PRICE ANALYSIS OF PEANUTS AND NUTS MARKETS

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

December 2021

Major Subject: Agricultural Economics

ABSTRACT

PRICE ANALYSIS OF PEANUTS AND NUTS MARKETS 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

This thesis uses monthly time series data on producer's price index of almonds, pecans, walnuts and peanuts in the United States. Main questions this thesis addresses are: Are almonds, peanuts, pecans, and walnuts markets integrated? What is a causal structure between observed products?

To answer these questions, we use two methods. First, develops a vector autoregressive model to discover relationship between observed products. Second, uses directed acyclic graphs to understand how innovations from each market are conveyed to other markets in contemporaneous time and to find current market pattern from raw data.

Results provided in this thesis may benefit producers by explaining price fluctuations and their possible future values. The following information is a summary of our study. Our result of causal graph from difference of producer price index unfolds the following way. Peanut is a leader and walnut is a follower. Almond and pecan are in between where peanut causes pecan and from pecan causal arrow goes to almond with a further route to walnut. Causal model from vector autoregressive model indicated that almond does not interact with others while new information from walnut will affect peanut negatively and new information from pecan will positively affect peanut. Our findings from forecast variance error decomposition from vector autoregressive model showed us that observed products explain a small percentage of errors in a range from one to six percent meaning that influence between them is insignificant. We also compared forecast values of vector autoregressive model with other simpler model such as Naïve, Autoregressive Integrated Moving Average, Seasonal Naïve and Exponential Smoothing.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work is supervised by a thesis dissertation committee consisting of the Professor Senarath Dharmasena and Professor Marco Palma of the Department of Agricultural Economics and Professor Piña, Jr. of the Department of Agricultural Leadership, Education and Communications. Special thanks for expertise and insight of Dr. Dharmasena.

Funding Source

This work was done without outside financial support.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

1.1 **Introduction**

The nuts industry in the United Stated has been growing and is experiencing higher demands than ever before. Primary reasons for demand uptick include increase in the population over of the past 30 years and public awareness of nuts health benefits (Ros, 2010). Along with this change in demand and other features, price of nuts has been changing over the past several decades. Price of nuts did not fluctuate much during the early mid-20th century: however, for the past 30 years nuts prices have been volatile. Due to novel forecasting techniques, that can accommodate such created demand for nut products by consumers level and the effect of that on nut prices, producers and processors of nuts must pay close attention to nut prices at the various stages of the agricultural value channel.

Farmers as well as nut processing industry are connected and depend on each other since farmers output and costs of processing of nut products may fluctuate affecting price of other nut products. Producers and processors of nut industry are connected and dependent on one another, since producers output and cost of processing of nut products may fluctuate, disturbing the price of other nut products. Understanding how new market information from a certain nut product affects another nut products may influence producers' revenue from nut sales. Due to this, the cost associated with nut processing industry will respond to such challenges to adjust for such changes in the cost of processing. Therefore, understanding market structure and causal patterns may help producers to anticipate market trends and reduce unexpected cost.

This study aims to fill the void providing producers more information about volatility of the market so that they could make a better decision. The data used in this study is producer price index of various nut products from publicly available sources such as U.S. Department of Agriculture (USDA) and Federal Reserve Economic Data (FRED).

An important remark about peanut must be made. Peanut is considered to be legume, but in this study, we talk about peanut as a nut product to shorten explanations.

1.2 **Objective**

The overarching objective of this research is to study the behavior of price of various nuts in the United States at the producer level which would help to determine nut price volatility and market integration patterns and develop price forecasts.

Specific objectives are as follows: using monthly prices of peanuts, walnuts, pecans and almond over the period December 1999 through August 2021, this thesis plans:

1) Explore price volatility among various nut products.
2) Study long-run price response of various nut products
3) Study contribution of new information in the nut market to the price of new information in the nut market to the price of various nut products.
4) Study market integration pattern among various nut products and new information generated in the nuts market.

CHAPTER 2

REVIEW OF NUTS DATA AND LITERATURE

2.1 **U.S. Nut Facts Overview**

It has been proven that nuts have most of the vital components for healthy living of humans. Some medical studies advocated for nut consumption due to their favorable impact on health outcomes (Ros, 2010). It was thought that nuts and seeds could cause weight gain due to their high energy density. However, this has been proven false. The ingestion of nuts can help to control satiety and increase thermogenesis. These foods are high in monounsaturated fatty acids (MUFA) and polyunsaturated fatty acids (PUFA), protein, fibers, vitamins, minerals, bioactive compounds, antioxidant potential, and nutrients. Nut intake has been shown to have benefits for health, including preventing or treating certain chronic diseases risk factors such as oxidative stress and changes in glycemic metabolism and lipid metabolism (de Souza, R., Schincaglia, R. M., Pimentel, G. D., & Mota, J. F, 2017). Many of the bioactive components in plant food are still not fully understood and characterized, i.e., carotenoids, phenolic acids, phytosterols, and polyphenolic compounds such as flavonoids and stilbenes (Janet C. King, Jeffrey Blumberg, Linda Ingwersen, Mazda Jenab, Katherine L. Tucker, 2008).

Understanding of health benefits boosts the demand for nut products. According to the United States Department of Agriculture, (USDA, Tree Nuts 2019 Export Highlights) for the last five decades, annual consumption has increased by 2.31 pounds per person from 1970 to 2016. Among those nuts, almonds have experienced the largest growth by 1.35 pounds per person. Figure 2.1 reflects nut production in the United States. Nut industry has shown a steady growth which might have been affected by the global demand for nuts. The USDA report reveals that even though the US market has a competitive advantage in nut production, it may face some challenges from the World Trade Organization (WTO) or regulations from importing countries (USDA, Tree Nuts 2019 Export Highlights).

2.2 **Nuts Profile**

Almonds (*Prunus dulcis*) are a tree nut native to Iran and surrounding countries. It is viewed as one of the nutritious food products; therefore, it is widely used for direct consumption as well as an intermediary good for other products. California produces more than 80% of the world's almonds**.** Tree nuts are harvested on a stretch of 400 miles of California, covering approximately the same size as a state of Rhode Island.

California's one of the most valued agricultural product was almonds in 2016. They accounted for $5.2 billion or about 11% of agricultural output of California. Production has more than tripled in recent years, from 703 million lbs. in 2000 to 2.27 billion lbs. in 2017. Prices grew over that time, due in part to overseas demand (USDA, 2018 California Almond Objective Measurement Report). Almond acreage that has been planted recently has replaced cotton and other traditional cash crops (California Department of Water Resources County Land Use Surveys, 2014-2016). On top of that, consumer demand and the industry's growth has largely benefitted from technological advances. During the harvesting process, farmers used tree shakers to gather their crops, the use of this machines is extensive facilitating reduction of labor input and costs. This also mitigate problem of labor shortages allowing producers to plan ahead.

According to the Almond Board of California, it showed that in 2014 industry indirectly supported more than 80 thousand jobs and employed over 20 thousand people directly. Indirect effects added $11 billion to California's Gross State Product (GSP). Almonds are essential export crop of California. In 2016, almost a quarter (22%) of California's agricultural exports were almonds. This accounts almost 4.5 billion worth of almond to foreign countries. However, in 2018 a major event happened when China decided to impose a 50 percent tariff on almonds as part of the China-United States trade dispute. Due to higher than market price levels, some Chinese companies have taken to importing almonds from Australia and African producers as a result. (Craymer, Lucy. "U.S. Almond Farmers Are Reeling from Chinese Tariffs", Wall Street Journal, 2018).

There are at least three major concerns to the production of almond. They are bees, water and waste.

1)Bees

Almond cultivation requires cross-pollination. Almonds can be pollinated by a variety of insects, but commercial almond farming relies heavily on honeybees. To ensure successful pollination, commercial almond growers might rent hives during the blooming season. In 2006, California almond growers started to experience losses from colony collapse disorder. This is a poorly understood phenomenon that causes a decline in bee populations. Although this resulted in higher pollination costs for many growers and a high demand for almonds, it also created an incentive for bees to be transported from other U.S. states to California. Since then, the state has seen a partial recovery in its bee population and now accounts for more than half all-U.S. bee colonies.

2)Water

California is know to have issue with water and particularly was hit by severe droughts in 2011 and 2017. Due to this droughts, state's almond growers suffered a severe production loss. Almond production itself carries economics and environmental concerns. In 2015, almond growers consumed about 10% of all state’s water supply. (Gonzales, "How Almonds Became a Scapegoat For California's Drought", NPR, 2018). Also, almond acreage has increased by 14% between 2007 and 2014, while almond irrigation has increased by 27% (Pickett, "In The Midst Of Drought, California Farmers Used More Water For Almonds", Forbes, 2018). Critics point out that California's 6,000 almond growers use 35 times as much water than the 466,000 Sacramento residents. Many almond farmers increased groundwater pumps to supplement the reduced state water supply. This can unsustainably drain aquifers and lead to land subsidence (Kasler, Dale; Reese, Phillip; Sabalow, Ryan. "California almonds, partly blamed for water shortage, now dropping in price", The Sacramento Bee, 21 December 2018.) The drought caused a decrease in almond production, which led to higher prices and a drop in consumer demand. Many farmers replaced older, less productive almond trees with more water-efficient varieties to compensate. Some farmers are concerned about the future supply of almonds because these trees will not be productive for half a century (Bjerga, "California Almonds Are Back After Four Years Of Brutal Drought", Bloomberg, 2018).

3)Waste products

The 2015 and 2016 crop years saw the California almond industry produce over 1.5 million metric tons of hulls, and more than 0.5 million tons of shells. These byproducts were traditionally used as livestock feed, bedding, and fuel for co-generation plants. A new way of integrations of almonds byproducts or waste in other industries such as automotive, food and pharmaceutical are currently being tested. (Bees And Almonds: How Are Almond Trees Pollinated, 2018)

One of the recent tests indicated that bioenergy feedstock has a potential application. Biochar made from almond shells and could be used to make plane and automobile tires that are more resistant to temperature changes. Biochar can also help to create stronger, biodegradable plastic products such as flowerpots and garbage bags. Another sustainability initiative is "whole orchard recycle". In this initiative almonds trees are undergoing crushing at the terminal stages of their life; the main reason of such initiative is to transfer nutrients of old trees to the soil with a further absorption of younger generation of almond trees. This also improves the soil's capacity to hold water (Davis, "The Billion-Dollar California Almond Industry's Blossoming Future", PasteMagazine, 2017).

Peanuts (*Arachis hypogaea)* native to the tropics and subtropics, are believed to have originated from South America. This crop has two classifications as both a grain legume and as an oil crop due to its high oil content. However, in this paper we put peanut in nut category to simplify explanations. It has similar nutritional values as almond and walnut. Peanuts grow in three major regions in the United States: The Southeast (Alabama Florida, Georgia Mississippi, South Carolina), and the Southwest (New Mexico Oklahoma, Texas, and Virginia, North Carolina (NASS, 2015). They thrive in subtropical and tropical climates (American Peanut Council 2014). Peanuts, which are edible seeds from a peanut plant, grow above ground and mature underground. Peanuts were considered a South regional food until the Civil War. Then, technological advances led to increased demand for peanut butter, peanut oil, roasted and salted peanuts.

Peanuts are usually planted after the last frost, which is typically in April or May. Harvesting period lasts around four or five months from planting. Therefore, the best time to market fresh green peanuts (not dried) is usually September or October. Peanuts that have been processed or dried have a longer shelf life and a longer marketing period (National Peanut Board, 2015).

Most peanut crops are processed before they reach customers, however fresh peanuts (also called boiling peanuts) can also be purchased at the harvest time. Fresh peanuts are highly perishable due to level of moisture, thus long-term storage is not applicable. Therefore, fresh peanuts cannot be sold online due to their short data of expiration. Although some customers prefer these fresh peanuts due to their esthetic look. They have bright hulls and usually show minimal no damage (Boiled Peanut World, 2013).

Crop producers must do thorough and careful planning before they market their peanut crop "fresh." Production process of fresh peanuts does not differ from average peanut crops. Some differences may include a different method of harvesting and post harvesting practices, and the varieties used. (Wright, 2014).

There are four types of peanuts are grown in America: Runner, Virginia, Spanish, and Valencia.

Main type of peanut that is commercially profitable since it is used in peanut butter is the Runner. This type accounts for three quarters of the nation's total acreage of peanuts. It is mainly found in the Southeast. The Virginia variety, which is primarily grown in Virginia and North Carolina to make gourmet snacks, accounts for 15% of the U.S. crop. Spanish peanuts are a common crop in Texas and Oklahoma. They provide 4% of the nation's peanut crop. One percent of the crop is Valencia peanuts, which are almost exclusively grown in New Mexico. These peanuts can be roasted and used to make all-natural peanut butter (National Peanut Board 2014).

The total U.S. peanut production was more than 7.2 billion pounds in 2017. This is an increase of 5.6 billion pounds from the previous year. Peanut yields increased slightly in 2017 to 4,072 pounds an acre (NASS 2018, 2017).

Peanuts are also processed to add some value in them. There are many products which peanuts can be turned into:

Roasted Peanuts

Roast peanuts are a popular snack, but they can also be used in nut mixtures, as well as being used to make candies (peanut-brittles), and other products (cookie or ice cream). (Hampton Farms, 2016).

Peanut Butter

Peanut butter is another well-know product which is made from roasted peanuts. Most of the peanut that are being processed go into peanut butter production.

Peanut Oil

Peanut oil is one of the essential ingredient in restaurants due to the feature that peanut oil has. It does not adsorb flavors and high smoke points. This oil is used in salads dressings and roasting vegetables. (The Peanut Institute, 2016).

Peanut Flour

Defatted roasted peanut flour can be used as a source of protein that is gluten-free. It can be used as a flour to thicken soups, breads, pastries, coat meats and fish (The Peanut Institute, 2016).

Biodiesel

Peanuts are high in oil (almost 50%) when compared with other oilseed crops. Peanuts can potentially produce 120 to 150 gallons per acre of biodiesel with a yield of 3,000 lbs, a 70-grade (70% of the peanut's weight in the shell), and 50% oil. Growers can produce as much as 3,500 to 4,500 pounds of peanuts an acre if they use good management practices. This means that there is potential for even greater biodiesel production per acre (University of Florida 2010, 2010). Peanut oil yield is higher than soybeans' 48 gallons per annum, but lower than rapeseed's 127-160gallons/annum (Herkes, 2014).

Pecans (*Carya illinoinensis*) are highly susceptible to a wide range of diseases, pests, and physiological disorders that can hamper their growth and efficiency. Primary growers in the US are Texas, Georgia, and New Mexico which account for more than 80% of global supply. Pecans can be grown commercially in 15 southern states: Alabama, Arkansas, Arizona, California, Florida, Georgia, Kansas, Louisiana, Missouri, Mississippi, North Carolina, New Mexico, Oklahoma, South Carolina and Texas (Wells, 2009). All varieties are derived directly from the native U.S. pecan, which has been grown wildly in North America for many years. To ensure superior quality, U.S. pecan growers developed new cultivars using non-GMO methods.

Pecans are a popular snack in the U.S., Canada and Mexico. Food processors have been focusing on individual consumers and offering multiple pecan snacks.

Most consumers love snacking on salted, unsalted and barbecue pecans. Innovative product categories such as maple glazed pecans and butter-roasted pecans are emerging in the market (Persistence Market Research, 2020). Pecan-based confectionery products like pecan candies and pecan chocolate, pecan pralines as well as honey glazed pecans and milk chocolate pecans are gaining popularity in the confectionery industry (Persistence Market Research, 2020).

Pecans can be grown in orchards or groves of trees. Both can live many years if they are well-cared for by experienced growers. A pecan tree takes 7-10 years to begin to produce full quantities of nuts. Once the production process begins, however, the tree can produce nuts for many years, sometimes even more than 100 (Pyzner, Bollich, 2006)

Walnut (*Juglans*) is native to Iran. It is confirmed that they have a higher saturation of monounsaturated fatty acids compared with other tree nuts. It is not proven that walnuts have a positive effect on health, because of lack of medical research in this field it is still controversial whether walnuts do indeed beneficially affect our bodies but have imputed to them healing properties. (Njike, V. Y., Ayettey, R., Petraro, P., Treu, J. A., & Katz, D. L. (2015). Main walnut state producer in the U.S. is California.

