PRICE ANALYSIS OF THE NUTS MARKET IN THE UNITED STATES

A Thesis

by

BEMBYA BEMBEEV

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Senarath Dharmasena
Co-Chair of Committee,	Marco Palma
Committee Members,	Manuel Pina Jr.
Head of Department,	David J. Leatham

May 2021

Major Subject: Agricultural Economics

ABSTRACT

PRICE ANALYSIS OF THE NUTS MARKET IN THE UNITED STATES

Bembya Bembeev
Department of Agricultural Economics
Texas A&M University

Research Advisor: Dr. Senarath Dharmasena
Department of Agricultural Economics
Texas A&M University

The purpose of this study is to determine a price model with open-source government data using monthly data for nut products in the United States. Nut products are used for direct consumption as well as intermediary goods for other products. Therefore, determination of a model for these highly volatile nut products is important for various reasons. Particularly, this research plans to analyze peanuts, almonds, walnut, pecans with a further plan to forecast prices. There are several ways this research attempted to forecast with ARMA, ARIMA, ECM. In addition, this paper is trying to develop a graphical causality model of producer level prices of aforementioned nut products to uncover the causality pattern in the nuts market using cutting-edge machine learning approaches (directed acyclic graphs).

CONTRIBUTORS

This work is supervised by a thesis dissertation committee consisting of the Professor Dharmasena and Professor Palma of the department of agricultural economics and Professor Pina of the department of agricultural leadership, education and communications.

Special thanks to expertise and insight of Dr. Dharmasena.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

Page

CHAPTER 1

1.1 **Problem Statement**

Tree nuts industry is experiencing higher demands than ever before. One the primary reasons for demand uptick are: increase of population and public awareness of tree nuts health benefits. Partially or mainly due to these causes, prices for tree nuts became volatile. Early and mid of 20s century tree nuts fluctuations were barely noticeable and could be predicted using rudimentary forecasting techniques. While for the past 30 years we can see sporadic changes caused by modern trends making it impossible to use previous forecasting techniques.

1.2 **Objective**

Tree nuts are widely used product, thus unpredictable price behavior of tree nuts may adversely affect producers and consumers. Forecasting future values as well as finding connection between the markets and causal structure may help determine our expectation about potential of these markets for manufacturers and clients. In addition to my main objective, no research has been concentrated on forecast for the past couple of years.

In this thesis, we consider monthly prices of Peanut, Walnut, Pecan and Almond over the period December 1999 through March 2021. This study aims to forecast peanuts, almonds, walnut, pecans and determine causality between these products. Initial attempts to forecast these products were created with help of autoregressive integrated moving average (ARIMA) model different ARIMA to each data set. However, during the research process better forecast methods were found. Error Correction Models (ECM) are specifically adapted for multivariate time series. Dynamic linear regression model (DLR) or vector autoregressions (VARs), error correction models(ECMs) are a powerful and practical way of defining the dynamic multivariate interactions characteristic of economic data. (Alogoskoufis, George & Smith, Ron. (1991). On Error Correction Models: Specification, Interpretation, Estimation. Journal of Economic Surveys). In other words, ECM is well-suited for multivariate data while ARIMA for univariate. Modelling ECM is beneficial in our case since we capable to capture dependency between series and adjust them accordingly. Therefore, this study focuses on using VARs modelling instead of ARIMA.

Initially research attempted to use VECM (Vector Error Correction Model), however, due to infeasibility of VECM research shifted to VAR model.

The main objective is to determine a proper model to forecast CPI of peanuts, almonds, walnut, pecans using VAR and determine causality between these tree nuts products.

CHAPTER 2

REVIEW OF TREE NUTS DATA AND LITERATURE

2.1 **US Tree Nuts Facts Overview**

It is a proven that nuts have most of the vital components for a healthy living of humans. A majority of medical scientist including Dr. Emilio Ros in his research article "Health Benefits of Nut Consumption" advocated for tree nut consumption due to their beneficial impact on health outcomes and provided several refences to other well-known organizations in his research pointing out validity of his view. Understanding of health benefits boosts the demand for nut products. According to USDA, (USDA, Tree Nuts 2019 Export Highlights) for the last five decades annual consumption has increased by 2.31 pound per person from 1970 to 2016. Among those tree nuts almond has experienced the largest growth by 1.35 pound per person. The U.S. tree nut industry showed a steady growth which might be affected by the global demand for tree nuts. The report of USDA as well includes that even though US market has a competitive advantage in tree nut production it may face some challenges from World Trade Organization (WTO) or regulations from importing countries (USDA, Tree Nuts 2019 Export Highlights).

2.2 **Tree Nuts Profile**

The Almond (*Prunus dulcis*) is a tree nut native to Iran and surrounding countries. It is view as one the nutritious product therefore it is widely used for direct consumption as well as intermediary good for other products. Almonds mostly produced in United

States, Spain, Iran Morocco, Turkey, Italy, Australia. It grows only in a certain microclimate with warm and dry summer and mild and wet winters (Mediterranean climate). In US it grows mostly in California (around 80%) but also Arizona, Georgia, Nevada, New Mexico, Utah, Texas and many other states.

The Peanut (*Arachis hypogaea)* native to tropics and subtropics. This crop has two classification as both grain legume and as oil crop due to high oil content. It has similar nutritional values as almond and walnut. Grows in three major regions in US: Southeastern region(Alabama, Georgia, Florida), Southwestern region(New Mexico, Oklahoma, Texas) and Eastern (Virginia, North Carolina, South Carolina). From all states above, Georgia and Alabama are leaders in peanut production.

The Pecan (*Carya illinoinensis*) has many variations of pronunciation, but the most common is /pɪˈkɑːn/. Pecan is highly susceptible to a wide range of diseases, pests, and physiological disorders that can hamper their growth and efficiency. Pecans has few research paper confirming their positive influence, thus further research is needed to provide any further information regarding their nutritional benefits to human bodies. Primary growers in US are Texas, Georgia, and New Mexico.

The Walnut (*Juglans*) is native to Iran. It is confirmed that it has a higher saturation compared with other tree nuts of monounsaturated fatty acids. It is not proven that Walnuts have positive effect on health, due to lack of medical research in this field it is still controversial whether walnuts do indeed beneficially affect our bodies and have imputed to them healing properties. Walnut mostly produced in California. Not to be confused with black walnuts which originally from eastern North America.

2.3 **Data**

All 4 datasets ending in November 2020 were taken from publicly accessible websites of Federal Reserve Economic Data (FRED) and United States Department of Agricultural Statistics Service (USDA). Almond data is available only from 1991 while Peanut has a record number of observations from all tree nut section of USDA, starting from 1947. Pecans has start date from 1999 and Walnuts one year earlier from 1998. Data contains monthly producer price index of nuts products. Preliminary analysis of the data shows that data of peanut, almond, walnut and pecans are non-stationary according to Dickey-Fuller Test (ADF). Data shows no seasonality leading to an assumption that all fluctuations have no seasonal pattern. Unit Root test procedure as well supports ADF test assumption and suggests that 1 differencing is required to make this data stationary.

Table 2.3 provides descriptive statistics of MSTL procedure for monthly aggregated data level data (265 months from 1999 to 2020). The price of different tree nut product varies, and we can notice from the table 2.3 that peanuts is the lowest in terms of price and pecan is the highest. The most purchased tree nut product is peanut, followed by almond (Cheng, 2017).

Table 2.1 Descriptive Statistic of MSTL

Peanut		Trend		Seasonal12		Remainder	
Min:	40.1	Min:	75.02	Min:	-7.26338	Min:	-39.8411
Median:	116	Median:	115.22	Median:	-0.15047	Median:	-0.25244
Mean:	116.9	Mean:	116.84	Mean:	0.01159	Mean:	0.00412
Max:	216.2	Max:	169.49	Max:	8.34914	Max:	42.95378
Almond		Trend		Seasonal12		Remainder	
Min:	93.3	Min:	100.9	Min:	-10.6377	Min:	-69.0563
Median:	182.5	Median:	192.4	Median:	-0.17743	Median:	0.979
Mean:	217	Mean:	216.9	Mean:	-0.02759	Mean:	0.1044
Max:	463.6	Max:	429.9	Max:	11.32488	Max:	60.8757
Pecan		Trend		Seasonal12		Remainder	
Min:	225.8	Min:	256.8	Min:	-22.6475	Min:	-146.313
Median:	638.2	Median:	604.7	Median:	-4.64561	Median:	-0.9774
Mean:	621.5	Mean:	621.8	Mean:	0.00623	Mean:	-0.3244
Max:	956.4	Max:	903.3	Max:	56.77532	Max:	211.5288
Walnut		Trend		Seasonal12		Remainder	
Min:	106.6	Min:	109.7	Min:	-9.6433	Min:	-51.0995
Median:	2O1.5	Median:	214.4	Median:	0.04854	Median:	-0.84247
Mean:	208	Mean:	207.8	Mean:	0.10693	Mean:	0.07772
Max:	365.2	Max:	350.5	Max:	16.29035	Max:	47.30291

2.4 **Literature Review**

In extant literature, a few research materials could be found on forecasting tree nuts showing a great potential for a further research. One of the first reports found during the analysis of the tree nut industry was article “Pecan Production and Price Trends”. In this academic work there was an attempted to forecast price by determining Demand and Supply equations. (Pecan production and price trends, Carl Shaffer, 1996). After that another effort was done in “Forecasting Price Relationships among U.S Tree Nuts Prices”, where forecaster determined the model using ARIMA procedure. (Mohammed Ibrahim and Wojciech J. Florkowski, 2009). The authors discussed several approaches to determine cointegration, they started with a technique developed by Engle and Granger (1987), but ultimately decide to stop on Johansen Cointegration Procedure (Johansen, 1988; Johansen and Juselius, 1990; Luppold and Prestemon, 2003) since it is more powerful. This feature was noted in presented research process and according to Eigen test there is no cointegration in our data set, this result was supported by Johannsen test as well.

Most of previous work did not set a goal to forecast tree nuts products therefore it is harder to find such material. Although quite a few papers share a common objective such as elasticity of demand or own-price elasticity. In “Short Run and Long Run dynamics in the Demand of U.S Tree Nuts” study Dr. Awondo estimated two types of models static and dynamic Almost Ideal Demand System (AIDS) to investigate the long run and in the short run behavior of U.S. consumers. He explains that Almond and

Pecans are more sensitive to short run own price change than long run while the opposite effect is happening in Walnuts. (Sebastain N. Awondo, Esendugue Greg Fonsah, 2014).

A similar work was done before by Dr. Guo Cheng who made a deep review of demand of tree nuts products and found that 10% of the price is not sensitive while the rest 90% is affected by various factors which are described in his paper. (Guo Cheng, Senarath Dharmasena, and Oral Capps, Jr., 2017). In his conclusion he stated that consumer can easily substitute tree nut products leading to a problem in this research that modelling solely of historical price is not a proper way to get the results.

