the Augmented Dickey Fuller (ADF) test. Table 1 presents the result of the unit root tests for the four variables.

Table 1 Unit root test.

Variable	Augmented Dickey Fuller (ADF)
	Level
CIC	3.7415
EXC	-1.2673
EXT	-1.3326
RM	1.0471
	Difference
CIC	-2.4268***
EXC	-6.1676***
EXT	-3.4600***
RM	-4.4321***

*The Null hypothesis is that series is non-stationary, or contain a unit root. The rejection of the null hypothesis based on the Mackinnon critical values. Note: *** denotes significant at 1% significance level.*

From the Table 1, the result confirms that the series becomes stationary after the first difference of the series. Thus, it can be observed that all the variables (CIC, EXC, EXT, RM) exhibit non-stationary series which are integrated of the first order, *I(1)*.

Outlier detection in Time Series data

An outlier is a data point that is significantly far away from the majority of the data and can have an influence on estimating the parameters. Scatter plot of each model is implemented by plotting the residuals against the fitted values to investigate that the outliers do exist in the data. Figure 2 illustrates the scatter plot for CIC, EXC, EXT and RM models.

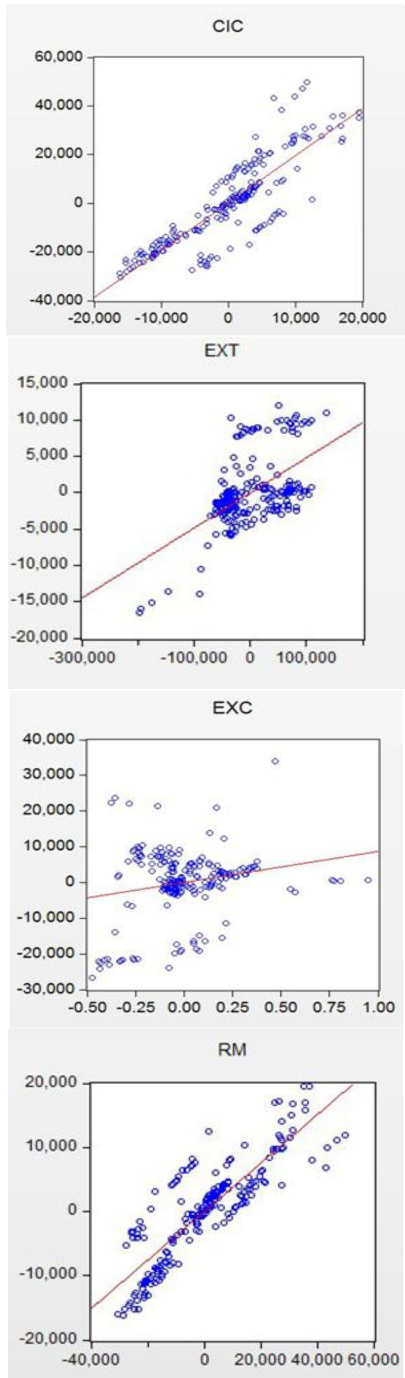


Fig.2 Scatter plot for CIC, EXC, EXT and RM models.

Figure 2 indicates the residuals represented by dots and the fitted values represented by straight line where it portrays that most of the dots in each models are a way off to the side from the fitted line of the points. Based on the plots, it can be summarized that outliers do exist in CIC, EXC, EXT and RM model.

Vector autoregressive Model

HQ criterion is used in purpose of determining the optimal lag number model, where Khim and Liew (2004) proved in their work that the HQ gave better results on a sample more than 60 observations. Since this study had 204 observations, the optimal lag number is determined to HQ criteria which is 1, as shown in Table 2.

Table 2 Selected VAR model using HQ

Lag	HQ Value
0	66.2563
1	52.9225*
2	52.9662

Forecast accuracy measurement

The performance of the forecast models is evaluated by using MAPE. Table 3 shows the forecasting performance of the Robust VAR via filtering that are LMS, LTS and LQD and imputation (Mean and Median) techniques as appeared in Table 4. The best model with a lowest error measurement is chosen for forecasting the economic growth.

Table 3 MAPE values for filtering techniques.

Indicator	LMS	LTS	LQD
CIC	2.1497	2.0225	1.1678
EXC	11.1449	9.6190	8.3425
EXT	2.6329	2.6315	2.6046
RM	1.0182	0.9999	0.9480

Table 3 shows that the LQD has the lowest error for CIC, EXC, EXT and RM. This was followed by LTS and then LMS.

Table 4 MAPE values for imputation techniques.

Indicator	Mean	Median
CIC	2.5824	2.4411
EXC	12.6691	11.2953
EXT	3.9888	3.1169
RM	3.4434	1.8307

For imputation in Table 4, median assigned the lowest error when compared to mean.

Comparative study

The error measurement based on VAR by LQD technique and median replacement is compared. Results are presented in Table 5.

