

Comparison of Stationarity on Ljung Box Test Statistics for Forecasting

Jayeola Dare^{1,*}, Aye O. Patrick¹ and David O. Oyewola²

¹Department of Mathematical Sciences, Adekunle Ajasin University, Akungba Akoko, Ondo State, Nigeria

e-mail: dare.jayeola@aau.edu.ng; darchid2002@yahoo.com; ayepatricks@gmail.com

²Department of Mathematics & Computer Science, Federal University, Kashere P.M.B 0182, Gombe, Nigeria

e-mail: davidakaprof01@yahoo.com; davidoyewole@fukashere.edu.ng

Abstract

The movements in Asset prices are very complex, therefore seem to be unpredictable. However, one of the main challenges of the econometric models is to get the best data for forecasting in order to present accurate results. This paper investigates the performance of stationary and non-stationary data on Ljung Box test statistics, to check the fitness of the data for forecasting. In the paper three assets (Groundnut, sorghum and soya bean) are used, tests are conducted for Ljung box statistics; Correlogram, Histogram Normality and Heteroscedasticity test. It is observed that stationary data are better for forecasting than non-stationary data in this research.

1. Introduction

Forecasting is fundamental to the risk management process in order to price assets derivatives, hedging strategies and estimating the financial risk of a firm in portfolio. In recent years, Autoregressive Conditional Heteroscedasticity (ARCH) type models have become popular as a means of capturing observed characteristics of financial returns like thick tails and volatility clustering. These models use time series data on returns to model conditional variance. An alternative way to estimate future volatility is to use options prices, which reflect the market expectations of volatility. Analytical option pricing models can be used to back out implied volatility over the remaining life of the option

Received: December 10, 2021; Accepted: January 24, 2022

2020 Mathematics Subject Classification: 62-XX.

Keywords and phrases: ARCH, asset, forecast, stationary.

*Corresponding author

Copyright © 2022 the Authors

given the observed market price. In the construction of volatility forecasts, energy market participants would like to know which model produces the most accurate forecasts, as well as, whether the complex time series models enhance any significant volatility information beyond what contained in option prices.

[4] compared the relative information content and predictive power of implied volatility and ARCH family forecasts for asset futures. A similar study by [7] studied analysis for financial management and efficiency of options market in predicting volatility. [3] examined the prediction of financial volatilities for crude oil, gold and natural gas markets. Finance is focused on intertemporal decision making under uncertainty and so forecasts of unknown future outcomes is integral to several areas of finance asset pricing requires forecasts of future cash. Risk management relies on forecasts of variances and covariance of returns on portfolios that frequently comprise large numbers of assets. Countless studies in corporate finance analyse firms capital budgeting decisions which in turn depend on projected cash flows and firms forecasts of the costs and benefits of issuing debt and equity. A large literature in banking analyses the possibility of runs which reflects investors' forecasts of both a bank solvency and liquidity as well as their expectation of other agents (depositors) decisions on whether to run or stay put. While economic and financial forecasting share many methods and perspectives, some important features help differentiate the two areas.

Competitive pressures and market efficiency mean that the signal-to-noise ratio in many financial forecasting problems particularly predictability of asset returns is very low compared to standard forecasting problems in macroeconomics in which the presence of a sizeable persistent component makes forecasting easier.

The presence of weak predictors with low predictive power and the resulting importance of parameter estimation error is therefore the norm rather than the exception in financial forecasting. The possibility of readily trading on price forecasts makes the scope for feedback effective from forecasts to actual outcomes stronger in finance than in other areas of economics. Model instability is therefore particularly important to financial forecasting; overstating and issues related to data mining have increasingly become a concern in financial forecasting due to the ease with which numerous forecasting models can be fitted to a given data set and the difficulty of generating new and genuinely independent data sets on which to test the forecasting performance. In particular, how should the performance of a forecasting model be evaluated when this model is selected as the best performer among a larger set of competing specifications? This situation

generates a multiple hypothesis testing problem that, if not accounted for, can lead to findings of spurious predictability patterns and serious distortions in inference while volatility forecasting also features prominently in forecasting of macroeconomic variables. The risk management is concerned with forecasting the correlations between very large sets of variables and so gives rise to high dimensional forecasting problems. Moreover, access to high frequency data, sampled every few seconds during trading sessions for the most liquid assets, means that measures of realized variances can be constructed and used to forecast future risks. This type of data does not, as yet, have obvious counterparts in economics where measurements tend to be conducted at a lower frequency.