Walnuts are a rounded seeded fruit. After full ripening, the husk can be removed to reveal the wrinkly walnut skin. The husk and shell will become hard as they ripen. The brown seed coat, which is rich in antioxidants, protects the seed kernels. They are commonly known as shelled nuts. The antioxidants protect the oily seed from oxygen and prevent rancidity.

Walnuts trees do not grow leaves until the middle of spring, and they do not usually start to produce nuts until about halfway through the season. They also secrete chemicals into the soil to prevent other vegetation from growing. (University of Delaware, "Invasive Plant Secretes Acid To Kill Nearby Plants And Spread.", 2007).

Besides positive effects of walnuts described above this nut also contains magnesium, phosphorus and vitamin B6, as well as manganese and copper. The following plant compounds are found in walnuts: catechin, phytic acid, ellagic acid, melatonin. (Claudia Sánchez-González, Carlos J Ciudad, Véronique Noé, Maria Izquierdo-Pulido, 2017)

A few researchers have concluded the positive health benefits of walnuts. One of those studies has shown that walnuts can reduce risk of developing heart disease. They lower LDL cholesterol and improve blood vessel function. This will help to decrease the chance of plaque buildup. Walnuts contain several components that may have anticancer properties, including: phytosterols, gamma-tocopherol, omega-3 fatty acids and various antioxidant polyphenols (Fatima, Showkat, Hussain, 2018)

2.3 **Literature Review**

Only a few articles were found in the extant literature of forecasting nuts showing a great potential for further research. One of the first reports found during the analysis of the nut industry was the article “Pecan Production and Price Trends.”. In this academic work there was an attempt to forecast price by determining demand and supply equations. (Shaffer, 1996). Another effort was done by Ibrahim and Florkowski where forecasters determined the model using the ARIMA procedure. (Ibrahim, Florkowski, 2009). The authors discussed several approaches to determine cointegration. They started with a technique developed by Engle and Granger (1987), but ultimately decided to use the Johansen Cointegration Procedure (Johansen, 1988; Johansen and Juselius, 1990; Luppold and Prestemon, 2003).

Most of the previous work did not set a goal to forecast nuts products. Therefore, it was harder to find such material, although quite a few papers share a common objective such as elasticity of demand or own-price elasticity. Another study estimated two types of models, static and dynamic. Almost Ideal Demand System (AIDS) to investigate the long run and short run behavior of U.S. consumers. It 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 Cheng, Dharmasena, Capps who made a deep review of demand of nuts products and found that 10% of the price is not sensitive while the rest 90% is affected by various factors. (Cheng, Dharmasena, Capps, 2017). In conclusion, it was stated that consumers can easily substitute nut products leading to a problem in this research where modelling solely on historical price is not a proper way to get the results due to interaction of nuts between themselves. Thus, a causal component needs to be taken into consideration.

Therefore, an important component missing in this approach is Causal Inference. The Causal Inference based on the Structural Causal Model (SCM) was developed (Judea Pearl,1995,2000) in economics and social science (Goldberger,1973; Duncan,1975). Despite slightly complicated nature of the Causal Model, it simply provides us knowledge whether a certain market is affected by another market. In the research paper described in the paragraph above (Cheng, Dharmasena, Capps, 2017), they found that almond, pecans and cashew are substitutes, however the study did not go further by implementing a causality modelling as well as previous research (Ibrahim, Florkowski, 2009).

In the work of Dharmasena and Kim (2018) a causal model of pecan state producers was built. They researched dependency of the pecan market between southern producer states. Another study by Hawkins and Dharmasena (2019) conducted causal analysis of peanuts. Both papers used Directed Acyclic Graphs (Pearl,2009). Kim and Dharmasena (2018) used Greedy Equivalence Search (GES) algorithm in statistical software TETRAD (Glymour et al., 2014) as did Hawkins and Dharmasena (2019). Their analysis contributed to this research which adapted their approach to the objectives of this research. However, the above researchers concentrated on a finding causal model in a particular product, while no research has been done finding causal model among nuts.

CHAPTER 3

DATA AND METHODOLOGY

3.1 **Data**

Monthly price (at producer level) of peanuts, walnuts, pecans and almonds with a base year 1991 over the of period December 1999 through July 2021 were used in this study and were taken from publicly accessible websites of the Federal Reserve Economic Data (FRED) and U.S. Department of Agricultural Statistics Service (USDA). In the absence of complete data on actual prices of various nuts of the producer level, this study uses monthly producer price index (PPI) of various nut products as a proxy for actual prices. Preliminary analysis shows peanuts, almonds, walnuts 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 one differencing is required to make this data stationary.

Table 2.1 provides descriptive statistics of multiple seasonal decomposition procedure for monthly aggregated data level data (265 months from 1999 to 2021). The prices of different nuts product vary, and it is visible that peanuts are the lowest in terms of price and almond is the highest. The most purchased nut product is peanuts, followed by almonds (Cheng, 2017).

In this paper an attempt was made to create a Causal Model. Results indicated slight dependency. Computing forecast error variance decomposition showed that variability in the almond series explain this behavior by 99.00383% while the rest is explained by peanuts, pecans and 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

3.68%, peanut 2.02% and walnut 0.01% a quite high number compared to previous results. A slightly different situation happened in the walnut's dataset. Walnut variance explained itself by 95.07%, almond 0.37%, peanut 3.72%, pecans 0.83%. Further research found that according to the Eigen value test there is no cointegration in our set, this result was supported by the Johannsen test as well (Johansen, 1988; Johansen and Juselius, 1990; Luppold and Prestemon, 2003).

3.2 **Non-Stationarity**

In a time series, data are said to be non-stationary if data does not return to its historical mean for the long period of time or the data is non-mean-reverting. However, if data returns to its historical mean frequently and mean reverting, such data series is said to be mean-stationary. Also, if a time-series is non-stationary, such series is said to have a "unit root" (or said to have an infinite variance of error). If such series is used in regression analysis, for example, development of an auto-regressive model to forecast prices, AR(p) model, estimated coefficients in such model may not found to be significant due to infinite variance innovation with estimated coefficient. Hence, this might lead to various spurious statistical testing with regards to the significance of the estimated coefficient.

There are several ways to determine stationarity. The first and the easiest way to get an idea whether your data is stationary is to graph it and observe whether it contains any trend or extreme fluctuations. However, sometimes the graph may look stationary when it is not. Therefore, looking at the correlogram or conducting a unit root test, it is possible to check assumptions about observed dataset. A different way to test for stationarity is to look at the autocorrelation function (ACF) and the partially autocorrelated function (PACF). Observing a gradually going down (decay) pattern of lags then we can assume that that our data is non-stationary and needed to be differenced. Following this formula, we can calculate an ACF plot which is usually computed by statistical software.

(1) $$\hat{p}_k = \frac{y_k}{y_0}$$

y_k is cov(y_i, y_{i+k}) and y_0 is the variance of the stochastic process

At the same time, we can find what lag is significant based on the ACF plot which could be then verified with Akaike or Schwarz information criterion.

Moreover, we can apply a unit root test. To determine whether a time series has a unit root or is non-stationary we conduct several tests. The most used and well-known test is the Dickey-Fuller test or DF test (Dickey & Fuller, 1979). Another widely used testing is Phillips-Perron test (PP) (Phillips & Perron, 1998). The ADF is preferred test since ADF make the error (e_t) smaller by incorporating lags of dependent variable in the right-hand side of the equation.

In this work, the ADF test is used to test for unit roots of each nut price series. The number of differences taken to make the non-stationary time series stationary is d, ie integration of order d, I(d). I(1) is a non-stationary time series in levels and I(0) is stationary time series in levels.

Dickey-Fuller (DF) test is shown using the following equation:

(2) $$\Delta X_t = a_0 + a_1 X_{t-1} + e_t, \text{ where } \Delta X_t = (X_t - X_{t-1})$$

X_{t-1} is the first lag of X, Δ is difference operator, a_0 is a constant, a_1 is

and e_t is the random error e_t ~ N(0,1)

The null hypothesis is that the series is non-stationary or series I(1). The alternative hypothesis is stationary series or I(0).

There is also Augmented Dickey-Fuller (ADF) test which test for stationarity. The main difference between ADF and DF test is that DF test is used for basic models such as AR(1) and if we are testing more complicated models such as ARMA(p,q) with unknown orders then we use ADF test

(3) $$\Delta y_t = a + Bt + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \cdots + \delta_{p-1} \Delta y_{t-p+1} + \varepsilon_t$$

where a is a constant B is coefficient on a time trend and p is a lag order of autoregressive process, δ is a unit root

3.3 **Vector Autoregressive Model**

Before explaining the Vector Autoregressive Model (VAR model), it could be beneficial to explain Autoregressive Model. The following explanation of the VAR modelling is based on Hamilton (1994). In the formula below c is a constant, ϕ is a parameter and y is a value conditional on t. Since Vector AR is explained in matrix form, 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 is represented as

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

C is a constant, Φ is a coefficient

where E(e_t)=0

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

A Vector Autoregressive model of order 1:

(5) $$y_t = v + B_1 y_{t-1} + e_t$$

v is a constant, B is time-invariant (k × p)-matrix, e_t is a k-vector of error terms.

The above model will have the following matrix:

(6) $$B=\begin{bmatrix} B_1 & B_2 & \dots & B_{p-1} & B_p \\ I_k & 0 & \dots & 0 & 0 \\ 0 & I_k & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & \dots & I_k & 0 \end{bmatrix}, y = \begin{bmatrix} y_1 \\ \vdots \\ y_p \end{bmatrix}, v = \begin{bmatrix} v \\ 0 \\ \vdots \\ 0 \end{bmatrix} \text{ and } e = \begin{bmatrix} e \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

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).

(7) $$Y_t = B_{10} + B_{11}y_{t-1} + \cdots + B_{1p}y_{t-p} + \gamma_{11}X_{t-1} + \cdots + \gamma_{1p}X_{t-p} + e_{1t}$$
$$X_t = B_{20} + B_{21}y_{t-1} + \cdots + B_{2p}y_{t-p} + \gamma_{21}X_{t-1} + \cdots + \gamma_{2p}X_{t-p} + e_{2t}$$

B and γ parameters are estimated using Ordinary Least Square method (OLS). (Hanck, Arnold, Gerber, and Schmelzer, 2020*)*.

Another important detail is the lag length of each variable. In the following paragraphs, there will be a discussion on how far back in time we need to go. If too many lags are taken, the results will be at a cost of degrees of freedom and too few we will encounter with an issue of autocorrelation.

One of the ways to determine an optimal lag length of VAR variables is to use information criteria, Bayesian Information Criterion (BIC). The smallest BIC(p) result is used to determine the optimal lag length of a variable. BIC is shown below:

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

$(\sum_i)$ is a (k×k) covariance matrix of the VAR errors. An important condition about covariance is that $(\sum_i)$ must be positive. K is a number of coefficients, p is a number of restrictions (lag order), T is a number of observations.

Alternatively, Akaike information criterion (AIC) also can be used to determine the optimal lag length of variables in VAR models.

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

3.4 **Causality Graphs**

Contemporaneous causality relationships among innovation term in VAR (new information) was studied using causality graphs, i.e., Directed Acrylic Graphs.

Directed Acyclic Graphs (DAG) is a widely known approach to detect confounding variables that require conditioning for estimation of causal effects. DAG contains directed edges linking nodes and their paths. A sequence of nodes connected by the arrows (nodes) is called a path. A directed path follows the edges in the direction which is indicated by the nodes (assuming there are three nodes Z, F and H where Z causes F and F causes H, then we have (Z→F→H)). A directed graph, the graph is defined as asymmetrical relationship where arrows are directing the relationship. In a direct graph, the graph in Z→F→H set is followed so that Z is directed to F and F is directed to H. For an indirect graph where Z→F→H Z indirectly causes H where F is a mediate or intermediate variable. To summarize, defining a sequence is done first and it is proceeded to recognize an indirect causal relationship. The next word in DAG abbreviation is acyclic which basically self-explains that there is no cycle. A cyclical graph is a graph where at least one of the nodes after a sequence of directed edges returns to the same node. A path of three variables would be $Z \rightarrow F \rightarrow H$ and $H \rightarrow Z$ creating a cycle. In this study the attempt was to understand causal effect thus only acyclic graphs were reviewed. An acyclic graph is where a $Z \rightarrow F \rightarrow H$ and H does not cause Z in contemporaneous time.

Three main three structures in a DAG model:

1) Common Cause or Causal Fork (Figure 3.1)
2) Causal Chain (Figure 3.2)
3) Collider or Inverted Causal Fork (Figure 3.3)

Figures 3.1, 3.2 and 3.3 showing causal structures of those three cases of DAG, which is used to explain the structures. The following material is derived from work of Dharmasena, Bessler and Capps (2016). Let the letter "ρ" represent a correlation between variables. For a causal chain structure, where A causes B and B causes C, we are going to have $\rho(A,C)\neq 0$ unconditional correlation A and C is not equal to zero, $\rho(A,C|B)=0$ conditional correlation A and C given B is equal to zero, new information from B makes A and C independent. For instance, consumption of soda may increase a person's weight and increased weight could develop into type 2 diabetes. So consumption of soda and type 2 diabetes will correlate $\rho(A,C)\neq 0$, but conditioning on weight, soda consumption, and type 2 diabetes will become independent $\rho(A,C|B)=0$.

In a common cause or causal fork where A causes B and C, we are going to have $\rho(B,C) \neq 0$ and $\rho(B,C|A)=0$. For example, common symptoms of a disease have symptoms such as coughing and headache $\rho(A,C)\neq 0$, so these symptoms will correlate. If a person has a headache, then it is likely that a person also has a coughing symptom. However, if we condition only on people who have certain disease such as a flu then $\rho(B,C|A)=0$ both of our symptoms are independent given a flu.

In collider or inverted fork where A and C causes B we have $\rho(A,C)=0$ and $\rho(A,C|B) \neq 0$. Here will be an entirely different dependence structure. In this case, A and C are independent (not correlated) and conditioning on a common effect C, we have a dependence (correlation). For example, time spent studying and intelligence can define what grade a student will get. So time and intelligence are independent $\rho(A,C)=0$, but conditioning on a grade we get a correlation between time and intelligence $\rho(A,C|B) \neq 0$.

DAG uses variance-covariance matrix from a set of variables to explore possible relationship among the set of variables and develop casualty structure. The algorithm used for this research was Fast Greedy Equivalent 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.

It cannot be expected to find one that can fully estimate one particular DAG. This is due to the fact that a DAG search algorithm is developed to find an equivalence search class. For example, A causes B is equivalent to A cause by B, probabilistically. Usually to have the same fit, estimation of equivalence class or a set of models is required.

There are two types of algorithms: constraint-based algorithm and score-based algorithm. The main difference is that constraint-based algorithm looks at node-to-node level or in other words at individual relationship. While in score-based the algorithm tries to identify whether entire model fits or not. Score-based algorithms fit several DAGs, scores them, and chooses the best one. There are many scores which could be used to estimate DAG; one of them could Bayesian or Akaike information criteria. In this paper we used score-based algorithm

In both cases, we can use prior knowledge to improve results of an algorithm. It can be achieved by forcing an edge to be included or excluded in the DAG, in other words we allow or force an effect to be there. In Blacklisting, a complete opposite of Whitelisting, we are forcing an edge to be excluded from the DAG model or no effect.

