Vol. 4 (Iss. 2) 2023, pp. 1054-1062 African Journal of Empirical Research https://ajernet.net ISSN 2709-2607

Application of the Vector Autoregressive Model Incorporating New Measurements Using the Bayesian Approach

Michael Musyoki[1]
David Alilah[2]
David Angwenyi[3]

[1]michaelmusyoki816@gmail.com*
[2]dalila@mmust.ac.ke
[3]dangwenyi@mmust.ac.ke

[1]PhD Candidate, [2]Senior Lecturer, [3]Lecturer, [1,2,3] Department of Mathematics, Masinde Muliro University of Science and Technology, Kenya

* Corresponding Author

ABSTRACT

In this paper, an application of the updated vector autoregresive model incorporating new information or measurements is considered. We consider secondary data obtained from the Kenya National Bureau of statistics, Statistical Abstract reports from 2000-2021 which is on monetary value marketed at current prices from crops, horticulture, livestock and related products, fisheries and forestry. A VAR(1) model is fitted to the data and then the model updated to incorporate the measurements. From the results, it is found that the updated model performs well on the simulated data based on the values of the root mean square error obtained.

Keywords: Bayesian Approach, Measurements, Root Mean Square Error, Vector Autoregressive

1. INTRODUCTION

The Vector Autoregressive (VAR) models were developed by the macroeconometrician Christopher Sims in 1980 where the main objective was to model the joint dynamics and causal relations among a set of macroeconomic variables and dominate time series econometrics modeling, Sims (1980). The joint dynamics includes how each variable in the model is explained by the past history of every variable and how the innovations may be correlated (Box et. al. 2015; Quenouille, 1957; Sims, 1980; Stock and Watson, 2001; Tiao and Box, 1981). Recently, the vector autoregressive models have gained much application in a wide range of disciplines such as medicine, epidemiology, economics, biology and macroeconomics among others. For instance, an application of the VAR models is given by Roush et. al. (2017) which is on prediction of gross domestic product using autoregressive models. They constructed a vector autoregressive model of order 4, that is VAR(4), model by selecting few macroeconomic indicators and predicted the Gross domestic product. The study relied on extensive database of historical economic data by the Federal Reserve Bank of St.Louis and found that the results from the model matched with historical data an implication that the model predicted consistently.

Yashavanth et. al. (2017) forecasted prices of coffee seeds using vector autoregressive (VAR) time series model in India. The VAR model was applied to model and forecast monthly wholesale price of clean coffee seeds in different coffee consuming centers namely: Bengaluru, Chennai and Hyderabad. After achieving stationarity, model selection was done based on the Akaike Information Criterion and VAR(2) model was selected. The model was also compared with univariate ARIMA models after which the study concluded that the VAR models fitted better than the ARIMA models based on the forecast accuracy measures. In addition, the study argues that when the ARIMA models are not available, then the VAR model can be used which makes use of the information available from other series when the series are cointegrated. More application of the VAR models is as seen in the works of (Abdullah, 2022; Hamzah et al. 2020; Hossain et. al., 2018; Kalliovirta et. al., 2019; Khairan et. al., 2022; Elalaoui et. al., 2021; Quenouille, 1957; Saheed et. al., 2021).

Despite the fact that the VAR models have been applied extensively in many areas due to their ability to perform well, there is a concern of what happens in the event that new information is obtained. In this paper, we consider fitting a VAR model to the data on crops, horticulture, livestock and related products, fisheries and forestry and then afterwards, update the model to incorporate new information or measurements.

The rest of the paper is structured as follows: first we have research methods in Section II, the fitted VAR model in section III, update of the model in section IV and then the conclusions given in section V.