Data points are often nonstationary or have means, variances, and covariances that change over time. Nonstationary behaviors can be trends, cycles, random walks, or combinations of the three. Nonstationary data, as a rule, are unpredictable and cannot be modelled or forecasted. The results obtained by using nonstationary time series may be spurious in that they may indicate a relationship between two variables where one does not exist. In order to receive consistent, reliable results, the nonstationary data needs to be transformed into stationary data. In contrast to the nonstationary process that has a variable variance and a mean that does not remain near, or returns to a long-run mean over time, the stationary process reverts around a constant long-term mean and has a constant variance independent of time. The remaining parts of this paper is organized as follow: section two reviews literature, section three describes the data section four explains the methodology, results and discussions are considered in section five while section six concludes the paper.

2. Literature Review

Economists have long thought that forecasts are potentially useful as decision aids, and have devoted considerable efforts to develop and assess forecasting methods [1]. Forecasts can provide decision makers with technical and market support to help execute policies. In stock markets, forecasts are typically made for the prices of assets commodity outputs. Less work has been done on forecasting the primary inputs needed to produce the commodities. However, most important variable inputs like crude oil and gold, since crude oil and gold are used as measurement in stock market [8]. With the recent price volatility in the fuel market, making wrong decisions in fuel purchasing can have a significant impact on the bottom line for farming firms or fuel providers. While the ability to anticipate short term fuel prices may be useful, very little work has been done to

evaluate the ability to forecast asset price. The dearth of research on this topic requires examining the energy forecasting literature in order to design an approach to forecast asset prices.

[15] in his work titled, Bias Correction effect of the AIC for selecting variables in Normal Multivariate Linear Regression models. He considered two criteria; The Akaike Information Criterion, AIC and Takeuchi Information Criterion TIC. In his paper he also compared the performance of AIC and Akaike Information Criterion Component, AICC. Both criteria may be viewed as estimators of the expected Kullback Leibler information. The bias of AIC and AICC are studied in the under fitting case, where none of the candidate models includes the true model [12]. Both normal linear regression and autoregressive candidate models are considered. The bias of AICC is typically smaller, often dramatically smaller, than that of AIC. A simulation study in which the true model is an infinite order auto regression shows that, even in moderate sample sizes, AICC provides substantially better model selections than AIC. [11] in their work titled; Forecasting of crude oil price, said as most important strategic resource around the world crude oil is the key commodity for the world economy. Therefore forecasting crude oil price has always been considered as a very challenging task which drew the interest of researchers. The price of oil is essentially determined by its supply and demand [3].

There appeared to be benefits from disaggregation and for searching for new causal variables. Greater volatility of oil prices could rekindle the integration of petroleum operations designed to save on transaction costs incurred in reducing uncertainty. More sophisticated, and hence more expensive, market instruments might be required to hedge risks. More complex contracts may need to be written among market participants compared with other commodities. Greater price volatility could account for the way the petroleum industry is in the vanguard of developing and applying modern asset valuation techniques [2]. [10] work titled: Applied Analysis Concept Development on Financial Aspect. The role of asset price, which explained their survey on forecasting output and inflation. Although they mention some historical precedents, their review focused on developments within the past fifteen years. The section concludes with an attempt to draw some general conclusions from their literature.

