COINTEGRATION OF AGRICULTURAL FOOD PRICES IN NAIROBI COUNTY

BY

SAMUEL GAKUBIA NJUGUNA

I56/80011/2012

SCHOOL OF MATHEMATICS

COLLEGE OF BIOLOGICAL AND PHYSICAL SCIENCES

UNIVERSITY OF NAIROBI

A Project Submitted in Partial Fulfillment for the Degree of Master of Science in Social Statistics

JUNE, 2015

DECLARATION

Candidate:

I, the undersigned, declare that this research project is my original work and to the best of my knowledge has not been presented for the award of a degree in any other University.

Signature:…………………………..   Date:……………………………….

Samuel Gakubia Njuguna

Reg. No.: I56/80011/2012

Supervisors:

This research project has been submitted with our approval as the University supervisors.

Signature:…………………………..   Date:……………………………….

Dr. Isaac C. Kipchirchir

Signature:…………………………..   Date:……………………………….

Dr. John Ndiritu

School of Mathematics,
University of Nairobi,
P.O Box 30197-00100,
NAIROBI.

ACKNOWLEDGEMENT

First, I heart fully thank God for good health, energy and opportunity to study and carry out this research project.

Second, I acknowledge and thank my supervisors Dr. Kipchirchir and Dr. Ndiritu for their guidance, patience and assistance throughout the study.

Special appreciation goes out to all my lecturers for their time and guidance through the academic work leading to the achievement of MSc Degree in Social Statistics.

I am also grateful to members of the MSc Social Statistics class of September 2012 whom we discussed and exchanged academic ideas during the study period.

DEDICATION

I dedicate this work to my wife Jacintah and daughters; Consolata and Jennifer who always give me the joy of life, support and encouragement. I also dedicate to my parents Mr and Mrs Njuguna who with love and selfless efforts have moulded me to be a complete and successful person in life.

ABSTRACT

Uncertainty in agricultural food markets generate such price volatility that, in calculating measures of market-wide inflation for guiding monetary policy, a “core” measure of inflation that removes food prices is preferred. This prompts the search for knowledge, awareness and deep understanding of the trends, stationarity, co-integration and causality in the agricultural food prices in the markets, what this study achieved objectively.

Twenty five food commodities were grouped into three categories; carbohydrates, proteins and vitamins and minerals. The average monthly price per kilogram was computed for each category, yielding sixty data points for each category. The data was analyzed using the R statistical software to investigate their cointegration and causality.

Based on Akaike Information Criterion, the optimal lag for the price time series was found to be one. Based on this lag and using the Augmented Dickey Fuller test, the Carbohydrate prices were found to be non-stationary; therefore they were differenced then de-trended so as to achieve stationarity. The protein and Vitamins and minerals prices were found to be stationary.

Using the Johansen test of cointegration, it was found that there exist two cointegrated models of the food prices. Further, the Granger tests of causality indicated significant unidirectional causality among the food prices. The prices of proteins cause those of carbohydrates and those vitamin and mineral foods. The prices of vitamin and minerals cause those of carbohydrates but not those of protein foods. The prices of carbohydrate foods do not cause those of either proteins or vitamin and mineral foods. Therefore, the recommended model is that which uses the prices of protein and vitamin and mineral foods to explain the prices of the carbohydrate foods. That is, $Carb = f\,(Prot, Vit.Min)$.

LIST OF ACRONYMS AND ABBREVIATIONS

ADF Test	Augmented Dickey Fuller Test
AIC	Akaike Information Criterion
ARDL	Auto Regressive Distributed Lags
Carb	Carbohydrates
CPI	Consumer Price Index
df	Degrees of Freedom
ECM	Error Correction Model
FPE	Final Prediction Error
G-7	Great Seven countries
GDP	Gross Domestic Product
HQ	Hannan-Quinn
I(0)	Integrated of order zero (Stationary)
I(1)	Integrated of order one (Non-stationary)
Kg	Kilogram
Ksh	Kenya Shilling
Ms Excel	Microsoft Excel
Prot	Proteins
$\mathbf{Q_{skew}}$	Quartile Coefficient of Skewness
$\mathbf{R^2}$	Coefficient of Determination
SIC	Schwarz Information Criterion
SE	Stock Exchange
Var	Variance
VAR	Vector Auto Regression
VECM	Vector Error Correction Model
Vit.Min	Vitamins and Minerals

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1.1 Background Information

According to World Bank report (2005), about two thirds of the world's poor population is mainly concentrated in rural areas, which are predominantly agriculture-oriented. Therefore in respect of poverty eradication and raising the welfare standards of the population, more focus should be put on agricultural activities. In Kenya, agriculture is and has been a cornerstone of the economy contributing thirty-five per cent of the Gross Domestic Product (GDP), offering up to eighty per cent of job opportunities and constituting forty per cent of the export earnings. However, in 2006, almost seventy-five percent of working Kenyans made their living by farming, compared with eighty percent in 1980. About one-half of Kenya's total agricultural output is non-marketed subsistence production. Therefore, it cannot be argued that farming is the most important economic sector in Kenya, although less than eight percent of the land is used for crop and food production, and less than twenty percent is suitable for cultivation.

Whilst agricultural food markets rarely behave in precise conformity with the theories of price expounded by economists, it remains the case that the understanding of the market behavior and the ability to predict it, is paramount for any farmer who wishes to maximize the revenues and profits from farm produce. At the same time, farmers and other stake holders in agriculture must be capable of applying alternative approaches when deterministic economic models prove inconsistent with the realities of a complex agricultural marketplace better explained by probabilistic behavioral models.

To this extent, proper monitoring and understanding of the cointegration in the agricultural food prices in the Kenyan markets is critical to the agricultural stake holders for the stability and improvement of the GDP, hence the livelihood of the Kenyan citizens. From 2006 to present, Kenya and the world at large has experienced two agricultural food price booms, where the prices of many, if not all, agricultural food commodities rose dramatically. Between 2006 and 2008, the price of wheat tripled, while the price of corn more than tripled. Commodity price changes such as those observed since 2006 may be more strongly correlated as documented by Silvennoinen and Thorp (2010) because of fundamental supply and demand relationships. Commodities may be substitutes or complements in production or consumption while some commodities may be used as inputs in the production process for other commodities. In addition to these microeconomic relationships, demand and supply for some commodities may respond to common macroeconomic shifters related to aggregate demand, exchange rates, and interest rates.

In general, commodity markets have undergone major changes in the last fifteen years. The popularity of commodity-related financial instruments, such as commodity indices, led many observers to consider that the commodity markets are more deeply connected to the financial market. While more participants in the commodity markets may induce a better risk sharing, the financialization process is also criticized for causing socially undesirable price volatility. While most markets experience unexpected shifts in supply and demand, few markets face the vast array of uncertainties present in agricultural markets. Financial traders of crop-based assets, seeking diversified protection from such uncertainties, create structured securities that pool multiple food crops into unified assets, which then are traded in secondary markets. However, understanding risk built into such structured securities requires a detailed understanding of the cointegration of the agricultural food prices.

1.2 Statement of the Problem

Uncertainties surrounding the prices of agricultural food commodities in Kenya may be attributed to sources such as temperature, floods, drought, demand and supply forces, political atmosphere, disease and infestation. In addition, agricultural commodity markets face uncertainties due to changes in demand and supply, seasons and geographical locations. Such uncertainties affect not only agricultural food prices, but also a myriad of downstream markets, from animal feed to industrial products, that directly or indirectly use agricultural commodities as inputs (Moss, 1992). In fact, uncertainty in agricultural food markets generates such price volatility that, in calculating measures of market-wide inflation for guiding monetary policy, a "core" measure of inflation that removes food prices is preferred. This prompts the search for knowledge, awareness and deep understanding of the trends, stationarity, co-integration and causality in the agricultural food prices in the markets.

1.3 Objectives of the Study

The general objective is to investigate and test the degree of cointegration among the agricultural food prices Kenya. The specific objectives are;

i. Investigate the stationarity of the agricultural food prices in Nairobi County using the ADF unit-root test.

ii. Test the cointegration in the agricultural food prices in Nairobi County using the Johansen test of cointegration.

iii. Test the existence and direction of causality in the agricultural food prices in Nairobi County using the Granger causality test.

1.4 Significance of the Study

Testing for cointegration in the agricultural food prices is crucial in setting up econometric models and in performing inference in food market analysis. Once the agricultural food prices are classified as stationary or non-stationary, cointegrated or non-cointegrated, causal or non-causal, econometricians, statisticians and others will be in a position to sort out long-run and short-run effects in their models, and to set up models where statistical inference will be more meaningful.

Therefore, the findings obtained from this study will go a long way in informing and facilitating development of better, meaningful and statistically significant economic models on food prices that will yield better and appropriately adjusted forecasts and estimates of the future behaviors of such prices. This will in effect shield the farmers, farm workers, agricultural food consumers and the nation at large from unanticipated acute market downturns of the agricultural food prices.