II. RESEARCH METHODS

This study considers the VAR modeling technique for model fitting. The identification or fitting of an ordinary VAR model involves model specification, estimation of model parameters and model checking to test whether the model is adequate. The order, p, of VAR is chosen which minimizes the Schwartz and Hannan-Quinn criteria as outlined by Lutkepohl (2005). The Schwartz criterion is given by

$$SC(p) = \ln|\widehat{\Sigma}_u(p)| + \frac{\ln T}{T} pv^2$$

On the other hand, the Hannan-Quinn criterion is given by

$$HQ(p) = \ln|\widehat{\Sigma}_u(p)| + \frac{2\ln\ln T}{T} pv^2$$

where, for both criteria, $\hat{\Sigma}_u$ is the estimated white noise covariance matrix, T is the sample size and v is the number of time series components. The criteria compares the residuals of the models and estimates the relative information loss of representing the original data using each of the model. In addition, the criteria weighs the quality of fit (covariance of residuals) against the complexity (number of free parameters) and therefore the model with least criterion value is considered as seen in Roush et. al. (2017). The parameters of a fitted VAR model can be estimated by ordinary least squares estimation method under the assumptions that error term has mean of zero, the variables are stationary and no outliers. The developed model is then subjected to diagnostic checking for its adequacy and this involves checking whether the residuals are white noise, normally distributed and uncorrelated. Afterwards, the model is used to forecast. To update the model, we consider the algorithm given by

Algorithm 1 Algorithm for Generalized updated VAR(p) model

1: Predict the state: $\hat{Y}_{t|t-1} = A_{1,t-1}\hat{Y}_{t-1|t-1} + \cdots + A_{p,t-p}\hat{Y}_{t-p|t-p}$
2: Predict the error covariance: $\hat{S}_{t|t-1} = A_{1,t-1}S_{t-1}A^T_{1,t-1} + \cdots + A_{p,t-p}S_{t-p}A^T_{p,t-p} + Q$
3: Compute the gain: $K_t = \frac{S_{t|t-1}P_t^T}{|P_t S_{t|t-1}P_t^T + R}$
4: Update the state: $\hat{Y}_{t|t} = A_{1,t-1}\hat{Y}_{t|t-1} + K_t\left(X_t - P_t\hat{Y}_{t|t-1}\right)$
5: Update the error covariance: $\hat{S}_{t|t} = S_{t|t-1} - K_t P_t S_{t|t-1}$

We consider secondary data obtained from the Kenya National Bureau of statistics, Statistical Abstract reports from 2000-2021 which is on monetary value marketed at current prices (Ksh. Million) from crops, horticulture, livestock and related products, fisheries and forestry.

III. THE FITTED VAR MODEL

The secondary data obtained was entered into Excel and saved under CSV format.
It was then read in R software for analysis. First, a time series plot of the variables is done which is as given in Figure 1.