Thus, an important component missing in this approach is Causal Inference. The Causal Inference based on the Structural Causal Model (SCM) and was developed (Judea Pearl,1995,2000) in economics and social science (Goldberger,1973; Duncan,1975). Despite slightly complicated nature the Causal Model simply provides us knowledge whether a certain market affected by another market. In the article of Dr. Guo Cheng found that almond, pecans and cashew are substitutes, however he did not go further by implementing a causality modelling as well as Dr. Mohammed Ibrahim and Dr. Wojciech J. Florkowski.

An introduction of the world to Structural Causal Model (SCM) happened in 1990s (Pearl,1995) and was adopted by healthcare workers (Greenland et al.,1999; Glymour and Greenland,2008), yet its full potential was not utilized (Pearl, 2009). Pearl in his work “Causal inference in statistics” highlighted versatility of SCM and laid out at least 6 main ramifications of Causal Modelling. In this paper an attempt was made to create a Causal Model, results indicated slight dependency. Computing forecast error variance decomposition showed that almond explain his behavior by 99.00383% while the rest is explained by peanuts, pecans, walnuts. Peanuts had almost the same results 99.21364%. At the same time a different story unfolds in pecans and walnuts datasets. Pecan’s variance was explained by pecans by 99.2%, by almonds 3.680528%, peanut 2.028304% and walnut 0.01758713% a quite high number compared to previous results. A slightly different situation happened in walnut dataset. Walnut variance explained itself by 95.07036%, almond 0.37920432%, peanut 3.7200072%, pecans 0.8304314%.

CHAPTER 3

METHODOLOGY

3.1 **Stationarity**

In a research process many methods of forecasting were discussed. Initially this plan of this research was considered to include only ARIMA methodology to forecast. ARIMA process ability to predict future values based on previous values. This method is widely used by many economists and even central banks (George Athanasopoulos and Rob J. Hyndman, 2013). A second considered option to forecast was Exponential Smoothing. Exponential smoothing was proposed in late 50s of the 20 century (Brown, 1959; Holt, 1957; Winters, 1960) and showed a great promise.

As a first step to start working on data is make stationary. A property of stationary data is that mean and auto covariance of the series do not depend on time.

$$Y_t = Y_{t-1} + e_t$$

where, e_t is error term. Subscript t indicates time; therefore, we can say that Y_t is a constant forecast value conditional on t. Following the formula to find current Y_t we need to know previous value Y_{t-1} add error e_t to find exact value of current period.

To determine whether our time series reached stationarity we apply several tests. The most used and well-known test is Dickey-Fuller test or DF test (Dickey & Fuller, 1979) or another widely used testing is Phillips-Perron test (PP) (Phillips & Perron, 1998). These tests recognized as unit root tests for a series to be stationary. However, in our work we are planning to use DF test only.

If a differenced time series is integrated, we can denote this series by writing I(d), where d is order number of integrations of the series. So, to show a non-stationary series we can write I(1) while a series considered to be stationary if it has I(0).

Dickey-Fuller (DF) test regress $\Delta X_t = a_0 + a_1 X_{t-1}$. Our in DF test null hypothesis is that our series is non-stationary or series I(1) and alternative hypothesis is I(0) series or stationary.

3.2 **Autoregressive Integrated Moving Average (ARIMA)**

Since the data of tree nuts from USDA is non-stationary, we need to difference it until we get a stationarity of time series. This stationary data is then used to predict price by utilizing autoregressive integrated moving average **(**ARIMA) model and autoregressive moving average (ARMA) model.

Autoregressive moving average (ARMA) model can be represented as

$$\left(X_t - a_1 X_{t-1} - a_2 X_{t-2} - \cdots a_p X_{p-t}\right) = \varepsilon_t - \theta\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \cdots \theta_q \varepsilon_{t-q}$$

Or similarly we can write:

$$(1 - \sum_{i=0}^{p} a_i L^i) X_t = (1 - \sum_{i=0}^{p} \theta_i L^i)\varepsilon_t$$

where L is lag operator, a_i is parameter of autoregressive model, $\theta\varepsilon_t$ parameters of the moving average model and ε_t is error term.

ARIMA model can be written as

$$(1 - \sum_{i=0}^{p} a_p L^p)(1 - L)^d X_t = (1 - \sum_{i=0}^{p} \theta_q L^q)\varepsilon_t$$

The only difference of ARIMA with ARMA is that we are using integration here which is denoted here as $(1 - L)^d$

Using the above ARIMA model, it is possible to forecast future price. Based on the above formula this research was planning to analyze behavior of tree nuts prices.

3.3 **Exponential Smoothing**

An alternative way to forecast is Exponential Smoothing, this method is quite straight forward. The main principles behind this method are weighted averages of past observations, where weight of more recent observations has more impact on forecast then old observations. The formula below show how exponential smoothing method works.

$$s_0 = x_0$$

$$s_t = ax_{t-1} + (1-a)s_{t-1}$$

s_t – exponential smoothing forecast

a – is smoothing constant. (smoothing constant can take values between $0 < a < 1$)

x_{t-1} – is previous period of observation

s_{t-1} – is previous period of forecast

3.4 **Error Correction Model**

The course of action changed since discovery of ECM model. In the book "Time Series Analysis", Hamilton describes VAR modelling methodology. The formula below c is a constant, ϕ is a parameter and y is a value conditional on t. Since Vector AR is closely related to matrixes, It is necessary to mention that c is vector (n×1) of constant, ϕ_j is (n×n) matrix of autoregressive coefficients for j=1, 2.. p and the last term here is error e_t which has the (n×1) vector. (Hamilton, 1994)

An Autoregressive model could be represented as

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} \ldots + \phi_p y_{t-p} + e_t,$$

where E(e_t)=0

E(e_t, e_τ)=$\begin{cases} \sigma^2 \; for \; t = \tau \\ 0 \; otherwise \end{cases}$

While a Vector Autoregressive model of a p-lag order can be shown as below, where we regress two variables X_t and Y_t (k=2).

$$Y_t = B_{10} + B_{11} y_{t-1} + \cdots + B_{1p} y_{t-p} + \gamma_{11} X_{t-1} + \cdots + \gamma_{1p} X_{t-p} + e_{1t}$$

$$X_t = B_{20} + B_{21} y_{t-1} + \cdots + B_{2p} y_{t-p} + \gamma_{21} X_{t-1} + \cdots + \gamma_{2p} X_{t-p} + e_{2t}$$

The constant of the above equations B and γ parameters are estimated using Ordinary Least Square method (OLS). *(Hanck, Arnold, Gerber, and Schmelzer, 2020)*

One the ways to determine an optimal lag length is information criteria BIC. As in a single equation model we follow the same principle for multiple equation model such as written above VAR(p) by choosing a smallest BIC(p) result. Calculation of BIC should proceed as the following equation:

$$BIC(p) = log\left[\det\left(\sum\nolimits_i\right)\right] + k(kp+1)\frac{\log(T)}{T}$$

($\sum_i$) is a (k×k) covariance matrix of the VAR errors. An important condition about covariance is that ($\sum_i$) must be positive.

Another way econometrician chooses the right model is Akaike information criterion (AIC).

$$AIC(p) = log\left|\left(\sum\nolimits_i\right)\right| + \frac{2k^2 i}{T}$$

3.5 **Decomposition**

There are several decomposition methods such as Classical (includes additive decomposition and multiplicative decomposition), Seasonal Extraction in ARIMA Time Series (SEATS), X11(US Census Bureau and Statistics Canada), Seasonal and Trend decomposition using Loess (STL). The decomposition method which was used here was STL(advanced version of STL is called MSTL) is based on additive and multiplicative decomposition.

Additive decomposition includes four steps:

1) Compute T (trend) using the following formula: m (seasonal period) is an even number, it is possible to compute trend-cycle component $T = 2 * m - MA$ where m is seasonal period, MA is moving average.

   If m is odd number, then formula changes a bit to $T = m - MA$

2) $y_t - T_t$ detrending series. Where y is historical values and T is trend

3) To determine seasonality for each season component, we can just average the detrend values for that season. This will give us S_t (season)

4) Remainder computed using this formula $R_t = y_t - T_t - S_t$ or actual values subtracting trend and seasons

Multiplicative decomposition also includes four steps:

1) The same first step as in additive formula
2) Detrend data with $\frac{y_t}{T_t}$
3) Same step as in additive decomposition
4) Remainder had slightly different formula than in additive decomposition

$$R_t = \frac{y_t}{T_t S_t}$$

3.6 **Directed Acyclic Graphs**

Directed Acyclic Graphs (DAG) are a well-known approach for identifying causal assumptions and inform variable selection. DAG involves variance-covariance matrix from a set of variables. The algorithm used for this research was Fast Greedy Search (FGES). The following material is derived from Center for Causal Discovery (CCD). This algorithm receives as input a set of data of continuous variables, thoroughly searches over selected causal Bayesian network structures, and outputs the highest scoring model it finds. This model is designed to help researchers form a hypothesis for testing in their work.

FGES algorithm has the following requirements for data:

1) Data should be formatted in tables where columns represent variable and rows observation, besides that each variable in a sample is continious.
2) No missing values are presented in a table.

3) First row with names of the variables should be unique and placed in order. Data and variables names should be separated by a delimiter.
4) No Linear dependence in the data.
5) No variables that have zero covariance.

FGES has also recommendations for parameters, which will may alter behavior of this algorithm.

1) Penalty discount – specification of a complexity penalty parameter c that is presented in the BIC equation.
2) Depth specifies maximum number of edges into a node. Small depth value tends to decrease time required for search.
3) Disable heuristic speedup FGES by default applies a heuristic speedup.
4) Knowledge settings specify prior knowledge about the causal graph structure
5) Thread executes by default the algorithm will run in a
6) parallel fashion using as many threads as are needed and available on the system.
7) Graphml - is graphical description of causal and search path.

FGES outputs most probable CBN structure it finds, according to BIC measure. This algorithm detects causal relationship. An arrow → represents a direct causal while undirected edges such as a line – indicated that there is a causal arc, but direction cannot be determined.

FGES algorithm also has assumptions which should be satisfied:

1) samples in the data are independent and identically distributed.

2) Causal Markov condition holds. This condition variable is independent of its non-effects, given its direct causes (parents).

3) Causal faithfulness condition holds.

4) The causal faithfulness condition holds (Spirtes, 2010).

5) there are no missing data.

6) there are no hidden confounders of the measured variables. That is, none of the measured variables have a common hidden variable as one of their direct causes (relative to the variable set).

7) there is no selection bias.

8) there are no feedback cycles among the measured variables.

3.7 **Impulse Response Functions**

An impulse response functions (IRF) were also applied in this work and their main purpose is to how a model's variable acts when a shock happens in other variable or variables of the model.

