VECTOR AUTOREGRESSIVE MODEL OF MAIZE PRODUCTION IN NORTHERN REGION OF GHANA

ABSTRACT

Agricultural growth plays a crucial role in the Comprehensive African Agriculture Development Programme (CAADP, 2009) agenda. The program recognizes that increasing agricultural productivity is essential for reducing poverty, meeting food production targets, and lowering production costs and food prices for the impoverished.

This study aimed to develop two types of models. The first model employed a vector autoregressive (VAR) approach, which involved regressing the production of maize in one district against the production of maize in other districts at various lags. The second model utilized a VAR framework where the maize production in each district was regressed against the production of maize in other districts and the corresponding climate conditions at different lag periods.

The available data spanned from 1968 to 2018 and were recorded on an annual basis. Six climatic variables were included in the analysis. A lag order of 3 was selected for the models. The results of the autocorrelation test using the portmanteau test indicated no serial autocorrelation across all lag periods. The test for normality revealed that the residuals followed a normal distribution. Additionally, there was no evidence of heteroscedasticity in the data.

Furthermore, a Granger causality test was conducted on the selected districts to explore causal relationships. Variance decomposition analysis was performed to assess the variance relation in the data and understand the contribution of different factors. Based on adjusted R-squared, mean absolute error (MAE), and root mean squared error (RMSE) values, the models that incorporated climatic variables were found to be the most suitable for forecasting maize production in the selected districts. The VAR model, which captures the interdependencies between the variables, was utilized in this analysis. All variables in the VAR model were treated symmetrically, meaning that their relationships were considered equally important.

KEYWORDS: serial autocorrelation, test for normality, heteroskedasticity, Granger causality, portmanteau test, vector autoregression (VAR)

1. Introduction

One of the main goals of the Comprehensive African Agriculture Development Programme (CAADP, 2009) is the growth of agriculture. This is due to its ability to play an important role in reducing poverty and food insecurity. The agricultural sector constitutes over 30 percent of Ghana's GDP [1]. Crop yield in Ghana is generally low, ranging from 20 to 60 percent below the expected production. For example, cassava is expected to yield 28.0 Mt/ha but only yields 12.4 Mt/ha [1]. Maize is another crop that falls short of its potential, producing 1.7 Mt/ha instead of 6.0 Mt/ha. Ghana can enhance its agriculture sector by adopting strategies from the Asian Green Revolution. Maize production in Ghana faces various challenges, including inadequate use of improved varieties and poor soil fertility management. Maize is an important crop to ensure global food security. The interest in maize production has increased due to its demand as a biofuel, food, and feed crop in many countries [21,22]

Maize is considered an important crop for Ghana's agricultural sector and food security. It is a versatile crop cultivated by predominantly smallholder resource-poor farmers across all agroecological zones in Ghana under rain-fed conditions [2,3].

The adoption of new technologies and the effective use of existing ones without degrading natural resources can improve agricultural productivity [4]. Mechanization of farm operations is a crucial step in expanding production effectiveness [5]. Approximately 40 percent of farmers use some form of mechanization [1]. Tractor usage for land preparation reduces technical inefficiencies by ensuring timely land preparation. Maize is one of the most important cereal crops, widely used as food for humans, animal feed, and various industries [6]. Maize faces challenges in Sub-Saharan Africa, such as climate and soil factors, which can reduce production by about 80 percent annually due to drought stress [7]. The cereal that receives the most fertilizer application in Sub-Saharan Africa is maize [10,11]. The coastal savannah, forest zone, transition zone, and Guinea savannah zones are the main agroecological zones for maize cultivation in Ghana. While Ghana cultivates various cereal crops, maize accounts for 50 percent of the cereal crops produced. Post-harvest losses in Ghana range from 5 to 70 percent, often caused by storage pests such as weevils and Sitophilus zeamais (Mots.). The yearly land area used for maize cultivation in Ghana is approximately 650,000 ha [13]. Maize is cultivated alongside other cereal crops on the same piece of land in the coastal savannah and forest zones. Ghana produces over one million tons of maize annually. Maize production depends on the area planted and the yield, both of which have increased over the years. The transition zones have become important for maize cultivation due to favorable agroecological conditions, improved production technology, underutilized lands, and a good road transportation system [13]. Maize is not only one of the most important cereal crops produced in the country but also the most widely consumed staple crop. Maize production in Ghana has been increasing since 1965 [14, 15] and is primarily carried out by smallholder farmers. The main factors contributing to low maize productivity are the inadequate application of external inputs and low soil fertility. Maize production is not limited to Ghana; it takes place worldwide.

One of the major sources of calories is maize [1]. It has nearly replaced sorghum and other cereal crops in the northern part of the country. The average annual production of maize between 2007 and 2010 was 1.5 million metric tons [16]. The crop has been increasing its yield by 1.1% annually [17]. Maize is one of the thirty (30) cereal crops in the world that provides 90% of the caloric requirement. Ghanaians consume a significant amount of maize, as stated by MOFA [18].

The per capita consumption of maize is estimated to be about 42.5kg. Ghana sells approximately one million metric tons of maize annually, with most of the crop being consumed within households. The consumption of maize grains depends on cultural and traditional practices. Although maize is primarily consumed by humans, a large quantity is also used as animal feed, such as for poultry. Industrial processing accounts for only about 20% to 25% of the maize being marketed. The wholesale marketing of maize depends on location and transportation [19]. The use of improved seeds and fertilizers by farmers in the 1980s led to production gains of maize in sub-Saharan Africa [20]. One of the main methods used to improve food security is the reduction of post-harvest losses, as this can result in increased food prices and a decrease in farmers' income.