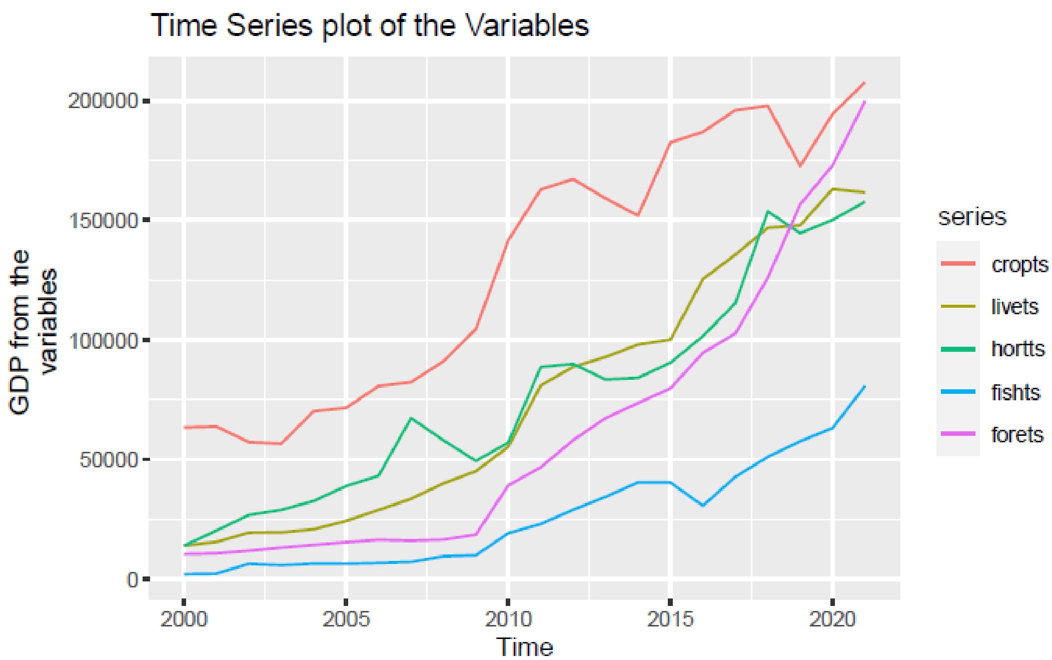

Figure 1: Time series plot of GDP from the variables, namely; crops, livestock and related products, horticulture, fisheries and forestry.

From figure 1, we observe that almost all the variables depict an increasing trend implying that the GDP from the variables has been increasing steadily over the years. Thus the series are non-stationary as confirmed by Dickey-Fuller test. We make the series stationary by applying differencing twice to the log of the variables. This gives the plot in Figure 2.

Figure 2: Time series plot of twice differenced log GDP from the Variables, that is, after applying second differencing to the log of the variables.

From Figure 2, it is observed that the variables appear stationary and can be adopted. The Augmented Dickey Fuller test also shows that the series are stationary. Using the lagselect function, we find that the Akaike information criteria (AIC), Hannan-Quinn (HQ) criteria, Schwartz Criteria (SC) and Final Prediction Error (FPE) selects order of the model as 1 i.e p = 1 implying VAR(1) model is a suitable model. The model is given by

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \\ y_{4,t} \\ y_{5,t} \end{pmatrix} = \begin{pmatrix} -0.3930 & -0.0627 & -0.1585 & -0.0901 & 0.0852 \\ 0.2784 & -0.4275 & 0.1043 & -0.0419 & 0.2829 \\ 0.6599 & 0.0578 & -0.2950 & -0.0195 & 0.4020 \\ -0.2265 & 0.0063 & 0.1889 & -0.6139 & 0.1257 \\ 0.1563 & -0.2264 & -0.0798 & -0.0194 & -0.4631 \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \\ y_{4,t-1} \\ y_{5,t-1} \end{pmatrix} + \begin{pmatrix} 0.0001487 \\ -0.001047 \\ -0.0008652 \\ 0.0001187 \\ -0.0000301 \end{pmatrix} + \begin{pmatrix} \omega_{1,t} \\ \omega_{2,t} \\ \omega_{3,t} \\ \omega_{4,t} \\ \omega_{5,t} \end{pmatrix} \quad (1)$$

The eigenvalues of the matrix of the penta-variate VAR model in equation 1 are obtained by solving for λ in the equation

$$det[\Phi_1 - \lambda I_5] = 0 \quad (2)$$

Where

$$\Phi_1 = \begin{pmatrix} -0.3930 & -0.0627 & -0.1585 & -0.0901 & 0.0852 \\ 0.2784 & -0.4275 & 0.1043 & -0.0419 & 0.2829 \\ 0.6599 & 0.0578 & -0.2950 & -0.0195 & 0.4020 \\ -0.2265 & 0.0063 & 0.1889 & -0.6139 & 0.1257 \\ 0.1563 & -0.2264 & -0.0798 & -0.0194 & -0.4631 \end{pmatrix} \quad (3)$$