There are some assumptions of DAG which cannot be violated. There are three assumptions:

1) Causal Sufficiency Assumption. It assumes that there are no common unobserved variables in the domain that are parent to a one or more observed variables of the domain. For example, having a common cause structure where A causes B and C then C and B are correlated and can be independent after conditioning on A. Thus, if A is not observed then condition on A is not possible. So having B and C in the model may include that B and C are correlated, but this conclusion is false due to correlation which is explain by A. In short, we can say no latent variables.
2) Markov Assumption. It assumes that in Bayesian network structure model A, any variable is independent of all its peers in A (those who didn’t cause or were caused by that variable), given its parents.
3) Faithfulness Assumption. It assumes that in Bayesian network graph A and probability distribution P are faithful to one another if and only if everyone and all independence

relations valid in P are those entailed by the Markov Assumption on A. For example, A causes B and causes C, at the same time C causes B. Let's also assume that there is positive correlation between A and B and also between C and B. In addition, there is weak correlation between A and C. So, putting condition on B we have A and C correlation equaling almost to zero simply due the way we parametrize the model. After that we might get change of a structure.

3.5 **Impulse Response Functions**

An impulse response functions (IRF) are developed to study the effect on variables due to one-time only shock of a given variable in the system.

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

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

The matrix ϕ_x could be interpreted as

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

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).

Assuming that every element of innovation e_t is changed by δ then this equation is possible:

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

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

Hamilton also discussed that one the fastest way to determine dynamic multipliers numerically is by simulation. To conduct the simulations, current setting are presented $y_{t-1} = y_{t-2} = \cdots = y_{t-p} = 0$. There is also a 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)

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

The equation above is called impulse response function.

3.6 **Forecast Error Variance Decomposition**

Let's again remind that y_t is dependent variable and e_t is a error or white noise which is iid (Independent and identically distributed). Here an assumption is made $\phi_x(L)e_t = \sum_{i=0}^{\infty} \phi_{x,i}\, e_{t-i}$ as it was introduced in Impulse Response Section (IRF) this coefficient $\phi_{x,i}$ is IRF of y to e.

Decomposition of the forecast errors due to innovations in z and other sources of variation as follows:

(14) $$f_{t+h|t-1} = \phi_{z,0} e_{t+h} + \cdots + \phi_{z,h} e_t + v_{t+h|t-1}$$

The last term $v_{t+h|t-1}$ is innovation term

According to the Sims (1980), explanation of the population share of the variances explained by the cotemporaneous and future innovations in e_t to total variations in $f_{t+h|t-1}$:

(15) $$s_h = \frac{Var(\phi_{z,0} e_{t+h} + \cdots + \phi_{z,h} e_t)}{Var(f_{t+h|t-1})}$$

This provides output of forecast error variance decomposition which is also referred as FEVD.

3.7 **Johansen Cointegration**

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

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

where T is the number of observations, λ is eigenvalues

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

CHAPTER 4

RESULTS AND DISCUSSION

4.1 **Initial Analysis**

In table 2.1 is shown the descriptive statistics of each price using the median, mean, coefficient of variation and standard deviation. These values were ranked as well from each nut product over the entire sample period. (December 1999 through July 2021).

Based on the analysis provide in the table 2.1, Almond is a leader in all parameters. It has the highest PPI mean as well as median, standard deviation (SD) and coefficient of deviation (CV). One of the primary reason of almonds being ranked as the highest nut with the highest PPI in the U.S. is due to long production process. Almond as well as walnut trees do not bring any profits in the initial stages of their life. Therefore, this information causes almond and walnut to be valued higher than other nuts. Walnut ranked the 2nd after almond in all statistical parameters. Another nut which faces challenges is pecan, it is one of the most susceptible to diseases nut from all other observed products in this paper. Challenges of growers are reflected on the price of Pecan which in their turn affecting PPI. Pecan parameters do not deviate significantly from walnut parameters. CV and SD of walnut and pecan are closer than peanut values.

For peanut, a completely different picture is presented by descriptive statistics. The lowest mean, median, CV and SD among the nuts is observed. It also surpasses all other observed nuts in terms of production volumes in US. Peanut market could have weak fluctuation due to slightly easier production processes (peanut is relatively easier to grow than other nuts) and availability since it is one of the most produced nuts in US. However, according to some observers, peanut industry suffers from a monopoly, where 80% of all raw peanut are being processed by two major companies. (Knox, “The Peanut Industry Has a Monopoly Problem—but Farmers Are Pushing Back”, Civil Eats; 15 January 2021).

4.2 **Plots of Historical Data**

Examination of the graph of each market has shown that all four observed nuts do not have any clear upward or downward trend. (Figure 4.1, Figure 4.2, Figure 4.3, Figure 4.4)

On the almond graph, there is a sharp uptick from 2009 till 2011. There might be several reasons for this increase. First, almond producers have been hit with increased costs due to the declining number of honeybees. During 2000s and beginning of 2010s, the bee population has declined at an alarming rate. Around 40% of bee colonies in US were under stress and it is well-known fact that bees are an essential part of almond producers. California almond producers have a long tradition of paying beekeepers from the US to help pollinate their orchards. This has become more costly as hive numbers have declined. As a result, almond farmers costs are increasing. Another reason was that according to forecasters in 2013, expectation of almond yield was higher while in reality supply dropped by 0.85 million tons. (“2013 California almond forecast”, USDA, 2013). The last reason could be explained by higher demand from abroad, China particularly imported more, from 30 million to 150 million tons. However, after 2015 we can see a decline which attributed to improved weather conditions. California had more precipitation compared with previous years leading to a better yield. However, since 2018 supply again went up to its past 2014 values. All these changes have a cyclical nature leading to the fact that current level of prices will likely to persist in the next 2 years.

Looking at the Peanut prices we observe some historically low prices in 2005 and steady upward trend since then to 2014. In the Chapter 2 it was mentioned that peanuts are susceptible to diseases. This in their turn leading to “roller coaster” behavior of prices. In 2009 in Georgia and Texas had an outbreak of salmonella at the peanut processing plant caused prices to go down a short period of time and again in 2010.Peanut processing plants in Texas and Georgia are responsible for a deadly Salmonella outbreak. Later in 2011, poor harvest led prices to record high values in a history ("Multistate Outbreak of Salmonella Typhimurium Infections Linked to Peanut Butter, 2008-2009” Center for Disease Control and Prevention (CDC), 2009). After that we can see that prices are normalizing, only in 2019 we saw again a short drop caused by a lawsuit (“Panhandle peanut farm alleges price-fixing in class-action lawsuit against Golden Peanut, Birdsong Corporation”, Dothan Eagle, September 2019).

Pecans historical pattern on the graph does not depict much trend. Some notable changes happened after the economic crisis of 2008. In 2008 there was a drop in supply leading to a higher price. In 2009, there is a spike which holds prices on a new level is attributed to global demand. China one the largest markets has developed a taste for this nut, hence increase in export demand. In 2011 droughts led to a short supply and spike of prices. After 2016 fluctuations are due to trade wars with China. Overall, we can say that pecan prices from 2010 to 2020 were relatively stable and current decline probably is connected to a lower purchasing power of population after 2020 pandemic which is still ongoing.

Walnuts are unique in this case. They show an increase despite decrease in other nut products in 2021. This anomaly could be connected to a consumer preference although futher analysis may be necessary to understand this. In 2018 we have a decline, one of the explanations for this event to occur might be production trends. As said before, walnuts are vulnerable to weather changes such as drought and diseases. In 2015 we can see a similar picture, a significant drop in prices which is due to a strong dollar which hampered exports leading to a decline in price and poor harvest (Walnut Market - Growth, Trends, COVID-19 Impact, and Forecasts (2021 - 2026).

4.3 **Stationarity Tests**

In methodology section we have discussed stationarity and its importance for VAR. A Dickey Fuller test was conducted for all 4 series of Almond, Pecan, Peanut and Walnut. The null hypothesis of DF test was that the series are non-stationary. The t-statistic value at 5% significance with critical level -2.89, which is determined by Dickey-Fuller (Dickey & Fuller, 1979). A series must be differenced to the point when it becomes stationary. In our study after first differencing, data was tested with DF and ADF which successfully rejected null hypothesis for calculated statistics less than 5% level.

ADF test of almond (-2.098) peanut (-2.4604), pecan (-2.1219), walnut (-2.5149) tells us that series is non-stationary in mean and new information influenced by recent values while historical mean plays less and less role in definition of current values. In other words, lagged level of series brings no useful influence in prediction of future values, when a series is non-stationary.

In the methodology part basics of autoregressive approach were discussed, from there we know that difference plays important role in autoregressive approach. Therefore, a first difference was taken. The results of first difference of the data showed that the series are stationary: almond (-6.1486), peanut (-7.9264), pecan (-6.7308), walnut (-7.1785). These results tell us that we have no unit root or that lagged level of series provides relevant information in forecasting change of the series.

4.4 **Result of Cointegration Test**

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, use of a Johansen Cointegration Procedure is required. 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).

4.5 **VAR Model**

Given that there is no cointegration between but prices series, the only option to forecast price is to use a estimated VAR model in the first difference. Akaike Information Criterion (AIC), Schwarz Criterion (SC), Hannan Quinn (HQ) and Akaike's Final Prediction Error Criterion (FPE) were used to determine the number of the lags of each series in the VAR model. Based on this, the optimal lag for the VAR model was chosen to be one (table 4.2). A seasonal dummy variable (month seasonal dummies) was included in the VAR model to account for seasonality in data. The VAR model results are printed in table 4.3, 4.4, 4.5, 4.6.

The variables used in the VAR model are defined as follows: AlmondL1 (almond lag one), PeanutL1 (peanut lag one), PecanL1 (pecan lag one), WalnutL1 (walnut lag one) and seasonal dummy variable sd1, sd2 and etc.

In almond(almo) estimation results for equation only 5 variables were significant at 95%: lag of almond(almo.l1), sd2,sd3,sd5,sd10. For peanut only 1 variable was considered to be significant: lag of peanut (pean.l1). For pecans almo.l1 was significant and for walnuts it was pean.l1 and sd1.

According to forecast error variance decomposition, which is shown in tables 4.7, 4.8, 4.9, 4.10, 98.9% of almonds in the short run is explained by new information generated from almonds. The almost the same values remain in the long run for almonds as well 98.7%. For peanuts about 96% of the error is explained by innovation in peanuts in a long run where another 2.7% comes from pecans and 96% in a short run, where another 2.2% comes from pecans. About 99% of pecan error is explained by pecan in short run, however it reduces up to 97% in the long run, where another 2.5% comes from almonds. For walnuts 97% of error is explained by itself in the short-run, however it reduces up 93% in the long-run, where another about 4.5% of error is explained by pecans in the long run and 2% by peanuts.

In the following paragraphs there will be a discussion of impulse response functions (IRF) results. Almond impulse for walnut response has higher magnitude compared to pecan and peanut responses. From error decomposition of almond (table 4.7) walnut has the lowest explanation of almond errors. Another interesting observation is that almond shock to walnut comes to equilibrium at the same rate as almond shock to pecan and peanut, while walnut shock to almond does not have similar pattern and expresses a bit more disturbance in the long run. Both of almond shock walnut and walnut shock to almond IRF graphs have 1 period delayed response. In addition, walnut shock to almond has a weaker magnitude than almond shock to walnut. For almond shock to pecan and peanut no delay is presented. Immediate reaction from almond to pecan and peanut indicates that markets are dependent on almond. Peanut reaction to almond shock is stronger than pecans to almond. It is also relatively quickly stabilizing as well as pecan IRF graph.

Peanut impulse response functions are similar in terms of pattern. Initial delay from a shock in a first period from peanut shock to pecan and almond with changing PPI shock pattern from positive to negative values with a gradual stabilization around zero at the 7th period. Peanut shock to pecan continues to fluctuate and stabilizes at 8th period. Other observed IRF graphs usually come to an equilibrium at 6th or 7th period. From forecast error decomposition table 4.8 this information is reflected; pecan explains 2.2% in the short run and increase to 2.7% in a long run of peanuts errors. In pecan's forecast error decomposition, peanut has an insignificant 0.05% of pecans errors. In the IRF of pecan shock to peanut the shock dies at 5th period. Therefore, based on these facts, a pecan has an impact on peanut much greater than peanut on pecan.

Walnut impulse response function is relatively similar. Delay in the 1st period and quick shock in the 2nd period with a stabilization around 5-7th period which is noticeable in forecast error decomposition as well. Shocks from other nuts is alike in magnitude, comparing forecast error decomposition tables we can notice that walnut explains only 93% of itself while other nuts are above 95%. It also has a delay on the 1st periods which leads to assumption that walnut is price follower since almost 7% of errors are being explained by other nuts.

Overall, pictures of IRF graphs are unique to each nut product and have unique patterns. However, there are similarities of peanut shock to almond, walnut, and pecan in terms of pattern. Similar to walnut shocks to almond and peanut. Walnut shock to pecan has a unique pattern. Delay in the first period leads to peak of a shock in 3rd period and equilibrium at 8 period. Another interesting pattern is shown with strong fluctuations from peanut shocks to almond and pecan with a decay at the 7th period. No shocks exceed 8th period meaning that there are no long-run relationships between these nut products.

To summarize impulse response function, most of the shocks are quick to stabilization at the 4-5th period (month). The highest magnitude is shown in peanut shock to pecan IRF graph followed by almond shock to walnut.

Figures 4.8, 4.9, 4.10, 4.11. are showing forecasted values based on VAR modelling. From the forecast, Almond series forecasts does increases slightly while actual values decreased. One of the reasons for that is that almond demand abroad dropped. Peanut forecast exhibit more fluctuation in the series which may 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 PPI change scale Pecan pattern is similar in graph but has the highest fluctuation in terms of PPI change. Walnut PPI changes are following historical walnut trend and at the same time following impulse response function. In 10/2019 there is a potential shock in PPI change which is correlated with Pecan, Peanuts and Almond growth. Besides that, Walnut starts with negative PPI change while most of the other nuts indicate positive PPI change indicating a direct link and casualty on Walnut industry. Pecan graph also indicate affect according to DAG model casualty on Peanut price. In 9/2020 Pecan once again showed an impact on Peanut PPI by decreasing it. Peanut PPI in its turn showing weak influence on Almond. According to DAG model lagged Pecan and lagged Almond both influence Almond PPI which is visible on the forecast as well. Decline of Pecans PPI in 5/2021 caused slight growth for Almonds PPI which is depicted in DAG model as well. Almond has some connection but judging by PPI change any influence on Almond PPI is relatively weak. VAR forecasting of Almond remains almost flat due to its historical pattern.

4.6 **Contemporaneous and Forecasted Behavior of Nuts**

To determine the contemporaneous causal relationship of innovation from VAR model, a Directed Acrylic Graphs (DAG) was created with help of software Tetrad version 6.3.4 (tetrad-gui-6.3.4) with FGES algorithm. A correlation matrix of errors (innovations) from VAR was developed. Also, the DAG is applied to raw data to determine the causality structure determining nut prices.

Correlation matrix of residuals from VAR:

	almond	*peanut*	*pecan*	*walnut*
almond	1.0000000	−0.06563	−0.06383	−0.0003948
peanut	−0.0656269	1.00000	0.15419	−0.1336238
pecan	−0.0638301	0.15419	1.00000	0.0647620
walnut	−0.0003948	−0.13362	0.06476	1.0000000

After examination of correlation matrix, there are several things we should pay attention to. There is no significant correlation in almond PPI. Peanuts and pecans have the highest positive correlation while walnut has almost no influence on almonds. Another high after peanut and pecan but negative correlation is observed in walnut and peanut relationship. Looking closely, almond have negative correlation with all nut products. Values for peanut and pecan around 6% while walnut almost does not correlate with almond. Overall, no strong correlation is observed. Correlation matrix will be discussed with causal model and forecast decomposition in detail later in this section. However, explanation of DAG model is needed here.

In the following paragraphs, there will be a discussion of DAG model as well as its explanation. In our DAG model, we denoted Peanut lagged values "Peanut1", Pecan lagged values is denoted "Pecan1", Walnut lagged values is denoted as "Walnut1", Almond lagged values denoted as "Almond1". We used FGES algorithm with a penalty discount at 1 for differenced data and at 0.2 for residuals.