2. Methodology

The vector autoregressive model (VAR) is considered the most appropriate time series model to utilize when attempting to assess the degree of interaction that exists among the variables under consideration. It represents an extension of the univariate time series model and is widely employed in analyzing multivariate time series data. The VAR model serves as a typical example of a multivariate time series model, allowing for a comprehensive understanding of the relationships and dependencies among the variables

Consider a system of p time series variables, denoted as $y_t = [y_{1t}, y_{2t}, \dots \dots \dots \dots . y_{pt}]^I$, where t represents the time index. The P-dimensional VAR(P) model can be written as:

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

where:

- y_t is a p-dimensional vector of variables at time t,
- c is a p- dimensional vector of constants,
- $A_{1,}A_2, \dots \dots \dots \dots A_P$ are coefficient matrices of dimensions $P \times P$, which represent the contemporaneous relationships between the variables,
- $y_{t-1}, y_{t-2}, \dots \dots \dots \dots \dots y_{t-p}$ are lagged values of the variables,
- e_t is a p-dimensional vector of error terms at time t.

To provide a more detailed representation, we can expand the VAR(p) model for each variable j in the system:

$$y_{jt} = c_j + A_{j1} * y_{1,t-1} + A_{j2} * y_{2,t-1} + \dots \dots \dots + A_{jp} * y_{p,t-1} + e_{jt}$$

where:

- y_{jt} is the jth jth variable at time t,
- c_j is the constant term specific to the jth variable,
- A_{ji} represents the coefficient matrix associated with the lagged values of the jthvariable from all p variables,
- $y_{i,t-1}$ represents the lagged values of the ith variable from all p variables,
- e_{jt} is the error term specific to the jth variable at time t.

The VAR model captures the dynamic interdependencies among the variables by including lagged values as predictors. The coefficients A_{ji} represent the contemporaneous relationships between the variables, indicating how each variable responds to changes in its own lagged values and the lagged values of other variables in the system.

Estimating the VAR model involves estimating the coefficient matrices $A_1, A_2, \dots \dots \dots A_p$ and the error terms e_t. Various techniques, such as ordinary least squares (OLS), maximum likelihood estimation, or Bayesian methods, can be employed for parameter estimation. Once the model is estimated, it can be used for forecasting future values of the variables and analyzing the relationships between them.

It's important to note that the VAR model assumes stationarity of the variables and may require data preprocessing techniques, such as differencing or transformation, to satisfy this assumption. Additionally, model selection techniques, such as information criteria or hypothesis testing, can be used to determine the appropriate lag order p for the VAR model.

The steps involved in modeling a VAR model were thoroughly examined. These steps include lag order selection, testing for autocorrelation, testing for normality, testing for heteroskedasticity, conducting the Granger causality test, and performing variance decomposition analysis.

There is a relationship between the number of lags and the degrees of freedom. Specifically, there exists an inverse relationship, indicating that as the lag increases, the degrees of freedom decrease. The lag with the minimum AIC is selected. If this criterion is not met, a likelihood ratio (LR) test is applied. (Johansen 1995).

To compare a model m_o with another model m_A , when model m_o is nested inside model m_A we can use a likelihood ratio test. The idea of nesting simply means that the less general model m_o (representing the null hypothesis) can be obtained by constraining some of the parameters of the more general model m_A (representing the alternative hypothesis)

$$LR = -2\ln\left(\frac{L(\widehat{\theta_o}/x)}{L(\widehat{\theta_A}/x)}\right) \quad (8)$$

Where $\hat{\theta}_o$ and $\hat{\theta}_A$ are the MLEs for the parameters of the respective models. If the model $\hat{\theta}_o$ is true and sample size large, then the distribution of *LRT* is chi-square with degrees of freedom equal to the number of additional parameters in model $m_A$vs. model m_o . This is related to the profile likelihood method used for computing approximate confidence intervals. Both are based on a likelihood ratio having an asymptotic chi-square distribution.

The Akaike Information Criterion (AIC) and the Schwarz Information Criterion (BIC) serve as measures of model fit, allowing us to choose the most suitable model among a set of estimated ones. We consider the model with the lowest AIC or BIC value as the best fit. These statistics are calculated using specific mathematical formulas.

$$AIC = \ln\frac{RSS}{n-k}+\frac{2}{n}k \quad (9)$$

$$BIC = \ln\frac{RSS}{n-k}+\frac{k}{n}\ln n \quad (10)$$

The residual sum of squares (RSS) represents the sum of squared residuals in the regression model, while 'n' denotes the sample size and 'k' signifies the number of parameters in the model. In contrast, the Hannan-Quinn Criterion can be expressed using the following equation or
Relationship.

$$.HQC = -2L_{max} + 2k\ln\ln n \quad (11)$$

The log-likelihood, denoted as L_{max} , is a measure of the goodness of fit in a statistical model. The parameter 'k' represents the number of parameters in the model, while 'n' corresponds to the number of observations in the dataset. The relation that exists between maize and climate variables in a VAR model can be deliberated by observing the impact of the effect of the variable of interest.