[14] in his work titled Forecasting method in finance; he reviewed and highlighted some of the key challenges in financial forecasting problems along with opportunities arising from the unique features of financial data. He analysed the difficulty of establishing predictability in an environment with a low signal-to-noise ratio, persistent

predictors, and instability in predictive relations arising from competitive pressures and investors learning. He discussed approaches for forecasting the mean, variance, and probability distribution of asset returns. Finally, He discovered how to evaluate financial forecasts while accounting for the possibility that numerous forecasting models may have been considered, leading to concerns of data mining. However, development on this study is still ongoing.

The presence of non-stationarity and its treatment complicate the measurement and use of vector autoregressive (VAR) models. Near the unit circle, conventional estimation procedures can underestimate the parameter space. Differencing, a standard approach for reducing non-stationarity, can distort multivariate interactions and cause forecasts to diverge appreciably from actual values [5]. Methods for forecasting with multivariate autoregressive models in the presence of non-stationarity are in their infancy [13]. Several approaches have appeared that are applicable to the non-stationarity problem, including estimation in differences, use of Bayesian VARs that shrink the parameter space to the first-differenced framework, and use of error correction models. No single empirical approach for treating the non-stationarity problem has been clearly articulated. The literature comparing the forecasting effectiveness of these approaches in the presence of non-stationarity is limited. However, this paper wishes to compare the fitness of both stationarity and non-stationarity for forecasting.

3. Data

The sample data for this research was collected from the monthly data of groundnut, sorghum and soya bean from yahoo finance DataStream. The data however spans from 2000 to 2010, the data used for the analysis of this study are stationary and non-stationary data.

4. Research Methodology

The approach used for this research is based on quantitative approach and this is because it involves the collection and analysis of numerical data. We shall use the approach of Q-statistic:

Q-Statistic Test

Q-Statistic is a test statistics output by either the Box-Pearce test or in a modified version which provides better sample properties by the Ljung Box test. The Ljung-Box test is widely applied in econometrics and other applications of time series analysis. The Q-statistic formula is given as follows:

$$Q = n(n + 2) \sum_{i=1}^h \frac{\rho_i^2}{n - 1} \quad (4.1)$$

where n is the sample size, ρ_i is the sample autocorrelation at lag i and h is the number of lags being tested.

Jarque Bera Statistics of Normality Test

The Jarque Bera (JB) test for normality (also known in statistics as the D'Agostino-Pearson or Bowman-Shenton test) is one of the most popular goodness of fit tests. The method that is often used in econometrics, that has been suggested and used for testing whether the distribution underlying a sample is normal, is the Bowman and Shenton

$$JB = n \left[\frac{(\text{skewness})^2}{6} + \frac{(\text{kurtosis} - 3)^2}{24} \right] \quad (4.2)$$

where n is the sample size. The Equation 4.2 was subsequently derived by Bera and Jarque as the Lagrangian Multiplier (LM) test against the Pearson family distributions. The JB statistics has an asymptotic chi-square distribution with two degree of freedom.

Measures of Skewness and Kurtosis

A fundamental task in many statistical analyses is to characterize the location and variability of a data set [9]. For univariate data Y_1, Y_2, \dots, Y_n the formula for skewness is as follows:

$$S_1 = \frac{\sum_{i=1}^n \frac{(Y_i - \hat{Y})^3}{n}}{S^3} \quad (4.3)$$

where \hat{Y} is the mean, S is the standard deviation and n is the sample size. The Equation 4.3 for skewness is referred to as the Fisher Pearson coefficient of skewness. The adjusted Fisher Pearson coefficient of skewness is:

$$S_2 = \frac{\sqrt{n(n-1)}}{n-2} \frac{\sum_{i=1}^n \frac{(Y_i - \hat{Y})^3}{n}}{S^3} \quad (4.4)$$

Equation 4.4 is an adjustment for sample size. The adjustment approaches 1 as n gets large.