FGS algorithm which is optimized and parallelized version of FGES algorithm developed by Meek (Meek, 1997). Methodological approach used by FGS is Bayesian Information Criterion (BIC) to score models to approximate the maximum likelihood of the data provided in a graph. In other words, it takes a natural logarithm of the marginal likelihood and approximates it. (Center for Causal Discovery, 2021)

"A CBN structure is a directed acyclic graph in which nodes represent variables and arcs represent direct causation among the nodes, where the meaning of direct is relative to the nodes in the CBN". (Center for Causal Discovery, 2021)

$$BIC = 2\times \ln P\ (data\ \Theta|, M)-c\times k\times \ln(n) \tag{17}$$

"Where M denotes a CBN structure, denotes the value of the CBN parameters (e.g., coefficients and error terms in a regression model) that maximize the data, k are the number of parameters in the CBN, n is the number of samples (cases), and c is a constant that is 1 in the traditional definition of BIC" (also referred as a penalty discount). Also, the marginal likelihood of the data given a graph structure M: P(data | M) (Center for Causal Discovery, 2021)

To better understand DAG model following quick review is provided. There are three structures in DAG named: Common Cause, Causal Chain and Collider. Let's imagine three components X, Y, Z and assume X is a common cause for Y and Z. We will have an inverted V-shape in our graph. X causes Y and Z. For a chain structure X would cause Y and Y would cause Z. For a collider structure we have an opposite of common cause. Y and Z causes X. Knowing this we can say that almond is a collider which is caused by its own lag and lag of pecan.

Looking at the causal relationship of differenced data, a current market pattern is shown in the Figure 4.6. This graph basically explains current market interaction of nut products based on difference of PPI data. From the graph, it is visible that peanut is a primary cause for other nut products. In other words, peanut is a price leader among these nut products. After peanut impact on pecan, the causal chain goes to almond and ends up in walnut series. This figure 4.6 also accounts for the past values of observed nuts. From the figure, it is visible that lag of pecan (pecan1) is one of the main epicenters of information which affects other markets. Lag of pecan causes almond and walnut, there is also a connection to its own market. The same applies to lag of walnut which affects current walnut price and lag of almond which affects current almond prices. This information about lags impact on current values is obvious and will not be discussed in detail. From here, a conclusion can be made based on what was said above. Peanut is a price leader that sends signals through pecan market and determines prices of nuts. From peanut new information is being processed by pecan market and creates an impulse to almond market which receives it and passes this signal to walnut. Moreover, past values of pecan can affect almond market in addition to existing signal from current values pecan market. For walnut market, in addition to almond influence past values of pecan as well have an impact.

To summarize discussion of figure 4.6 peanut producers are not affect by any observed nut products. Peanut products volatility determines where market will be. Pecan producers must keep a close eye on a situation on peanut market since this market in the only sources of disturbance on their market. For almond producers, one must consider peanut and pecan market situation to make a proper decision. Lastly walnut producers must account for all observed in study nut markets since their product is mainly being caused by almond, past values of pecan. Walnut producers may also pay attention to peanut market since this market is a price leader and any shock from that market will echo in their price as well.

Figure 4.5 provides contemporaneous causal relationship of innovations from VAR. Observation of the figure 4.5 provides us following interpretation. A collider structure of peanut being caused by walnut and pecan tells us how a new information will affect peanut. Any new information from walnut market will move peanut market by negative value (-0.1020). Moreover, new information from pecan will move peanut market by positive value (0.2690). Almond in contemporaneous causal relationship of innovations from VAR is not connected, meaning that VAR model could not determine almond integration in nut market. However, this may not reflect real situation. Additional data may be necessary to perform more analysis. The reminder here is that analysis was performed on producer price index therefore should not be interpreted in terms of dollar price.

4.7 **Forecast Validation**

We used 250 monthly observations for nut products (December 1999 to June 2021). For the purpose of model estimation, we split our dataset into training and test set to validate our forecasting attempts. We compared forecasting results of our models with each other using Mean average present error (MAPE), Mean Average Deviation (MAD), Mean Squared Error (MSE) and Root Mean Squared Error (RMSE)

(18) $$MAPE = \frac{1}{n}\sum_{t=1}^{\infty}\frac{|A_t - F_t|}{|A_t|}$$

Where A is actual value, F is forecasted value and n is number of observations.

(19) $$\text{MAD}=\frac{1}{n}\sum_{t=1}^{\infty} | x_i - \bar{x} |$$

Where n is number of observations, x is current observation and $\bar{x}$ is mean of x

(20) $$\text{MSE}=\frac{1}{n}\sum_{t=1}^{\infty}(x_i - \hat{x})^2$$

Where x hat is predicted value and x is observed value

(21) RMSE=square root of MSE

Based on the formulas provided above we compared our forecast models such as naïve, seasonal naïve, exponential smoothing (ETS), Autoregressive integrated moving average (ARIMA), Vector Autoregressive model (VAR). Results are provided in the table 4.11.

Based on the table 4.11 we can notice that a simple random walk model (Naïve) outperformed all other models by having the lowest errors on average across all models. Only once a random walk model was outperformed by ARIMA model in pecan series, but the error is not that significant.

CHAPTER 5

5.1 **Discussion and Conclusion**

In this study a monthly PPI of nut products were used over the period December 1999 through July 2021. Our results show that almond is the highest valued product from all observed nuts products. From the previous Chapters, it was mentioned that U.S. is a leader in almond production and highly dependent from exports and weather. In addition to that, various issues such as shortage of water contribute to high values of almond. From statistical analysis, it was found that almond, is a leader in median, mean, standard deviation (SD) and coefficient of deviation (CV) values, is followed by walnut which also has similar production process as almond but grown in a different region which reduces his values a bit compared with almond PPI. Lastly, after walnut goes pecan. Pecan is 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. The cheapest product after all other observed nut products is peanut. Peanuts statistical values are significantly lower than almonds, pecans and walnuts.

Stationarity test indicated that our series are non-stationary. A difference of four series solved this problem and made further forecasting with an autoregressive approach possible. In addition, our results showed that these four nuts are not tied together in any long run cointegration relationship. This was confirmed by results of cointegration test which was also backed by statistical test. Leading to a fact that we cannot predict long run relationship which will hold these markets together. Cointegration test produced negative results leading that vector error correction model is not possible leaving us with vector autoregressive option of forecasting only. Impulse response function and forecast error decomposition were derived from vector autoregressive analysis. From that analysis we discovered that walnut is a price follower. All its values had a delay from shocks from other nuts following with a strong magnitude response with a quick stabilization. Causal graphical analysis in differenced series also confirmed this information by showing us that walnut is a price sink which consumes information from almond, past values of pecan and its own lagged values and does not causing any effect on others. For peanut, an opposite picture is shown, peanut is a price leader affecting other nuts. Impulse response function as well confirms this by showing no delays from peanut shock and strong fluctuations after a peanut shock to other nut products. In addition to that, magnitude of peanut shocks on average are higher than any other observed nut product. Pecan and almond are in between of peanut and walnut causal relationship. So peanut disturbance in a price affects pecan then almond and lastly walnut through pecan and almond. Impulse response function confirms this by showing delay in the first period of walnut series. These delays tell us that walnut awaits information from other nut products before fluctuations. The differenced series used penalty discount at 1 indicated that walnut is a follower while peanut is a leader. Residual had a penalty discount at 0.2 showed us the following information. A causal model in differenced data from producer price index (PPI) indicated that peanut influences all other nut markets. Peanut has a direct influence on pecan market and indirect on almond and walnut markets. Pecan is influenced only by peanut and has a direct impact on almond which is also affected by past pecan values. For a walnut, past pecan values also play a role with almond series. From contemporaneous causal relationship of innovations from VAR the main findings is that new information from walnut will move peanut market by negative value (-0.102) and from pecan will move peanut market by positive value (0.269) in terms of producer price index. All the above information can be narrowed to the statement peanut sets the atmosphere in the nuts market. Therefore, understanding situation on a peanut market may give a clue of how situation may affect other nut markets. Although from forecast error decomposition it was observed that error account no more that 1-6%, but it is still possible to claim that peanut influence on other nuts exists, despite the fact it is a small one 1-2%. Impulse response function review also noted that long run relations are not presented, previously it was observed that after 6-8 months after shock, PPI tended to go back to its initial values.

This study also conducted a forecast validation by calculating MAPE. VAR model despite higher errors in MAPE than Naïve predicts and explains nuts better since it accounts for effect between each nut series which we explained previously and understood that causality is presented and causes disturbance. Another reason why naïve is generally performs better than any other forecasting models is due its nature. Naïve has simple straight-line forecast which is close to mean values leading to less errors.

In extant literature, small number of studies were conducted on forecasting of nuts product using VAR modelling. The purpose of this work is to help nut producers to navigate in this market with aid of forecasting of nut values by utilizing VAR modelling and building a causal model. Using USDA and FRED data, this study analyzed and constructed forecast error variance decomposition, which allowed us to understand how nut products interact with each other as well as potential future values of nuts.

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", 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.

United States Department of Agriculture, USDA, “2018 California Almond Objective MeasurementReport,https://www.nass.usda.gov/Statistics_by_State/California/Publications/Specialty_and_Other_Releases/Almond/Objective-Measurement/201807almom.pdf, accessed in August 2021.

California Department of Water Resources County Land Use Surveys, San Francisco County Land Use Survey 2014-2016, https://data.ca.gov/dataset/san-francisco-county-land-use-survey-2014, October 28, 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, S., Duncan, 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.

De Souza, R., Schincaglia, R. M., Pimentel, G. D., & Mota, J. F. (2017). “Nuts and Human Health Outcomes: A Systematic Review.” *Nutrients*, *9*(12), 1311. https://doi.org/10.3390/nu9121311

King, J., Blumberg, J., Ingwersen, L., Jenab, M., Tucker, K., “Tree Nuts and Peanuts as Components of a Healthy Diet”, *The Journal of Nutrition*, Volume 138, Issue 9, September 2008, Pages 1736S–1740S, https://doi.org/10.1093/jn/138.9.1736S

Kasler, Dale; Reese, Phillip; Sabalow, Ryan. "California almonds, partly blamed for water shortage, now dropping in price", The Sacramento Bee, 21 December 2018.

Ramsey, J.D., R. Sanchez-Romero, C. Glymour. Non-Gaussian methods and high-pass filters in the estimation of effective connections. *NeuroImage*, Volume 84, pp.986-1006., 2014.

John R. Pyzner, Bollich, Patricia A. “How long will it take for pecan trees to begin producing?”, 2006.

Njike, V. Y., Ayettey, R., Petraro, P., Treu, J. A., & Katz, D. L. (2015). “Walnut ingestion in adults at risk for diabetes: effects on body composition, diet quality, and cardiac risk measures.”, *BMJ open diabetes research & care*, *3*(1), e000115. https://doi.org/10.1136/bmjdrc-2015-000115

Claudia Sánchez-González, Carlos J Ciudad, Véronique Noé, Maria Izquierdo-Pulido “Health benefits of walnut polyphenols: An exploration beyond their lipid profile”, 2017 https://pubmed.ncbi.nlm.nih.gov/26713565/.

University of Delaware. "Invasive Plant Secretes Acid To Kill Nearby Plants And Spread." ScienceDaily. www.sciencedaily.com/releases/2007/10/071012084128.htm (accessed September 18, 2021).

Food and Agriculture Organization of the United Nations Statistics Division (FAOSTAT), 2013.

Hawkins, H., Dharmasena S. “Price Discovery and Integration of the Peanut Markets in the United States” Vol. 50, Issue 2, November 2019.

Kim, Y., Dharmasena S., “Price Discovery and Integration in U.S. Pecan Markets” Vol. 49, Issue 1, March 2018.

King JC, Blumberg J, Ingwersen L, Jenab M, Tucker KL, J Nutr, “Tree nuts and peanuts as components of a healthy diet.”, 2008. ;138(9):1736S-1740S. doi: 10.1093/jn/138.9.1736S. PMID: 18716178.

Craymer, Lucy. "U.S. Almond Farmers Are Reeling from Chinese Tariffs", Wall Street Journal, 2018.

Gonzales, "How Almonds Became a Scapegoat For California's Drought", NPR, 2018.

Pickett, "In The Midst Of Drought, California Farmers Used More Water For Almonds", Forbes, 2018.

Bjerga, "California Almonds Are Back After Four Years Of Brutal Drought", Bloomberg, 2018.

Davis, "The Billion-Dollar California Almond Industry's Blossoming Future", PasteMagazine, 2017.

Peanut Products, The Peanut Institute, https://peanut-institute.com/, 2016.

American Peanut Council, Membership Report 2013-2014, https://www.peanutsusa.com/phocadownload/AnnualReport/2014%20Membership%20Report.pdf

Boiled Peanut World, 2013, “Boiled Peanuts, Steamed Peanuts, Braised Peanuts”, https://www.boiled-peanut-world.com/

D.L Wright, B.L. Tillman, “The Peanut Industry – Peanuts: a Brief History, American Peanut Council”, University of Florida, 2014.

Crop Production Annual Summary, National Ag Statistics Service (NASS), USDA, https://usda.library.cornell.edu/concern/publications/k3569432s, 2018.

Food and Agriculture Organization of the United Nations Statistics Division (FAOSTAT), 2013. Click Item as Groundnuts with shell, Area as United States, and From Year 2013 To Year 2013.

Global Ag Trade System, Foreign Ag Service (FAS), USDA, https://apps.fas.usda.gov/gats/default.aspx?publish=1, 2015.

Herkes, J., “Rapeseed and Canola for Biodiesel Production”, Extension.org, 2014.

National Peanut Board, “How peanut grow?”, https://www.nationalpeanutboard.org/peanut-info/how-peanuts-grow.htm, 2015.

Orchant, R., “Boiled Peanuts: The Southern Snack We Adore”, The Huffington Post, 2014.

Hampton Farms, “Our Products”, https://www.agmrc.org/commodities-products/nuts/peanut-profile, 2016.

USDA – Farm Service Agency (FSA), Peanut Marketing Assistance Loans and Loan Deficiency Payments, https://www.fsa.usda.gov/Internet/FSA_File/peanuts.pdf, 2015.

National Peanut Board, “Peanut Types”, https://www.nationalpeanutboard.org/, 2014

Florida Cooperative Extension Service, Institute of Food and Agricultural Sciences, University of Florida, “Production of Biofuel Crops in Florida: Peanut”, 2010.

Wells, L., "Response of Pecan Orchard Soil Chemical and Biological Quality Indicators to Poultry Litter Application and Clover Cover Crops.", February 2011

Market Research Persistence, 2020, https://www.persistencemarketresearch.com/.

Fatima, T., Showkat, U. and Hussain, S.Z., “Nutritional and Health Benefits of Walnuts. Journal of Pharmacognosy and Phytochemistry”, 7, 1269-1271., 2018.

Shafer, Carl E., "Pecan Production And Price Trends 1979-1995," Faculty Paper Series 24019, Texas A&M University, Department of Agricultural Economics, 1996.