3. Data

Climate data was collected from the Council for Scientific and Industrial Research (CSIR), an agricultural research center located at Nyankpala. The data covers the period from 1968 to 2018 and includes measurements such as rainfall (mm), evaporation (kph), wind speed (km/hr), relative humidity (percentage), sunshine duration, and temperature (degrees Celsius). The maize data, obtained from the Ministry of Food and Agriculture (MOFA), represents maize production measured in metric tons (MT).

Several techniques were employed for the data analysis, including the vector auto-regressive model (VAR) and associated factors. The VAR model involves an initial set of random vectors, which is then extended to incorporate the initial vector of maize production from various districts, along with another set of vector variables representing climatic conditions.

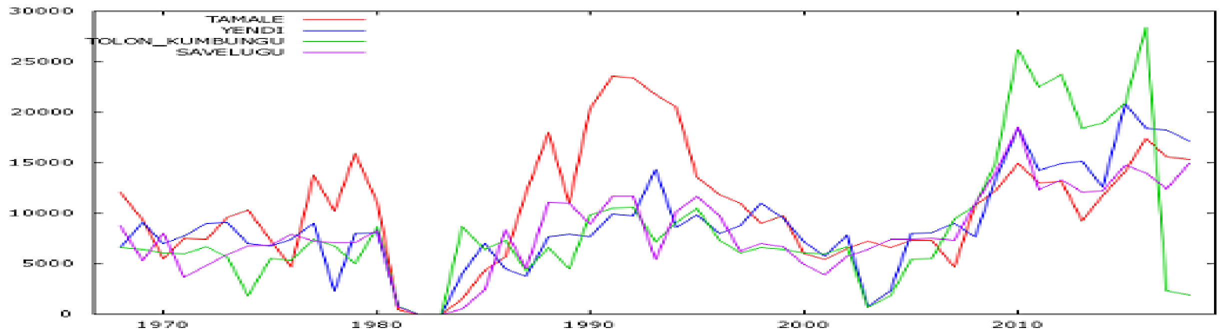

Figure 1: Time Series Plots of Maize Production in the Four (4) Selected Districts in Northern Region of Ghana.

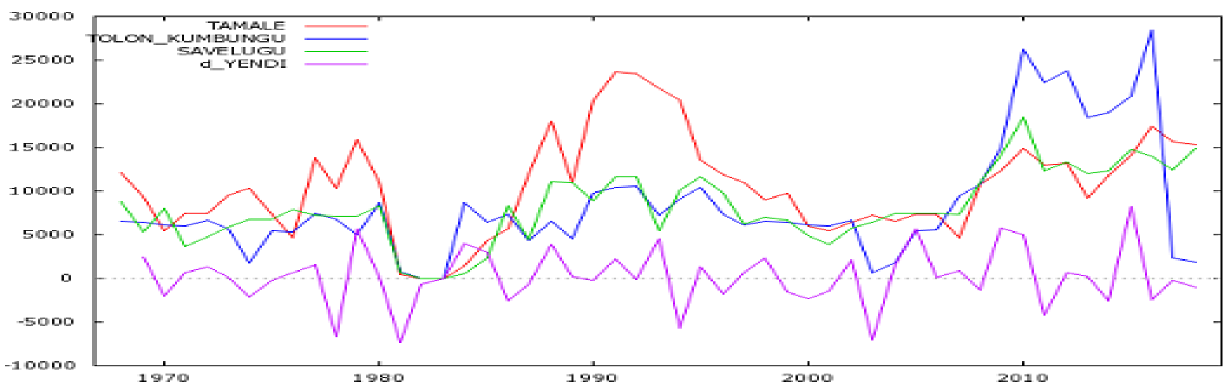

Figure 2: Plot of Maize Production in Tamale, TolonKumbungu, Savelugu and A First Difference of Maize Data in Yendi.

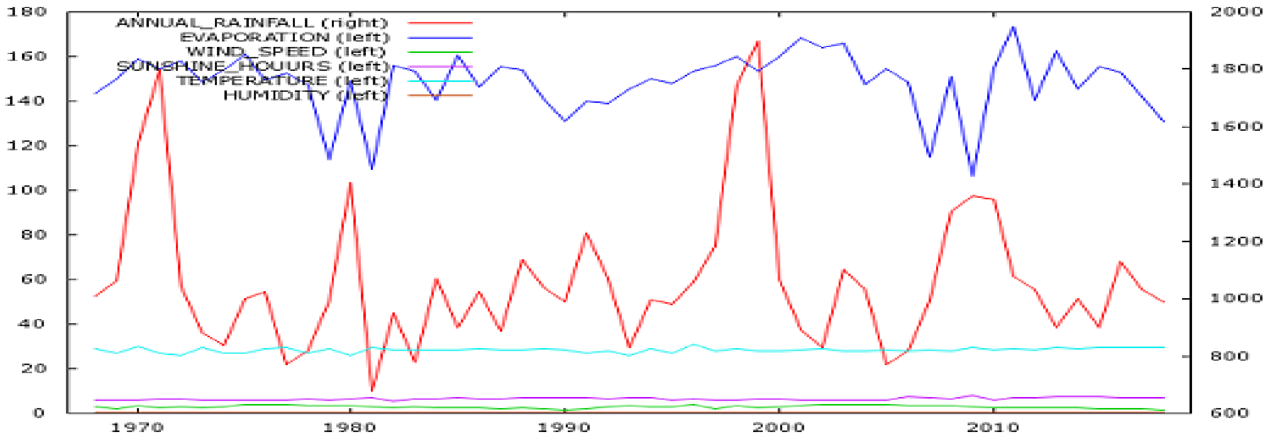

Figure 3: Plot of Original Time Series Data of Climate Variables (same graph) in Northern Region.