Kurtosis is a measure of whether the data are heavy-tailed or light-tailed relative to a normal distribution. That is, data sets with high kurtosis tend to have heavy tails or outlier [9]. Data sets with low kurtosis tend to have light tails or lack of outliers. The histogram

is an effective graphical technique for showing both the skewness and kurtosis of data set. For univariate data Y_1 the formula is as follows:

$$Kurtosis = \frac{\sum_{i=1}^n (Y_i - \hat{Y})^4}{S^4} \tag{4.5}$$

where \hat{Y} is the mean, S is the standard deviation and n is the number of data points. The kurtosis for a standard normal distribution is three.

The Breusch-Pagan test

A more formal, mathematical way of detecting heteroscedasticity is what is known as the Breusch-Pagan test. It involves using a variance function and using a χ^2 -test to test the null hypothesis that heteroscedasticity is not present against the alternative hypothesis that heteroscedasticity is present.

5. Research Findings and Discussions

The data used for this research are groundnut, sorghum and soya bean. The data used for the analysis of this paper are both stationary and non-stationary for comparison of their results. Thus, the Ljung box tests are conducted on the data.

Correlogram Testing

This test uses all the samples of residual Autocorrelation function (ACF), Partial Autocorrelation function (PACF) test. Table 5.1 shows the results of the assets for both stationary and non-stationary. The following hypothesis is used for the correlogram test:

- H_0 : There is no correlogram in the time series data
- H_1 : There is correlogram in the time series data

Table 5.1: Correlogram showing P-values for Stationarity and Non-stationarity of the Assets

Stationarity			Non-stationarity		
Groundnut	Sorghum	Soya bean	Groundnut	Sorghum	Soya bean
0.109	0.104	0.090	0.000	0.000	0.000
0.067	0.245	0.196	0.000	0.000	0.000
0.030	0.412	0.231	0.000	0.000	0.000
0.005	0.579	0.358	0.000	0.000	0.000
0.007	0.634	0.364	0.000	0.000	0.000

This test is to verify if there is correlogram in the time series data. Correlogram (also known as autocorrelation function, ACF plot) is a visual way to show serial correlation in data that change over time. It means error in time series data. The guideline is if p-value is less than 5% we reject null hypothesis otherwise we cannot reject null hypothesis. Looking at Table 5.1, p-values of groundnut gives 0.109, 0.067, 0.030, 0.005 and 0.007, the first two values are greater than 5% but the last three values are less than 5%. However, all the p-values of the two other assets are greater than 5% hence, we cannot reject null hypothesis. But for Non-stationarity, all the p-values for the three assets are 0.0000 which is less than 5%. It means, in this case we have to reject null hypothesis. There is correlogram in the non-stationary data.

Histogram and Normality test

Figures 5.1 to 5.6 and Table 4.6 present histograms and normality test of the three assets. Figures 5.1 to 5.3 represent histograms of stationary data while Figures 5.4 to 5.6 represent histograms of non-stationary data. Therefore, if the residuals are normally distributed, the histogram is bell-shaped and the Jarque-Bera statistic would be insignificant (greater than 5%) [6]. The normality tests conducted are presented in Figures 5.1 to 5.6.