APPENDIX A

TABLES

Table 2. 1 Descriptive Statistic

Nut Type	Mean	Rank by mean	Median	Rank by median	SD	Rank by SD	CV	Rank by CV
Peanut	73.82057	4	70.98121	4	14.42454	4	0.1954	4
Almond	251.2836	1	233.9	1	104.0581	1	0.4141062	1
Pecan	157.6674	3	158.4781	3	48.22074	3	0.3058383	3
Walnut	212.0744	2	204	2	77.12994	2	0.3636928	2

Table 4. 1 Cointegration Test

Number of Cointegrations	Trace test value	Cutoff value at 5%	Decision	Cutoff value at 1%	Decision
r=0	55.34	53.12	R	60.16	F
r=1	28.31	34.91	F	41.07	F
r=2	12.75	19.96	F	24.60	F
r=3	3.34	9.24	F	12.97	F

Table 4. 2 Lag Selection of VAR model[1]

Lag	1		2		3		4
AIC(n)	1.98E+01	AIC(n)	1.98E+01	AIC(n)	1.99E+01	AIC(n)	1.99E+01
HQ(n)	1.99E+01	HQ(n)	2.01E+01	HQ(n)	2.02E+01	HQ(n)	2.04E+01
SC(n)	2.01E+01	SC(n)	2.04E+01	SC(n)	2.07E+01	SC(n)	2.10E+01
FPE(n)	4.04E+08	FPE(n)	4.13E+08	FPE(n)	4.27E+08	FPE(n)	4.59E+08

Lag	5		6		7		8
AIC(n)	2.00E+01	AIC(n)	2.01E+01	AIC(n)	2.02E+01	AIC(n)	2.03E+01
HQ(n)	2.06E+01	HQ(n)	2.08E+01	HQ(n)	2.09E+01	HQ(n)	2.11E+01
SC(n)	2.13E+01	SC(n)	2.17E+01	SC(n)	2.20E+01	SC(n)	2.23E+01
FPE(n)	5.05E+08	FPE(n)	5.61E+08	FPE(n)	5.92E+08	FPE(n)	6.61E+08

Lag	9		10		11		12
AIC(n)	2.04E+01	AIC(n)	2.05E+01	AIC(n)	2.04E+01	AIC(n)	2.04E+01
HQ(n)	2.13E+01	HQ(n)	2.15E+01	HQ(n)	2.16E+01	HQ(n)	2.16E+01
SC(n)	2.27E+01	SC(n)	2.30E+01	SC(n)	2.32E+01	SC(n)	2.34E+01
FPE(n)	7.33E+08	FPE(n)	7.74E+08	FPE(n)	7.70E+08	FPE(n)	7.64E+08

[1] AIC – Akaike Information Criterion
HQ – Hannan Quinn
SC – Schwarz Criterion
FPE – Final Prediction Error

Table 4. 3 Almond VAR equation

Variables	Estimate	Std. Error	Pr(>|t|)
Almond. L1	-0.1652	0.06619	0.0133*
Peanut L1	0.08881	0.09455	0.3486
Pecan L1	0.16763	0.1488	0.2612
Walnut L1	-0.02238	0.06562	0.7334
Constant	1.67209	1.00458	0.0974
Sd1	-6.98769	5.04345	0.1673
Sd2	-12.2396	5.03446	0.0159*
Sd3	-12.3597	5.0309	0.0148*
Sd4	-9.3433	4.97695	0.0618
Sd5	-9.75442	4.94288	0.0497*
Sd6	-9.5008	4.96695	0.0571
Sd7	-8.87352	4.98803	0.0766
Sd8	-6.31186	4.94244	0.2029
Sd9	-7.60336	4.94513	0.1256
Sd10	-10.4099	4.99168	0.0382*
Sd11	-7.73327	5.08649	0.1299

Almond L1 is a lagged almond variable

Peanut L1 is a lagged peanut variable

Pecan L1 is a lagged pecan variable

Walnut is a lagged walnut variable

Sd is a seasonal dummy variable

* - significance band on p-value 0.05

Table 4. 4 Peanut VAR equation

Variables	Estimate	Std. Error	Pr(>|t|)
Almond. L1	-0.00105	0.028745	0.971
Peanut L1	-0.02988	0.041059	0.468
Pecan L1	-0.29041	0.06462	1.13e-05 ***
Walnut L1	-0.02599	0.028498	0.363
Constant	-0.0506	0.436266	0.908
Sd1	-1.15305	2.190258	0.599
Sd2	0.639871	2.186353	0.77
Sd3	-3.25664	2.18481	0.138
Sd4	-1.85667	2.161377	0.391
Sd5	-0.33973	2.146585	0.874
Sd6	0.84999	2.157036	0.694
Sd7	-0.52602	2.166189	0.808
Sd8	-3.71364	2.146392	0.085
Sd9	-2.99883	2.14756	0.164
Sd10	2.788825	2.167774	0.2
Sd11	-1.10406	2.208949	0.618

Almond L1 is a lagged almond variable

Peanut L1 is a lagged peanut variable

Pecan L1 is a lagged pecan variable

Walnut is a lagged walnut variable

Sd is a seasonal dummy variable

* - significance band on p-value 0.05

Table 4. 5 Pecan VAR equation

Variables	Estimate	Std. Error	Pr(>|t|)
Almond. L1	-0.10444	0.047277	0.0282 *
Peanut L1	-0.08065	0.067529	0.2336
Pecan L1	0.024064	0.10628	0.8211
Walnut L1	0.00047	0.04687	0.992
Constant	0.57879	0.717515	0.4207
Sd1	2.483337	3.602265	0.4913
Sd2	-4.01007	3.595842	0.266
Sd3	-1.76965	3.593305	0.6229
Sd4	-0.48782	3.554766	0.891
Sd5	-0.85287	3.530437	0.8093
Sd6	-2.72813	3.547625	0.4427
Sd7	-0.57935	3.562679	0.871
Sd8	-0.56245	3.530119	0.8736
Sd9	1.153519	3.53204	0.7443
Sd10	-0.59916	3.565285	0.8667
Sd11	3.996585	3.633005	0.2725

Almond L1 is a lagged almond variable

Peanut L1 is a lagged peanut variable

Pecan L1 is a lagged pecan variable

Walnut is a lagged walnut variable

Sd is a seasonal dummy variable

* - significance band on p-value 0.05

Table 4. 6 Walnut VAR equation

Variables	Estimate	Std. Error	Pr(>|t|)
Almond. L1	-0.03521	0.06705	0.60003
Peanut L1	0.01371	0.09577	0.88634
Pecan L1	0.42811	0.15073	0.00493 **
Walnut L1	0.02553	0.06648	0.70131
Constant	0.3746	1.01764	0.71314
Sd1	-13.942	5.10902	0.00687 **
Sd2	-0.96838	5.09991	0.84958
Sd3	-4.86231	5.09631	0.34109
Sd4	2.10372	5.04165	0.67689
Sd5	-2.50989	5.00715	0.61669
Sd6	3.23896	5.03153	0.52042
Sd7	0.22351	5.05288	0.96476
Sd8	-1.41176	5.0067	0.77823
Sd9	1.0029	5.00942	0.84151
Sd10	0.2704	5.05657	0.9574
Sd11	0.3904	5.15262	0.93967

Almond L1 is a lagged almond variable

Peanut L1 is a lagged peanut variable

Pecan L1 is a lagged pecan variable

Walnut is a lagged walnut variable

Sd is a seasonal dummy variable

* - significance band on p-value 0.05

Table 4. 7 Forecast Error Variance Decomposition for Almonds

Period	Almond	Pecan	Peanut	Walnut
1	1	0	0	0
2	0.989099	0.005733	0.004687	0.000481
3	0.987298	0.006303	0.005918	0.000481
4	0.987138	0.006336	0.006042	0.000484
5	0.987128	0.006337	0.00605	0.000484
24	0.987128	0.006337	0.006051	0.000484

Table 4. 8 Forecast Error Variance Decomposition for Peanuts

Period	Almond	Pecan	Peanut	Walnut
1	0.004074	0.02259825	0.9733275	0.00E+00
2	0.004059	0.02713913	0.9655347	3.27E-03
3	0.004055	0.02760675	0.9648607	3.48E-03
4	0.004066	0.02763185	0.9648105	3.49E-03
5	0.004068	0.02763277	0.9648071	3.49E-03
24	0.004068	0.02763279	0.9648069	0.003492

Table 4. 9 Forecast Error Variance Decomposition for Pecans

Period	Almond	Pecan	Peanut	Walnut
1	0.00431	0.99569	0	0
2	0.0238	0.97599	0.0002	4.22E-07
3	0.02511	0.97445	0.00044	5.84E-06
4	0.02514	0.97436	0.00049	5.88E-06
5	0.02514	0.97435	0.0005	5.99E-06
24	0.02514	0.97435	0.0005	6.01E-06

Table 4. 10 Forecast Error Variance Decomposition for Walnuts

Period	Almond	Pecan	Peanut	Walnut
1	1.86E-07	0.002794	0.020772	0.976434
2	2.17E-03	0.043334	0.019872	0.934628
3	2.23E-03	0.045531	0.019961	0.932273
4	2.24E-03	0.045622	0.019971	0.932172
5	2.24E-03	0.045625	0.019971	0.932169
24	2.24E-03	0.045625	0.019971	0.932169

Table 4. 11 Forecast Validation Exercise.[2]

MAPE	almond	peanut	pecan	walnut
ARIMA	0.15646	0.06529	0.05206	0.41539
Naïve	0.12779	0.03684	0.05338	0.15847
Snaive	0.17174	0.14443	0.09396	0.28631
VAR	0.2334	0.03844	0.09643	0.12481
ETS	0.16372	0.05616	0.06277	0.20544

MAD	almond	peanut	pecan	walnut
ARIMA	73.3071	5.97907	34.8399	62
Naïve	55.1667	3.84583	38.0458	30.3
Snaive	64.1833	10.2125	73.0125	67.3833
VAR	77.0677	3.8062	64.0825	30.3643
ETS	75.4501	5.39358	44.6538	36.974

MSE	almond	peanut	pecan	walnut
ARIMA	10465	64.4048	3483.62	5243.93
Naïve	5672.08	25.9904	4411.22	1704.99
Snaive	5919.27	272.596	8001.72	5837.46
VAR	9666.96	23.1665	7256.16	1679.65
ETS	9553.97	39.1582	4916.05	2344.27

RMSE	almond	peanut	pecan	walnut
ARIMA	102.299	8.02526	59.0222	72.415
Naïve	75.2132	5.09808	66.417	41.2916
Snaive	76.9368	16.5105	89.4523	76.4033
VAR	98.3207	4.81316	85.1831	50.9836
ETS	97.7444	6.25765	70.1146	48.4176

[2] ARIMA – Autoregressive moving average
Snaive- Seasonal Naïve
ETS -Exponential smoothing

APPENDIX B

Figures

Figure 2. 1 Utilized production

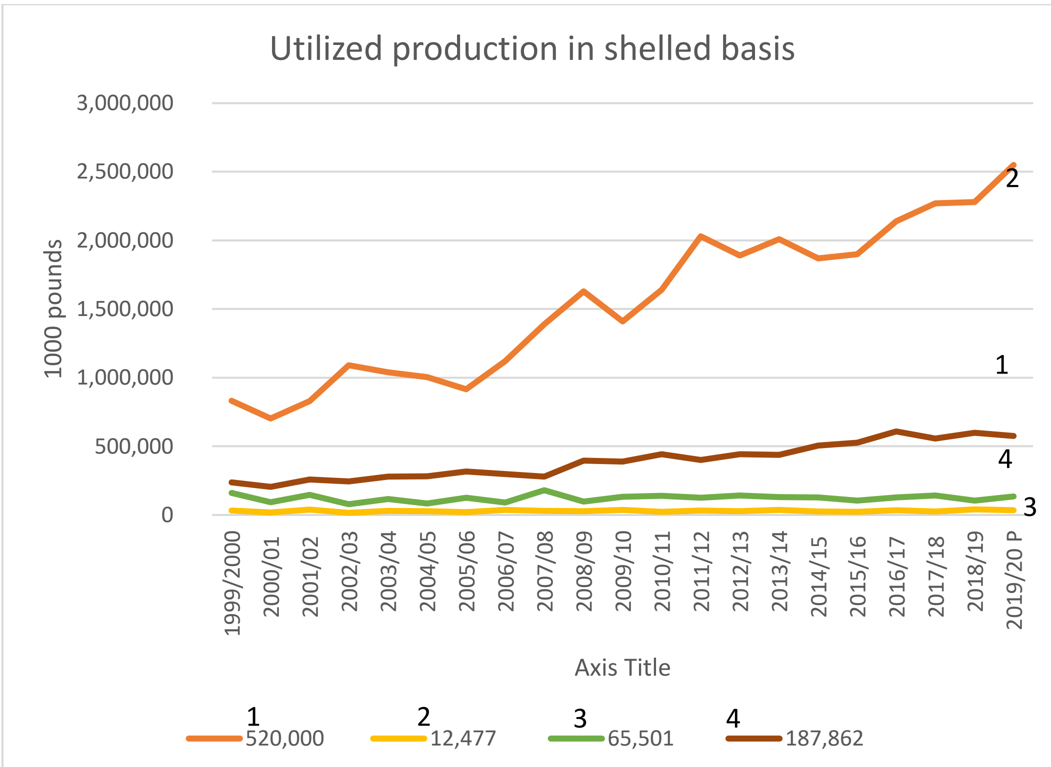
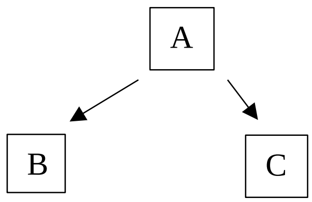


Figure 3. 1 Causal structures

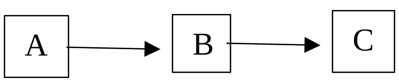
Common Cause



In this case we have that A causes B and C.

Figure 3.2

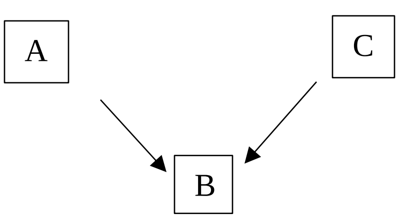
Chain



Here we have A causes B and B causes C.

Figure 3.3

Collider



The last model we have A and C causes B.