Equivalently, this is given by

$$det \begin{pmatrix} -0.3930-\lambda & -0.0627 & -0.1585 & -0.0901 & 0.0852 \\ 0.2784 & -0.4275-\lambda & 0.1043 & -0.0419 & 0.2829 \\ 0.6599 & 0.0578 & -0.2950-\lambda & -0.0195 & 0.4020 \\ -0.2265 & 0.0063 & 0.1889 & -0.6139-\lambda & 0.1257 \\ 0.1563 & -0.2264 & -0.0798 & -0.0194 & -0.4631-\lambda \end{pmatrix} = 0 \quad (4)$$

The eigenvalues are $\lambda_1 = -0.7078666 + 0i$, $\lambda_2 = -0.2623644 + 0.4698648i$, $\lambda_3 = -0.2623644 - 0.4698648i$, $\lambda_4 = -0.4799684 + 0.0770515i$ and $\lambda_5 = -0.4799684 - 0.0770515i$ whose moduli are 0.7078666, 0.5381524, 0.5381524, 0.4861138 and 0.4861138 respectively. All the eigenvalues have modulus less than one (lie within the complex unit circle) thus the model is stable.

We tested for Granger-Causality of the variables and found that Crops, Livestock, Horticulture, Fishing and Forestry do not Granger-cause the other variables. The test on normality of the residuals for the model in Equation 1 found that the residuals are normally distributed.

IV. UPDATING THE MODEL

Suppose we now combine the fitted model given in equation 1 with the measurement equation so that we have,

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \\ y_{4,t} \\ y_{5,t} \end{pmatrix} = \begin{pmatrix} -0.3930 & -0.0627 & -0.1585 & -0.0901 & 0.0852 \\ 0.2784 & -0.4275 & 0.1043 & -0.0419 & 0.2829 \\ 0.6599 & 0.0578 & -0.2950 & -0.0195 & 0.4020 \\ -0.2265 & 0.0063 & 0.1889 & -0.6139 & 0.1257 \\ 0.1563 & -0.2264 & -0.0798 & -0.0194 & -0.4631 \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \\ y_{4,t-1} \\ y_{5,t-1} \end{pmatrix} + \begin{pmatrix} u_{1,t} \\ u_{2,t} \\ u_{3,t} \\ u_{4,t} \\ u_{5,t} \end{pmatrix}$$
$$\begin{pmatrix} x_{1,t} \\ x_{2,t} \\ x_{3,t} \\ x_{4,t} \\ x_{5,t} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \\ y_{4,t} \\ y_{5,t} \end{pmatrix} + \begin{pmatrix} \eta_{1,t} \\ \eta_{2,t} \\ \eta_{3,t} \\ \eta_{4,t} \\ \eta_{5,t} \end{pmatrix} \quad (5)$$

We then subject equation 5 to algorithm 1 to update the model. Setting

$$Q = \begin{pmatrix} 0.0190578 & 0.001970 & 0.005758 & 0.0002899 & 0.009302 \\ 0.0019701 & 0.008454 & 0.005498 & 0.0043966 & 0.001251 \\ 0.0057579 & 0.005498 & 0.047239 & 0.0244832 & 0.004615 \\ 0.0002899 & 0.004397 & 0.024483 & 0.0982311 & 0.029322 \\ 0.0093015 & 0.001251 & 0.004615 & 0.0293222 & 0.046527 \end{pmatrix}$$

$$R = \begin{pmatrix} 0.001 & 0 & 0 & 0 & 0 \\ 0 & 0.001 & 0 & 0 & 0 \\ 0 & 0 & 0.001 & 0 & 0 \\ 0 & 0 & 0 & 0.001 & 0 \\ 0 & 0 & 0 & 0 & 0.001 \end{pmatrix}$$

And

$$S_0 = \begin{pmatrix} 0.001 & 0 & 0 & 0 & 0 \\ 0 & 0.001 & 0 & 0 & 0 \\ 0 & 0 & 0.001 & 0 & 0 \\ 0 & 0 & 0 & 0.001 & 0 \\ 0 & 0 & 0 & 0 & 0.001 \end{pmatrix}$$

in MATLAB, the plots in Figure 3 - Figure 7 are obtained for the five variables being referred to.