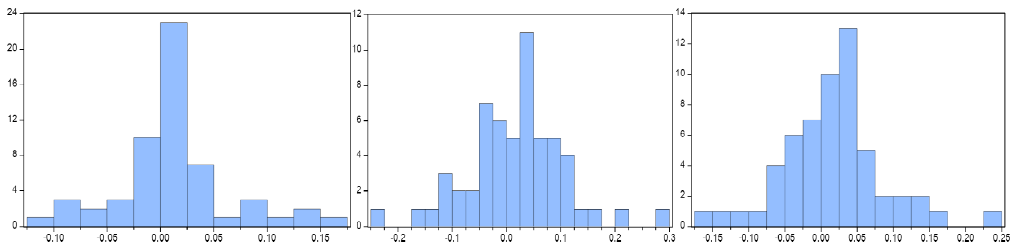


Figure 5.1: Histogram of Groundnut

Figure 5.2: Histogram of Sorghum

Figure 5.3: Histogram of Soya bean

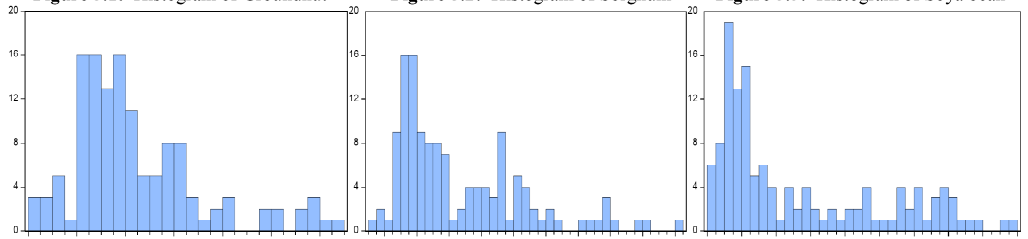


Figure 5.4: Histogram of Groundnut

Figure 5.5: Histogram of Sorghum

Figure 5.6: Histogram of Soya bean

Figures 5.1 to 5.6 present histograms of the three assets, looking at Figures 5.1 to 5.3 are bell shaped hence, the histograms are normally distributed. On the other hand, Figures

5.4 to 5.6 are not bell shaped but rather right tailed, therefore not normally distributed. This showed that the histograms of stationary data are normally distributed.

Table 5.2: Normality Test.

Normality Test	Stationarity			Non-stationarity		
	Groundnut	Sorghum	Soya bean	Groundnut	Sorghum	Soya bean
Skewness	0.5914	0.0351	0.0853	1.2418	1.2034	0.9060
Kurtosis	4.0453	4.2571	4.1430	4.2705	3.9401	2.4407
Jargue-Bera	5.9179	3.7648	3.1721	42.1529	36.1618	19.4787
Probability	0.0519	0.1522	0.2047	0.0000	0.0000	0.0000

This test is to confirm the normality of both stationarity and non-stationarity of the three assets. The guideline for asset to be normally distributed; skewness, kurtosis, Jargue-Bera must be greater than 5% and probability must equally greater than 5%. Looking at Table 5.2, Groundnut has skewness of 0.5914, kurtosis of 4.0453, Jargue bera of 5.9179 and probability 0.0519. Sorghum obtains skewness 0.0351, kurtosis 4.2571, Jargue bera 3.7648 and probability 0.1522. Lastly, Soya bean has skewness 0.0853, kurtosis 4.1430, Jargue bera 3.1721 and probability 0.2047. Relating all these values with the guideline given, they are very close; hence this confirmed that the stationary data are normally distributed. Nevertheless, the non- stationary data gave spurious results, the probabilities gave 0.0000 which is less than 5%, and this showed that non-stationary data are not normally distributed.

Heteroscedasticity Test

Another test is ARCH test or heteroscedasticity test used to check the significance of ARCH effect. The guideline is, if p-value is less than 5% reject null hypothesis otherwise we cannot reject null hypothesis. Table 5.3 presents the p-values for groundnut, sorghum and soya bean, for both stationary and non-stationary data. The following hypothesis is used:

H_0 : There is no ARCH effect in the time series data.

H_1 : There is ARCH effect in the time series data.

Table 5.3: ARCH Effect Test.

Assets	Stationarity	Non-stationarity
	Probability value	Probability value
Groundnut	0.1189	0.0003
Sorghum	0.0988	0.0004
Soya bean	0.2170	0.0038

Table 5.3 gives the probability values of stationary data; groundnut to be 0.1189; sorghum 0.0988 and soya bean 0.2170. Also, probability values of non-stationary data; 0.0003, 0.0004 and 0.0038 respectively. The probability values of stationary data are all greater than 5% while probability values of non-stationary data are all less than 5%. This implies that the stationary data have no ARCH effect but non-stationary data have ARCH effect.

