UNIVERSITY OF CAPE COAST

APPLICATION OF GENERALIZED PRINCIPAL COMPONENT ANALYSIS TO CLASSIFICATION OF PRICE LEVELS OF MARKETS IN GHANA

CHANIK KWAO-AGBOADO

2020

UNIVERSITY OF CAPE COAST

APPLICATION OF GENERALIZED PRINCIPAL COMPONENT ANALYSIS TO CLASSIFICATION OF PRICE LEVELS OF MARKETS IN GHANA

BY

CHANIK KWAO-AGBOADO

Thesis submitted to the Department of Statistics of the School of Physical Sciences, College of Agriculture and Natural Sciences, of the University of Cape Coast in partial fulfilment of the requirements for the award of Master of Philosophy degree in Statistics

NOVEMBER 2020

DECLARATION

Candidate's Declaration

I hereby declare that this thesis is the result of my own original research and that no part of it has been presented for another degree in this university or elsewhere.

Candidate's Signature:.. Date:..........................

Name: Chanik Kwao-Agboado

Supervisors' Declaration

We hereby declare that the preparation and presentation of the thesis were supervised in accordance with the guidelines on supervision of thesis laid down by the University of Cape Coast.

Principal Supervisor's Signature:................................... Date:........................

Name: Prof. Bismark K. Nkansah

Co-Supervisor's Signature: .... Date:........................

Name: Dr. Francis Eyiah-Bediako

ABSTRACT

The study determines the main dimensions along which to classify specific markets as extremely high-priced, high-priced, medium priced, low priced and those that are extremely low-priced. For this purpose, secondary data is obtained on prices of 19 food items in 91 Markets in non-consecutive years from 2008 to 2015 in all regions of the country from the Statistical, Research and Information Directorate of the Ministry of Food and Agriculture. This data constitutes a multivariate dataset with nineteen variables and 455 observations. Generalized Principal Component Analysis, which is a procedure for data classification and summarization, is considered appropriate for analyzing this high dimensional dataset. An appropriate index is designed to identify the various levels of the prices and eigen-analysis among others further informed the choice of the technique. It is found that there are indeed extreme priced markets in Ghana. The results reveal that the classification of the markets is not influenced by their location. Another important observation is that root and tubers and cereals are major components of local food items. In order to reduce the disparity in price levels there is the need for concentration to be given to the production of these two types of local food items for economic wellbeing of Ghanaians

KEY WORDS

Extreme Price Levels

Generalized Principal Component Analysis

Local Food Commodities

Market Classifications

Market Price

Matrix Concatenation

ACKNOWLEDGEMENTS

I am most especially grateful to my Principal Supervisor, Prof. Bismark K. Nkansah, of the Department of Statistics, University of Cape Coast for his instructions, patience and counsel throughout the stages of this work.

I am also grateful to my Co-Supervisor Dr. Francis Eyiah-Bediako, of the Department of Statistics, University of Cape Coast, for his various contributions to the success of this work. I am grateful to all my lecturers in the Department of Statistics. To my friends who were there to encourage me along the way, I appreciate you.

Finally, to all members of my family, I am grateful for the continued support throughout the periods of the studies.

DEDICATION

To my lovely wife Lucy and cherished daughter Kathryn Aseye.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

FAO	Food and Agricultural Organisation
IFAD	International Fund for Agricultural Development
MoFA	Ministry of Food and Agriculture
OECD	Organization for Economic Co-operation and Development
PNDC	Provisional National Defence Council
PPP	Public Private Partnership
SRID	Statistics Research and Information Directorate
UNCTAD	United Nations Conference on Trade and Development
UNEP	United Nations Environment Program