Finally, the findings of the study will act as a point of reference for future studies. The documented report will be available where it can easily be accessed and will in particular benefit future researchers seeking understand the cointegration and causality in the food prices.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

This chapter gives a survey of the existing literature about the cointegration and causality. It gives a summary of the theoretical and conceptual literature into the understanding of the concepts of cointegration. The chapter further reviews other researchers' works on cointegration in various variables and contexts. This is useful in refining the methodology and conducting an objective comparison of the approaches as well as the findings of other researchers and those applied and obtained respectively. The chapter further delves into the existing literature on cointegration in economic and financial variables such as prices and returns.

2.2 Theoretical Review

The theory of cointegration refers to the long-run equilibrium of an economic relationship, which means that although the time series may contain stochastic elements, they will move together in the long run and the difference between them will be determined by a certain relationship if these time series are cointegrated. In other words, the economic variables may have an independent course between them in the short-run, but on the other hand, if there are long paths (that is, if they are cointegrated), these must be taken into account through the specification of the error correction, in the examination of causal relationships between the economic variables. Cointegration naturally arises in economics and finance (Mills, 1999).

In economics, cointegration is most often associated with economic theories that imply equilibrium relationships between time series variables. The permanent income model implies cointegration between consumption and income, with consumption being the common trend. Money demand models imply cointegration between money, income, prices and interest rates (Alexander, 2001). Mills argues that growth theory models imply cointegration between income, consumption and investment, with productivity being the common trend. Purchasing power parity implies cointegration between the nominal exchange rate and foreign and domestic prices (Hayashi, 2000). Covered interest rate parity implies cointegration between forward and spot exchange rates. The expectations hypothesis of the term structure implies cointegration between nominal interest rates at different maturities. The equilibrium relationships implied by these economic theories are referred to as long-run equilibrium relationships, because the economic forces that act in response to deviations from equilibrium may take a long time to restore equilibrium. As a result, cointegration is modeled using long spans of low frequency time series data measured monthly, quarterly or annually.

In finance, cointegration may be a high frequency relationship or a low frequency relationship. Cointegration at a high frequency is motivated by arbitrage arguments (Tsay, 2001). The Law of One Price implies that identical assets must sell for the same price to avoid arbitrage opportunities. This implies cointegration between the prices of the same asset trading on different markets, for example. Similar arbitrage arguments imply cointegration between spot and futures prices, spot and forward prices, and bid and ask prices. Here the terminology long-run equilibrium relationship is somewhat misleading because the economic forces acting to eliminate arbitrage opportunities work very quickly. Cointegration is appropriately modeled using an asymptotic time series that is, large samples. Cointegration at a low frequency is motivated by economic equilibrium theories linking assets prices or expected returns to fundamentals (Cochraine, 2001). For example, the present value model of stock prices states that a stock's price is an expected discounted present value of its expected future dividends or earnings. This links the behavior of stock prices at low frequencies to the behavior of dividends or earnings. In this case, cointegration is modeled using low frequency data and is used to explain the long-run behavior of stock prices or expected returns.

In the Kenyan agricultural sector, one of the burning issues is about; the higher price gap between the price received by producer (farmer) and price paid by the consumer. The price of the commodity in the production areas or in local market is low, while the consumers' price in different market hubs particularly in capital city Nairobi is very high. This has been a big concern among producer, consumer, media and policy makers. Pokhrel (2010) argued that the farmers are forced to sell their products at whatever price is fixed by the traders because of inaccessibility of proper markets, perishable nature of their products, and lack of safe storage. If continuation of such problem exists in long-term, it would have a downbeat effect in agricultural production leading to impede in livelihood, and food and nutrition security.

Market cointegration has positive relationship with market efficiency and market competitiveness; as the market is cointegrated, it tends to be efficient, and competitive. In cointegrated markets, price differences between two markets are equal to the transaction costs between those two markets, provided that trade occurs. A price change in one market will be transmitted on a "one-to-one" basis to the other market either instantaneously or over a number of lags (Sanogo and Maliki, 2010). In general, the major factors influencing market cointegration are inefficient marketing services, lack of infrastructures, restriction to entry firms, and ineffective information services. The previous studies mainly focused on market price cointegration (Sanogo and Maliki, 2010; Shrestha, 2012), and farm-retail price spread. Notably, the average food prices of the food categories; carbohydrates, proteins and vitamins and minerals have not being studied, yet, general access of these by the nationals of a state guarantees food and nutrition security particularly in a developing economy where prices fluctuate erratically.

Mishra and Kumar (2013), reported that the longer the distance between agricultural food markets, the weaker the integration, and vice versa. In fact, no research had been conducted on the cointegration of the prices of the three categories agricultural food products. As such, this study was carried out to analyze the market price cointegration and causality of agricultural food prices in Kenya while using Nairobi County as a case study.

2.3 Empirical Review

Many studies have used the concept of VAR, cointegration and causality to investigate the link between various economic factors. For example, Cologni and Manera (2005) used a structural cointegrated VAR model to study the effects of oil price shocks on output and prices in G-8 countries. Their key finding was that for most of the countries considered, there seems to be an impact of unexpected oil price shocks on interest rates, suggesting a contracting monetary policy response directed to fight inflation.

Çelik and Akgül (2011) studied the relationship between CPI and oil prices in Turkey using the Vector Error Correction Model (VECM). Their study revealed that a 1% increase in fuel prices caused the CPI to rise by 1.26% with an approximate one year lag.

Rudra et al (2012) studied the cointegration of market prices of tomato and its effects to Nepalese farmers. They concluded that the price series of all the markets in the study areas were stationary, and revealed that Chitwan and Morang markets were well cointegrated with Kathmandu market. The price adjustment process in source markets was very fast, especially when the negative price shocks deviation in Kathmandu market, indicated that the tomato farmers in Chitwan and Morang districts were absolutely dependent on Kathmandu market. They suggested that it would be quite easy for source markets to adjust the negative price shock, and more difficult to adjust by farmers, eventually leading to the farmer getting hurt and discouraged.

Ansar and Asaghar (2013) analyzed the impact of oil prices on stock exchange and CPI in Pakistan and concluded that there was no strong relationship between oil prices, CPI and SE100 Index.

Engle (2002) studied the relationship between wholesale and retail prices of food products and Guthrie (1981) analyzed the relationship between general wholesale and retail price indicators. Both surveys suggested causality from lower to higher levels (retail prices). Larue (1991) searched for cointegrating relations and concluded that (contrary to the prevailing notion that output prices are more flexible than input prices) output prices are "weakly exogenous" in the model in the sense that although they are cointegrated, they don't respond systematically to the imbalance of input prices and retail prices.

Christos and Christos (2015) investigated the cointegration of farm production costs, producer prices and retail food prices in Greece. The maximum likelihood method of Johansen showed a long-term equilibrium relationship; that is a positive constant relationship between the three variables in the period 2000-2012. The short-term dynamic model showed that there is unidirectional causal relationship from output prices to food prices. When considered jointly, farm input and farm output prices again Granger-cause the retail food prices. In the long run there is unidirectional causality from output and input prices towards retail food prices.

Cunado et al (2003) analyzed the effect of oil prices on inflation and industrial manufacturing for several European countries for the period of 1960 to 1999. Their findings were that there is an asymmetric effect of oil price on production and inflation. Their findings suggest that there are expected differences in countries' responses to changes in global oil prices depending on their macroeconomic status, whether the country is an oil importer or exporter, and the monetary policies adopted by a given country in response to global oil prices and other trends like exchange rate variations.

Moss (1992) used cointegration analysis to investigate whether the prices received by producers and the price they pay move together in the long run. He found no cointegrating relationship, implying that the effect of margin's compression between input costs and reduced selling price (Cost-Price Squeeze) cannot be rejected in the long run. If the prices that farmers receive and the prices they pay are cointegrated, the Cost-Price Squeeze effect is not sustained in the long run. This means that under the presence of inflation input prices rise more than output prices, since farmers are price receivers and they are not able to pass higher input costs to consumers and thus they have to adjust the use of inputs and outputs, as the ratio of output / input decreases.

Niyimbanira (2013) analyzed the relationship between oil prices and inflation in South Africa. In his paper, he modeled inflation as a dependant variable which is driven by oil prices. Enders and Holt (2012) implemented methods for detecting multiple structural breaks in time series data and stressed that demand growth in emerging economies and the increased utilization of certain crop products contributed to price surges. McPhail (2011) argues on the basis of Granger-causality tests that food prices are the main drivers of oil prices. Serra et al (2011) estimated nonlinear cointegration models and found strong linkages between maize and energy prices which are related in a nonlinear manner.

Masih and Mashi (1996) used Johansen's cointegration analysis to study the relationship between energy use and gross domestic product (GDP) in a group of six Asian countries, including India, Pakistan, Malaysia, Singapore, Philippines and Indonesia. The results indicated that there were cointegration relationships in energy use and GDP among countries like India, Pakistan and Indonesia. However, no cointegration was found in the case of Malaysia, Singapore and the Philippines. The flow of causality was found to be running from energy to GDP in India and GDP to energy in Pakistan and Indonesia.

Yang (2000) considered the causal relationship between energy use and GDP in Taiwan. Using different measures of energy consumption, he found a bi-directional causality between energy and the GDP. This result contradicts that of Cheng and Lai (1997), who found that there is a unidirectional causal relationship from GDP.

