Forecasting Futures Trading Volume for Cotton using Vector Auto Regression and ARIMA Model

Mohanapriya M[1]*, Patil Santosh Ganapati[1]
[1]Department of Physical Sciences & Information Technology,Agricultural Engineering College&Research Institute, TamilNadu Agricultural University, Coimbatore-641003

ABSTRACT:

This study involves comparing the forecasting performance of the ARIMA model and Vector Auto Regression model. Monthly data of open price, close price, mean cash price, open interests, spot price, and trading volume of cotton futures are used for analysis. Trading volume is taken as a dependent variable. Data was checked for stationarity using Augumented Dickey-Fuller test based on the *p*-value. Trading volume was stationary at level and all other variables were stationary at first difference. The optimal ARIMA model was selected based on AutoCorrelation Function plot, Partial Auto Correlation Function plot, and lower Akaike Information Criteria , Bayesian Information Criteria values. The optimal model was found to be ARIMA (2, 0, 0). The optimal lag for VAR model was selected based on AIC and BIC. The forecast is obtained based on the coefficients of the model. The forecasted values were compared using RMSE, MAPE, MAE and Theil U Statistic. Based on the results, ARIMA is better at forecasting the futures trading volume of cotton than VAR.

Keywords: *Futures market; Volume; ARIMA;VAR;Cotton ; ADF test.*

INTRODUCTION

Forecasting is the process of making predictions of the future based on the present and past data and most commonly by analysis of trends. The forecasting technique is more accurate when the environment is not rapidly changing. Futures are derivative financial contracts that obligate the parties to transact an asset at a predetermined future date and price. The buyer must purchase or the seller must sell the underlying asset at the set price, regardless of the current market price at the expiration date.

A futures contract allows an investor to speculate on the direction of a security, commodity, or financial instrument. Futures are used to hedge the price movement of the underlying asset to help prevent losses from unfavorable changes.

Predicting the trading volume at particular times of day and using particular information, it is possible to explain the use of order execution strategies and the role of algorithmic trading, the setting of trading hours, and the growing role of global information flows in driving trade and price. Large hedgers and speculators seeking to minimize transactions costs may improve order execution through better volume forecasts.

Forecasting of futures prices of sugar was done by using ARIMA model (Manas Miraji., 2009).A neural network model is used to forecast monthly futures trading volume for barley, canola, flax, oats, rye and wheat (Kaastra., 1994). For forecasting futures trading market variables ARIMA model were used. (Kutsev, S.V, Smirnov, A.E, Sorocoumov.V.E., 1996). Indian economy is agro based, as agriculture constitutes the backbone of the rural livelihood. Cotton is one of the important fibers and cash crops of India and plays a dominant role in the industrial and agricultural economy of the country. It provides the basic raw material to the cotton textile industry. As India is the largest producer of cotton, forecasting futures trading volume of cotton is important.

MATERIAL AND METHODS

DATA:

The study was undertaken using secondary data. Futures trading volume of cotton was forecast using monthly data of open price, close price, spot price, open interest and mean cash price collected from National Commodity Derivatives Exchange limited (NCDEX) and Multi Commodity Exchange (MCX). Data was collected between October 2011 to December 2020.

*Corresponding author's e-mail: mohanapriyamuthukrishnan@gmail.com

ARIMA:

ARIMA is an acronym that stands for Auto-Regressive Integrated Moving Average. An ARIMA model is a class of statistical models for analyzing and forecasting time series data. It is a generalization of the simpler auto-regressive moving average and adds the notation of integration.

Assumptions:

➢ Data should be stationary – Stationary means that the properties of the series doesn't depend on the time when it is captured. A white noise series and series with cyclic behavior can also be considered as stationary series.

➢ Data should be univariate – ARIMA works on a single variable. Auto-regression is all about regression with the past values.

Steps:

i. Exploratory analysis: Autocorrelation analysis to examine serial dependence, used to estimate which value in the past correlates with the current value, which provides the p, d, q estimate for ARIMA models. Spectral analysis to examine cyclic behavior. Trend estimation and decomposition.

ii. Fit the model: To determine the order of the model to be fitted to the data, we need three variables: p, d, and q, which are non-negative integers that refer to the order of the autoregressive, integrated, and moving average parts of the model respectively. PACF and ACF are used to find out the values of p and q respectively. Maximum Likelihood Estimation (MLE) to estimate the ARIMA model. Finding Akaike's Information Criterion (AIC) and Schwarz Bayesian Information Criterion (BIC) for a set of models and investigating the models with the lowest AIC and BIC values. Along with AIC and BIC, we also need to watch those coefficient values closely and we should decide whether to include that component or not according to their significance level.

iii. Diagnostic measures

iv. Forecasting using an ARIMA model: The parameters of that ARIMA model can be used as a predictive model for making forecasts for future values of the time series once the best-suited model is selected for time series data.