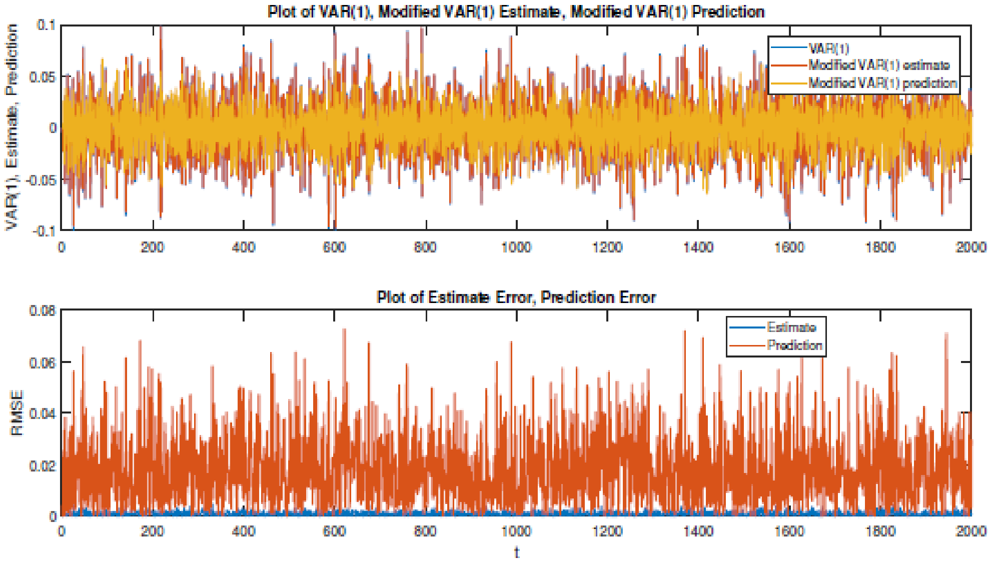

Figure 3: Fitted Pentavariate VAR(1) - Variable 1. Subplot one gives comparison of the VAR(1), modified VAR(1) estimate and modified VAR(1) prediction, denoted by the blue line, red line and the yellow line, respectively while the second subplot shows the errors between VAR(1) and the modified VAR(1) and between VAR(1) and the modified VAR(1) prediction as time evolves.