To make things less complicated, let's consider a Moving Average MA (∞) process:

$$y_t = c + e_t + \phi_1 e_{t-1} + \phi_2 e_{t-2} + \cdots$$

The matrix ϕ_x could be interpreted as

$$\frac{\partial y_{t+s}}{\partial e_t} = \phi_x;$$

According to Hamilton, the row I, column j element of ϕ_x defines the outcome of a one unit increase in the j variable's innovation at date t (e_{jt}) for the value of the i variable at the time t+s ($y_{i,t+s}$), holding all other innovations at all date constants (Hamilton, 1994). If we assume that every element of innovation e_t is changed by δ then we can write this equation:

$$\nabla y_{t+s} = \frac{\partial y_{t+s}}{\partial e_{1t}}\delta_1 + \frac{\partial y_{t+s}}{\partial e_{2t}}\delta_2 + \cdots + \frac{\partial y_{t+s}}{\partial e_{nt}}\delta_{tt} = \phi_s\delta,$$

Where δ = $(\delta_1, \delta_2 \ldots \delta_n)$

Hamilton also discussed that one the fastest way to determine dynamic multipliers numerically is by simulation. To conduct the simulations, we can set $y_{t-1} = y_{t-2} = \cdots = y_{t-p} = 0.$ We also need to set innovation term $(e_{jt} = 1)$ and all other term of e_j to zero. Value of y_{t+s} at date t+s of our simulation should match to out j column of the matrix ϕ_s. By implementing a separate simulation for impulses to each of the innovations j=1,2, …n, all other columns of ϕ_s can be computed. (Hamilton, 1994)

$$\frac{\partial y_{i,t+s}}{\partial e_{jt}}$$

The equation above is called impulse response function.

3.8 **Granger Causality**

Causality seeks a causal relationship between two variables. To implement this test first we need to state our hypotheses. For example, variable y does not granger-causes variable x. A second step would be choosing lags (explain what lag is?)

In addition, after we choose our lag, we need to find F-value.

$$y(t) = \sum_{i=1}^{\infty} a_i\, y(t-i) + c_1 + v_1(t)$$

$$y(t) = \sum_{i=1}^{\infty} a_i\, y(t-i) + \sum_{j=1}^{\infty} B_j\, x(t-j) + c_2 + v_2(t)$$

Considering these two equations for Granger Causality. Restricted top and Unrestricted bottom.

Similarly, we test these equations to see if y variable granger causes x variable.

$$x(t) = \sum_{i=1}^{\infty} a_i\, x(t-i) + c_1 + u_1(t)$$

$$x(t) = \sum_{i=1}^{\infty} a_i\, x(t-i) + \sum_{j=1}^{\infty} B_j\, y(t-j) + c_2 + u_2(t)$$

After this calculation we can at last find f-statistic using the formula below:

$$F = \frac{ESS_R - ESS_{UR}}{q} \times \frac{n-k}{ESS_{UR}}$$

After all the above calculations depending on the result we can reject or accept our null hypothesis. (Glen, 2016)

3.9 **Johansen Cointegration**

Johansen procedure is useful in determining if three or more series are cointegrated. In order to apply cointegration method, the series must be cointegrated. In 1988, Johansen introduced the following trace test statistics:

$$\lambda_{trace}(r) = -T\sum_{j=t+1}^{n} \log\,(1-\widehat{\lambda_j})\,\lambda$$

where T is the number of observations.

The series are not cointegrated if r is equal to zero, otherwise Vector Error Correction model can be applied for the observed data.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 **Initial Analysis**

In table 2.3 data section we have discussed properties of data using trend, season, remainder and data itself which provided descriptive statics of mean, median min and max.

Based on the analysis provide in the table 2.3, Pecan has the highest CPI mean as well as max and min value for that. Seasonal STL decomposition also shows high value for min and max which can be translated that Pecan data is highly seasonal but considering remainder part (unexplained by mstl) we can say that Pecan data is highly fluctuating data. In addition to that, Pecan is the most susceptible to diseases nut from all other observed tree nuts in this paper. It could explain higher CPI value compare with other researched nut products. Challenges of growers are reflected on the price of Pecan which in their turn affecting CPI. Almond has the second place after Pecan in term of volatility. Almond min and max value almost identical with Walnuts in raw data and seasonal sections. US is the world leader in production of Almond meaning the US is a price leader without any doubts. This could be one of the reasons of low seasonality changes from -10 to 11, which is identical to Peanuts. Peanuts is the cheapest nut observed has low seasonality and remainder values making him easier for forecasting purposes. Trend values also vary significantly lower than other products. Lastly, Walnuts also shows us well explained (low remainder) and relatively unseasonal data. However, trend values are still quite high.

Rule of any forecaster is to choose least complicated model if those simple models can provide similar results. The main reason to do so is that over time simple model can be adjusted relatively easily than more complicated one. Therefore, we started from a very simple models such as mean, naïve and seasonal naïve. We applied a Naïve method which could not provide statistically significant results (MAPE values are going to infinity). Similar was happening to mean and seasonal naïve methods. All these simple models failed to do any reasonable forecast. A step up from these models is exponential smoothing. This method was developed in late 1950s. The principle of exponential smoothing is weighted averages of past values where recent past observation has a higher weight than more older observations. This method also failed due to low statistical values.

A more advanced forecasting method is Autoregressive Integrated Moving Average (ARIMA). An initial step was to see whether data has any trend or seasonal components. There are several decomposition methods such as Classical (includes additive decomposition and multiplicative decomposition), Seasonal Extraction in ARIMA Time Series (SEATS), X11(US Census Bureau and Statistics Canada), Seasonal and Trend decomposition using Loess (STL). From all these methods STL or MSTL (enhanced version of STL) performed better. MSTL found that seasonality is presented in all four datasets but varied. For almonds seasonality had a clear pattern from 1990s to 2000s. From 2000s to 2010s seasonality is not clearly visible and it could be the reason why statistical software failed to determine seasonality. From 2010s we can see seasonal pattern again. There is also a clear upward trend with small fluctuation in the end. For peanuts, seasonality has various intensity, especially in 1980s where trend and seasonal component go up. For the rest of Peanut set there are some changes in every decade from 1960s to 2020s, but they all overshadowed by 1980s events. MSTL decomposition of Pecans has not constant seasonal pattern. Upward trend is present throughout the data. Remainder of MSTL decomposition is also quite big meaning that software failed to extract seasonality and trend. MSTL of Walnut also has big remainder highlighting the same issue with walnuts. Seasonality in Walnut data is absent in the first two quarters of the data. The last two has excessively high values, same information is reflected in trend section as well. Decomposition has shown that remainder in MSTL is quite high therefore we cannot heavily rely on it. However, there we noticed that seasonality indeed presented and needed to be accounted. In methodology section we have discussed stationarity and its importance for VAR. Almost the same importance stationarity plays in ARIMA model. We need to remove all non-constant seasonality as well trend patterns which will hamper or make it impossible to determine valid ARIMA model. A Dickey Fuller test was conducted for all 4 series of Almond, Pecan, Peanut and Walnut. The null hypothesis of DF test was that our series are non-stationary. The setting of the test at 5% significance with critical level -2.89. After first differencing data was tested with DF and ADF test which successfully rejected null hypothesis. After this ARIMA modelling is possible. One of the accelerated ways to test multiple ARIMA models of different ordering is ARIMA package in "R" which capable to test all possible combinations of ARIMA and find optimal results based AIC and BIC criteria. So, running that functions provide us an output of ARIMA (0.1.1) and ARIMA (1.0.1) for almond dataset. Analysis of residuals of those ARIMA functions showed us that residuals are centered around zero which is good and ACF plot tells us no lags cross significance level. Forecast based on these statistical results will probably be good. ARIMA combinations for pecans, walnuts and peanuts have similar properties. However, need to account for causality altered the course of actions. ARIMA is a great tool for forecasting time-series data, but we have a panel data which contradict principles of ARIMA forecasting. We could forecast splitting panel data in time series data but in this case we would not be able to see any connections between the data, thus ARIMA forecasting attempts were abandoned. Causality indeed plays a role in price formation. Later in this section, we did an Error Decomposition of Residuals of VAR model which told us that there is some causality hidden in errors. Thus, choosing this method was incorrect due to fact that ARIMA does not account for causality between these price series. A more advanced method of forecasting is VECM (Vector Error Correction Model). This method based on cointegration. The VECM model requires at least 1 cointegration between series and to find that we can use Johansen Cointegration Procedure. The cointegration results shown in the Table 4.1. Test type: trace statistic, without linear trend and constant in cointegration. The series show no cointegration meaning that condition for applying VECM model is not satisfied. In other words, there is no long-run relationship between substitutes. The findings are consistent with earlier conclusion of research article “Forecasting Price Relationships among U.S Tree Nuts Prices” (Florkowski, Ibrahim, 2009), but inconsistent with Florkowski and Lai (1997).

Table 4.1 Cointegration Test

	test	10 pct	5 pct	1 pct
r=3	3.28	7.52	9.24	12.97
r=2	12.33	17.85	19.96	24.60
r=1	27.83	32.00	34.91	41.07
r=0	53.15	49.65	53.12	60.16

4.2 **VAR modelling**

Given the information that VECM is not applicable, that leaves only one option to forecast which is VAR model. Akaike Information Criterion (AIC), Schwarz Criterion (SC), Hannan Quinn (HQ) and Akaike's Final Prediction Error Criterion (FPE) all have proposed to choose 1 first. Based on this, Optimal lag for VAR modelling was found at the 1-st lag. Another specification for our model was seasonality. Differencing can remove or reduce seasonality which in case was reduced and the need to account for seasons was required. Plugging in this lag and seasons into our VAR model gives us the following output:

Table 4.2 Almond VAR equation

Almond equation	Estimate	Std. Error	T value	Pr(>|t|)
Almond. L1	-0.15961	0.07060	-2.261	0.0247
Peanut L1	0.13304	0.10983	1.211	0.2270
Pecan L1	-0.02130	0.02483	-0.858	0.3917
Walnut L1	-0.02823	0.06771	-0.417	0.6771
Constant	1.10892	1.05188	1.054	0.2929
Sd1	-1.27038	5.24516	-0.242	0.8088
Sd2	-5.75235	5.24018	-1.098	0.2734
Sd3	-6.88854	5.26222	-1.309	0.1918
Sd4	-3.45010	5.19931	-0.664	0.5076
Sd5	-3.66661	5.16121	-0.710	0.4781
Sd6	-3.56926	5.18674	-0.688	0.4920
Sd7	-3.28118	5.21452	-0.629	0.5298
Sd8	-0.49028	5.16679	-0.095	0.9242
Sd9	-1.69330	5.16574	-0.328	0.7434
Sd10	-4.13231	5.14644	-0.803	0.4228
Sd11	-2.05120	5.23683	-0.392	0.6956