Belloumi (2009) applied the Johansen's cointegration procedure to examine the causal relationship between per capita energy consumption and per capita gross domestic product for Tunisia during the 1971-2004 period. In order to test for Granger-causality in the presence of cointegration among variables, a vector error-correction model (VECM) was used instead of a vector autoregressive (VAR) model. His results indicate that the per capita gross domestic product and per capita energy consumption for Tunisia are related by one cointegrating vector and there is a long-run bi-directional causal relationship between the two series and a short-run unidirectional causality from energy to GDP. The source of causation in the long-run was found to be the error correction terms in both directions. Hence an important policy implication resulting from this analysis is that energy can be considered as a limiting factor to GDP growth in Tunisia. It was argued that Tunisia's economy is energy dependent and is relatively vulnerable to energy shocks.

Asufu (2000) tested the cointegration relationship between energy use and income in four Asian countries using cointegration, Johansen's procedure and error-correction analysis. He found that cointegration runs from energy to income in India and Indonesia, and that there is bidirectional causality in Thailand and the Philippines.

CHAPTER 3: METHODOLOGY

3.1 Introduction

In this chapter the methodology that was used in undertaking the study is outlined. The chapter details the data used and the analysis techniques that were employed in the study. These techniques include the methods of optimal model selection, assessment of the deterministic and stochastic trends of the time series price data, model selection, the unit root tests of stationarity and the hypothesis tests of cointegration and causality. R Statistical software and Ms Excel were used for the analysis.

3.2 Data

Secondary monthly wholesale price data for 25 agricultural food commodities was acquired from the Ministry for Agriculture, Livestock and Fisheries; Agribusiness and Market Development Department. Such data are usually collected by extension officers in Kenya for agricultural commodities. The researcher targeted wholesale prices for the said 25 agricultural food commodities over the period of 5 years from January 2010 to December 2014. The data covered the key markets in Nairobi, spread all over the county to ensure representativeness. The time series data was measured in Kenya shillings per Kilogram (Ksh/Kg).

The food commodities were grouped into three food categories; carbohydrates, proteins and vitamins and minerals. The average monthly price per kilogram was then computed for each category, yielding 60 data points for each category. As such, the study did optimal model selection, investigated the deterministic and stochastic trends, stationarity and cointegration of the three food categories.

3.3 Stationarity of a Time Series

A time series is said to be stationary if its mean and variance are constant over time that is, time invariant. Such a time series will tend to return to its mean (mean reversion) and fluctuations around this mean will have a broadly constant amplitude. This implies that a stationary process will not drift too far away from its mean value because of the finite variance. Consequently, the probability structure of a stationary time series is independent of time, hence uniform over time.

On the other hand, a non stationary time series will have a time varying or time varying variance or both. The white noise process $\varepsilon_t \sim N(0, \sigma^2)$ such that it is *iid* (independently and identically distributed) with $Cov(\varepsilon_t, \varepsilon_{t-s}) = 0$.

Examples of non-stationary processes are; pure random walk, deterministic trend, random walk with a drift, random walk with a drift and deterministic trend.

3.3.1 Pure Random Walk Process

A pure random walk process is defined by the equation;

$$Y_t = Y_{t-1} + \varepsilon_t, \qquad (3.1)$$

where ; Y_t is the observation at time *"t"*

Y_{t-1} is the observation at time *"t-1"*

ε_t is the white noise.

Random walk predicts that the value at time *"t"* will be equal to the last period value plus a stochastic (non-systematic) component that is a white noise, which means ε_t is independent and identically distributed with mean *"0"* and variance *"σ^2"*. Random walk can also be named a process integrated of some order, a process with a unit root or a process with a stochastic trend. It is a non mean reverting process that can move away from the mean either in a positive or negative direction. Another characteristic of a random walk is that the variance evolves over time and goes to infinity as time goes to infinity; therefore, a random walk cannot be predicted.

Solving equation (3.1) recursively, we obtain;

$$Y_t = Y_0 + \sum_{i=1}^{t} \varepsilon_i \qquad (3.2)$$

Thus, $E(Y_t) = Y_0$ and $Var(Y_t) = t\sigma^2$ since the white noise; $\varepsilon_t \sim iid\ N(0, \sigma^2)$. Pure random walk process is therefore non-stationary.

3.3.2 Random Walk with a Drift

A random walk with a drift process is defined by the equation;

$$Y_t = \alpha + Y_{t-1} + \varepsilon_t \qquad (3.3)$$

where α is a constant (drift).

If the random walk model predicts that the value at time *"t"* will equal the last period's value plus a constant, or drift (α), and a white noise term (ε_t), then the process is random walk with a drift. It also does not revert to a long-run mean and has variance dependent on time.

Solving equation (3.3) recursively, we obtain;

$$Y_t = n\alpha + Y_0 + \sum_{i=1}^{t} \varepsilon_i t. \tag{3.4}$$

Where "n" is the number of observations.

Thus, $E(Y_t) = \alpha t + Y_0$ and $Var(Y_t) = t\sigma^2$. Random walk process with a drift is therefore non-stationary

However, the differenced series;

$$\Delta Y_t = \alpha + \varepsilon_t \tag{3.5}$$

where $\Delta Y_t = Y_t - Y_{t-1}$, is stationary since $E(\Delta Y_t) = 0$ and $Var(\Delta Y_t) = Var(\varepsilon_t) = \sigma^2$. Random walk process with a drift is therefore difference-stationary.

3.3.3 Deterministic Trend

A deterministic trend is defined by the equation;

$$Y_t = \alpha + \beta t + \varepsilon_t \tag{3.6}$$

where; β is a constant (slope).

Often a random walk with a drift is confused for a deterministic trend. Both include a drift and a white noise component, but the value at time "t" in the case of a random walk is regressed on the last period's value (Y_{t-1}), while in the case of a deterministic trend it is regressed on a time trend (βt). A non-stationary process with a deterministic trend has a mean that grows around a fixed trend, which is constant and independent of time.

From equation (3.6) it is observed that $E(Y_t) = \alpha + \beta t$ (non-stationary) and $Var(Y_t) = \sigma^2$ (Stationary). Therefore, the process $Y_t = \alpha + \beta t + \varepsilon_t$ is non-stationary with a deterministic trend. However, the de-trended series (subtract mean from the series) is stationary.

3.3.4 Random Walk with Drift and Deterministic Trend

A random walk with a drift and deterministic trend process is defined by the equation;

$$Y_t = \alpha + Y_{t-1} + \beta t + \varepsilon_t \qquad (3.7)$$

This is the type of a non-stationary process that combines a random walk with a drift component (α) and a deterministic trend (βt). It specifies the value at time *"t"* by the last period's value, a drift, a trend and a stochastic component.

In this case, the process is also non-stationary. The differenced series;

$$\Delta Y_t = \alpha + \beta t + \varepsilon_t \qquad (3.8)$$

is still time-varying hence non-stationary too! Therefore, de-trending of the differenced series is necessary to make it stationary.

Figure 3.1 shows examples of the different types of non-stationary time series from simulated data.

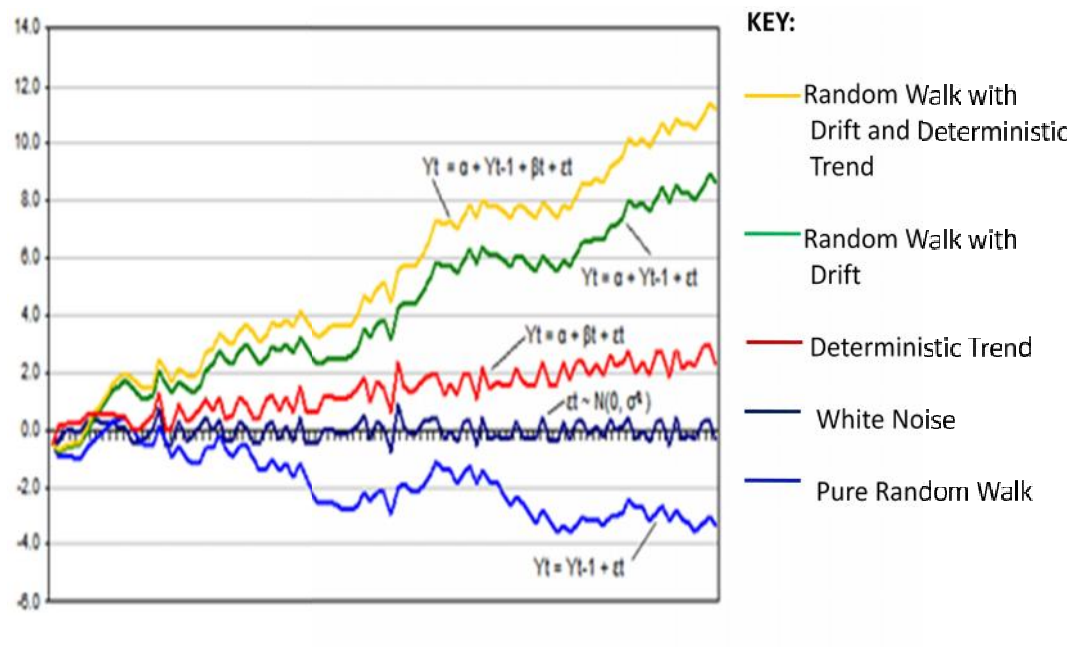

Figure 3.1: Types of non-stationary processes (*Source: Investopedia.com fig. 2)*

3.4 The Unit Root Test of Stationarity

Following Hendry and Juselius (2000), data can be unit root that is, *I(1)*. Such data cannot be used to investigate relationships between the variables because of spurious regression and OLS estimates become invalid. However, data showing such properties can be made stationary by first differencing. If a series is such that its first difference is stationary (and has positive spectrum at zero frequency) then the series has an exact (or pure) unit root (Granger and Swanson, 1996).