Figure 4. 1 Almond PPI

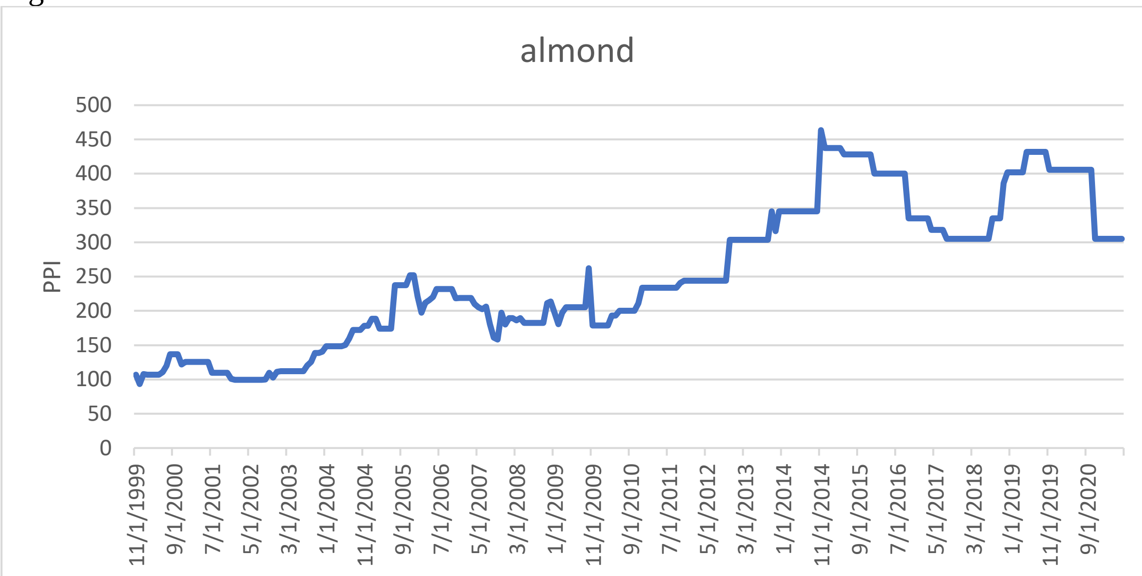


Figure 4. 2 Pecan PPI

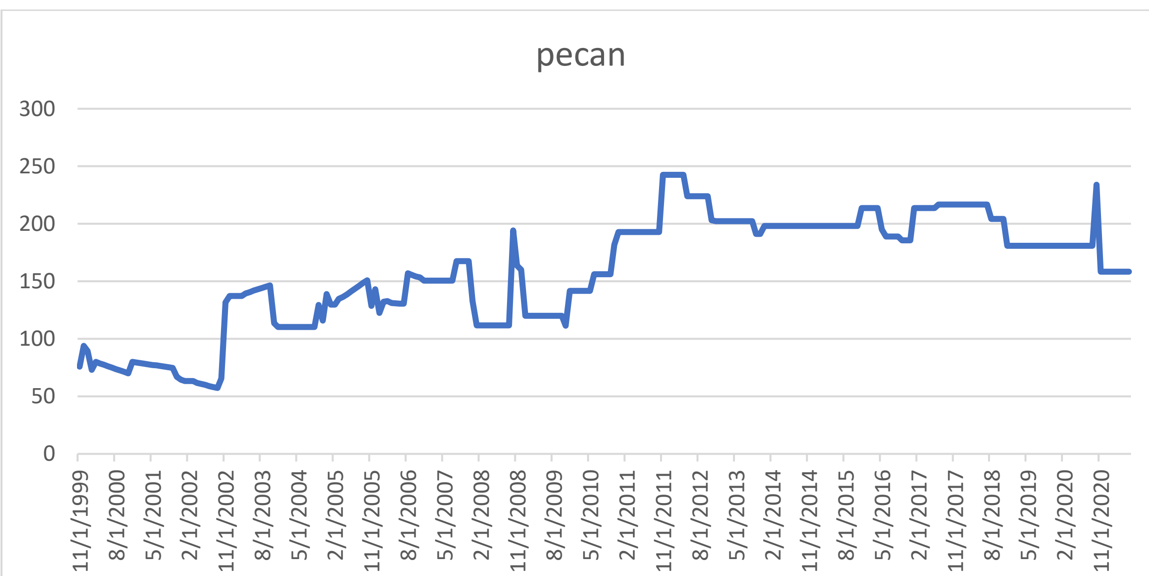


Figure 4. 3 Peanut PPI

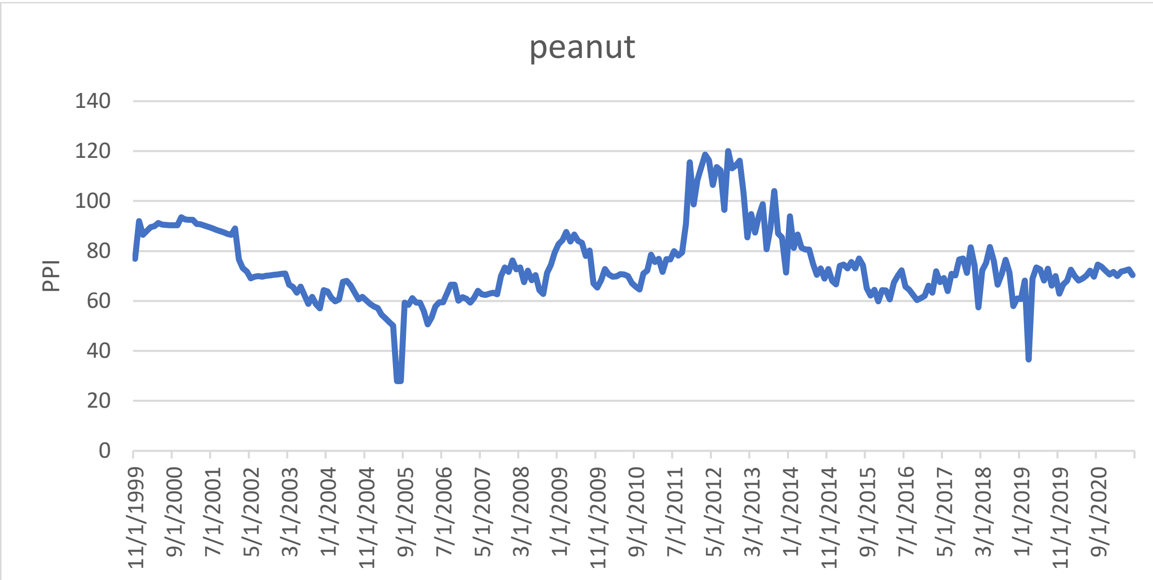


Figure 4. 4 Walnut PPI

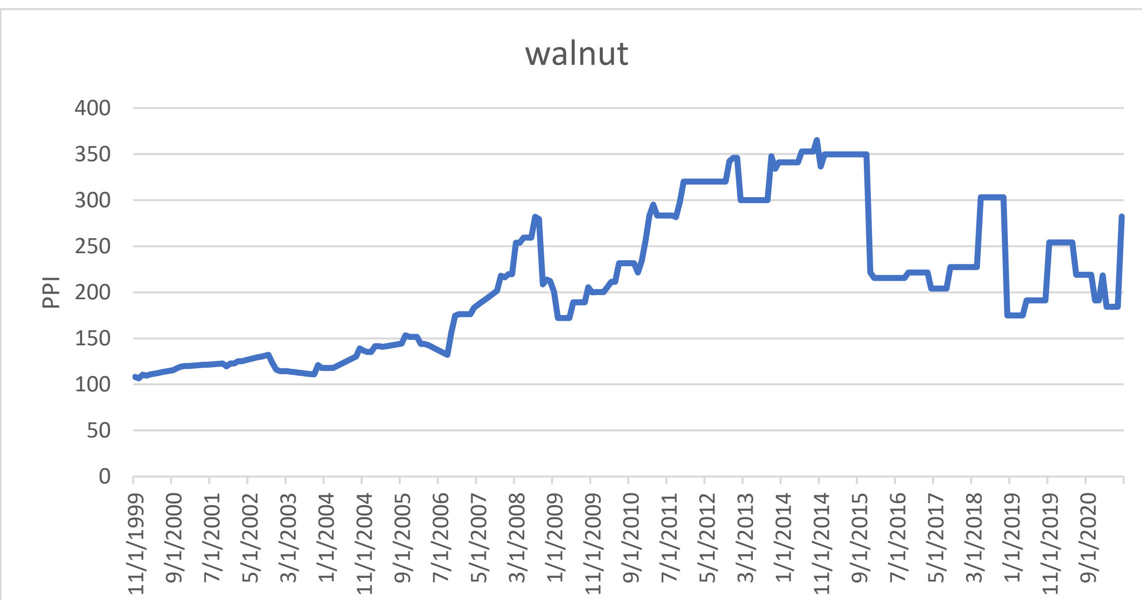


Figure 4. 5 Contemporaneous causal relationship of innovations from VAR

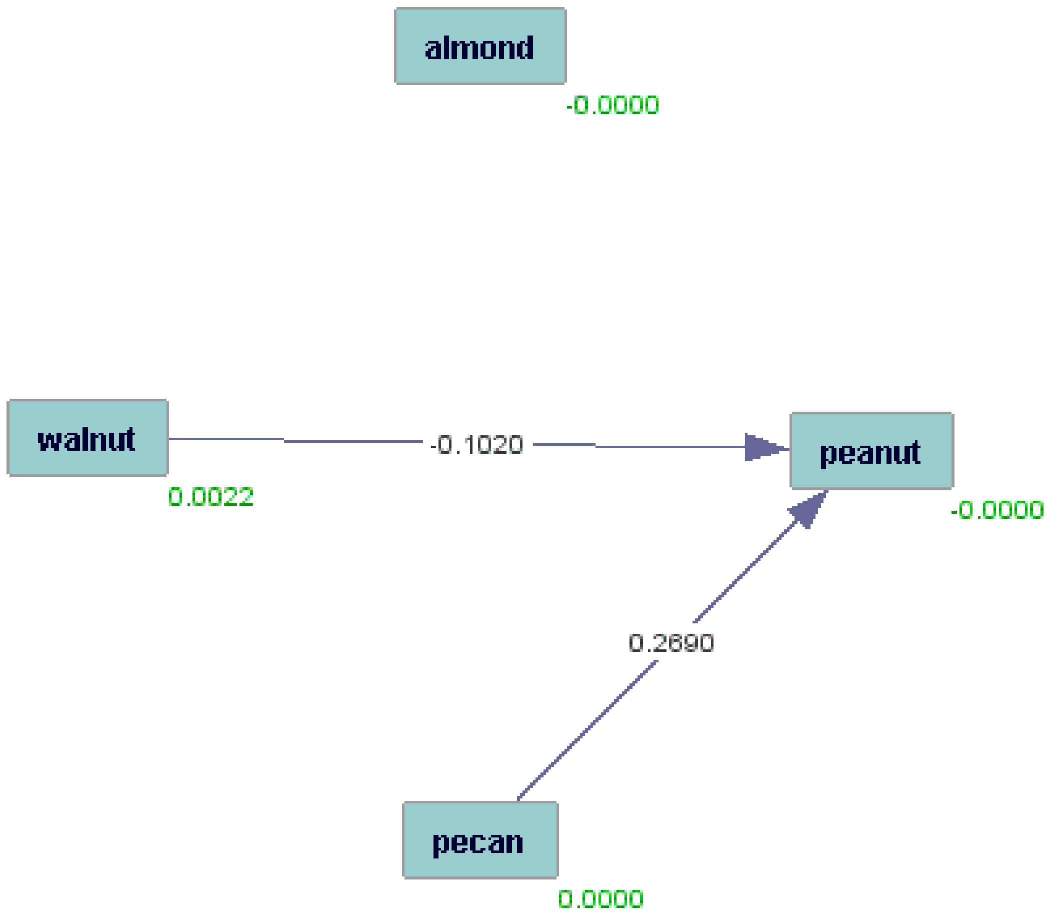


Figure 4. 6 Causal relationship of differenced data

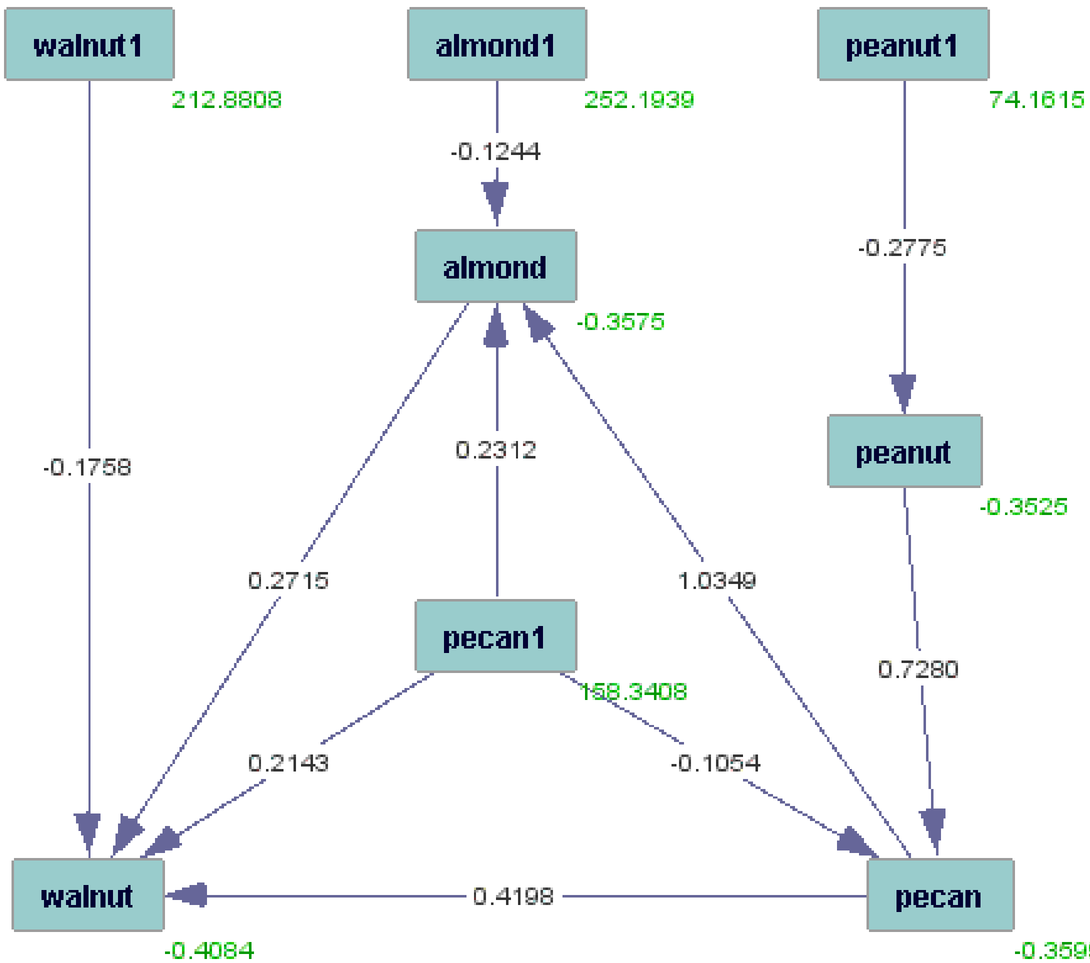
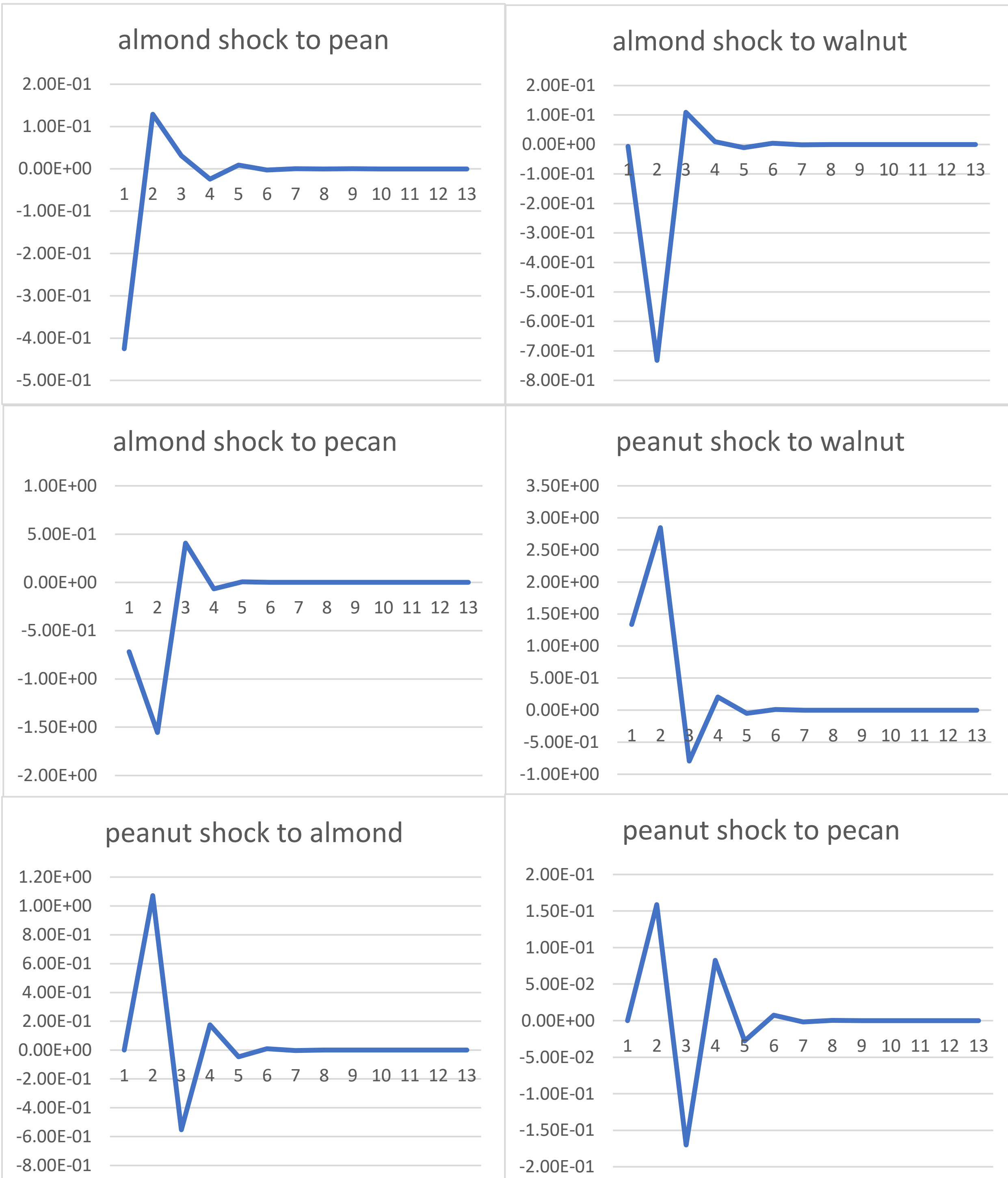
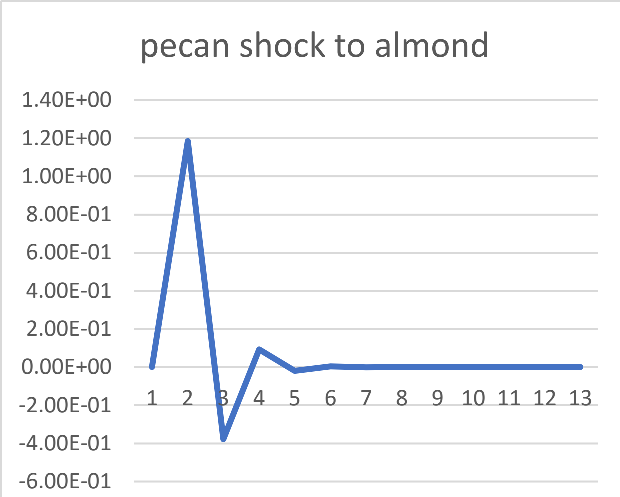
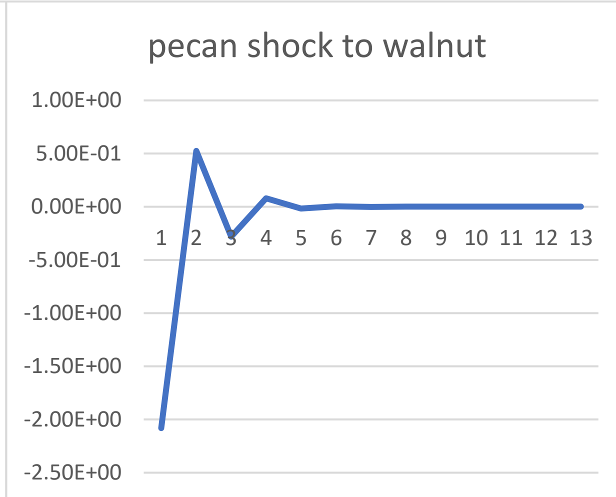
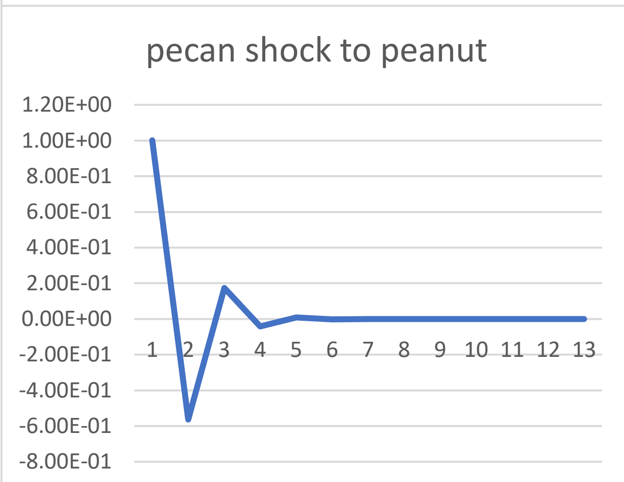