Table 4. 3 Peanut VAR equation

Peanut equation	Estimate	Std. Error	T value	Pr(>|t|)
Almond. L1	-0.004107	0.039910	-0.103	0.918
Peanut L1	-0.301929	0.062090	-4.863	2.12e-06
Pecan L1	-0.010011	0.014034	-0.713	0.476
Walnut L1	-0.034821	0.038276	-0.910	0.364
Constant	-0.074186	0.594636	-0.125	0.901
Sd1	-0.796453	2.965139	0.269	0.788
Sd2	1.592369	2.962323	0.538	0.591
Sd3	-3.569639	2.974783	-1.200	0.231
Sd4	-2.145354	2.939219	-0.730	0.466
Sd5	-0.128634	2.917680	-0.044	0.965
Sd6	1.714771	2.932113	0.585	0.559
Sd7	-0.084186	2.947821	-0.029	0.977
Sd8	-4.461329	2.920839	-1.527	0.128
Sd9	-3.743711	2.920241	-1.282	0.201
Sd10	3.979564	2.909332	1.368	0.173
Sd11	-0.753976	2.960431	-0.255	0.799

Table 4.4 Pecan VAR equation

Pecan equation	Estimate	Std. Error	T value	Pr(>|t|)
Almond. L1	-0.39127	0.19992	-1.957	0.05152
Peanut L1	0.17102	0.31103	0.550	0.58294
Pecan L1	-0.19738	0.07030	-2.808	0.00541
Walnut L1	-0.03913	0.19174	-0.204	0.83848
Constant	2.03076	2.97870	0.682	0.49606
Sd1	18.77733	14.85324	1.264	0.20741
Sd2	-4.18609	14.83913	-0.282	0.77812
Sd3	1.65540	14.90155	0.111	0.91164
Sd4	8.31717	14.72340	0.565	0.57268
Sd5	7.20757	14.61550	0.493	0.62237
Sd6	-0.33689	14.68780	-0.023	0.98172
Sd7	6.98635	14.76649	0.473	0.63657
Sd8	8.13471	14.63132	0.556	0.57875
Sd9	14.75821	14.62833	1.009	0.31407
Sd10	8.63262	14.57368	0.592	0.55419
Sd11	33.61680	14.82965	2.267	0.02431

Table 4.5 Walnut VAR equation

Walnut equation	Estimate	Std. Error	T value	Pr(>|t|)
Almond. L1	-0.04506	0.06753	-0.667	0.50528
Peanut L1	0.31443	0.10506	2.993	0.00306
Pecan L1	-0.01281	0.02375	-0.540	0.58998
Walnut L1	0.03344	0.06476	0.516	0.60615
Constant	0.36599	1.00611	0.364	0.71636
Sd1	-15.16161	5.01695	-3.022	0.00279
Sd2	-2.69171	5.01218	-0.537	0.59175
Sd3	-6.90449	5.03327	-1.372	0.17144
Sd4	-0.02721	4.97309	-0.005	0.99564
Sd5	-4.28493	4.93665	-0.868	0.38629
Sd6	1.14241	4.96107	0.230	0.81808
Sd7	-3.62692	4.98765	-0.727	0.46784
Sd8	-3.26549	4.94199	-0.661	0.50941
Sd9	-0.95668	4.94098	-0.194	0.84664
Sd10	-1.72628	4.92252	-0.351	0.72613
Sd11	-1.76808	5.00898	-0.353	0.72442

4.3 **Forecast Error Variance Decomposition**

At a later stage of the research a forecast error decomposition was done

Table 4.6 Forecast Error Variance Decomposition for Almonds

almo	pean	peca	waln
1	0	0	0
0.992474	0.002166	0.004859	0.000501
0.99162	0.002664	0.005215	0.000501
0.991567	0.002702	0.005229	0.000502
0.991565	0.002704	0.005229	0.000502
0.991565	0.002704	0.005229	0.000502
0.991565	0.002704	0.005229	0.000502
0.991565	0.002704	0.005229	0.000502
0.991565	0.002704	0.005229	0.000502
0.991565	0.002704	0.005229	0.000502
0.991565	0.002704	0.005229	0.000502
0.991565	0.002704	0.005229	0.000502
0.991565	0.002704	0.005229	0.000502
0.991565	0.002704	0.005229	0.000502
0.991565	0.002704	0.005229	0.000502
0.991565	0.002704	0.005229	0.000502
0.991565	0.002704	0.005229	0.000502
0.991565	0.002704	0.005229	0.000502
0.991565	0.002704	0.005229	0.000502
0.991565	0.002704	0.005229	0.000502
0.991565	0.002704	0.005229	0.000502
0.991565	0.002704	0.005229	0.000502
0.991565	0.002704	0.005229	0.000502
0.991565	0.002704	0.005229	0.000502

From table 4.6 the first period is completely explained by almonds. In the second period we can see that almonds are denoted as “almo” is explained by 99.2474%, pecan by 0.4859% denoted as “peca” and pean denoted as “pean” explained by 0.2166% the rest is explained by walnuts. In the last 24th period we can see that almond share slightly decreased compared to the second period since we have 99.1565%. Small adjustment for peanut 0.2704%, for pecan 0.5229% and walnut 0.0502%. Based on the above information we can say that almond prices are mostly independent.

Table 4.7 Forecast Error Variance Decomposition for Peanuts

almo	pean	peca	waln
0.007275	0.992726	0	0
0.00748	0.986975	0.002067	0.003478
0.007471	0.986583	0.002308	0.003638
0.007482	0.986556	0.002317	0.003645
0.007484	0.986554	0.002317	0.003645
0.007484	0.986554	0.002317	0.003645
0.007484	0.986554	0.002317	0.003645
0.007484	0.986554	0.002317	0.003645
0.007484	0.986554	0.002317	0.003645
0.007484	0.986554	0.002317	0.003645
0.007484	0.986554	0.002317	0.003645
0.007484	0.986554	0.002317	0.003645
0.007484	0.986554	0.002317	0.003645
0.007484	0.986554	0.002317	0.003645
0.007484	0.986554	0.002317	0.003645
0.007484	0.986554	0.002317	0.003645
0.007484	0.986554	0.002317	0.003645
0.007484	0.986554	0.002317	0.003645
0.007484	0.986554	0.002317	0.003645
0.007484	0.986554	0.002317	0.003645
0.007484	0.986554	0.002317	0.003645
0.007484	0.986554	0.002317	0.003645
0.007484	0.986554	0.002317	0.003645
0.007484	0.986554	0.002317	0.003645

Computed forecast error variance decomposition for Peanuts in table 4.7 have different results than almond. From the first period peanuts are partially explained by almonds which signify weak causality. We can see that for second period a small fraction is explained by pecans and walnuts which continues till the end of 24th period.

Table 4.8 Forecast Error Variance Decomposition for Pecans

almo	pean	peca	waln
0.004534	0.023731	0.971735	0.00E+00
0.024436	0.023109	0.952454	1.25E-08
0.025824	0.023175	0.950995	6.74E-06
0.025862	0.023196	0.950936	6.77E-06
0.025862	0.023198	0.950934	6.79E-06
0.025862	0.023198	0.950933	6.79E-06
0.025862	0.023198	0.950933	6.79E-06
0.025862	0.023198	0.950933	6.79E-06
0.025862	0.023198	0.950933	6.79E-06
0.025862	0.023198	0.950933	6.79E-06
0.025862	0.023198	0.950933	6.79E-06
0.025862	0.023198	0.950933	6.79E-06
0.025862	0.023198	0.950933	6.79E-06
0.025862	0.023198	0.950933	6.79E-06
0.025862	0.023198	0.950933	6.79E-06
0.025862	0.023198	0.950933	6.79E-06
0.025862	0.023198	0.950933	6.79E-06
0.025862	0.023198	0.950933	6.79E-06
0.025862	0.023198	0.950933	6.79E-06
0.025862	0.023198	0.950933	6.79E-06
0.025862	0.023198	0.950933	6.79E-06
0.025862	0.023198	0.950933	6.79E-06
0.025862	0.023198	0.950933	6.79E-06
0.025862	0.023198	0.950933	6.79E-06

Pecans errors have more connections than previous almond and peanuts errors. Pecans is influenced by peanuts 2.3731% and almond 0.4534% at the first period. At the last 24th period we have that almost 5% coming from non-pecan errors.

Table 4.9 Forecast Error Variance Decomposition for Walnuts

almo	pean	peca	waln
1.86E-07	0.002794	0.020772	0.976434
2.17E-03	0.043334	0.019872	0.934628
2.23E-03	0.045531	0.019961	0.932273
2.24E-03	0.045622	0.019971	0.932172
2.24E-03	0.045625	0.019971	0.932169
2.24E-03	0.045625	0.019971	0.932169
2.24E-03	0.045625	0.019971	0.932169
2.24E-03	0.045625	0.019971	0.932169
2.24E-03	0.045625	0.019971	0.932169
2.24E-03	0.045625	0.019971	0.932169
2.24E-03	0.045625	0.019971	0.932169
2.24E-03	0.045625	0.019971	0.932169
2.24E-03	0.045625	0.019971	0.932169
2.24E-03	0.045625	0.019971	0.932169
2.24E-03	0.045625	0.019971	0.932169
2.24E-03	0.045625	0.019971	0.932169
2.24E-03	0.045625	0.019971	0.932169
2.24E-03	0.045625	0.019971	0.932169
2.24E-03	0.045625	0.019971	0.932169
2.24E-03	0.045625	0.019971	0.932169
2.24E-03	0.045625	0.019971	0.932169
2.24E-03	0.045625	0.019971	0.932169
2.24E-03	0.045625	0.019971	0.932169
2.24E-03	0.045625	0.019971	0.932169