Basically, root is a characteristic of a time series data that oftenly causes problems in statistical inference of the time series models. A time-dependent process has a stochastic trend if its characteristic equation has a root of 1 that is, a unit root exists. The characteristics of a stochastic process (unit-root process) include;

i) Movements that cause permanent effects to the time series such that they do not decay as in the case of a stationary process

ii) A stochastic process (unit-root process) has a variance that is dependent on time, and diverges to infinity

iii) Once a time series is found to have a unit root, it can be differenced to make it stationary. For instance, if Y_t is *I(1)*, then the series given by; $\Delta Y_t = Y_t - Y_{t-1}$ is *I(0)* that is, stationary and bears a deterministic trend.

A unit root test tests whether a given time series process/variable has a deterministic or stochastic trend (stationary or non-stationary) using an autoregressive model.

3.4.1 Dickey–Fuller test

This is a test of whether a variable has a unit root (random walk) or not. It was developed by statisticians David Dickey and Wayne Fuller in 1979.

Consider a simple AR (1) model;

$$Y_t = \rho Y_{t-1} + \varepsilon_t \tag{3.9}$$

where Y_t is the variable of interest, *t* is the time index, ρ is a coefficient, and ε_t is the error term. A unit root is present if $\rho = 1$. The model would be non-stationary in this case. By subtracting Y_{t-1} from both sides, the regression model can be written as;

$$Y_t - Y_{t-1} = \rho Y_{t-1} - Y_{t-1} + \varepsilon_t$$

$$\Delta Y_t = (\rho - 1) Y_{t-1} + \varepsilon_t$$

$$= \mu Y_{t-1} + \varepsilon_t \tag{3.10}$$

where; $\mu = \rho - 1$ and ΔY_t is the first difference operator.

This model can be estimated and testing for a unit root is equivalent to testing $\mu = 0$. In fact, the DF test tests the null hypothesis that $\mu = 0$ .Since the test is done over the residual term rather than raw data, it is not possible to use standard t-distribution to provide critical values. Therefore this statistic has a specific distribution simply known as the Dickey–Fuller table. There are three main versions of the test

Test for a unit root; equation (3.10)

Test for a unit root with drift;

$$\Delta Y_t = \alpha + \mu Y_{t-1} + \varepsilon_t \qquad (3.11)$$

Test for a unit root with drift and deterministic time trend:

$$\Delta Y_t = \alpha + \beta t + \mu Y_{t-1} + \varepsilon_t \qquad (3.12)$$

Each version of the test has its own critical value which depends on the size of the sample. In each case, the null hypothesis is that there is a unit root, $\mu = 0$. The tests have low statistical power in that they often cannot distinguish between true unit-root processes *($\mu = 0$)* and near unit-root processes (is close to zero) which results to the "near observation equivalence" problem.

The intuition behind the test is that if the series is stationary (or trend stationary), then it has a tendency to return to a constant (or deterministically trending) mean. Therefore large values will tend to be followed by smaller values (negative changes), and small values by larger values (positive changes). Accordingly, the level of the series will be a significant predictor of next period's change, and will have a negative coefficient $\mu < 0$. If, on the other hand, the series is integrated, then positive changes and negative changes will occur with probabilities that do not depend on the current level of the series; in a random walk, where a person is now does not affect which way the person will go next.

It is notable from equation (3.11) that if a unit root exists;

$$\Delta Y_t = \alpha + \varepsilon_t \qquad (3.13)$$

since $\mu = 0$.

Solving equation (3.13) recursively for Y_t, it is observed that;

$$Y_t = Y_0 + \sum_{i=1}^{t} \varepsilon_i + \alpha t \qquad (3.14)$$

This will have a deterministic trend coming from αt and a stochastic intercept term coming from $Y_0 + \sum_{i=1}^{t} \varepsilon_i$, resulting in a stochastic trend.

The choice of the three main versions of the test should be used is not a minor issue. The decision is important for the size of the unit root test (the probability of rejecting the null hypothesis of a unit root when there is one) and the power of the unit root test (the probability of rejecting the null hypothesis of a unit root when there is not one). Inappropriate exclusion of the intercept or deterministic time trend term leads to bias in the coefficient estimate for μ leading to the actual size for the unit root test not matching the reported one.

If the time trend term is inappropriately excluded with the term estimated, then the power of the unit root test can be substantially reduced as a trend may be captured through the random-walk with drift model. On the other hand, inappropriate inclusion of the intercept or time trend term reduces the power of the unit root test, and sometimes that reduced power can be substantial. Use of prior knowledge about whether the intercept and deterministic time trend should be included is of course ideal but not always possible. When such prior knowledge is unavailable, various testing strategies have been suggested, for example, by Dolado (1990) and by Enders (2004), who suggested the use of the Augmented Dickey–Fuller test (ADF) extension. The ADF extension removes all the structural effects (autocorrelation) in the time series and then tests using the same procedure as the DF test.

3.4.2 Augmented Dickey-Fuller (ADF) test

Consider the general AR (p) process given by;

$$Y_t = \alpha + \rho_1 Y_{t-1} + \rho_2 Y_{t-2} \ldots \rho_p Y_{t-p} + \varepsilon_t \tag{3.15}$$

If this is the process generating the data but an AR (1) model is fitted, say;

$$Y_t = \alpha + \rho_1 Y_{t-1} + v_t \tag{3.16}$$

Then;

$$v_t = \rho_2 Y_{t-2} \ldots \rho_p Y_{t-p} + \varepsilon_t \tag{3.17}$$

and the autocorrelations of v_t and v_{t-k} for $k > 1$ will be non-zero because of the presence of the lagged terms. Thus an indication of whether it is appropriate to fit an AR (1) model can be aided by considering the autocorrelations of the residual from the fitted models. In the ADF test, the DF test is extended to autoregressive processes of order greater than one. Consider the simple AR (2) process;

$$Y_t = \alpha + \rho_1 Y_{t-1} + \rho_2 Y_{t-2} + \varepsilon_t \qquad (3.18)$$

This can also be written as;

$$Y_t = \alpha + \rho_1 Y_{t-1} + \rho_2 Y_{t-1} - \rho_2 Y_{t-1} + \rho_2 Y_{t-2} + \varepsilon_t$$

$$= \alpha + (\rho_1 + \rho_2) Y_{t-1} - \rho_2 (Y_{t-1} - Y_{t-2}) + \varepsilon_t. \qquad (3.19)$$

Subtracting Y_{t-1} from both sides;

$$Y_t - Y_{t-1} = \alpha + (\rho_1 + \rho_2) Y_{t-1} - \rho_2 (Y_{t-1} - Y_{t-2}) - Y_{t-1} + \varepsilon_t$$

$$\Delta Y_t = \alpha + (\rho_1 + \rho_2 - 1) Y_{t-1} + \rho_2 (Y_{t-2} - Y_{t-1}) + \varepsilon_t$$

$$\Delta Y_t = \alpha + \mu Y_{t-1} + \theta_1 \Delta Y_{t-1} + \varepsilon_t \qquad (3.20)$$

$$where\ \mu = \rho_1 + \rho_2 - 1 \qquad and \quad \theta_1 = \rho_2.$$

This implies that for a general differenced AR (p) process;

$$\Delta Y_t = \alpha + \mu Y_{t-1} + \theta_1 \Delta Y_{t-1} + \cdots + \theta_{p-1} \Delta Y_{t-p} + \varepsilon_t$$

that is,

$$\Delta Y_t = \alpha + \mu Y_{t-1} + \sum_{i=1}^{p} \theta_i \Delta Y_{t-i} + \varepsilon_t \qquad (3.21)$$

Including the trend component βt, the above model becomes;

$$\Delta Y_t = \alpha + \beta t + \mu Y_{t-1} + \theta_1 \Delta Y_{t-1} + \cdots + \theta_{p-1} \Delta Y_{t-p} + \varepsilon_t$$

that is,

$$\Delta Y_t = \alpha + \beta t + \mu Y_{t-1} + \sum_{i=1}^{p} \theta_i \Delta Y_{t-i} + \varepsilon_t \qquad (3.22)$$

Where α is a constant, "β" is the coefficient on a time trend and p the lag-order of the autoregressive process. It is observed that setting $\beta = 0$ corresponds to modeling a random walk with a "drift". By including lags of the order "p" the ADF formulation allows for higher-order autoregressive processes. This implies that the lag length "p" has to be determined when applying the test. Several approaches are available to determine the lag length and among them is the Akaike Information Criterion (AIC). Others include; Hannan-Quinn (HQ) criterion, Schwarz Information Criterion (SIC) and Final Prediction Error (FPE) criterion.