6. Conclusion

This paper investigates the performance of stationary and non-stationary data on Ljung Box test statistics, to check the fitness of the data for forecasting. Three tests of Ljung box statistics were conducted which include; Correlogram, Histogram and Normality and Heteroscedasticity test. Moreover, all these tests conducted, the results of both stationarity and non-stationarity were compared. The result of Correlogram test showed that the p-values of stationary data were greater than 5% but the p-values of non-stationary data were less than 5% which showed that stationary data have no Correlogram but non-stationary data have. Also, the histogram and normality test conducted implied that stationary data are normally distributed but non-stationary data are right tailed. The last test conducted was heteroscedasticity, to verify the presence of ARCH effect in the data. The result showed that, stationary data have no ARCH effect while non-stationary data have.

In conclusion, the analysis of this study, to test the fitness of both stationary and non-stationary data showed vividly that stationary data is better for forecasting for better accuracy.

References

- [1] E. Allaj, The Black-Litterman model: a consistent estimation of the parameter tau, *Financial Markets and Portfolio Management* 27(2) (2013), 217-251. <https://doi.org/10.1007/s11408-013-0205-x>
- [2] D. M. Al-Najjar, Modelling and estimation of volatility using ARCH/GARCH models in Jordan's stock market, *Asian Journal of Finance & Accounting* 8(1) (2016), 152-167. <https://doi.org/10.5296/ajfa.v8i1.9129>
- [3] S. Beidas-Strom and A. Pescatori, Oil price volatility and the role of speculation, *IMF Working Paper No. 14/218*, 2014. <https://doi.org/10.2139/ssrn.2561232>
- [4] M. H. Davis and S. Lleo, A simple procedure for combining expert opinion with statistical estimates to achieve superior portfolio performance, *The Journal of Portfolio Management* 42(4) (2016), 49-58. <https://doi.org/10.3905/jpm.2016.42.4.049>
- [5] R. F. Engle and B. S. Yoo, Forecasting and testing in co-integrated systems, *J. Econometrics* 35(1) (2018), 143-159. [https://doi.org/10.1016/0304-4076\(87\)90085-6](https://doi.org/10.1016/0304-4076(87)90085-6)
- [6] A. Ghasemi and S. Zahediasl, Normality tests for statistical analysis: a guide for non-statisticians, *International Journal of Endocrinology and Metabolism* 10(2) (2012), 486-489. <https://doi.org/10.5812/ijem.3505>
- [7] R. C. Haggins, J. Koski and T. Mitton, *Analysis for Financial Management*, New York: McGraw Hill Education, 2016.
- [8] H. Ince and T. B. Trafalis, A hybrid forecasting model for stock market prediction, *Economic Computation and Economic Cybernetics Studies and Research* 51 (2017), 263-280.
- [9] H. Y. Kim, Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis, *Restor. Dent. Endod.* 38(1) (2013), 52-54. <https://doi.org/10.5395/rde.2013.38.1.52>
- [10] S. Krylov, Applied analysis concept development: financial aspect, *European Journal of Accounting Auditing Research* 3(5) (2015), 30-54.
- [11] H. Manel and A. Chaker, Forecasting crude oil price using artificial neural network, *A Literature Survey Economics Bulletin* 35(2) (2015), 1339-1359.
- [12] M. Schulmerich, Y. M. Laporcher and C. H. Eu, *Applied Asset and Risk Management*, Springer, 2015. <https://doi.org/10.1007/978-3-642-55444-5>
- [13] J. H. Stock and M. W. Watson, Variable trends in economic time series, *J. Econ. Perspect.* 2(1) (2015), 147-174. <https://doi.org/10.1257/jep.2.3.147>

- [14] A. Timmermann, Forecasting method in finance, *Annual Review of Financial Economics* 10 (2018), 449-479. <http://dx.doi.org/10.1146/annurev-financial-110217-022713>
- [15] H. Yanagihara, K. Kamo, S. Imori and M. Yamamura, A study on the bias-correction effect of the AIC for selecting variables in normal multivariate linear regression models under models misspecification, *REVSTAT - Statistical J.* 15(3) (2017), 299-332.

This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted, use, distribution and reproduction in any medium, or format for any purpose, even commercially provided the work is properly cited.