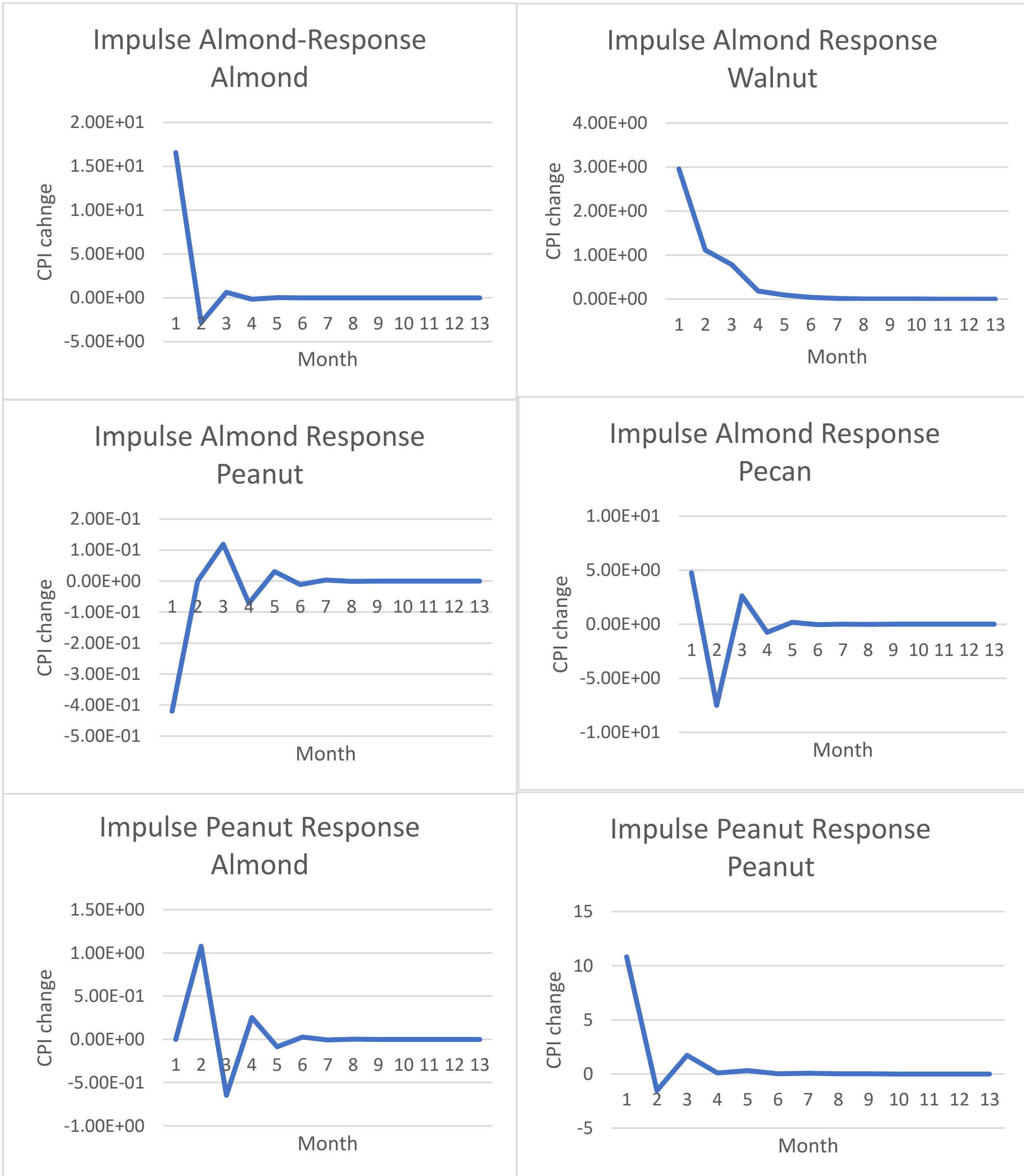
Walnuts are the most dependent from other series based on error decomposition. 2.3566% from non-walnuts in the first period and 6.7831% in the last period. Peanuts impact on Walnuts in the first period had only 0.2794% and closer to the end 4.5625%. Pecans shows an opposite picture from 2.0772% in the first period to 1.9971% in the last period.

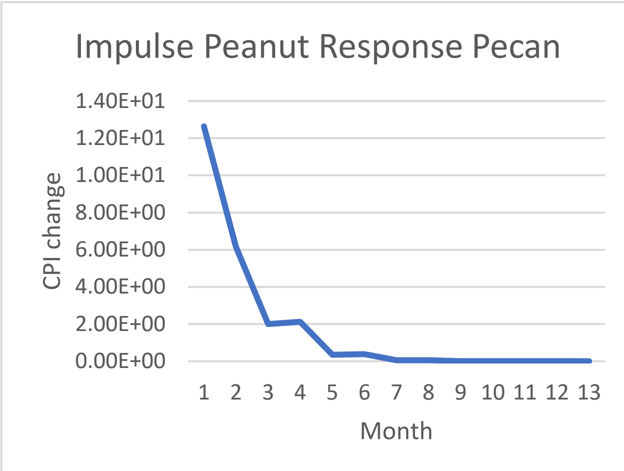
4.4 **Impulse Response Function**

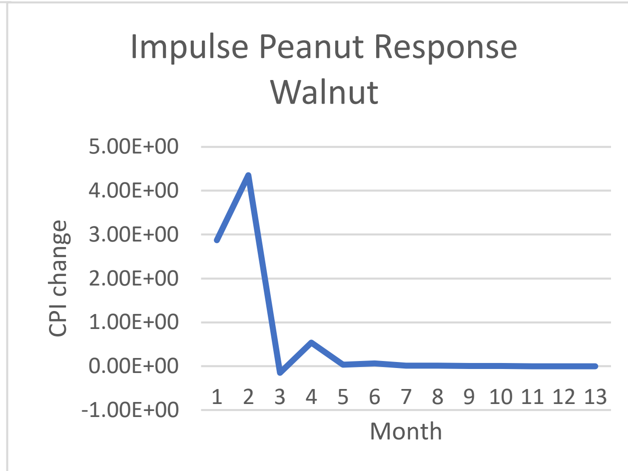
Impulse response functions (IRF) are provided below in the Figure 4. 1 Impulse Response Function. Most of the information provided in IRF support results made in Forecast Error Variance Decomposition. As an example, we can take a look at Almonds impulse to Walnuts response, the graph shows initial shock which gradually declines and dies in 6 or 7 period (months). This information is supported in Walnuts Error Decomposition (Table 4.9). Walnuts errors declines in the first month and stabilizes in the 4 period. We can see identical pattern but different magnitude for peanut and pecans. The shock gradually dies in 6 and 7 period. Another interesting pattern here is pecans shock and almond response as well as pecans shock and Peanut response. The series fluctuate in the beginning which is reflected in forecast error variance decomposition as well. Almond shock to Peanut and Pecans have inverted relations for first period but then mostly similar direction after that. Lastly Walnut interaction with other tree nuts has similar graph figure. Shock leads to negative values in the 2 period and fluctuation to equilibrium in 5^{th} period. All of the tree nuts according to error variance decomposition and IRF are going to equilibrium level at 5^{th} or 6^{th}. Therefore, there is no long run relationship between these nut products. Price market is relatively quick to stabilize themselves.

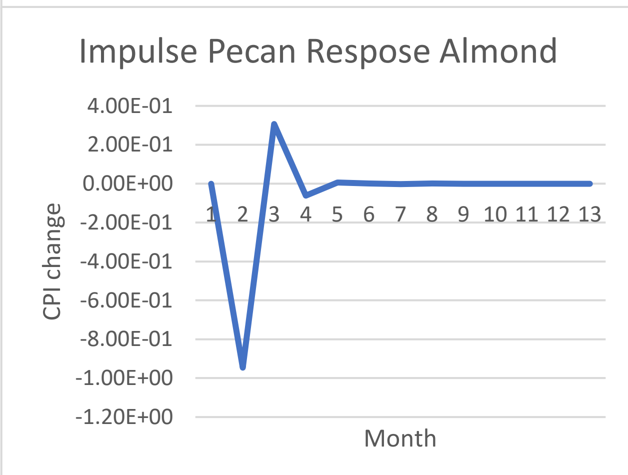
Impulse response function, in my view does not need for further explanation since all the information is conveniently depicted in the figure 4.2. It is easier for reader to look at the pictures than read several paragraphs of explanation of IRF results.

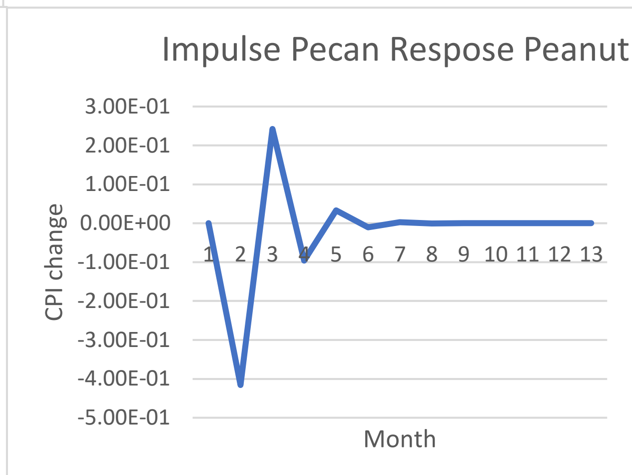
Figure 4. 1 Impulse Response Function

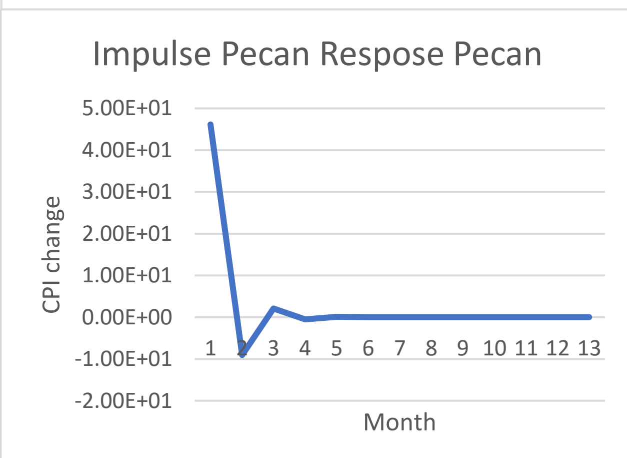


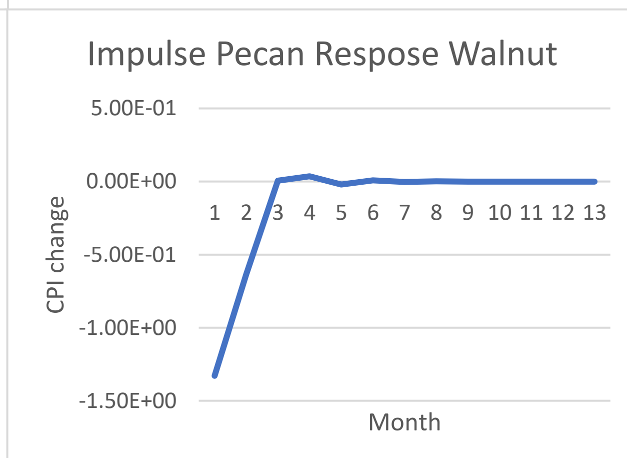

Impulse Peanut Response Pecan
1.40E+01
1.20E+01
1.00E+01
8.00E+00
6.00E+00
4.00E+00
2.00E+00
0.00E+00
CPI change
1 2 3 4 5 6 7 8 9 10 11 12 13
Month


Impulse Peanut Response Walnut
5.00E+00
4.00E+00
3.00E+00
2.00E+00
1.00E+00
0.00E+00
-1.00E+00
CPI change
1 2 3 4 5 6 7 8 9 10 11 12 13
Month