The ADF unit-root test is performed to test the hypotheses;

Null, $H_0: \mu = 0$; There exists a unit-root (Non-stationary)

Alternative, $H_1: \mu < 0$; There does not exist a unit-root (Stationary)

The ADF test statistic is a negative number. The lower the test statistic, the stronger the rejection of the null hypothesis at the specified significance level.

The ADF test statistic is computed as;

$$ADF_\tau = \frac{\mu}{SE(\mu)} \tag{3.23}$$

The computed statistic is then compared with the relevant critical value for the ADF Test. If the test statistic is less than the critical value, the null hypothesis is rejected hence a unit root does not exist. This test is non symmetrical so an absolute value of the test statistic may not be considered.

3.5 Cointegration

Two or more time series are cointegrated if they share a common stochastic drift. If two or more time series are individually integrated, but some linear combination of them has a lower order of integration, then the series are said to be cointegrated. Granger C., Newbold P. (1974) showed that, in the early days, many economists used linear regressions on non-stationary time series data, which is a dangerous approach that could yield spurious correlation. The possible presence of cointegration must be taken into account when choosing a technique to test hypotheses or build a model concerning the relationship between two variables having unit roots that is, integrated of at least order one. A superior method of testing for cointegration is the Johansen test of cointegration.

3.5.1 Johansen Test of Cointegration

Johansen's procedure builds cointegrated variables directly on maximum likelihood estimation instead of relying on OLS estimation. This procedure relies heavily on the relationship between the rank of a matrix and its characteristic roots. Johansen derived the maximum likelihood estimation using sequential tests for determining the number of cointegrating vectors. His method can be seen as a secondary generation approach in the sense that it builds directly on maximum likelihood instead of partly relying on least squares. In fact, Johansen's procedure is nothing more than a multivariate generalization of the Dickey-Fuller test. Consequently, he proposes two different likelihood ratio tests namely; the trace test and the maximum Eigen value test.

This procedure is a vector cointegration test method. It has the advantage over the Engle-Granger and the Phillips-Ouliaris methods in that it can estimate more than one cointegration relationship, if the data set contains two or more time series. It is subject to asymptotic properties, that is, large samples. If the sample size is too small then the results will not be reliable and one should use Auto Regressive Distributed Lags (ARDL).

Johansen's methodology takes its starting point in the Vector Auto Regression (VAR) of order $\boldsymbol{p}$ given by;

$$\boldsymbol{Y_t = \rho_1 Y_{t-1} + \rho_2 Y_{t-2} \ldots + \rho_p Y_{t-p} + \varepsilon_t} \qquad (3.24)$$

Where; $\boldsymbol{Y_t}$ is an $n \times 1$ vector of variables that are integrated of order one,

$\boldsymbol{\varepsilon_t}$ is an $n \times 1$ vector of white noise and

$\boldsymbol{\rho_i}$ $(i = 1,2, \ldots p)$ are $n \times n$ coefficient matrices.

Subtracting Y_{t-1} from both sides;

$$\boldsymbol{\Delta Y_t = \phi_1 \Delta Y_{t-1} + \phi_2 \Delta Y_{t-2} \ldots + \phi_{p-1} \Delta Y_{t-p+1} - \prod Y_{t-p} + \varepsilon_t} \qquad (3.25)$$

Where;

$\boldsymbol{\phi_1} = \boldsymbol{\rho_1} - \mathbf{I}$, $\boldsymbol{\phi_2} = \boldsymbol{\rho_1} - \boldsymbol{\phi_1}$, $\boldsymbol{\phi_3} = \boldsymbol{\rho_1} - \boldsymbol{\phi_2}$, $\boldsymbol{\phi_4} = \boldsymbol{\rho_1} - \boldsymbol{\phi_3}$ and $\boldsymbol{\prod} = \boldsymbol{I} - \boldsymbol{\phi_1} - \boldsymbol{\phi_2} - \boldsymbol{\phi_2} \ldots - \boldsymbol{\phi_p}$ where $\boldsymbol{I}$ is the n x n Identity matrix.

The matrix $\boldsymbol{\phi}$ determines the extent to which the system is cointegrated. It is called the *impact matrix*.

Consider the first equation of the system to be:

$$\Delta Y_{1t} = \omega_{11}\Delta Y_{t-1} + \omega_{12}\Delta Y_{t-2} \ldots + \omega_{1p-1}\Delta Y_{t-p+1} - \pi_1 Y_{t-p} + \varepsilon_{1t} \qquad (3.26)$$

where ; ω_{1j} is the first row of ϕ_j (j = 1, 2, …. p-1) and π_1 is the first row of $\prod$.

Here, ΔY_{1t} is stationary, ϕ_j (j = 1, 2, …. p-1) are all stationary, ε_t is assumed to be stationary, and so for a meaningful equation, $\pi_1 Y_{t-p}$ must be stationary. If none of the components of Y_t are cointegrated, they must all be zero. If they are cointegrated, then all the rows of $\prod$ must be cointegrated.

The matrix $\prod$ has order $n \times n$ and rank n (that is, n linearly independent rows or columns) forming a basis for n-dimensional vector space. Therefore, all $n \times 1$ vectors can be generated as linear combinations of its row. Any of these linear combinations of the rows would lead to stationarity, meaning that Y_{t-p} has stationary components if the rank of $\prod$ is $r < n$. To detect the number of cointegrating vectors, Johansen proposed two likelihood ratio tests:

i) *The trace test* :- This tests ; H_0: $r \leq n$ against H_1: $r > n$. The test statistic is given by;

$$J_{trace} = -n \sum_{i=r+1}^{n} ln\,(1 - \hat{\lambda}_i) \qquad (3.27)$$

ii) *The maximum Eigen value*:- This tests H_0: $r \leq n$ against H_1: $r = n+1$. The test statistic is given by;

$$J_{max} = -n ln\,(1 - \hat{\lambda}_{r+1}) \qquad (3.28)$$

where; T is the sample size and $\hat{\lambda}_i$ is the i^{th} largest canonical correlation.

In both cases, the computed test statistic is compared with the critical statistic, and the null hypothesis rejected if; $\textit{test statistic} \geq \textit{critical statistic}$. The asymptotic critical values can be found in Johansen and Juselius (1992) and are also given by most econometric software packages.

3.6 Causality Test

Causality test is a statistical hypothesis test for determining whether one time series is useful in forecasting another. Ordinarily, regressions reflect "mere" correlations, but causality in economics could be reflected by measuring the ability of predicting the future values of a time series using past values of another time series. A time series X is said to cause Y if it can be shown, usually through a series of t-tests and F-tests on lagged values of X (and with lagged values of Y also included), that those X values provide statistically significant information about future values of Y.

3.6.1 Granger Causality Test

Granger causality is a statistical concept of causality that is based on prediction. According to Granger causality, if a variable X_1 Granger-causes a variable X_2, then past values of X_1 should contain information that helps predict X_2 above and beyond the information contained in past values of X_2 alone. Its mathematical formulation is based on linear regression modeling of stochastic processes (Granger 1969). More complex extensions to nonlinear cases exist, however these extensions are often more difficult to apply in practice. Granger causality was developed in 1960s and has been widely used in economics since the 1960s.

If a time series is a stationary process, the test is performed using the level values of two (or more) variables. If the variables are non-stationary, then the test is done using first (or higher) differences. The number of lags to be included is usually chosen using an information criterion, such as the AIC or the SIC. Any particular lagged value of one of the variables is retained in the regression if it is significant according to a t-test and it and the other lagged values of the variable jointly add explanatory power to the model according to an F-test. Then the null hypothesis of no Granger causality is not rejected if and only if no lagged values of an explanatory variable have been retained in the regression.

The Granger causality test assumptions are;

i) The future cannot cause the past. The past causes the present or the future.
ii) A cause contains unique information about an effect not available elsewhere.

X_t is said not to Granger-cause Y_t if for all $h > 0$;

$$F(Y_{t+h}|\Omega_t) = F(y_{t+h}|\Omega_t - X_t) \tag{3.29}$$

Where F is the conditional distribution and $\Omega_t - X_t$ is all the information in the universe except series X_t.

The whole distribution F is generally difficult to handle empirically and we turn to conditional expectation and variance. It is defined for all $h > 0$ and not only for $h = 1$. Causality at different h does not imply each other. They are neither sufficient nor necessary. Ω_t contains all the information in the universe up to time "t" that excludes the potential ignored common factors problem. Instantaneous causality $\Omega_{t+h} - X_{t+h}$ and feedback are difficult to interpret unless it has additional structural information.