4. Test for stationarity

Statistical tests such as the Augmented Dickey-Fuller (ADF) test or the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test can be employed.

The hypotheses for the Augmented Dickey-Fuller (ADF) test are:

Null hypothesis (H0): The maize production time series is non-stationary because there is a unit root (if p-value > 0.05)

Alternative hypothesis (H1): The maize production time series is stationary because there is no unit root (if p-value ≤ 0.05)

Table 1 Probability values in different region

Variable	Probability values	Conclusion
Tamale	0.04786	Stationary
Yendi	2.46e-11	Stationary
Tolonkumbungu	0.02688	Stationary
Savelugu	0.04485	Stationary

Source: Author's computation from the data, 2019

Null hypothesis (H0): The Climate Conditions time series is non-stationary because there is a unit root (if p-value > 0.05)

Alternative hypothesis (H1): The Climate Conditions time series is stationary because there is no unit root (if p-value ≤ 0.05)

Table 2 Probability values in different variability parameters

Variable	Probability values	Conclusion
Annual rainfall	9.713e-05	Stationary
Evaporation	1.172e-07	Stationary
Wind speed	0.0008704	Stationary
Sunshine(hours)	1.233e-05	Stationary
Temperature	8.641e-11	Stationary
Humidity	0.0002549	Stationary

Source: Author's computation from the data, 2019

5. Lag Order Selection

Table 3 Results of Lag Order Selection

Lags	Loglik	p(LR)	AIC	BIC	HQC
1	-2218.74921		99.095711	103.425844*	100.725169
2	-2115.46101	0.00000	98.955788	107.222405	102.066571
3	-1944.73354	0.00000	95.946108*	108.149209	100.538217*

Source: Author's computation from the data, 2019

The portmanteau test is specifically designed to detect the presence of autocorrelation or serial correlation in the residuals, which implies that the model does not adequately capture the temporal dependencies in the data. An LMF (Langrage Multiplier-F) test for autocorrelation was conducted

Null Hypothesis (H_0): The residuals of the model under consideration are independently and identically distributed (if Portmanteau Test p-value > 0.05).

Alternate Hypothesis (H_1): The residuals of the model under consideration exhibit autocorrelation or serial correlation (if Portmanteau Test p-value ≤ 0.05).

Table 4 Data statistical results

Lags (TAM)	Multiple R-square(R2)	Df	Portmanteau Test p-value.	p-value
1	0.6405	1900	1	1.1e-05
2	0.7263	1800	0.999	1.57e-03
3	0.8842	1700	0.1518	2.71e-03
Lags (YEN)	Multiple R-square(R2)	Df	Portmanteau Test p-value	p-value
1	0.1559	1900	1	0.7851
2	0.5006	1800	0.999	0.2644
3	0.6242	1700	0.1518	0.655
Lags (TOL)	Multiple R-square(R2)	Df	Portmanteau Test p-value	p-value
1	0.6697	1900	1	2.6e-06
2	0.7775	1800	0.999	1.7e-04
3	0.8431	1700	0.1518	1.7e-02
Lags (SAV)	Multiple R-square(R2)	Df	Portmanteau Test p-value	p-value
1	0.7240	1900	1	1.2e-07
2	0.8368	1800	0.999	4.9e-06
3	0.9102	1700	0.1518	5.09e-04

Source: Author's computation from the data, 2019

6.Test for Normality

The Jarque-Bera test on normality was conducted on the data as shown below.

Null Hypothesis (H_0): The data in question follows a normal distribution. In this case, for the Jarque-Bera test, the null hypothesis is that the skewness and kurtosis of the data are consistent with those expected in a normal distribution (if $p > 0.05$)

Alternative Hypothesis (H_1): The data does not follow a normal distribution. In other words, the skewness and kurtosis of the data deviate significantly from the expected values under a normal distribution assumption (if $p \leq 0.05$)

Table 5 Jarque-Bera test results

	Jarque-Bera test	Skewness	Kurtosis
Df	20	10	10
Chi-squared	16.59	8.5728	8.0169
p-value	0.6794	0.5731	0.6272

Source: Author's computation from the data, 2019

7. Test for Heteroskedasticity

An ARCH (Autoregressive Conditional Heteroskedasticity) is conducted on the Data set.

Null Hypothesis (H_0): There is no heteroskedasticity present in the regression model. It posits that the error variance is constant across all levels of the independent variables (if $p > 0.05$)

Alternative Hypothesis (H_1): There is heteroskedasticity in the regression model (if $p \leq 0.05$)

Table 6 Result of chi-square test

Chi-squared	Df	p-value
330	1200	1

Source: Author's computation from the data, 2019

8. Granger causality test

Null Hypothesis (H_0): The past values of Variable X do not have a significant causal effect on the future values of Variable Y (if $p > 0.05$)

Alternative Hypothesis (H_1): The past values of Variable X do have a significant causal effect on the future values of Variable Y (if $p \leq 0.05$)

Table 7 Granger causality test result

causality between	Town	P-value	conclusion
Tamale	Yendi	0.3	Tamale does granger cause Yendi
Tamale	Tolon	0.3194	Tamale does granger cause Tolon
Tamale	Savelugu	0.01145	Tamale does not granger cause Savelugu
Yendi	Savelugu	0.04774	Yendi does not granger cause Savelugu
Yendi	Tolon	0.04172	Yendi does not granger cause Tolon
Yendi	Savelugu	0.4784	Tolon granger cause Savelugu