Impulse Pecan Respose Almond
4.00E-01
2.00E-01
0.00E+00
-2.00E-01
-4.00E-01
-6.00E-01
-8.00E-01
-1.00E+00
-1.20E+00
CPI change
1 2 3 4 5 6 7 8 9 10 11 12 13
Month


Impulse Pecan Respose Peanut
3.00E-01
2.00E-01
1.00E-01
0.00E+00
-1.00E-01
-2.00E-01
-3.00E-01
-4.00E-01
-5.00E-01
CPI change
1 2 3 4 5 6 7 8 9 10 11 12 13
Month


Impulse Pecan Respose Pecan
5.00E+01
4.00E+01
3.00E+01
2.00E+01
1.00E+01
0.00E+00
-1.00E+01
-2.00E+01
CPI change
1 2 3 4 5 6 7 8 9 10 11 12 13
Month


Impulse Pecan Respose Walnut
5.00E-01
0.00E+00
-5.00E-01
-1.00E+00
-1.50E+00
CPI change
1 2 3 4 5 6 7 8 9 10 11 12 13
Month

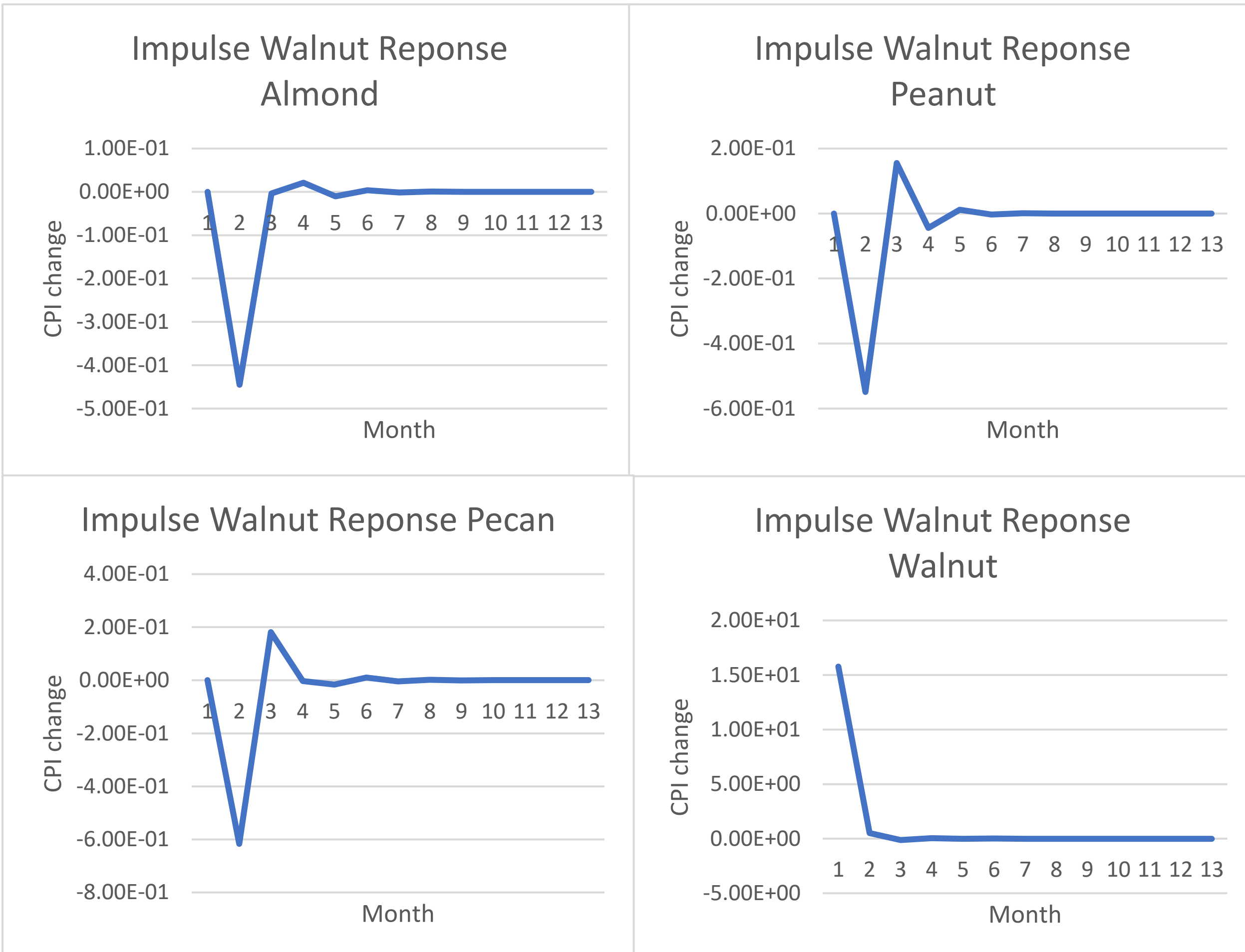
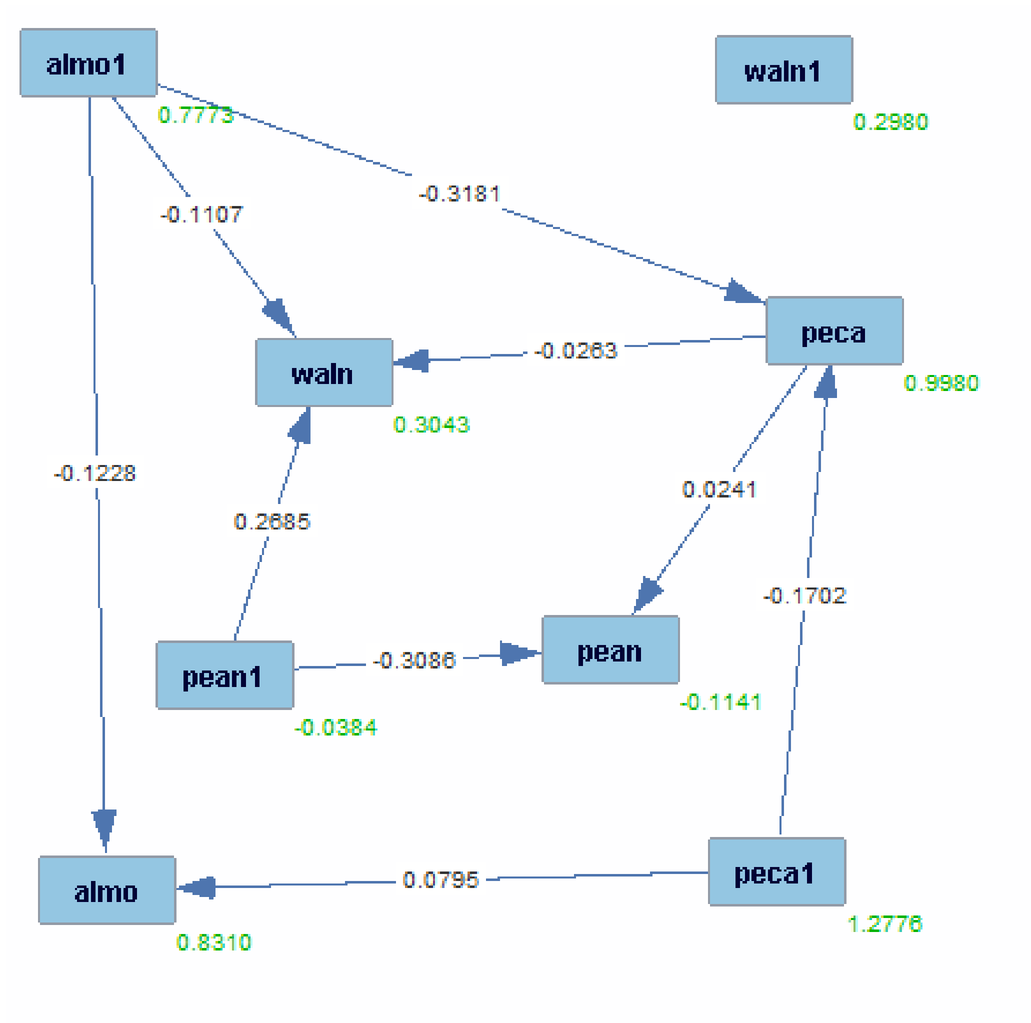


Figure 4. 1 demonstrated Impulse Response Function of all four tree nut product interactions between themselves.