VAR (Vector Auto Regression):

Vector Auto Regression (VAR) is a multivariate time series forecasting algorithm that is used when two or more time-series influence each other. It is bidirectional, i.e., the variables influence each other. All the variables in a VAR are treated symmetrically by including for each variable an equation explaining its evolution based on its lags and the lags of all the other variables in the model. VAR method of modeling offers several advantages over traditional alternatives based on structural equation system

$$y_{1,t} = c_1 + \Phi_{11,1} y_{1,t-1} + \Phi_{12,1} y_{2,t-1} + e_{1,t}$$

$$y_{2,t} = c_2 + \Phi_{21,1} y_{1,t-1} + \Phi_{22,1} y_{2,t-1} + e_{2,t}$$

Where $e_{1,t}$ and $e_{2,t}$ are white noise processes that may be contemporaneously correlated. The coefficient $\Phi_{ii,\ell}$ captures the influence of the ℓ_{th} lag of variable y_i on itself, while the coefficient $\Phi_{ij,\ell}$ captures the influence of the ℓ_{th} lag of variable y_j on y_i.

First the data are checked for stationarity using Augmented Dickey-Fuller test at 5% level of significance. If the series is non- stationery, we take the first difference of the data to make them stationary. Then fit a VAR model. In both case, models were estimated equation by equation using the principle of least squares. Coefficientwas estimated and the forecast was obtained.

RESULTS AND DISCUSSION

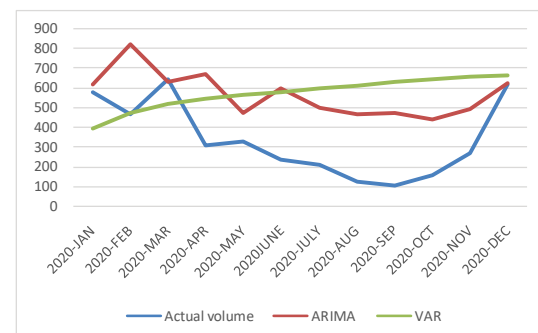For forecasting futures trading volume, the methodologies ARIMA and VAR were used.

Figure 1. Actual vs forecast

ARIMA:

ARIMA is used to understand the past values or predict future data in a series. This model uses univariate past values of trading volume to predict futures trading volume. For ARIMA, the data should be stationary. ARIMA uses stationary data to make predictions about the future. To check stationarity ADF test is used. Based on the ADF test, the data seems to be stationary as *p*-value is 0.001326 taken from table 1. So the data is already stationary at
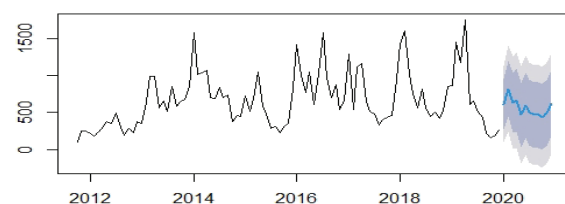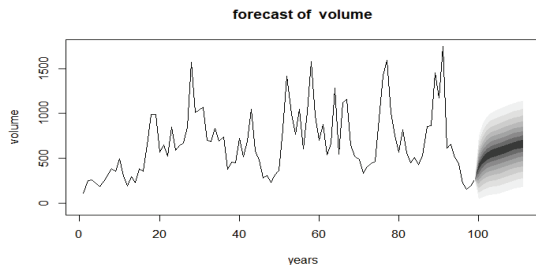

Figure 2. ARIMA forecast

level, then d=0. *p* is the order of the "Autoregressive" to (AR) term. It refers to the number of lags of y to be used as predictors. The required number of AR terms can be found out using the Partial Auto-Correlation Function (PACF) plot. The order of AR term will be equal to many lags that crosses the significance limit in the PACF plot.

Figure 3. VAR forecast

MA terms are technically, the error of the lagged forecast. Auto Correlation Function tells how many MA terms are required to remove an auto correlation in the stationarized series. Based on the ACF and PACF plot, AIC, BIC, the model is identified as ARIMA (2, 0, 0).

Table 1. ADF test results

Variables	p-value at level	p-value at first difference
Volume	0.0132*	0.01*
Open price	0.4732	0.01*
Close price	0.4942	0.01*
Mean cash price	0.0774	0.01*
Open interests	0.4709	0.01*
Spot price	0.4708	0.01*

The parameters of the best model were estimated. Model is written with its significant parameters as using table 2,

$$y_t = 643.9327 + 0.591091y_{t-1} + 0.073361y_{t-2}$$

Table 2. ARIMA model

Aspects	Estimates	Standard error	Z-value	p-value
Intercept	643.93267	83.114466	7.7475	0.01*
AR1	0.591091	0.100034	5.9089	0.4695
AR2	0.073361	0.101416	0.7234	0.01*

The forecast is obtained by back forecasting are presented in Table 3.

VAR:

Vector Auto Regression is a statistical model that captures the relationship between multiple quantities as they change over time.So it is used to forecast multivariate time series. The volume is taken as the dependent variable and open price, close price, mean cash price, open interests, and spot price are taken as the independent variable.