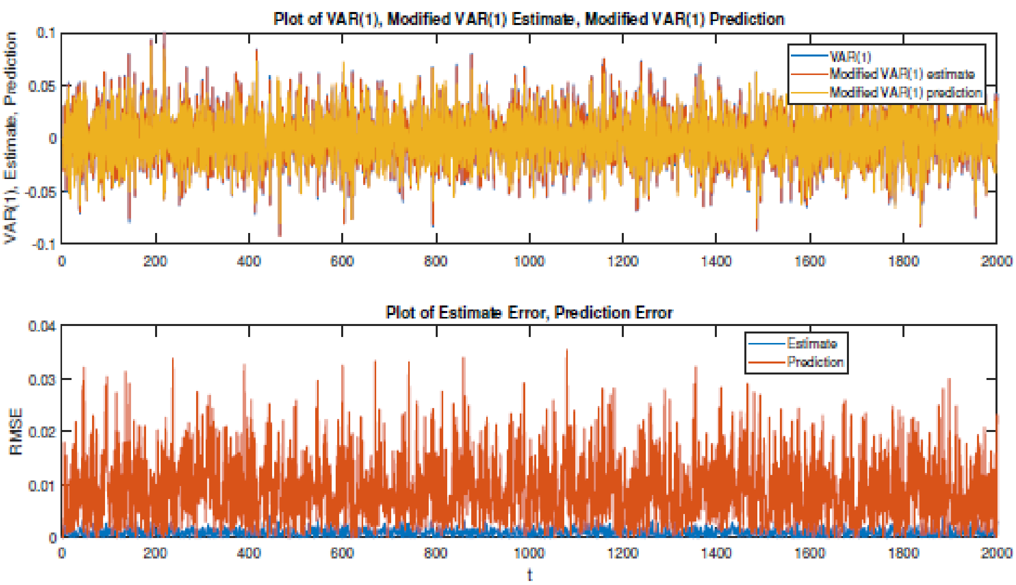

Figure 4: Fitted Pentavariate VAR(1) - Variable 2. The first subplot gives comparison of the VAR(1), modified VAR(1) estimate and modified VAR(1) prediction, denoted by the blue line, red line and the yellow line, respectively while the second subplot the errors between VAR(1) and the modified VAR(1) and between VAR(1) and the modified VAR(1) prediction.

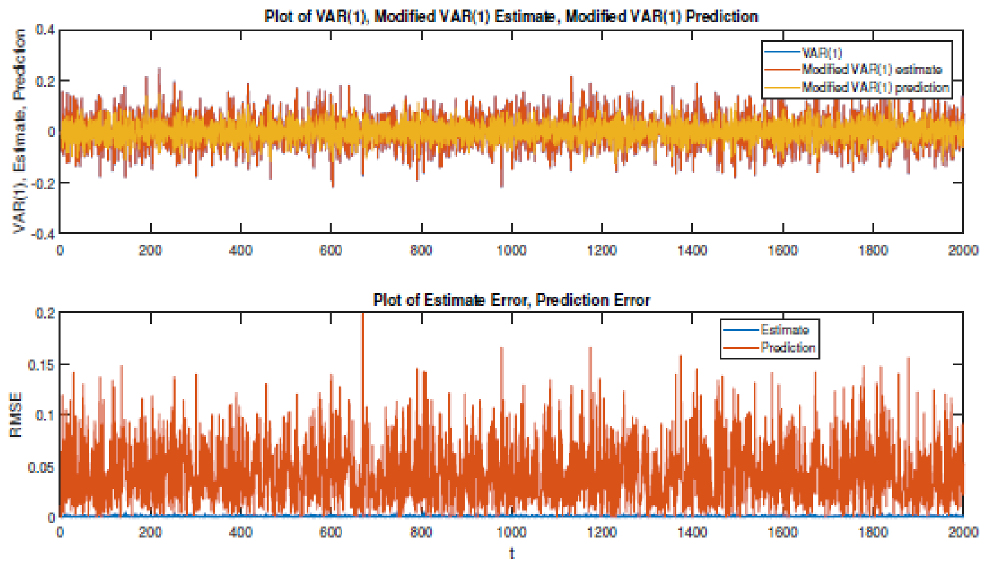

Figure 5: Fitted Pentavariate VAR(1) - Variable 3. In subplot one, comparison of the VAR(1), modified VAR(1) estimate and modified VAR(1) prediction, denoted by the blue line, red line and the yellow line, respectively is shown, while the second subplot displays the errors between VAR(1) and the modified VAR(1) and between VAR(1) and the modified VAR(1) prediction.