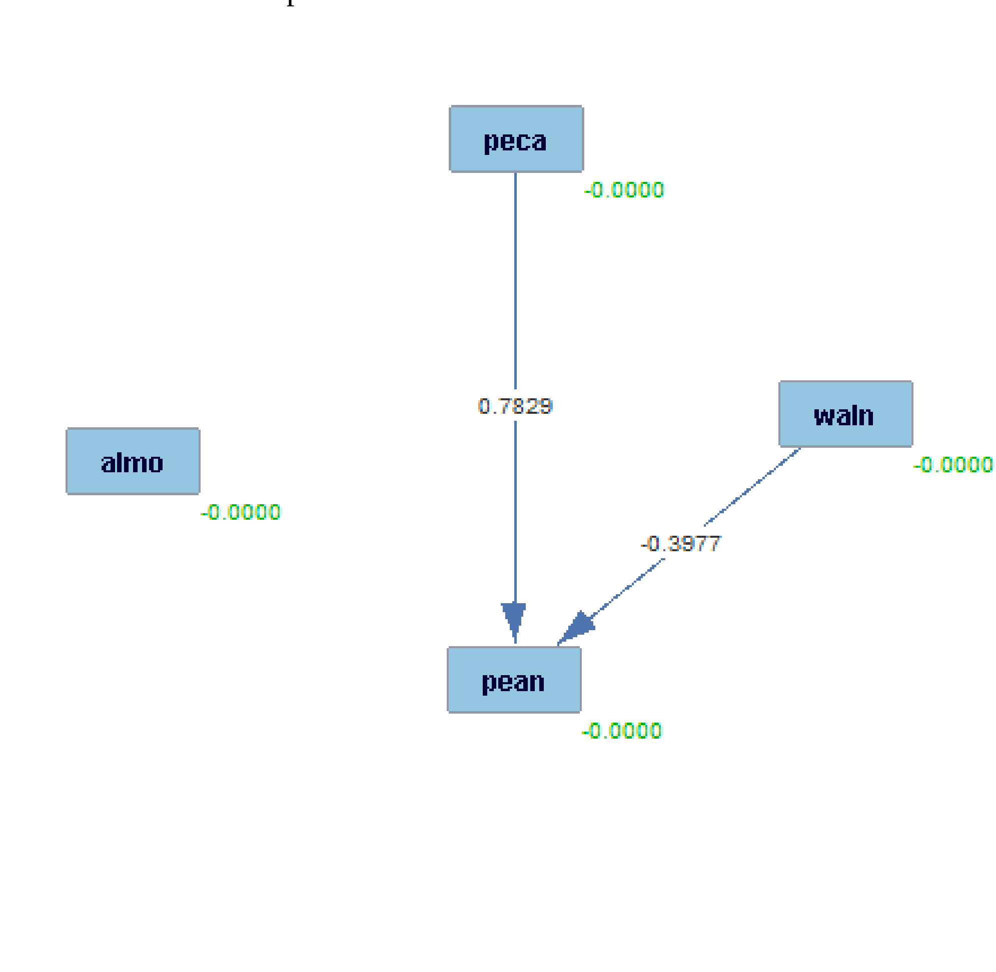
4.5 **DAG model**

Figure 4. 2 Causal relationship difference



From Figure 4.2 we can see causal relationship of differenced data. Peanut is denoted as “pean” lagged values of it is “pean1”, Pecan is denoted as “peca” its lagged version “peca1”, Walnut is denoted as “waln” and lag of it “waln1”, Almond denoted as “almo” and lagged version as “almo1”. Lagged data of almond denoted in Figure 4.2 has several causal arrows to pecan walnut and almond as well as their causal magnitude lagged almond has the strongest negative magnitude to pecan which means that pecan depended on lagged almond values and its own lagged values. Almond price does not have any effect, but through its lagged values it has the strongest influence on Walnut and Pecans values. Another fact is that Walnuts do not show any connections to its lagged values. Walnut is affected by lagged Almond, Pecan and lagged Peanut prices. Thus, walnut is follower of the tree nut industry since walnut lagged values do not have any connecting to the market at all. The same could be told about Peanuts current price has no effect on anything while its lagged affecting only Walnuts. There is also a relatively strong positive influence coming from pecan to peanuts as well as walnuts, in addition pecans lagged values also affecting almond price which puts pecans to a leader position on this market. Pecan and its lagged values have a direct impact on almond peanuts and walnut price according to Figure 4.2.

Figure 4. 3 Causal relationship residuals



Causal residual graph in the figure 4.3 (taken from VAR model residuals) has a slightly different view on causality here. Walnut and Pecans are affecting Peanut while in the Figure 4.2 the causality was going in the opposite direction for walnut. However, the same direction for Pecan and Peanut. Almond is isolated from the rest of the data which does not contradict previous graph.

4.6 **Forecasting**

Figure 4. 4 Almond Forecast

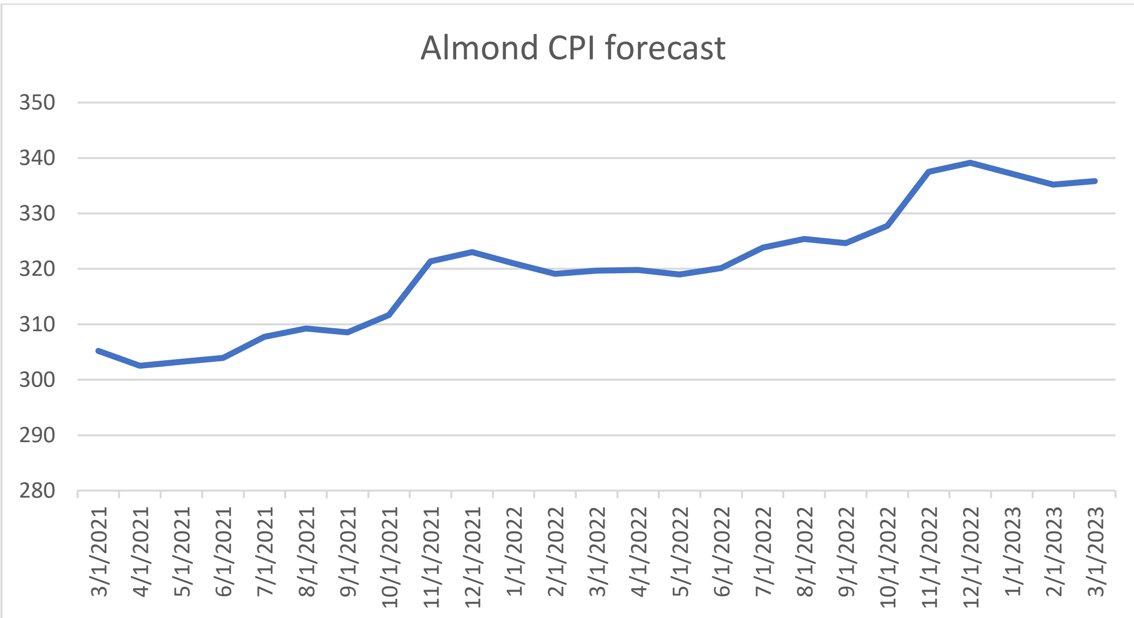


Figure 4. 5 Peanut Forecast

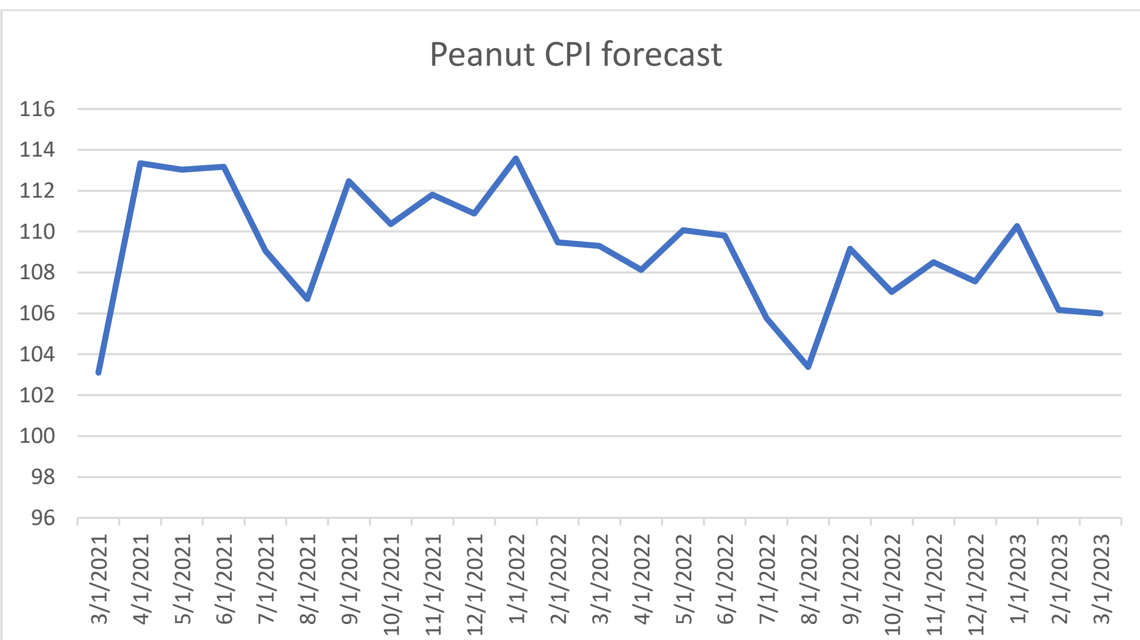


Figure 4. 6 Pecan Forecast

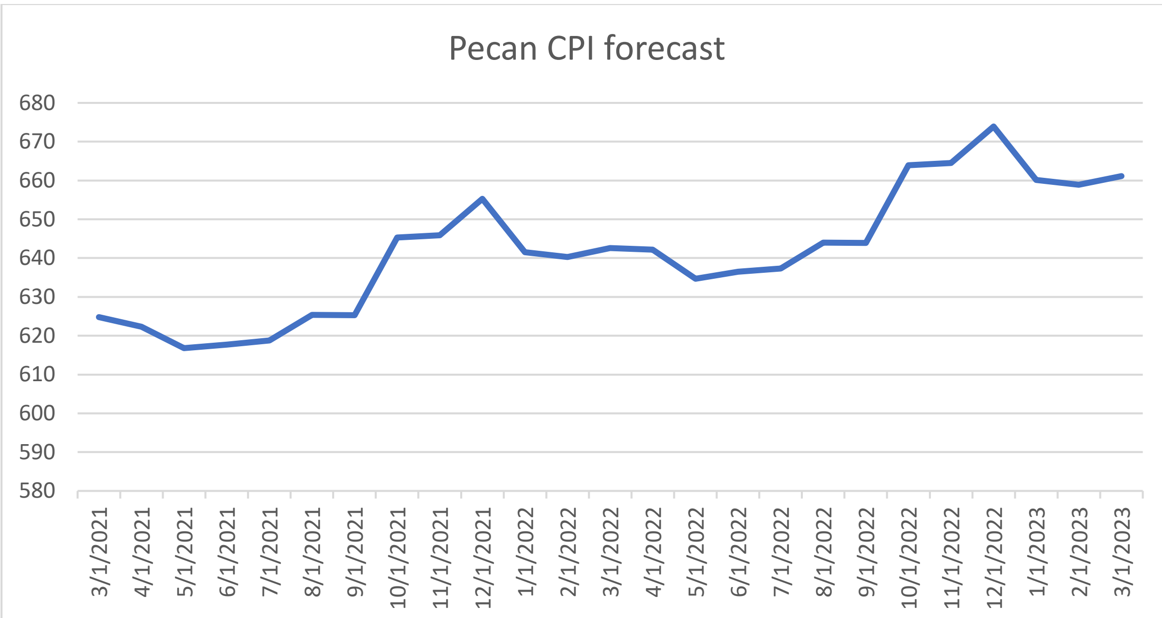
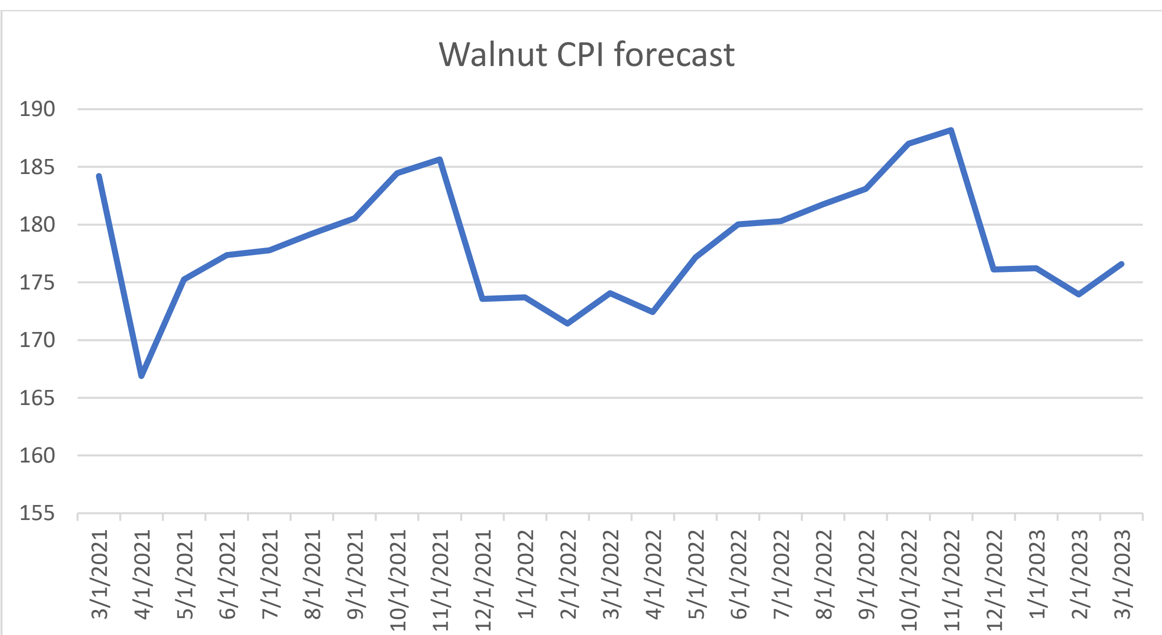