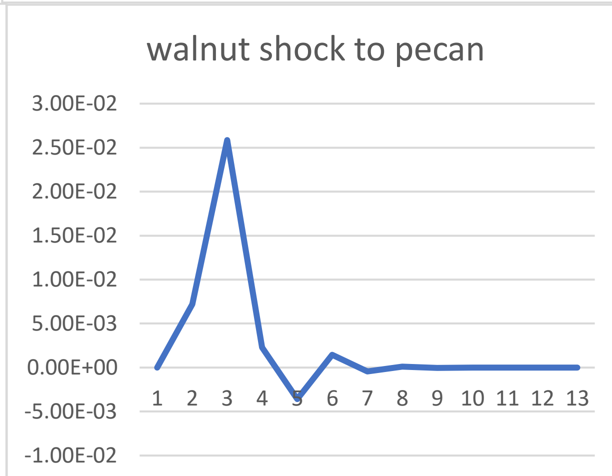
Figure 4. 7 Impulse Response Functions of all four nut products

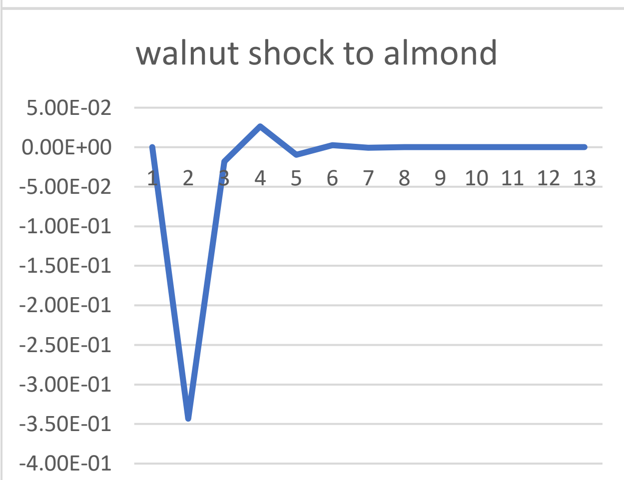



pecan shock to almond
1.40E+00
1.20E+00
1.00E+00
8.00E-01
6.00E-01
4.00E-01
2.00E-01
0.00E+00
-2.00E-01
-4.00E-01
-6.00E-01
1 2 3 4 5 6 7 8 9 10 11 12 13


pecan shock to walnut
1.00E+00
5.00E-01
0.00E+00
-5.00E-01
-1.00E+00
-1.50E+00
-2.00E+00
-2.50E+00
1 2 3 4 5 6 7 8 9 10 11 12 13


pecan shock to peanut
1.20E+00
1.00E+00
8.00E-01
6.00E-01
4.00E-01
2.00E-01
0.00E+00
-2.00E-01
-4.00E-01
-6.00E-01
-8.00E-01
1 2 3 4 5 6 7 8 9 10 11 12 13


walnut shock to pecan
3.00E-02
2.50E-02
2.00E-02
1.50E-02
1.00E-02
5.00E-03
0.00E+00
-5.00E-03
-1.00E-02
1 2 3 4 5 6 7 8 9 10 11 12 13


walnut shock to almond
5.00E-02
0.00E+00
-5.00E-02
-1.00E-01
-1.50E-01
-2.00E-01
-2.50E-01
-3.00E-01
-3.50E-01
-4.00E-01
1 2 3 4 5 6 7 8 9 10 11 12 13

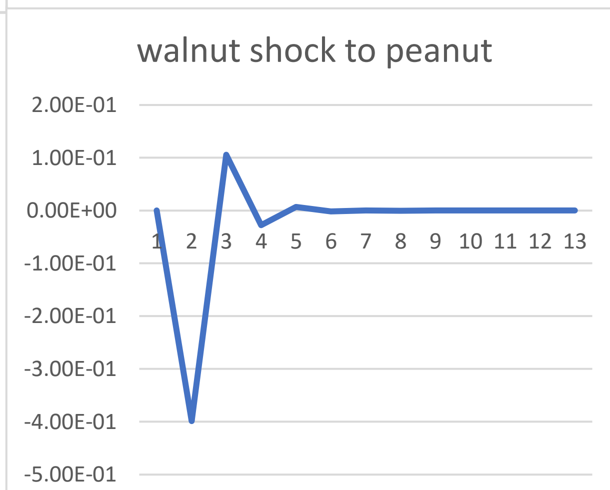

walnut shock to peanut
2.00E-01
1.00E-01
0.00E+00
-1.00E-01
-2.00E-01
-3.00E-01
-4.00E-01
-5.00E-01
1 2 3 4 5 6 7 8 9 10 11 12 13

Figure 4. 8 VAR model Almond forecast

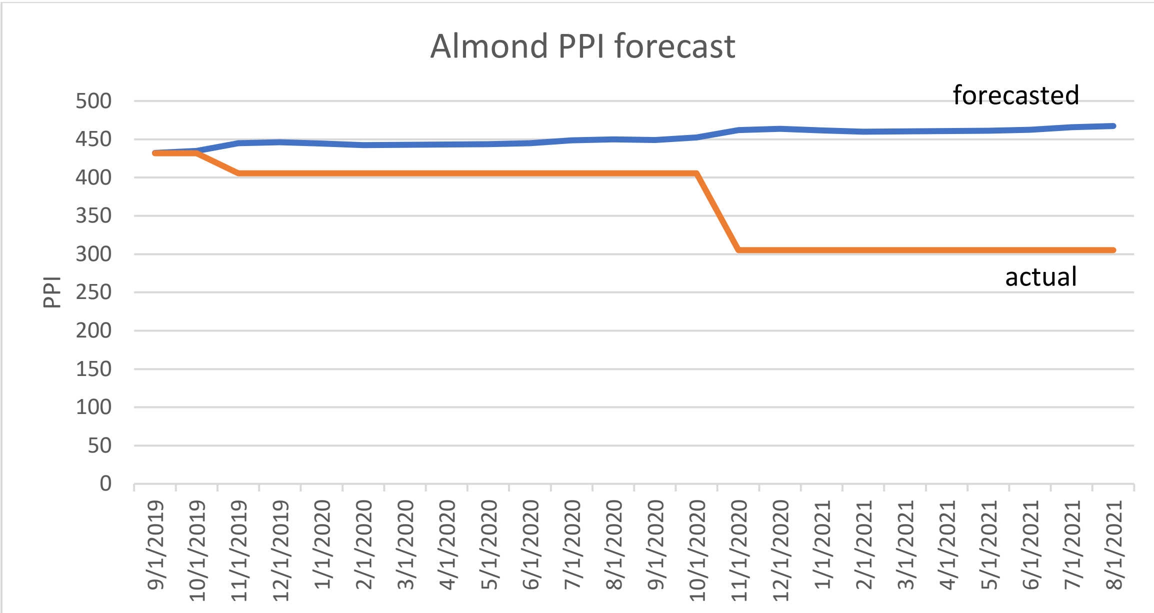


Figure 4. 9 VAR model Peanut forecast

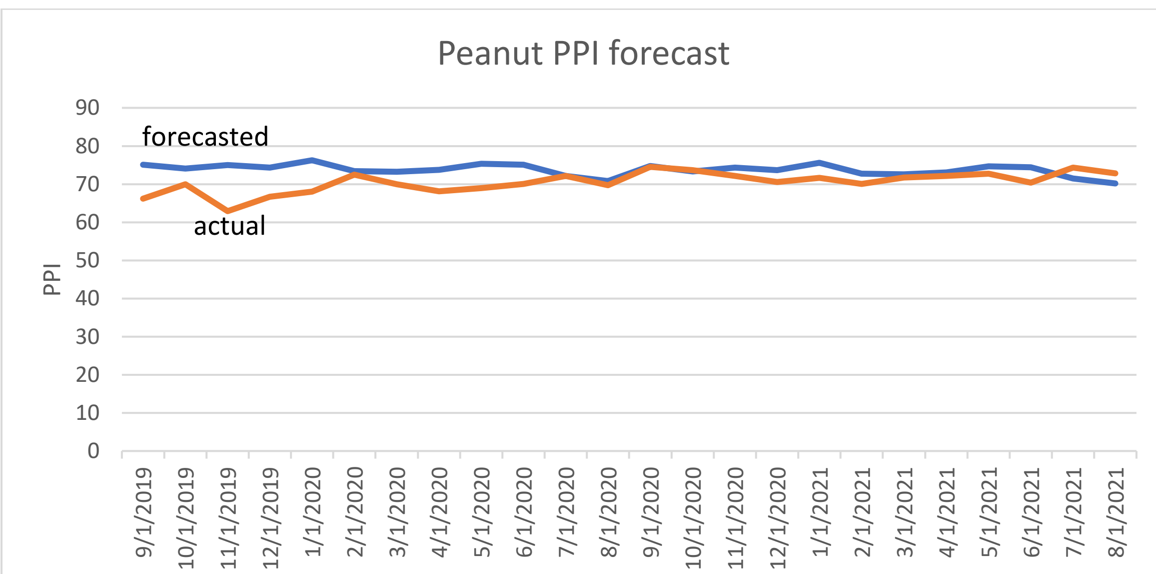


Figure 4. 10 VAR model Pecan forecast

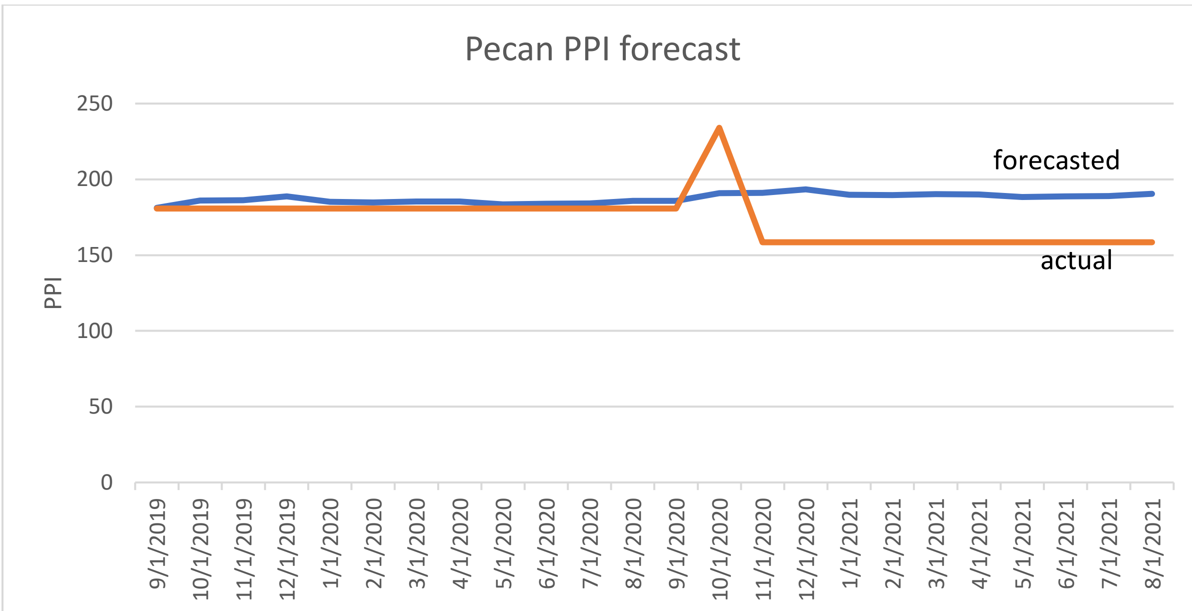
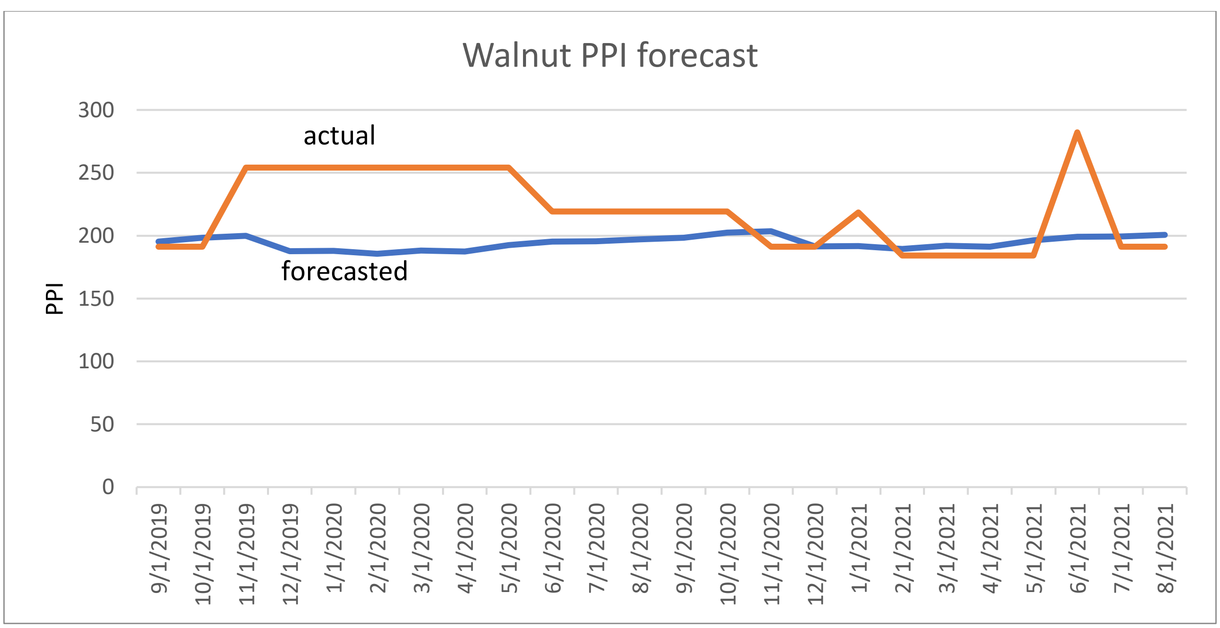