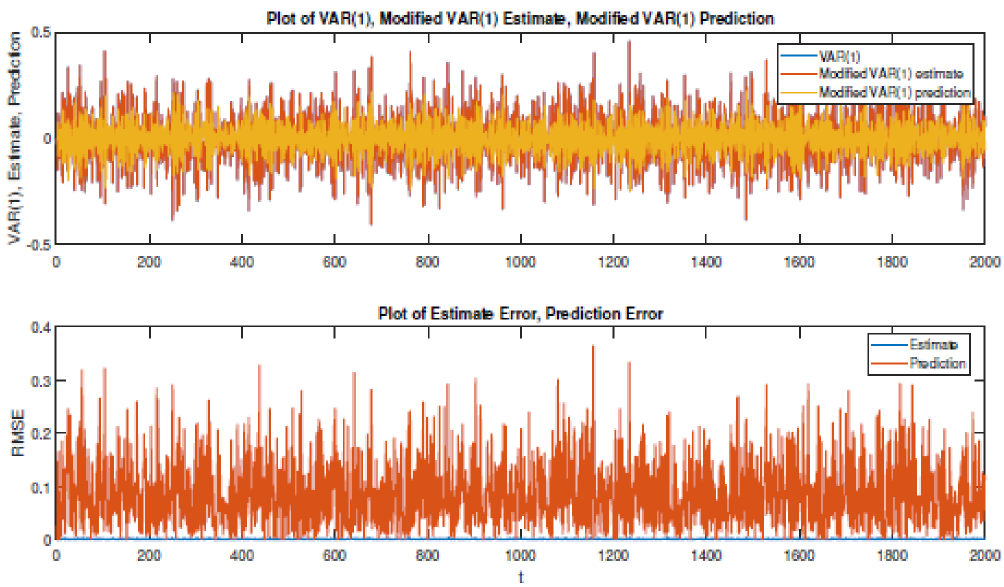

Figure 6: Fitted Pentavariate VAR(1) - Variable 4. Here, the first subplot gives comparison of the VAR(1), modified VAR(1) estimate and modified VAR(1) prediction, denoted by the blue line, red line and the yellow line, respectively while the second subplot shows the errors between VAR(1) and the modified VAR(1) and between VAR(1) and the modified VAR(1) prediction.

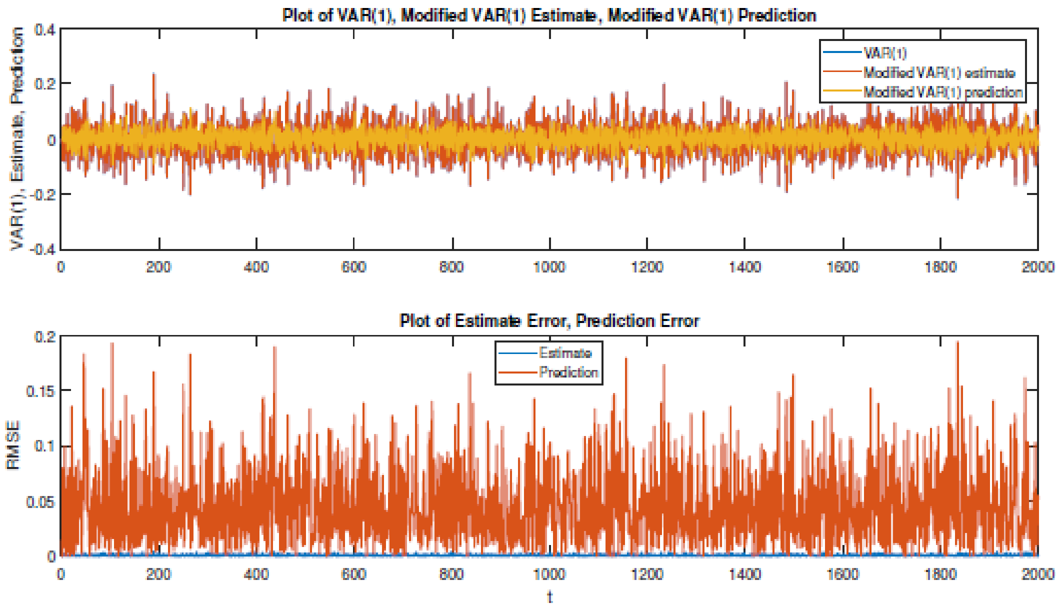

Figure 7: Fitted Pentavariate VAR(1) - Variable 5. In the first subplot, we have comparison of the VAR(1), modified VAR(1) estimate and modified VAR(1) prediction, denoted by the blue line, red line and the yellow line, respectively while in the second subplot, we have the errors between VAR(1) and the modified VAR(1) and between VAR(1) and the modified VAR(1) prediction.

From Figures 3 - 7, which represent the first, second, third, fourth and fifth variables, that is Crops, Livestock, Horticulture, Fishing and Forestry respectively, it can be observed that the values of the root mean square error are fairly small an indication that the updated model performs well.