A refined definition may be given as; X_t does not Granger-cause Y_{t+h} with respect to information J_t if;

$$E(Y_{t+h}|J_t, X_t) = E(Y_{t+h}|J_t) \tag{3.30}$$

For a bivariate system y_t, x_t defined by;

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} A_{11}(B) & A_{12}(B) \\ A_{21}(B) & A_{22}(B) \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{xt} \end{bmatrix}$$

$$= \begin{bmatrix} \emptyset_{11}(B) & \emptyset_{12}(B) \\ \emptyset_{21}(B) & \emptyset_{22}(B) \end{bmatrix} \begin{bmatrix} \varepsilon_{yt-1} \\ \varepsilon_{xt-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{xt} \end{bmatrix} \tag{3.31}$$

X_t does not Granger-cause Y_t if $\phi_{12}\ (B) = 0$ or $\phi_{12,i}\ (B) = 0$ for all $i = 1,2 \ldots$ This condition is equivalent to $A_{12,i} = 0$ for all $i = 1, 2 \ldots p$. In other words, this corresponds to the restrictions that all cross-lags coefficients are all zero which can be tested using the Wald statistics.

Considering a bivariate AR(1) process, such that A_{ij}(B) = A_{ij} $,i,j = 1, 2;$ there are four possible causal directions between x and y;

Feedback $H_0 : X \leftrightarrow Y$ *(Bidirectional)*

$$H_0 = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

Independent, $H_1 : X \perp Y$ *(No causality)*

$$H_1 = \begin{pmatrix} A_{11} & 0 \\ 0 & A_{22} \end{pmatrix}$$

X causes Y but Y does not cause X, H_2: X $\twoheadrightarrow$ Y

$$H_2 = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}$$

Y causes X but X does not cause Y, $H_3 : Y \twoheadrightarrow X$

$$H_3 = \begin{pmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{pmatrix}$$

CHAPTER 4: DATA ANALYSIS AND RESULTS

4.1 Introduction

This chapter presents data analysis findings, presentations and interpretations. The data used for analysis was secondary data providing monthly wholesale prices for 25 agricultural food commodities over the period of 5 years. The data covered the key markets in Nairobi County and was recorded in Kenya shillings per Kilogram (Ksh/Kg). The food commodities were grouped into three food categories; carbohydrates (Carb), proteins (Prot) and vitamins and minerals (Vit.Min). The average monthly price per kilogram was then computed for each category, yielding 60 data points for each category.

4.2 Descriptive Analysis

Table 4.1: Descriptive Statistics

	Carb	**Prot**	**Vit.Min**
Min.	30.91	55.95	30.48
1st Qu.	35.03	61.08	37.28
Median (2nd Qu.)	38.11	65.73	39.42
Mean	38.07	65.63	39.46
3rd Qu.	39.9	69.74	42.09
Max.	47.53	78.52	46.97
Q_{Skew}	-0.26489	-0.0739	0.1101871

The summaries indicate that the carbohydrates and the Vitamins and Mineral foods share a range of a minimum of sh. 30.48 per Kg to a maximum of sh. 47.53, with mean prices of around sh. 39.00 per Kg. The protein food prices take a different range of sh. 55.95 to a maximum of sh. 78.52 with a mean of sh. 65.63 per Kg.

Further, the Quartile coefficients of skewness indicate pronounced negative skewness for the carbohydrate food prices, moderate negative skewness for the protein food prices and significant positive skewness for the vitamin and mineral food prices.

Table 4.2: Variation

	Mean	Std. Dev	Coeff. of Var
Carb	38.07	3.835952	0.1007605
Prot	65.63	5.36783	0.0817893
Vit.Min	39.46	3.8478	0.0975114

Carbohydrate and vitamin and mineral food prices have the highest coefficients of variation at 10.08% and 9.75% respectively. The observation is that the protein foods have the highest prices, but these prices are the most stable prices in the markets. Those of carbohydrates, vitamin and mineral foods keep on fluctuating from time to time bringing about significant time series movements, hence price volatility within the ranges specified previously in this chapter.

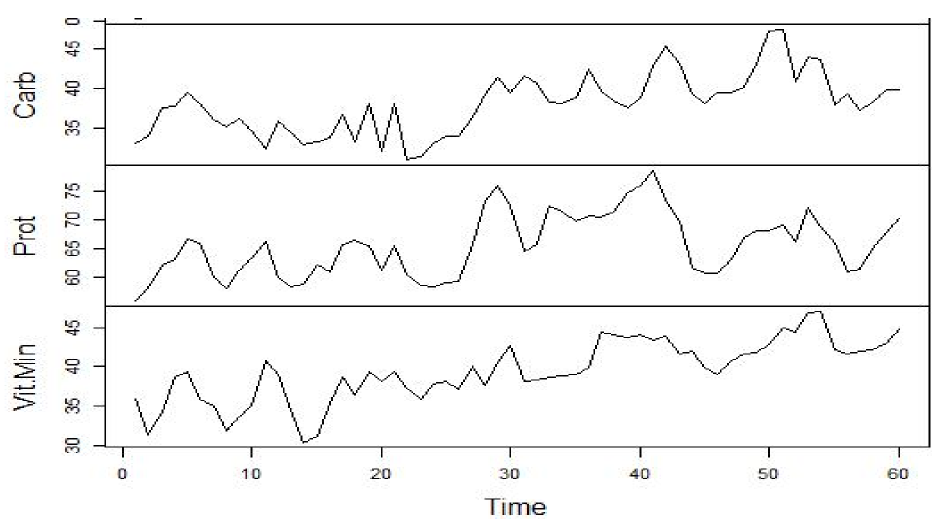

Figure 4.1: Time Plots

With the months as a control variable, it is evident from the time series plots that there exist significant volatility and instability, in all the three types of food prices. These results to the irregular periodic upward and downward movements observed in the plots. As such, the time series are observed to yield random walk with drift processes.

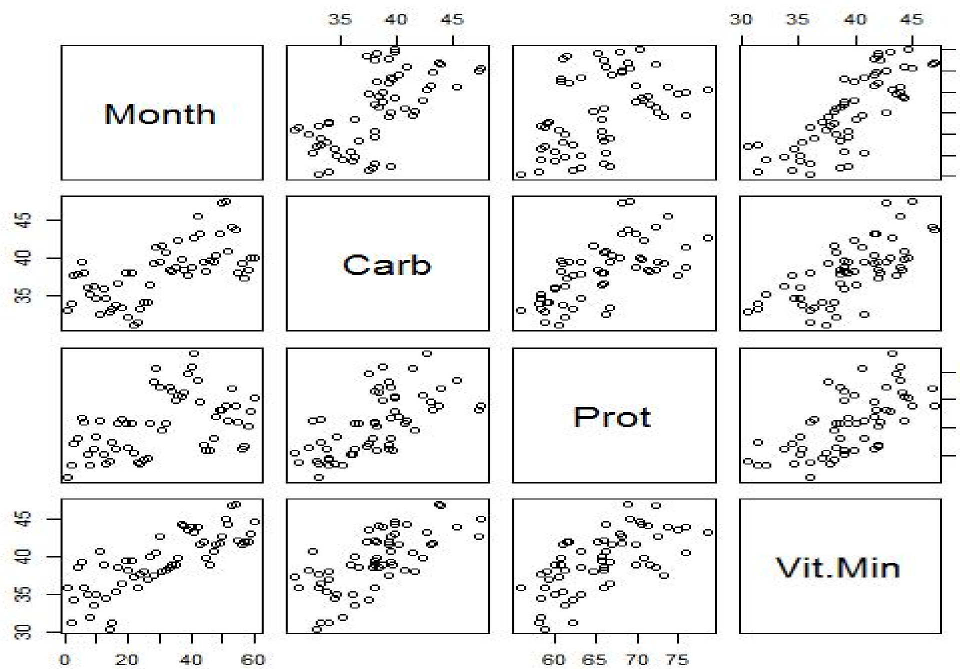

Figure 4.2: Scatter plots

The scatter plots indicate general upward trends for all food prices.

Using the Ordinary Least Squares (OLS) method, linear regression of the food prices (dependent) against time (independent) resulted into the following models;

$$Price(Carb) = 33.8243 + 0.1391 * Time \quad ; Adjusted\ R^2 = 39.09\% \tag{4.1}$$

$$Price(Prot) = 61.4279 + 0.1377 * Time \quad ; Adjusted\ R^2 = 18.69\% \tag{4.2}$$

$$Price(Vit.Min) = 34.1308 + 0.1746 * Time \quad ; Adjusted\ R^2 = 62.16\% \tag{4.3}$$

The linear regression models confirmed weak linearity of the time series price data for all the three food categories, with the vitamins and minerals' prices indicating a relatively strong linear fit with adjusted $R^2 = 62.16\%$. Proteins indicated the weakest linear fit with adjusted $R^2 = 18.69\%$. Consequently, the deterministic linear trend cannot be ignored, hence added to the random walk with a drift models to yield random walk with drift and deterministic trend models.