Source: Author's computation from the data, 2019

9. VAR model checking

Let YLW denotes a set of vectors of maize production in each selected district at various lags.

we have EVAR (3) and therefore $Y_{LW} = Y_{JL} = \begin{pmatrix} y_{jt-1} \\ y_{jt-2} \\ y_{jt-3} \end{pmatrix}$. For our research Y_{JL} is a 12 by 1 vector,

since each district have coefficients up to lag 3. Also, W_{LC} denote set of vectors of climate

conditions at various lags. For the purpose of our research, we can say that $W_{LC} = W_{jt} = \begin{pmatrix} w_{jt-1} \\ w_{jt-2} \\ w_{jt-3} \end{pmatrix}$ w_{jt} is an 18 by 1 vector. The model below shows the VAR model of maize production

in northern region without climate factors

$$\begin{pmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \\ y_{4t} \end{pmatrix} = \begin{pmatrix} 2844 \\ -713 \\ 1039 \\ 1279 \end{pmatrix} + \begin{pmatrix} 0.75^{***} & -0.099 & 0.156 & 0.121 \\ 0.158 & -0.633^{**} & 0.076 & 0.139 \\ -0.120 & 0.298 & 0.523^{*} & 0.457 \\ 0.337^{**} & 0.026 & -0.016 & 0.465\ ^{*} \end{pmatrix} y_{1t-1} +$$

$$\begin{pmatrix} -0.017 & -0.204 & -0.019 & 0.418 \\ -0.104 & -0.500\ ^{*} & -0.171 & 0.192 \\ -0.1300 & -0.33 & -0.09 & 0.242 \\ -0.194 & -0.169 & 0.188 & 0.022 \end{pmatrix} y_{2t-2} +$$

$$\begin{pmatrix} -0.097 & 0.111 & -0.164 & -0.310 \\ -0.202 & -0.079 & -0.08 & 0.057 \\ -0.056 & 0.126 & -0.267 & 0.156 \\ -0.107 & 0.184 & 0.146 & 0.024 \end{pmatrix} y_{3t-3} + \begin{pmatrix} 3940 \\ 3197 \\ 4937 \\ 2308 \end{pmatrix}$$

10. Interpretation of the Model

The model analyzed above is a VAR model that does not include climate variables. Specifically, focusing on the model equation for Tamale at lag 1, it exhibits an R-squared value of 59.13%, as derived from Table 7. This signifies that 59.13% of the overall changes in the structure can be accounted for by Tamale at that particular lag. Similarly, the R-squared value for Yendi is 9.221%, indicating that 9.221% of the total alterations in the structure can be explained by Yendi at that lag. For Tolon, the R-squared value is 57.72%, denoting the proportion of variations it can clarify within the system at that lag. Furthermore, at lag 1, Savelugu has an R-squared value of 66.2%.

Additionally, the significance levels at lag 1 are as follows: Tamale (significant at 1%), Yendi (significant at 5%), Tolon (significant at 10%), and Savelugu (significant at 10%).

Shifting to lag 2, the R-squared value for Tamale increases to 61.28%, representing the percentage of variations it can explain at that specific lag in the model. Yendi remains significant at 10% and has an R-squared value of 28.87%, signifying the percentage of variation it can account for at lag 2. In the same vein, Tolon has an R-squared value of 60.12%, while Savelugu's R-squared value rises to 72.83%, both at lag 2.

Moving to lag 3, Tamale is capable of explaining 66.35% of the variation in the model, while Yendi can explain 32.32% of the total variations at the same lag. Tolon can explain 61.39% of the variations within the model, and 77.93% of the variations observed in the model can be explained by Savelugu.

Furthermore, the mean absolute error (MAE) for Tamale is calculated as 3000.791, and the root mean square error (RMSE) is determined as 4180.076. As for Yendi, the MAE is found to be 2443.37, and the RMSE is 3213.12. Without climate variables, Tolon exhibits an MAE of 3264.72 and an RMSE of 4758.60. Similarly, Savelugu, without the climate variables, has an MAE of 1730.584 and an RMSE of 2173.363.

$$\begin{pmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \\ y_{4t} \end{pmatrix} = \begin{pmatrix} -245500 \\ 3823 \\ -105700 \\ -80120 \end{pmatrix} + \begin{pmatrix} 0.507^{*} & -0.033 & 0.073 & -0.324 \\ 0.152 & -0.613^{*} & -0.156 & 0.369 \\ -0.164 & 0.239 & -0.005 & 0.956 \\ 0.411^{**} & 0.099 & -0.175 & 0.129 \end{pmatrix} y_{1t-1} +$$

$$\begin{pmatrix} 0.357 & -0.012 & -0.342 & 0.312 \\ -0.395 & -0.534 & -0.147 & 0.361 \\ -0.518 & -0.789 & -0.005 & 0.572 \\ -0.168 & -0.246 & 0.276 & 0.243 \end{pmatrix} y_{2t-2} + \begin{pmatrix} -0.265 & 0.076 & -0.390 & 0.299 \\ -0.061 & -0.327 & -0.164 & 0.109 \\ 0.239 & -0.003 & 0.088 & -0.565 \\ -0.147 & -0.024 & -0.55 & 0.040 \end{pmatrix} y_{3t-3} +$$