V. CONCLUSIONS

In this paper we focused on showing the concept of updating VAR models. We considered secondary data obtained from the Kenya National Bureau of statistics, Statistical Abstract reports from 2000-2021 which was on monetary value marketed at current prices from crops, horticulture, livestock and related products, fisheries and forestry. A VAR(1) model was then fitted for the mentioned variables. The model was then updated to incorporate new information. The plots of the Root Mean Square Error (RMSE) shows that the updated model performs well as indicated by small values of RMSE in the update and in the prediction steps.

Data Availability

The data used to support the study's findings is secondary data available upon request from the author or obtained from the Kenya National Bureau of Statistics (KNBS) website on quarterly GDP reports and some simulated data.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

REFERENCES

Abdullah T. L. (2022). Forecasting time series using Vector Autoregressive Model, *Int. J. Nonlinear Anal. Appl.*, 13(1), 499-511.

Box G. E. P., Jenkins G. M., Reinsel G. C., & Ljung G. M. (2015). Time Series Analysis: Forecasting and Control, 5th Edition, ISBN: 978-1-118-67502-1.

Elalaoui, O., Fadlaoui, A., Maatala, N., & Ibrahimy, A. (2021). Agriculture and GDP Causality Nexus in Morocco: Empirical Evidence from a VAR Approach. *International Journal of Agricultural Economics*, *6*(4), 198-207. https://doi.org/10.11648/j.ijae.20210604.17

Hamzah L. M., Nabilah S. U., Usman M., & Wamiliana, V.E. (2020). Dynamic Modeling and Forecasting of Data Export of Agricultural Commodityby Vector Autoregressive Model, *Journal of Southwest Jiaotong University*, 55(3), 1 - 10.

Hossain, M., Al Amin, A. A., & Islam, A. H. M. S. (2018). Modeling and forecasting of climatic parameters: univariate SARIMA versus multivariate vector autoregression approach. *Journal of the Bangladesh Agricultural University*, *16*(1), 131–143. Retrieved from https://www.banglajol.info/index.php/JBAU/article/view/36494

Kalliovirta, L., Niskanen, O., & Heikkila, A. M. (2019). *Forecasting Milk prices with VAR Models - Application to Farm Gate Price in Finland*. Available at SSRN: https://ssrn.com/abstract=3473862 or http://dx.doi.org/10.2139/ssrn.3473862

Khairan R., Firuz K. and Aswani K. C. (2022) Forecasting COVID-19: Vector Autoregression-Based Model. *Arabian Journal for Science and Engineering*, *47*, 6851–6860.

Lutkepohl H. (2005). *New Introduction to Multiple Time Series Analysis*. Berlin, Heidelberg, New York: Springer.

Quenouille, M. H. (1957). *The Analysis of Multiple Time-Series*. London: Griffin.

Roush, J., Siopes, K., & Hu, G. (2017). “Predicting gross domestic product using autoregressive models.” IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA), 317-322. 10.1109/SERA.2017.7965745.

Saheed B. A., Kola Y. K., Adewole O. G., Muideen O. A., Saheed O. J., Muhammed T. M., Ekundayo S. O., Olakiitan I. A., Magdalene P., & Oluwole J. O. (2021) Vector Autoregressive Modeling of Crop Production Index – Permanent Cropland Relationship in Nigeria, *Annals: Computer Science Series*, 19(1), 92 - 97.

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

Stock H. J.S., & Watson W. M. (2001). *Vector Autoregressions*. https://faculty.washington.edu/ezivot/econ584/stck watson var.pdf

Tiao G. C. and Box G. E. P. (1981). Modeling Multiple Time Series with Applications, *Journal of the American Statistical Association*, *76*(376) 802-816.

Yashavanth B. S., SinghK. N., Paul K. A. and Paul K. R. (2017) Forecasting prices of coffee seeds using Vector Autoregressive Time Series Model. *Indian Journal of Agricultural Sciences*, *87*(6), 754-758.