Figure 4. 11 VAR model Walnut forecast



APPENDIX C

Input Programs Used in the Thesis (in R)

```
data <- read_csv("updated.csv")

almor <- ts(data$almond, start = c(1999,11), frequency = 12)

peanr <- ts(data$peanut, start = c(1999,11), frequency = 12)

pecar <- ts(data$pecan, start = c(1999,11), frequency = 12)

walnr <- ts(data$walnut, start = c(1999,11), frequency = 12)


overall <-cbind(almor, peanr, pecar, walnr)

cor(overall, method = c("pearson"))



fit_mstl_almo<-mstl(almor)

autoplot(fit_mstl_almo)+ggtitle("mstl decomposition of almond")

#


#

fit_mstl_pean<-mstl(peanr)

autoplot(fit_mstl_pean)+ggtitle("mstl decomposition of peanut")

#

fit_mstl_peca<-mstl(pecar)

autoplot(fit_mstl_peca)+ggtitle("mstl decomposition of pecan")

#
```

```
fit_mstl_waln<-mstl(walnr)
autoplot(fit_mstl_waln)+ggtitle("mstl decomposition of walnut")

#Taking difference series
almo<-diff(almor)
pean<-diff(peanr)
peca<-diff(pecar)
waln<-diff(walnr)

#Testing for stationarity
adf.test(almo)
adf.test(pean)
adf.test(peca)
adf.test(waln)
#creating subsets
train.almo <-subset(almo, end=length(almo)-24)
train.pean <-subset(pean, end=length(pean)-24)
train.peca <-subset(peca, end=length(peca)-24)
train.waln <-subset(waln, end=length(waln)-24)

test.almo <-subset(almo, start=length(almo)-23)
test.pean <-subset(pean, start=length(almo)-23)
test.peca <-subset(peca, start=length(almo)-23)
```

```
test.waln <-subset(waln, start=length(almo)-23)


autoplot(overall)


###########


pp.test(almo)

pp.test(pean)

pp.test(peca)

pp.test(waln)

#confirming stationarity of new subsets

adf.test(train.almo)

adf.test(train.pean)

adf.test(train.peca)

adf.test(train.waln)

pp.test(test.almo)

pp.test(test.pean)

pp.test(test.peca)

pp.test(test.waln)


##################################################

#Everything is stationary

##################################################
```

```
cb <-cbind(train.almo, train.pean, train.peca, train.waln)

colnames(cb) <- cbind("almo","pean", "peca", "waln")

#

tt <-cbind(test.almo, test.pean, test.peca, test.waln)

colnames(tt) <- cbind("almo.t","pean.t", "peca.t", "waln.t")

##################################################

#var model

##################################################

lagselect <- VARselect(cb, lag.max = 15, type = "const")

lagselect$selection

lagselect$criteria

####################################################

#adding seasonality to VAR

####################################################

Var_model1 <- VAR(cb, p = 1, type = "const", season = 12, exog = NULL)

Var_model1

summary(Var_model1)


# Forecast of VAR

forecast <- predict(Var_model1, n.ahead = 24)

forecast


###################################################
```

```
accuracy(almo.f, almo.a)

accuracy(pean.f, pean.a)

accuracy(peca.f, peca.a)

accuracy(waln.f, waln.a)



##############################

#Error decomposition

###############################

fore_nut <- fevd(Var_model1, n.ahead = 24)

fore_nut

plot(fore_nut)


#######################################

#extracting data for DAG model########

#######################################

residuals(Var_model1)

#######################################


##################

plot(forecast)

print(forecast)
```

```
fanchart(forecast, names = "almo", main = "Fanchart for almo", xlab = "Horizon", ylab =
"almo")
fanchart(forecast, names = "pean", main = "Fanchart for pean", xlab = "Horizon", ylab = "pean")
fanchart(forecast, names = "peca", main = "Fanchart for peca", xlab = "Horizon", ylab = "peca")
fanchart(forecast, names = "waln", main = "Fanchart for waln", xlab = "Horizon", ylab = "waln")
forecast
# forecast issue testing
###########


Serial1 <- serial.test(Var_model1, lags.pt = 12, type = "PT.asymptotic")
Serial1
#no serial correlation from test


Norm1 <- normality.test(Var_model1, multivariate.only = TRUE)
Norm1

####
#didnt pass normaility
#

Stability1 <- stability(Var_model1, type = "OLS-CUSUM")
```

```
plot(Stability1)

####

#no structural breaks was found

#Impulse Response Function


almo_irf <- irf(Var_model1, impulse = "almo", response = "almo", n.ahead = 12, boot = TRUE)

plot(almo_irf, ylab = "almo", main = "almo's shock to almo")

pean_irf <- irf(Var_model1, impulse = "almo", response = "pean", n.ahead = 12, boot = TRUE)

plot(pean_irf, ylab = "pean", main = "almo's shock to pean")

peca_irf <- irf(Var_model1, impulse = "almo", response = "peca", n.ahead = 12, boot = TRUE)

plot(peca_irf, ylab = "peca", main = "almo's shock to peca")

waln_irf <- irf(Var_model1, impulse = "almo", response = "waln", n.ahead = 12, boot = TRUE)

plot(waln_irf, ylab = "waln", main = "almo's shock to waln")

print(almo_irf)

print(pean_irf)

print(peca_irf)

print(waln_irf)

#######

#repeating

########

almo_irf <- irf(Var_model1, impulse = "pean", response = "almo", n.ahead = 12, boot = TRUE)

plot(almo_irf, ylab = "almo", main = "pean's shock to almo")

pean_irf <- irf(Var_model1, impulse = "pean", response = "pean", n.ahead = 12, boot = TRUE)
```

```
plot(pean_irf, ylab = "pean", main = "pean's shock to pean")
peca_irf <- irf(Var_model1, impulse = "pean", response = "peca", n.ahead = 12, boot = TRUE)
plot(peca_irf, ylab = "peca", main = "pean's shock to peca")
waln_irf <- irf(Var_model1, impulse = "pean", response = "waln", n.ahead = 12, boot = TRUE)
plot(waln_irf, ylab = "waln", main = "pean's shock to waln")
print(almo_irf)
print(pean_irf)
print(peca_irf)
print(waln_irf)
########

almo_irf <- irf(Var_model1, impulse = "peca", response = "almo", n.ahead = 12, boot = TRUE)
plot(almo_irf, ylab = "almo", main = "peca's shock to almo")
pean_irf <- irf(Var_model1, impulse = "peca", response = "pean", n.ahead = 12, boot = TRUE)
plot(pean_irf, ylab = "pean", main = "peca's shock to pean")
peca_irf <- irf(Var_model1, impulse = "peca", response = "peca", n.ahead = 12, boot = TRUE)
plot(peca_irf, ylab = "peca", main = "peca's shock to peca")
waln_irf <- irf(Var_model1, impulse = "peca", response = "waln", n.ahead = 12, boot = TRUE)
plot(waln_irf, ylab = "waln", main = "peca's shock to waln")
print(almo_irf)
print(pean_irf)
print(peca_irf)
print(waln_irf)
```

```
#########
almo_irf <- irf(Var_model1, impulse = "waln", response = "almo", n.ahead = 12, boot = TRUE)
plot(almo_irf, ylab = "almo", main = "waln's shock to almo")
pean_irf <- irf(Var_model1, impulse = "waln", response = "pean", n.ahead = 12, boot = TRUE)
plot(pean_irf, ylab = "pean", main = "waln's shock to pean")
peca_irf <- irf(Var_model1, impulse = "waln", response = "peca", n.ahead = 12, boot = TRUE)
plot(peca_irf, ylab = "peca", main = "waln's shock to peca")
waln_irf <- irf(Var_model1, impulse = "waln", response = "waln", n.ahead = 12, boot = TRUE)
plot(waln_irf, ylab = "waln", main = "waln's shock to waln")
print(almo_irf)
print(pean_irf)
print(peca_irf)
print(waln_irf)
#####
#analysis of supply
#####
############################################################
############################################################
############################################################
```

```
#Output of VAR

VAR Estimation Results:

=========================

Endogenous variables: almo, pean, peca, waln

Deterministic variables: const

Sample size: 231

Log Likelihood: -3926.248

Roots of the characteristic polynomial:

0.1903 0.1903 0.06721 0.06721

Call:

VAR(y = cb, p = 1, type = "const", season = 12L, exogen = NULL)


Estimation results for equation almo:

=====================================

almo = almo.l1 + pean.l1 + peca.l1 + waln.l1 + const + sd1 + sd2 + sd3 + sd4 + sd5 + sd6 + sd7
+ sd8 + sd9 + sd10 + sd11


        Estimate Std. Error t value Pr(>|t|)

almo.l1  -0.17051    0.06718  -2.538   0.0119 *

pean.l1   0.06334    0.11640   0.544   0.5869

peca.l1   0.02425    0.02416   1.004   0.3167

waln.l1  -0.02271    0.06601  -0.344   0.7312

const     1.55706    1.01946   1.527   0.1281
```

sd1 -6.73086 5.07345 -1.327 0.1860

sd2 -12.12081 5.06035 -2.395 0.0175 *

sd3 -12.09065 5.06249 -2.388 0.0178 *

sd4 -9.43843 5.00193 -1.887 0.0605 .

sd5 -9.76931 5.05284 -1.933 0.0545 .

sd6 -10.71084 5.04216 -2.124 0.0348 *

sd7 -8.79909 5.08304 -1.731 0.0849 .

sd8 -6.00177 5.02959 -1.193 0.2341

sd9 -7.57986 5.03116 -1.507 0.1334

sd10 -10.41092 5.01549 -2.076 0.0391 *

sd11 -7.29990 5.12891 -1.423 0.1561

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 15.41 on 215 degrees of freedom

Multiple R-Squared: 0.07668,Adjusted R-squared: 0.01226

F-statistic: 1.19 on 15 and 215 DF, p-value: 0.2808

Estimation results for equation pean:

=====================================

pean = almo.l1 + pean.l1 + peca.l1 + waln.l1 + const + sd1 + sd2 + sd3 + sd4 + sd5 + sd6 + sd7 + sd8 + sd9 + sd10 + sd11

```
         Estimate Std. Error t value Pr(>|t|)
almo.l1  0.001548   0.040550   0.038  0.96957
pean.l1 -0.249154   0.070261  -3.546  0.00048 ***
peca.l1 -0.012107   0.014584  -0.830  0.40735
waln.l1 -0.036707   0.039847  -0.921  0.35798
const   -0.320398   0.615357  -0.521  0.60313
sd1     -1.845534   3.062385  -0.603  0.54738
sd2      0.831933   3.054476   0.272  0.78560
sd3     -4.887986   3.055766  -1.600  0.11116
sd4     -2.595014   3.019216  -0.859  0.39102
sd5     -2.398437   3.049945  -0.786  0.43251
sd6      0.273559   3.043496   0.090  0.92846
sd7     -0.999882   3.068176  -0.326  0.74483
sd8     -5.285606   3.035907  -1.741  0.08311 .
sd9     -4.674529   3.036857  -1.539  0.12521
sd10     4.009066   3.027397   1.324  0.18682
sd11    -1.922651   3.095859  -0.621  0.53523
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Residual standard error: 9.302 on 215 degrees of freedom

Multiple R-Squared: 0.1425,Adjusted R-squared: 0.0827

F-statistic: 2.382 on 15 and 215 DF, p-value: 0.003363

Estimation results for equation peca:

=====================================

peca = almo.l1 + pean.l1 + peca.l1 + waln.l1 + const + sd1 + sd2 + sd3 + sd4 + sd5 + sd6 + sd7 + sd8 + sd9 + sd10 + sd11

	Estimate	Std. Error	t value	Pr(>|t|)	
almo.l1	-4.188e-01	1.904e-01	-2.200	0.0289	*
pean.l1	5.755e-02	3.299e-01	0.174	0.8617	
peca.l1	-8.076e-02	6.848e-02	-1.179	0.2395	
waln.l1	-3.191e-04	1.871e-01	-0.002	0.9986	
const	2.255e+00	2.889e+00	0.780	0.4360	
sd1	9.859e+00	1.438e+01	0.686	0.4937	
sd2	-1.584e+01	1.434e+01	-1.104	0.2706	
sd3	-6.984e+00	1.435e+01	-0.487	0.6269	
sd4	-1.992e+00	1.418e+01	-0.141	0.8884	
sd5	-3.494e+00	1.432e+01	-0.244	0.8075	
sd6	-1.100e+01	1.429e+01	-0.770	0.4422	

sd7 -2.830e+00 1.441e+01 -0.196 0.8444

sd8 -2.152e+00 1.426e+01 -0.151 0.8801

sd9 4.931e+00 1.426e+01 0.346 0.7298

sd10 -2.383e+00 1.422e+01 -0.168 0.8670

sd11 1.579e+01 1.454e+01 1.086 0.2787

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 43.68 on 215 degrees of freedom

Multiple R-Squared: 0.06048,Adjusted R-squared: -0.005065

F-statistic: 0.9227 on 15 and 215 DF, p-value: 0.5396

Estimation results for equation waln:

===================================

waln = almo.l1 + pean.l1 + peca.l1 + waln.l1 + const + sd1 + sd2 + sd3 + sd4 + sd5 + sd6 + sd7 + sd8 + sd9 + sd10 + sd11

	Estimate	Std. Error	t value	Pr(>|t|)	
almo.l1	-0.030261	0.068337	-0.443	0.65834	
pean.l1	0.344867	0.118408	2.913	0.00396	**
peca.l1	0.002078	0.024577	0.085	0.93270	

waln.l1	0.026428	0.067152	0.394	0.69430	
const	0.338931	1.037032	0.327	0.74412	
sd1	-14.166084	5.160893	-2.745	0.00657	**
sd2	-1.062158	5.147564	-0.206	0.83672	
sd3	-5.091946	5.149738	-0.989	0.32388	
sd4	2.195204	5.088142	0.431	0.66658	
sd5	-3.477037	5.139929	-0.676	0.49947	
sd6	3.348854	5.129060	0.653	0.51451	
sd7	0.179870	5.170651	0.035	0.97228	
sd8	-1.511617	5.116271	-0.295	0.76793	
sd9	1.109739	5.117872	0.217	0.82854	
sd10	0.274769	5.101930	0.054	0.95710	
sd11	0.013755	5.217304	0.003	0.99790	

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 15.68 on 215 degrees of freedom

Multiple R-Squared: 0.1069, Adjusted R-squared: 0.04464

F-statistic: 1.716 on 15 and 215 DF, p-value: 0.04938

Covariance matrix of residuals:

	almo	pean	peca	waln
almo	237.5093	-12.228	-45.33	0.1042
pean	-12.2275	86.535	64.70	7.6744
peca	-45.3294	64.700	1907.93	-91.7324
waln	0.1042	7.674	-91.73	245.7668

Correlation matrix of residuals:

	almo	pean	peca	waln
almo	1.0000000	-0.08529	-0.06734	0.0004314
pean	-0.0852908	1.00000	0.15923	0.0526244
peca	-0.0673377	0.15923	1.00000	-0.1339613
waln	0.0004314	0.05262	-0.13396	1.0000000