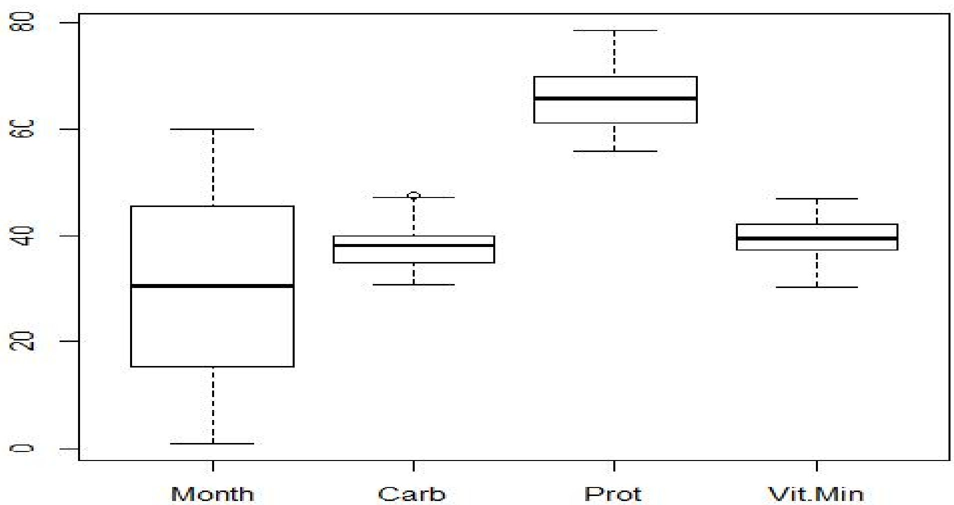

Figure 4.3: Box Plots - Test for normality of the food prices

With the "month" as a control variable, it is clear that the all the food prices are approximately normally distributed.

4.3 Optimal lag selection

The Akaike Information Criterion (AIC) was conducted to determine the optimal lag (***p***) for the Augmented Dickey-Fuller unit root test. However, the results of the Hannan-Quinn (HQ) criterion, Schwarz Information Criterion (SIC) and Final Prediction Error (FPE) criterion were utilized for confirmation and control.

Table 4.3: Lag Selection Criteria

	Lag Number (p)				
	1	**2**	**3**	**4**	**5**
AIC(n)	5.357168	5.423812	5.584734	5.676014	5.857725
HQ(n)	5.568873	5.76254	6.050486	6.268788	6.577522
SC(n)	5.904623	6.299739	6.789134	7.208887	7.719071
FPE(n)	212.444	228.1598	270.7321	301.9751	372.568

According to all criteria used in the test, the optimal lag number is $p = 1$. This was obtained as the optimal for a VAR with a constant, a VAR with trend and a VAR with both as deterministic regressors.

4.4 Unit Root, Cointegration and Causality Tests

Using the Augmented Dickey-Fuller unit root test with lag $p=1$, stationarity of the food price data was tested giving the outcome in table 4.4. The critical values in the function **`ur.df()`** are drawn from Dickey and Fuller (1981) and Hamilton (1994).

Table 4.4: Augmented Dickey Fuller (ADF) Unit Root Tests

Variable	Deterministic terms	Lags	Test value	Critical values		
				1%	**5%**	**10%**
Carb	const, trend	1	-3.0419	-4.04	-3.45	-3.15
ΔCarb	const	1	-2.3958	-3.51	-2.89	-2.58
Prot	const, trend	1	-4.0465	-4.04	-3.45	-3.15
ΔProt	const	1	-3.623	-3.51	-2.89	-2.58
Vit.Min	const, trend	1	-5.154	-4.04	-3.45	-3.15
ΔVit.Min	const	1	-5.154	-3.51	-2.89	-2.58

It is observed that the test statistics for proteins and vitamins and minerals are greater than the critical values at all the three significance levels. As such, the null hypothesis of existence of a unit root is rejected. This implies that the proteins and vitamin and minerals price data have no unit roots that is, *I (0)* therefore they are stationary with deterministic trends.

On the other hand, the test values for carbohydrate prices are less than the critical values at all the three significant levels. Consequently, the null hypothesis is not rejected meaning that a unit root exists. This implies that the carbohydrate price series is non stationary. As such, it is an integrated time series of order one that is, *I (1)*. Due to the non-stationarity, the time series was differenced and then de-trended in order to achieve stationarity which is a requirement for the Johansen test of cointegration and the Granger test of causality.

Table 4.5: Johansen Test of Cointegration Results

Test type: trace statistic , with linear trend						
Null hypothesis (H_0)	**Alternative hypothesis (H_1)**	**Test statistic**	**10%**	**5%**	**1%**	**Results**
r <= 2	r > 2	5.45	6.5	8.18	11.65	Fail to reject H_0
r <= 1	r > 1	34.44	15.66	17.95	23.52	Reject H_0
r = 0	r > 0	73.95	28.71	31.52	37.22	Reject H_0
Test type: maximal Eigen value statistic, with linear trend						
Null hypothesis (H_0)	**Alternative hypothesis (H_1)**	**Test statistic**	**10%**	**5%**	**1%**	**Results**
r <= 2	r = 3	5.45	6.5	8.18	11.65	Fail to reject H_0
r <= 1	r = 2	28.99	12.91	14.9	19.19	Reject H_0
r = 0	r = 1	39.51	18.9	21.07	25.75	Reject H_0

The results of the Johansen tests of cointegration, both trace and maximal Eigen values tests indicate that there are two cointegrated models of the prices of the three food categories. This implies that two combinations of the prices of the food categories move together (that is, are associated) in the long run. As such, the prices do meet the requirements for the ECM or VECM; hence these may be developed and will have significant economic indications as far as movements and co-movements of food prices in Nairobi County are concerned. However, it must be noted that there does not exist one cointegrated model of the three categories of food prices. This prompted causality tests to identify the price that does not cause movements in the other prices that is, identify where there is unidirectional causality and where there is bidirectional causality.

Table 4.6: Granger-Causality test results

Cause Variable	Response Variable	Null Hypothesis (H_0)	F-test Statistic	Df_1	Df_2	P-value	Decision
Carb	Prot	Carb do NOT cause Prot	1.3556	1	112	0.2468	Fail to reject H_0
	Vit.Min	Carb do NOT cause Vit.Min	1.3060	1	112	0.2556	Fail to reject H_0
Prot	Carb	Prot do NOT cause Carb	6.8241	1	112	0.0102	Reject H_0
	Vit.Min	Prot do NOT cause Vit.Min	6.9179	1	110	0.0098	Reject H_0
Vit.Min	Carb	Vit.Min do NOT cause Carb	5.4125	1	112	0.0218	Reject H_0
	Prot	Vit.Min do NOT cause Prot	0.6761	1	110	0.4127	Fail to reject H_0

The Granger causality test results give small F-test statistics and large p-values (greater than 0.05 significance level) where carbohydrate prices are the "cause". This led to failure to reject the null hypothesis (H_0), hence the conclusion that; prices of carbohydrates do not Granger-cause those of proteins or those of vitamins and minerals.

The results yielded large F-test statistics and small p-values (less than 0.05 significance level) where protein prices are the "cause". This led to the rejection of the null hypothesis (H_0), hence the conclusion that; prices of proteins Granger-cause those of Carbohydrates and as well those of vitamin and minerals.

Lastly, where the prices of vitamins and minerals were considered as the "cause", a large F-test statistic and a small p-value were obtained for the carbohydrate prices, implying that the vitamin and minerals prices Granger-cause those of carbohydrates. However, the F-test statistic is small and the p-value large in the test for protein prices. This implies that prices of vitamin and minerals do not Granger-cause those of proteins.

In conclusion, unidirectional causality is observed in the food prices as discussed above, and the two cointegrated models can now be determined as;

$$Carb = f(Prot, Vit.Min) \tag{4.4}$$

$$Vit.Min = f(Prot) \tag{4.5}$$

Table 4.7: Correlation Matrix

	Carb	**Prot**	**Vit.Min**
Carb	1	0.6454785	0.6944917
Prot	0.6454785	1	0.5521237
Vit.Min	0.6944917	0.5521237	1

Correlation analysis results in the correlation matrix above indicate significant direct linear relationships between paired price time series. However, it is observed that vitamin and Mineral prices are better predictors of the carbohydrate prices than the protein prices. The protein prices are correlated to those of the vitamin and minerals, but the correlation is not strong enough to imply collinearity if the two prices are used as predictors. As such, the whole model (equation 4.4) will be preferred over the reduced model (equation 4.5).

CHAPTER 5: CONCLUSION AND RECOMMENDATIONS

5.1 Introduction

This chapter gives a summary of the findings and interpretations as per the data analysis in the previous chapter, the implications and thus the conclusions. Further, the chapter outlines the suggested recommendations for action based on the findings of the study. In addition, the chapter suggests areas of further study and research. Finally, the chapter enumerates the key challenges and unanticipated limitations that were encountered through the study.

5.2 Conclusions

Food prices in Nairobi markets range from a minimum of Ksh 30.48 to a maximum of Ksh 78.52 per Kg with a mean of Ksh 47.72 per Kg. This is associated with a coefficient of variation of 9.34% indicating significant general variation and volatility of the prices. The least expensive food category is that of carbohydrates with a mean price of Ksh 38.07 per Kg, while the most expensive are the protein foods with a mean price of Ksh 65.63 per Kg. these are associated with 10.08% and 8.18% coefficient of variations respectively. This implies that the protein prices, though high, they are more stable relative to those of the carbohydrates. The prices of protein foods as well as those of the vitamins and minerals are moderately normal with very slight negative and positive skewness respectively.

Food prices in Nairobi County form time series that exhibit random walk with drift and deterministic trend. Further, the price time series are non stationary for the carbohydrate foods but stationary for proteins and vitamins and mineral foods.