Table 5 Comparison of all results

Indicator	Median	LQD
CIC	2.4411	1.1678
EXC	11.2953	8.3425
EXT	3.1169	2.6046
RM	1.8307	0.9480

Table 5 indicates that the LQD has the lowest error measurement compared to Median. Therefore, the robust VAR by LQD filtering is selected as an alternative model to forecast the economic growth in Malaysia.

CONCLUSION

The forecasts of economic growth measured by using the robust VAR model have been developed, through the filtering techniques and imputation techniques. It is found that the best model judged by MAPE were Robust VAR LQD model that had the lowest error measurement, followed by VAR LTS, VAR LMS, VAR Median and the last one was VAR Mean model. Thus, VAR LQD is chosen to predict the economic growth with highest certainty of accuracy measurement**.**

ACKNOWLEDGEMENT

This work was partly supported by the Universiti Tun Hussein Onn Malaysia under the Centre for Graduate Studies (CGS).

REFERENCES

Agalega, E., & Antwi, S. 2013. The impact of macroeconomic variables on gross domestic product: Empirical evidence from Ghana. *International Business Research*, 6(5), 108–116.

Almonem, M. 2015. Robust methods in regression analysis: Comparison and Improvement. *Journal of Science*, 38–53.

Anscombe, F. J. 1960. Rejection of outliers. *Technometrics*, 123–147.

Berk, R. A. 1990. A primer on robust regression. In Fox, J. & Long, J. S.**,** (Eds.) *Modern Methods of Data Analysis Newbury Park*, CA: Sage Publications, 292-323.

Birkes, D. & Dodge, Y. 1993. *Alternative Methods of Regression*, John Wiley & Sons, Canada, 28-35.

Bruce, L. B., Richard, T. O & Anne B. K. 2005. Forecasting, time series, and regression: An applied approach. *Thomson Brooks/Cole*, 111-125.

Burke, S., & Wycombe, H. 2012. Robust statistics & non-parametric methods. *Statistics and Data Analysis*, 19–24.

Cankaya, S., & Abacı, S. H. 2015. A comparative study of some estimation methods in simple linear regression model for different sample sizes in presence of outliers. *Journal of Agriculture-Food Science and Technology*, 3(6), 380–386.

Chen, C., & Liu, L.-M. 1993. Forecasting time series with outliers. *Journal of Forecasting*, 12(1), 13–35.

Fan, L. & Jerome, R. 2017. An empirical comparison of multiple imputation methods for categorical data. *Journal of the American Statistician*, 162-170.

Fried, R., Bernholt, T., & Gather, U. 2006. Repeated median and hybrid filters. *Computational Statistics and Data Analysis*, 50(9), 2313–2338.

Fried, R. 2007. On the robust detection of edges in time series. *Computational Statistics & Data Analysis*, 52, 1063–1074.

Gaspar, S., Cardoso, A., & Alberto, J. 2017. The traditional Energy-growth nexus : A comparison between sustainable development and economic growth approaches. *Ecological Indicators*, 75, 286–296.

Gupta, R., & Wohar, M. 2017. Forecasting oil and stock returns with a Qual VAR using over 150 years off data. *Energy Economics*, 62, 181–186.

Hampel, F. R., Ronchetti, E. M., Rousseeuw P. J., and Stahel, W.A. 1986. Robust statistics. *The Approach based on Influence Functions*, John Wiley and Sons, New York, 35-41.

Khim, V., & Liew, S. (2004). Which lag length selection criteria should we employ? *Universiti Putra Malaysia: Economics Bulletin*, 3(33), 1–9.

Khodaparasti, R. B., & Moslehi, S. 2014. Application of the Varma Model for sales forecast: Case of Urmia Gray Cement Factory. *Timisoara Journal of Economics and Business*, 7(1), 89–101.

Koivunen, V. 1995 A robust nonlinear filter for image restoration, *IEEE Transactions on Image Processing*, 4, 569-578.

Lanne, M., & Nyberg, H. 2015. Nonlinear dynamic interrelationships between real activity and stock returns. *Econometric Analysis of Time Series*, 85-101.

Nyberg, H. 2017. Forecasting US interest rates and business cycle with a nonlinear regime switching VAR model. *Journal of Forecasting*, 1–15.

Onwukwe, C., & Nwafor, G. 2014. A multivariate time series modeling of major economic indicators in Nigeria. *American Journal of Applied Mathematics and Statistics*, 2(6), 376–385.

Petrovska, M., & Naumovski, N. 2016. Forecasting Macedonian business cycle turning points using Qual VAR model. *Journal of Central Banking Theory and Practice*, 5(3), 1–18.

Reform, G. 2006. Robust filters for intensive care monitoring: Beyond the Running Median. *Econometrics*, 1–21.

Schumacker, R. E., Monahan, M. P., & Mount, R. E. 2002. A comparison of OLS and Robust Regression using S-Plus. *Multiple Linear Regression Viewpoints*, 28(2), 10–13.

Yu, Y., Zhu, Y., Li, S., & Wan, D. 2014. Time series outlier detection based on sliding window prediction. *Mathematical Problems in Engineering*, 2–14.