Figure 4. 7 Walnut Forecast



Figures 4.3, 4.5, 4.6, 4.7 are showing possible future values based on VAR modelling. From the forecast, Almond series has an initial shock followed by a quite stable forecast. Peanut forecast exhibit enormous fluctuation which can be happening due to previous shocks. Pecans forecast repeats Almond pattern with a few exceptions at the middle and at the end of the series. Although we need to pay attention to CPI change scale Pecan pattern is similar in graph but has the highest fluctuation in terms of CPI change. Walnut CPI changes are following historical walnut trend and at the same time following impulse response function. In 4/2022 there is a potential shock in CPI change which is correlated with Pecan, Peanuts and Almond growth. Besides that, Walnut starts with negative CPI change while most of the other tree nuts indicate positive CPI change indicating a direct link and casualty on Walnut industry. Pecan graph also indicate affect according to DAG model casualty on Peanut price. In 2/2023 Pecan once again showed an impact on Peanut CPI by decreasing it 5.5 CPI. Peanut CPI in its turn showing weak influence on Almond. According to DAG model lagged Pecan and lagged Almond both influence Almond CPI which is visible on the forecast as well. Decline of Pecans CPI in 10/2021 caused growth for Almonds CPI which is depicted in DAG model as well. Almond has some connection but judging by CPI change any influence on Almond CPI is relatively weak. VAR forecasting of Almond remains almost flat due to its historical pattern. This is backed by MSTL decomposition where Almond had higher remainder than any other Tree nut in the middle section of the data from 2000s to 2010s. Similar behavior is exhibited by Pecan CPI forecast a relatively flat forecast in the beginning of 2022 is due to that nature of historical data.

CHAPTER 5

CONCLUSION

Results of Forecasting CPI has shown that Pecan is the highest valued product from all observed tree nuts products we had. From the previous Chapters, we understood that Pecan is more prone to various illnesses adding additional expense for pecan producers. Its high volatility can also be explained by unexpected weather or biological threats to nut crops. Thus, economics principles stepping to ensure a right price at a given volume. The most stable nut product could be considered Almond and Peanut to some extent. US is an almond produce leader in the world. However, there are several things this forecast didn't include which is bee population which is closely correlates with almond production. Information about bee population colonies was not available as well their prices which made it impossible to account for such highly influential factor. Bee services are skyrocketing for the past decades. With this current trend Almond price will only grow in US. Another stable nut product could be considered Peanut. Although lately there are reports of record high temperature and droughts in pecan growing states, which can make this forecast totally invalid since weather from 1999 to 2021 was fluctuating and was considered to be record but every year a new record high numbers are posted meaning that next forecasting should be done in cooperation with people capable to model weather and provide some level of confidence to ensure a good forecasting of tree nuts products. Walnuts farmers dependent on world trade and in current geopolitical situation price fluctuation will be inevitable higher.

In extant literature, no studies were conducted on forecasting of tree nuts product using VAR modelling. The purpose of this work is to forecast tree nuts values by utilizing VAR modelling. Using USDA and FRED data, this study analyzed and constructed forecast error variance decomposition, which allowed us to understand how tree nut products interact with each other as well as potential future values of tree nuts. This research has utilized VAR modelling not to full extent meaning that further research may clarify or refute results of this paper. As it was noted by in “Causal inference in statistics” article VAR modelling has significantly advanced over the last 10-15 year (Pearl,1995). At this rate of progress, likelihood that better forecasting tools will be available in a near future are significantly high.

REFERENCES

Ros, E. “Health Benefits of Nut Consumption.” *Nutrients*. 2010;2(7):652-682. doi:10.3390/nu2070652. Internet address: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3257681/ June 24, 2010.

Pearl, J. “Causal inference in statistics: An overview” Computer Science Department University of California, Los Angeles, CA 90095 US. Statistics Surveys Vol. 3 (2009) 96–146 ISSN: 1935-7516 DOI:10.1214/09-SS057, September 2009.

Hamilton, D. James. “Time Series Analysis.”, ISBN: 9780691042893. Princeton University Press, 1 edition, January 11, 1994.

Hossain, Ahammad & Kamruzzaman, Md & Ali, Md. Vector Autoregressive (VAR) Modeling and Projection of DSE. Chinese Business Review. 14. 273-289. 10.17265/1537-1506/2015.06.001, July 2015.

Florkowski W.J. and Hubbard E.E. Structure and Performance of the Pecan Market. In: Santerre C.R. (eds) Pecan Technology. Springer, Boston, MA (1994).

Lin, Biing-Hwan, Frazao, E. and Allshouse, J. “U.S. Consumption Patterns of Tree Nuts”. Food Review/ National Food Review, AgEcon Search: Research in Agricultural and Applied Economics, 2001.

Cromwell, J. et. al. Multivariate Tests for Time Series Models, Issue 100. Sage University, 1994.

Granger, C. Investigating Causal Relations by Econometric Models and Cross-Spectral Methods. *Econometrica*, Volume 37, Issue 3, August 1969.

Hoover, K. Causality in Macroeconomics. Cambridge University Press, 2001.

Leamer, E. Vector Autoregressions for Causal Inference?, in K. Brunner – A.H. Meltzer (a cura di), Understanding Monetary Regime, Carnegie-Rochester Conference Series on Public Policy, 22, pp. 255-304, 1985.

Stephanie Glen. "Granger Causality: Definition, Running the Test" From StatisticsHowTo.com: Elementary Statistics for the rest of us! https://www.statisticshowto.com/granger-causality/ December 30, 2016.

Dharmasena, S. “Class Notes AGEC 447:” Directed acyclic graph, 2020.

Dharmasena, S. “Consumer Demand for Nut Products in the United States: Application of Semi-parametric Estimation of Censored Quadratic Almost Ideal Demand System (C-

QUAIDS) with Household-Level Micro Data.", Southern Agricultural Economics Association (SAEA), 2017 Annual Meeting, February 4-7, 2017, Mobile, Alabama.

Ibrahim, M. Florkowski, W. "Forecasting Price Relationships among U.S Tree Nuts Prices." Southern Agricultural Economics Annual Meeting, Association, Atlanta, Georgia, January 31-February 3, 2009.

Souza dos Santos, M., Gazel Yared, J., Cordeiro de Santana, A., Rebello, F., Santana Bezerra, A. "Production Evolution and Export Performance of Brazil Nut in the Brazilian Amazon." International Journal of Food and Agricultural Economics. ISSN 2147-8988, E-ISSN: 2149-3766. Vol. 7, No. 3, 2019, pp. 215-228, 2019.

Awondo, Sebastain Nde and Fonsah, Esendugue Greg, 2014. "Short Run and Long Run dynamics in the Demand of U.S Tree Nuts," 2014 Annual Meeting, Dallas, Texas 162465, Southern Agricultural Economics Association. February 1-4, 2014.

Guo, C., Dharmasena, S. and Capps, O. "The Taste for Variety: Demand Analysis for Nut Products in the United States." Southern Agricultural Economics Association Annual Meeting, Mobile, Alabama, February 2017.

Florkowski, W. J. and Lai, Y. "Cointegration between Prices of Pecans and Other Edible Nuts: Forecasting and Implications." WAEA Annual Meeting, July 13-16, 1997,Reno, Nevada, 1997.

Pearl, J. "Causal diagrams for empirical research." Biometrika, Oxford University Press, Issue 4, 669-688, Volume 82, December 1995.

Federal Reserve Economic Data, FRED, https://fred.stlouisfed.org/#, accessed in May 2021.

United States Department of Agriculture, USDA, https://www.ers.usda.gov/topics/food-markets-prices/consumer-and-producer-price-indexes/, accessed in May 2021.

Maxwell, D. "Optimal Structure Identification with Greedy Search." Chickering Microsoft Research One Microsoft Way Redmond, WA 98052, 2002.

Engle, R. and Granger C. W. J., "Co-Integration and Error Correction: Representation, Estimation, and Testing.", Granger Econometrica, Vol. 55, No. 2., pp. 251-276. March, 1987.

United States Department of Agriculture, USDA, "Tree Nuts 2019 Export Highlights.", https://www.fas.usda.gov/tree-nuts-2019-export-highlights, 2019.

United States Department of Agriculture, USDA, "Overview.", https://www.ers.usda.gov/topics/crops/ fruit-tree-nuts/, September 2, 2020.

Alogoskoufis, G. and Smith, R. “On Error Correction Models: Specification, Interpretation, Estimation.” Journal of Economics Survey, March 1991.

Johansen, S. “Statistical Analysis of Cointegration Vectors.”, Journal of Economic Dynamics and Control, Volume 12, Issues 2–3, Pages 231-254, June–September 1988.

Goldberger, A.S. and Duncan, O.D., “Structural equation model in the social sciences.”, Journal of Econometrics, vol. 4, issue 1, 97-97, 1976.

Luppold, W. and Prestemon, J., “Tests for Long-Run Relationships in Hardwood Lumber Prices. Forest Science.”, 49, 2003.

Johansen, S. and Juselius, K., “Maximum Likelihood Estimation and Inference on Cointegration — with Applications to the Demand for Money.”, Oxford Bulletin of Economics and Statistics, May 1990.

Greenland, S., Robins, J. and Pearl, J., “Confounding and Collapsibility in Causal Inference Statistical Science, Vol. 14, No. 1, pp. 29-46, February 1999.

Greenland S. and Pearl J., “Causal Diagrams.”, In: Lovric M. (eds) International Encyclopedia of Statistical Science. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-04898-2_162, 2011.

Dickey, D. and Fuller, Wayne, “Distribution of the Estimators for Autoregressive Time Series with a Unit Root.”, JASA. Journal of the American Statistical Association. 74. 10.2307/2286348, 1979.