Based on the Akaike Information criterion and other statistical information criteria, the optimal lag for the price time series is one**.** Based on this lag and using the Augmented Dickey Fuller test, the proteins and vitamins and minerals prices are stationary with deterministic linear trends, while the carbohydrate prices are non-stationary. Differencing and de-trending of these prices resulted into stationary data.

Based on the Johansen test of cointegration, there exist two cointegrated models of the of food prices. As such, ECM and VECM models may be developed for economic and statistical purposes. On the other hand, Granger tests of causality indicated that there exists significant causality among the food prices. The prices of carbohydrate foods do not Granger-cause those of either proteins or vitamin and mineral foods. The prices of proteins, Granger-cause those of carbohydrates and the vitamin and minerals foods.   The prices of vitamin and minerals Granger-cause those of carbohydrates and but not those of protein foods. As such, the price time series of one food category can be used to significantly forecast the future prices of another food category while taking into account these findings. This confirms the results of the Johansen test of cointegration guaranteeing more confidence on the ECM and VECM econometric models on these food prices.

5.3 Recommendations

It is recommended that further research may be done to develop and assess the ECM and VECM econometric models on the agricultural food prices taking into consideration the findings of this study in terms of stationarity, optimal lag, cointegration and causality in these agricultural food prices.

References

Alexander C. (2001). *Market Models: A Guide to Financial Data Analysis*, John Wiley and Sons.

Ansar I. and Asaghar M. N. (2013). The impact of oil prices on stock exchange and CPI in Pakistan. *IOSR Journal of Business and Management (IOSR-JBM). E-ISSN: 2278-487X*. Volume 7, Issue 6.

Asufu-Adjaye D. (2000). The relationship between energy consumption, energy prices and economic growth: time series evidence from Asian developing countries. *Energy economics*, vol. 22:615-625.

Belloumi M. (2009). Energy consumption and GDP in Tunisia: Cointegration and causality analysis. *Energy policy*, vol. 37:2745-2753.

Çelik T. and Akgül B. (2011). Changes in Fuel Prices in Turkey: An Estimation of the inflation effect using VAR Analysis. *Journal of Economics and Business*.

Cheng B. and Lai T. (1997). An investigation of cointegration and causality between energy consumption and economic activity in Taiwan. *Energy Economics*, vol. 19: 435-444.

Christos P. and Christos T. (2015). Farm production costs, producer Prices and retail food prices: A cointegration analysis. *AUA Working Paper Series No. 2015* vol.1:8-9.

Cochrane J. (2001). *Asset Pricing*. Princeton University Press, New Jersey.

Cologni A. and Manera M. (2005). Oil Prices, Inflation and Interest Rates in a Structural Cointegrated VAR Model for the G-7 Countries. *NOTA DI LAVORO 101.2005. IEM – International Energy Markets.*

Cunado J. et al (2003). Oil prices, economic activity and inflation: evidence for some Asian countries. *The Quarterly Review of Economics and Finance.*

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association.*

Dolado J. (1993). *Cointegration, error-correction, and econometric analysis of non stationary data*. Oxford University Press, London.

Enders W. (2004). Applied econometrics time series. *Wiley series in Probability and Statistics.*

Enders W. and Holt. M. (2012). Sharp Breaks or Smooth Shifts? An Investigation of the Evolution of Primary Commodity Prices. *American Journal of Agricultural Economics*, 94(3): 659–673.

Engle R. F. (2002). Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models. *Journal of Economics and Business*.

Engle R. F. and Granger C. W. (1987). Co-integration and error correction: Representation, estimation and testing. *Journal of Business and Economic Statistics*.

Granger C. (1981). Some Properties of Time Series Data and Their Use in Econometric Model Specification. *Journal of Econometrics*.

Granger C. and Newbold P. (1974). Spurious Regressions in Econometrics. *Journal of Econometrics.*

Granger C. and Swanson N. (1996). An Introduction to Stochastic Unit Root Processes. *Journal of Econometrics.*

Guthrie R. (1981). The relationship between wholesale and consumer prices. *Southern Economic Journal 47.*

Hayashi F. (2000). *Econometrics*. Princeton University Press. pp. 623–669. ISBN 0-691-01018-8.

Hendry D. and Juselius K. (2000). Explaining cointegration analysis: Part I. *Energy Journal* vol. 21, 1-42.

http://www.investopedia.com/articles/trading/07/stationary.asp#ixzz3cvT9raR0

Johansen S. (1988). Statistical analysis of cointegrated vectors. *Journal of Economic Dynamics and Controls.*

Johansen S. (1995). *Likelihood-based Inference in Cointegration Vector Autoregressive Models.* Oxford University Press.

Johansen S. and Juselius K. (1992). Testing structural hypothesis in a multivariate cointegration analysis of the PPP and the UIP for UK. *Journal of Econometrics*.

Khalifa H. and Sakka M. (2004). Energy use and output in Canada: a multivariate cointegration analysis. *Energy Economics*, vol. 25:225-238.

Larue B. (1991). Farm input, farm output and retail food prices. *Canadian Journal of Agricultural Economics,* vol.*39*: 335-353.

Masih A. and Mashi R. (1996). Energy consumption and real income temporal causality, results for a multi-country study based on cointegration and error correction techniques. *Energy Economics*, vol. 1:165-183.

McPhail L. (2011). Assessing the impact of US ethanol on fossil fuel markets: A structural VAR Approach. *Energy Economics*.

Mills T. (1999). *The Econometric Analysis of Financial Time Series*, Cambridge University Press, Cambridge.

Mishra R. and Kumar A. (2013). The spatial integration of vegetable markets in Nepal. *Asian Journal of Agriculture and Development*.

Moss C. (1992). The Cost-Price Squeeze in Agriculture: An Application of Cointegration. *Review of Agricultural Economics.*

Niyimbanira F. (2013). An overview of methods for testing short and long run equilibrium with time series data: Cointegration and error correction mechanisms. *Mediterranean Journal of Social Sciences*.

Phillips P. and Oualiaris S. (1998). Testing for cointegration using principal component methods. *Economics Dynamic and Control*, vol. 12:205-230.

Pokhrel D. (2010). Comparison of farm production and marketing cost and benefit among selected vegetable pockets in Nepal. *The Journal of Agriculture and Environment.*

Rudra B, Wen-Chi H. and Raju G. (2012). Market Price Cointegration of Tomato: Effects to Nepalese Farmers. *International Journal of Food and Agricultural Economics,* Vol 2 No. 2.

Sanogo I. and Maliki A. (2010). Rice market integration and food security in Nepal: The role of cross-border trade with India. *Food Policy.*

Serra T, Zilberman D, Gil J. and Goodwin, B. (2011). Nonlinearities in the U.S. corn-ethanol oil-gasoline price system. *Agricultural Economics*.

Shrestha R. (2012). Factors affecting price spread of rice in Nepal. *The Journal of Agrciculture and Environment*.

Silvennoinen A. and Thorp T. (2010). Modelling conditional correlations of asset returns: A smooth transition approach. *CREATES Research Paper 2012-09,* Arhus University.

Tsay R. (2001). Testing and Modelling Threshold Autoregressive Processes. *Journal of the American Statistical Association*, vol. 84.

World Bank Report: Available at: *http://data.worldbank.org/indicator/SI.POV.GINI*

Yang H. (2000). A note on casual relationship between energy and GNP in Taiwan. *Energy Economics*, vol. 22:309-317.

Appendix I: Data Organization

Daily price data is usually collected by extension officers in the Ministry for Agriculture, Livestock and Fisheries; Agribusiness and Market Development Department on the various products on sale from the major markets in Kenya. The data is usually collected in terms of Kenya shillings per standard unit of sale which could be a Kilogram, a 90 Kilograms bag among other units of sale.

Such data was acquired with permission from the relevant authorities covering Nairobi county markets for a period of five years starting 1st January 2010 to 31st December 2014. The data was sieved to remove the non-agricultural products as well as the inconsistent agricultural food products. This yielded the twenty-five agricultural food products considered in the study. The twenty-five food products were then classified into three categories; carbohydrate foods, protein foods and vitamin and mineral foods (see Appendix II).

The prices collected were all converted into Kenya shillings per Kilogram in order to ensure consistency of the units of measurement. The mean unit price per month was then computed as a summary measure for each product. This yielded a time series of sixty data points for each of the twenty-five food products. Finally, monthly mean prices were computed for the food categories reducing the data from a set of twenty-five price time series to a set of three price time series. This is the data that was used in the study.

Appendix II: Agricultural food products Considered

Category A: Carbohydrate foods

1. Dry Maize
2. Green Maize
3. Finger Millet
4. Sorghum
5. Red Irish Potatoes
6. White Irish Potatoes
7. Sweet Potatoes
8. Wheat

Category B: Protein foods

1. Beans Rosecoco
2. Beans Mwitemania
3. Green Gram
4. Cowpeas
5. Fresh Peas
6. Beans Dolichos *(Njahi)*
7. *Mwezi Moja*
8. Groundnuts

Category C: Vitamin and Mineral foods

1. Cabbages
2. Kales
3. Carrots
4. Tomatoes
5. Ripe Bananas
6. Passion Fruits
7. Oranges
8. Mangoes
9. Pineapples