$$\begin{pmatrix} 1.327 & -36.01 & 594.6 & 832.6 & 458.7 & 25410 \\ 1.775 & 10.480 & -0.002 & 538.4 & -0.001 & 12600 \\ -0.770 & 250.80 & 846.60 & 9324^{**} & 18.96 & -14180 \\ 3.724 & 72.190 & -106.20 & 2788^{*} & -486.8 & 20300 \end{pmatrix} w_{1t-1}$$

$$+ \begin{pmatrix} -0.024 & 187.6^{*} & 2588 & 5928 & 212.2 & 16370 \\ -3.203 & 75.11 & 1638 & 1256 & -946.9 & -2641 \\ 1.901 & -111.8 & 2746 & -2485 & -277.5 & 15890 \\ -3.334 & 131.2\_^{*} & 1144 & 4395^{*} & -693.7 & 7138 \end{pmatrix} w_{2t-2}$$

$$+ \begin{pmatrix} 1.728 & 184.9^{*} & -1018 & 9158^{**} & 843.5 & 42550 \\ 0.064 & -60.50 & -620.80 & 700.1 & 765.8 & 20410 \\ 1.018 & -167.84 & -84.40 & -3785 & 776.3 & 52610 \\ 1.104 & 14.35 & 1754 & 2212 & -210.6 & 7897 \end{pmatrix} w_{3t-3} + \begin{pmatrix} 3370 \\ 3472 \\ 4588 \\ 2147 \end{pmatrix}$$

11. Interpretation of the Model with climate conditions

The model presented above is a VAR model that incorporates climatic variables, as derived from Table 7 At lag 3, it exhibits an R-squared value of 88.42%. This implies that 88.42% of the total alterations in the structure can be explained by the model. Furthermore, the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) for this model are calculated as 3078.71 and 3826.59, respectively. Comparatively, this represents an improvement over the R-squared value of 66.35%, MAE of 3000.79, and RMSE of 4180.076 for the model without climatic variables at lag 3 (refer to Table 7).

Observing the RMSE of the model with climatic variables, we can see that it is smaller than the RMSE of the model without climatic variables. This justifies why the model with climatic variables is considered better, as a smaller RMSE indicates a better fit. The model for Tamale reveals that maize production in Tamale is dependent on its past production values as well as those of other districts at various time lags, ranging from lag 1 to lag 3. However, none of the coefficients in the model are significant at the 5% level, except for sunshine. This implies that the production of maize in other districts at different lags cannot be used to determine the current maize production in Tamale, even at lags up to 3.

Moving on to the equation for Yendi, it has an R-squared value of 62.42%, representing the proportion of the total alterations in the equation that can be accounted for by Yendi itself. The MAE and RMSE for Yendi are calculated as 2366.88 and 3168.55, respectively. Comparatively, this is an improvement over the R-squared value of 32.32% without climatic variables, which represents the amount of explained variations in its model. The RMSE for the model without climatic conditions was 3213.12, while the RMSE for Yendi with climatic conditions was 3168.55. This indicates that including climatic conditions could be a better approach for predicting future maize production in Yendi, as the RMSE with climatic conditions is smaller than that without climatic conditions. Furthermore, none of the maize production districts, including the climatic variables, are significant at the 5% level in the Yendi equation. This means that none of these districts, including the climatic variables, can be used to determine maize production in Yendi.

Examining the third equation, which represents maize production in Tolon, it has an R-squared value of 84.31%. The MAE and RMSE for Tolon are calculated as 3013.94 and 4458.81, respectively. None of the districts are significant at the 5% level, except for sunshine, which is significant at lag 1. This suggests that only sunshine can be used to predict maize production in Tolon, but only at lag 1 and not at lags 2 or 3. Consequently, we can conclude that the RMSE of Tolon with climatic conditions, which is 4458.81, provides a better fit as it is smaller than the RMSE without climatic conditions, which is 4758.60.

Lastly, the fourth equation represents maize production in Savelugu, with an R-squared value of 91.02%. The MAE and RMSE for Savelugu are calculated as 1753.39 and 2038

12. Summary and Findings

The primary objectives of the study were as follows:

1. To investigate the impact of climatic conditions on maize production in the districts of Tamale metropolitan, Savelugu/Nanton, Tolon/Kumbungu, and Yendi.
2. To examine the relationship that exists in the production of maize among the selected districts.
3. To develop an optimal model for forecasting maize production in the selected districts and towns, and subsequently forecast maize production for a five-year period.

As previously mentioned in the problem statement, climate variables play a significant role in influencing the production of all crops, including maize, in the northern region of Ghana. The region experiences relatively low rainfall, resulting in recurring issues of food insecurity and malnutrition. To address these challenges, the Vector Autoregressive (VAR) model was developed as a means to mitigate some of these problems. The model is categorized into two groups: one with climate variables and another without climate variables. While various studies have been conducted on forecasting maize production in the region, limited research has been done using the VAR model that includes climate variables. This thesis focuses specifically on studying maize production in the northern region of Ghana.

The time series data used in the study covers the period from 1968 to 2018. At the preliminary stage, the descriptive statistics of the data were analyzed to gain insights into its characteristics. To assess stationarity, the Autocorrelation Function (ACF) of the data was plotted, which provides an initial test for identifying any patterns over time. The stationarity of the data was further examined using the Augmented Dickey-Fuller (ADF) test. In cases where non-stationarity was detected, a first difference was applied to transform the data and achieve stationarity. Additionally, cross-correlation analysis was conducted to evaluate the extent of correlation or multicollinearity existing between the response variables at different time lags.