Table 3. Forecasted values

MONTH	Actual volume	ARIMA	VAR
2020-JAN	580.1522	613.3972	396.2459
2020-FEB	468.9357	820.8339	471.4638
2020-MAR	645.3306	633.0308	517.8416
2020-APR	304.875	666.7686	544.9387
2020-MAY	330.6842	470.1154	562.8362
2020JUNE	234.0476	594.7191	577.6469
2020-JULY	205.6788	500.9168	593.3506
2020-AUG	119.8847	467.2815	609.8445
2020-SEP	102.9286	468.9645	626.0673
2020-OCT	155.7168	438.7968	640.5651
2020-NOV	268.2287	489.434	652.3299
2020-DEC	618.0682	623.0171	661.043

For building a model in Vector Auto Regression, the data should be stationary. Data were checked for stationarity using Augmented Dickey-Fuller test at 5% level of significance.

Table 4. Model performance metrics

METRICS	ARIMA	VAR
RMSE	256.7533	332.6847
MAPE	86.8534	118.938
MAE	139.6793	162.5002
THEIL U STATISTIC	0.702617	0.864899

Volume was stationary at level and other dependent variables (open price, close price, mean cash price, Open Interests and spot price) were stationary at the first difference was based on p-value from table 1 . Optimal lag was selected based on Akaike Information Criteria (AIC).The optimal lag for the model is 2.

Table 5. ARIMA forecast

Month	Volume
2021-JAN	679.5
2021-FEB	975.5
2021-MAR	915.5
2021-APR	798.5
2021-MAY	526.5
2021-JUN	575.1
2021-JUL	450.2
2021-AUG	440.1
2021-SEP	393.7
2021-OCT	349.2
2021-NOV	407.9
2021-DEC	497.1

Coefficients of the model were estimated and the forecast was obtained, which are presented in Table 3. Forecasted futures trading volumes are presented in Table 5.

CONCLUSION

Forecasted values were compared using model performance metrics such as Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE) and Thiel's U statistic. Metrics values were compared for ARIMA and VAR using table 4 and it show that ARIMA was better than VAR as the metrics of ARIMA are lower than VAR.Therefore, the ARIMA is better in forecasting the futures trading volume of cotton than VAR.

REFERENCES

Assis, K. A., Amran and Y. Remali. 2010. Forecasting cocoa bean prices using univariate time series models. *RW-IRSSJ.,* **1(1):**71.

Bhardwaj, S. P., *et al.*, 2014. An empirical investigation of ARIMA and GARCH models in agricultural price forecasting. *EA* . ,**59(3)**: 415-428.

Carlton, Dennis W., 1984 Futures markets. Their purpose, their history, their growth, their successes and failures. *J Futures mark.,***4.3**: 237.

Fousekis, Panos, and Dimitra Tzaferi., 2019 Price returns and trading volume changes in agricultural futures markets: An empirical analysis with quantile regressions *JEA,* **19**:116.

Kaastra, Iebeling, and Milton S. Boyd., 1995. Forecasting futures trading volume using neural networks. *The J Futures mark (1986-1998)* **15.18**: 953

KumarMahto, Anil, Ranjit Biswas, and M. Afshar Alam., 2019 Short term forecasting of agriculture commodity price by using ARIMA: based on Indian market. *International Conference on Advances in Computing and Data Sciences*. Springer, Singapore.

Lo, Andrew W., and Jiang Wang., 2000 Trading volume: definitions, data analysis, and implications of portfolio theory." *RFS,* **13.2**:257-300.

Lo, Andrew W., and Jiang Wang , Stock., 2010 market trading volume. *Handbook of financial econometrics: Applications*. Elsevier, 241-342.

Martell, Terrence F., and Avner S. Wolf., 1987 Determinants of trading volume in futures markets. *J Futures mark (1986-1998)* **7.3**:233.

Ouyang, Hongbing, Xiaolu Wei, and Qiufeng Wu. , 2019 Agricultural commodity futures prices prediction via long-and short-term time series network. *Journal J.Appl.Econ.,* **22.1**:468-483.

Ragunathan, Vanitha, and Albert Peker., 1997 Price variability, trading volume and market depth: evidence from the Australian futures market. *Appl. Financial.Econ.,* **7.5**, pp. 447-454.

Sahoo, Pravakar, and Rajiv Kumar.,2009 Efficiency and futures trading-price nexus in Indian commodity futures markets .*Global Business Review* **10.2**:187-201

Shivakumar, N. M., and G. Kotreshwar., 2017 Modelling maize futures price-An empirical analysis based on ARIMA approach. *Asian J.Manag.,* **8.3**: 734-740.

Wang, George HK, Jot Yau, and Tony Baptiste., 1997 .Trading volume and transaction costs in futures markets. *J futures mark, (1986-1998)* **17.7**: 757.