Based on the Akaike Information Criterion (AIC) value of 95.946108*, lag 3 was selected as the optimal lag for the model. Furthermore, the Portmanteau test for autocorrelation and the Jarque-Bera test for

Normality were conducted to assess the adequacy of the model. The results of the models revealed that incorporating climate conditions led to improved R-squared values, accounting for over 20% of the total variations in the maize production districts compared to models without climate conditions. Specifically, the R-squared values for the two categories by district are as follows: Tamale had an R-squared value of 88.42% with climate conditions compared to 66.35% without climate conditions, Yendi had an R-squared value of 62.62% with climate conditions compared to 32.32% without climate variables, Tolon had an R-squared value of 84.31% with climate conditions compared to 61.39% without climate conditions, and Savelugu had an R-squared value of 91.02% with climate conditions compared to 77.93% without climate conditions. Based on these findings, it can be concluded that the VAR model with climate conditions is the most effective in forecasting maize production in the various districts, as it led to a 20% improvement in the R-squared values. The vector equations for each selected district were derived and used to predict maize production in each district with a 95% confidence interval (CI).

13. Conclusion

Theoretical concept was presented in chapter 3. This has to do with basic concept of time series analysis theory involving univariate and multivariate time series models were explicitly explained. Data collected includes climate conditions and production of maize in the selected districts specifically from 1968 to 2018.Statistical technique applied for the analysis is the vector auto regression because the nature of the data is multivariate in nature. Multiple software

which included Excel, Minitab, GRETL and R-Studio were the software that used for the analysis of the data at various stages of the analysis. The results of the models showed that, when we include climate conditions it gives us better R squares of over 22% of the R squares of each maize production in the various districts without the climate conditions. For the purpose of these research, the VAR model was explained in details. Generally, in arriving at a model, the following steps were involved, thus; model identification, Estimation, lag length selection, model diagnostic, Normality test, structural analysis, granger causality and forecasting. There has been an amount of academic work in the form of thesis that has been done individually on maize production using some of the above-mentioned time series models separately and in many of these works there is a mentioned that other models that can be looked at on the subject of forecasting maize production in the northern region. In this work we have gone to look at the VAR model. We realize that when the climate variables were not added to the model, the multiple R-square (R2) (This is the extent of variance in the response variable that can be explained by the predictor variable) were small. After the climate variables were added, the model performs better since we had an improve multiple R square (R2) and smaller root mean square error (RMSE). In our analysis, we realize that climate variables play a highly significant role in the forecasting maize production in northern region. We also think that forty years (40 years) was more than enough for the study and the analysis can be considered as accurate.

14. Recommendation

Based on the findings of the study conducted using data from 1968 to 2018, the following recommendations are proposed:

1. Regular Review and Updated Data: It is recommended to periodically review and update the data used in the study. This will ensure that the analysis remains relevant and aligned with the current challenges faced by farmers in maize production. By incorporating up-to-date data, the study can provide more accurate insights and support decision-making processes.

2. Further Research on VAR Models in Agriculture: The literature review conducted in this study suggests that there is potential for further research on the application of VAR models in agriculture. Specifically, exploring the extent to which climate variables influence crop diseases in Ghana would be beneficial.

Investigating the relationship between climate conditions and disease prevalence can help in developing strategies to mitigate the impact and improve crop management practices.

3. Inclusion of Climate Variables in Maize Production Forecasting: The study demonstrates the importance of including climate variables in the forecasting models for maize production in the northern region. To achieve more accurate forecasts, it is recommended to incorporate climate variables alongside lag values of the response variables. This will enhance the predictive power of the models and enable farmers and policymakers to make informed decisions.

4. Managing Correlated Maize Production: The study reveals that maize production in the selected districts exhibits correlation at specific lag values. To ensure stable and sustainable production, it is crucial to implement strict measures that prevent a decline in production in correlated districts. This could involve sharing best practices, coordinating efforts, and establishing mechanisms for collaboration and support among districts to mitigate any adverse effects on production.

5. National-Level Analysis of Climate Variables: Further studies should be conducted to assess the effect of climate variables on maize production across different regions of the country. This analysis will provide insights into whether maize production in one region is directly influenced by production in other regions. Understanding these interdependencies can inform policies and interventions aimed at enhancing overall maize production in the country.

6. Consideration of VAR Model Improvement: Future research should focus on examining factors that contribute to the improvement of the VAR model with climate conditions compared to the VAR model without climate conditions. Understanding the specific influences and mechanisms behind this improvement can provide valuable insights into enhancing the forecasting capabilities and overall performance of the VAR model.

By implementing these recommendations, stakeholders in the agricultural sector can benefit from more accurate and reliable forecasts, leading to improved decision-making, enhanced productivity, and ultimately addressing the challenges faced in maize production in the northern region of Ghana.

15.Acknowledgments

We would like to express our sincere gratitude to the following individuals and institutions for their invaluable contributions to this research:

Ministry of food and Agriculture for the data and the faculty and staff at the Department of Statistics and Actuarial Science, College of Science, Kwame Nkrumah University of Science and Technology (KNUST), Kumasi, Ghana, for providing a conducive academic environment.

16.Conflict of Interest

The authors declare that there is no conflict of interest related to this research. The study was conducted with academic and research integrity, and the authors have no financial or personal relationships that could influence the objectivity of the research or the interpretation of its findings. The primary focus of this work is to contribute to the understanding of maize production dynamics in the Northern Region of Ghana, without any external biases.

REFERENCES

1. .MoFA. (2011). Agriculture in Ghana. Facts and Figures (2010) Statistical, Research and Information Division, Ministry of Food and Agriculture, Ghana. Retrieved January 31, 2017, from http://mofa.gov.gh/site/wpcontent/uploads/2011/10/AGRICULTURE-IN-GHANAFF-2010.pdf.
2. MoFA, *Agriculture In Ghana: Facts And Figures 2018*, Ministry of Food and Agriculture., Statistics Research and Information Directorate, Accra, Ghana, 2018.
3. J. O. Fening, N. Ewusi-Mensah, and E. Y. Safo, “Short-term effects of cattle manure compost and NPK application on maize grain yield and soil chemical and physical properties,” *Agricultural Science Research Journal*, vol. 3, pp. 69–83, 2011.
4. Bhasin VK, 2002. Agricultural productivity, efficiency, and soil fertility management practices of vegetable growers in the Upper East Region of Ghana. Revised research report submitted to the SADAO Foundation Available at http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.468.6999&rep=rep1&type=pdf (Accessed 14 February 2017).
5. Kibaara, B. W. (2005). Technical efficiency in Kenyan’s maize production; an application of the stochastic frontier approach. A Master Thesis submitted to Colorado State University, USA.
6. Dowswell, Christopher & Paliwal, R. & Cantrell, Ronald. (2019). Maize in the Third World. 10.1201/9780429042171.
7. BOLANOS,J.&EDMEADES, G. O. (1993 a). Eight cycles of selection for drought tolerance in lowland tropical maize.I. Responses in grain yield, biomass, and radiation utilization. Field Crops Research 31, 233–252
8. NeSmith, D. S. and Ritchie, J. T. (1992). Effects of soil water deficit during tessel emergence on development and yield component of maize (Zea mays L.). Field Crops Research 28: 251-256.
9. Menkir, A. and Akintunde, A. O. (2001). Evaluation of the performance of maize hybrids, improved open polinated and farmers’ local varieties under well-watered and drought stress conditions. Maydica 46: 227-238.
10. Desai, G.M., and V. Gandhi. 1988. Fertilizer consumption in sub-Saharan Africa: An analysis of growth and profile of use. In T.B. Tshibaka and C.A. Baanante (eds.), Fertilizer Policy in Tropical Africa. Lomé, Togo: International Fertilizer Development Center and International Food Policy Research Institute.
11. Gerner, H., and G. Harris. 1993. The use and supply of fertilizers in sub-Saharan Africa. In H. van Reuler and W.H. Prins (eds.), The Role of Plant Nutrients for Sustainable Food Crop Production in Sub-Saharan Africa. Leidschendam, The Netherlands: VKP (Dutch Association of Fertilizer Producers).
12. Kelly, V., Jayne, T., and Crawford, E. (2005). Farmers’ Demand for Fertilizer in Sub-Saharan Africa. East Lansing: MSU

13. Morris M., Mekuria, M. and Gerpacio, R. (2003). Impacts of CIMMYT Maize Breeding Research. In Crop Variety Improvement and its Effect on Productivity. eds. Evenson R. E. and D. Gollin. CABI, Wallingford, UK
14. FAO Statistical Databases. 2008. FAOSTAT: Agriculture Data. Available online: http://faostat.fao.org.
15. Morris, M.L., Tripp, R. and Dankyi, A.A. 1999. Adoption and Impacts of Improved Maize Production Technology. A Case Study of the Ghana Grains Development Project, Economics Program Paper 99-01. Mexico, D.F., CIMMYT. Available online http://www.cimmyt.org/Research/economics/map/research_results/program_papers/pdf/ EPP%2099_01.pdf.
16. Randon M. and Ashitey E. 2011. Poultry and Products Brief Annual: Ghana. Global Agricultural Information Network (GAIN) Report. 10/7/2011. Retrieved December 14, 2012, from http://gain.fas.usda.gov/Recent%20GAIN%20Publications /Poultry%20and%20Products%20Brief%20Annual_Accra _Ghana_10-7-2011.pdf.
17. IFPRI (2014). International Food Policy Research Institute (Ghana Strategy Support Program, GSSP). Ghana Agricultural News Digest – September 8, 2014. New GSSP policy note prepared by Arhin, B.G.
18. MoFA. (2000). Food and Agriculture Sector Development Policy (FASDEP I). 55pp.
19. Amanor-Boadu, V. (2012). Maize Price Trends in Ghana (2007-2011). USAID-METSS Report: Ghana Research and Issue Paper Series # 01-2012, April 2012.
20. Melinda, S., Derek, B. and Jayne, T. (2011). Maize Revolutions in Sub-Saharan Africa. http://papers.ssrn.com/sol3/papers.cfm?abstract _id=1839 149. Accessed on 12/2/12.
21. Ahmad I, Ahmad B, Boote K, Hoogenboom G. Adaptation strategies for maize production under climate change for semi-arid environments. European Journal of Agronomy. 2020 Apr 1;115:126040.
22. Jia Q, Yang L, An H, Dong S, Chang S, Zhang C, Liu Y, Hou F. Nitrogen fertilization and planting models regulate maize productivity, nitrate and root distributions in semi-arid regions. Soil and Tillage Research. 2020 Jun 1;200:104636.