CHAPTER ONE

INTRODUCTION

Understanding the nature of prices of food items is important to governments, policy think tanks, individuals among other stakeholders. Extreme levels of prices of food items present economists with some difficulty in understanding the dynamics of market prices. It is on this backdrop that this study attempts to present a robust scientific understanding of the problem of extreme food prices. This study is significant in that, it uses a combination of multivariate statistical procedures (Kleinbaum & Kupper, 1978) to analyse the time dependent multivariate data and proffer interpretation that can help understand the consequences of price variations. In order to apply the techniques in the manner that is carried out in this work, the time dependent data is considered as a multiple multivariate dataset. In the process, it has become necessary to apply further computational techniques, such as the development of algorithms to facilitate the implementation of the techniques. The study also puts into perspective other studies done in the area using other statistical procedures. This study indeed alludes to some conclusions made from previous studies (Eyiah-Bediako, 2019) that made use of similar data and attempts to explore further. At the end of this study, it would be apparent the relevance of applying different techniques to such a high dimensional dataset that have already been examined in the literature. In the end, it is anticipated that the study would be an important addition to the literature.

Background to the Study

An important building block of strategic marketing is market classification. Most successful companies drive their businesses based on strategic marketing. Classifying markets can be seen as a tool for decision-making for marketing managers in the crucial task of selecting a target market for a given product and designing an appropriate marketing strategy (Tynan & Drayton, 1987). Paramount to marketing strategy is the theory of market pricing. Market forces of demand and supply determine the prices of commodities.

Market Price

We define the market price of a food commodity item as the short-term equilibrium price that is decided daily in the market because of the prevailing short-term to medium-term stable conditions. The equilibrium market price, also called the determined price, is the equilibrium price at which marketers sell the commodity. The price of commodities changes slightly from day to day in each market, meaning the prices of goods in the market are dynamic. There is also what is termed as a co-movement of prices of food items, where the prices of unrelated food items move together (Pindyck & Rotemberg, 1990). This affects the prices of food items. Globally from the start of the 21st century, commodity prices have seen an increase of 26% on average annually. For 2020, the U.S. Department of Agriculture predicts that food prices will increase between 1.5% and 2.5%. Dairy prices are expected to rise 1.5% to 2.5%, vegetable prices 0% to 1%, fresh fruit prices 1% to 2%, cereal and bakery prices 2% to 3%, beef and veal prices 0% to 1%, poultry prices will rise 0.5% to 1.5%, and pork prices 1.5% and 2.5%. In Africa, the prices of food commodity food have seen increases from the 1990s. Even though most economies in the sub region are largely dependent on Agriculture. The importation of some commodities have largely affected the prices of these food items across the continent especially in the West African sub region. This means that prices of commodities generally will witness some form of increase. Some short term and long-term factors affect the demand and supply of food commodities. Some short-term factors include extreme weather, diseases, droughts, catastrophes among others. Some long-term factors include; high cost of crude oil and extreme climate changes(Trostle, 2010).

Transportation is a fixed cost on price of food commodities(Chen, Kuo, & Chen, 2010; Chen, Wang, Zhang, & Zheng, 2019; Esmaeili & Shokoohi, 2011; Obadi & Korcek, 2014; Zhang & Reed, 2008). In 2008, food prices generally rose by 6.4%, which was the largest single year increase since 1984; this caused speculation that eventually led to high prices in 2009. Again in 2011, prices of food items rose by 4.8% which some observers say led to the Arab Spring rising. In 2012, global food prices increased by 2.5%. Each year sees its own increase in the prices of food items. This increase in the prices of food items lead to price volatility with its attendant extreme prices. We describe price volatility as the magnitude of price fluctuations or the risk of high, unexpected price changes. The consequences of extreme prices can intensify and contribute to broader social risks in terms of food security, human development, and political stability in any country. Hunger especially is indicative of a population’s limited physical access to food. It is mainly associated with inadequate dietary intake that usually predisposes children to a higher risk of disease infection through poor nutrition. The relatively small share of government expenditure on agriculture sometimes causes this rather high prevalence of hunger mainly due to limited access to the required quantity and quality of food.

Market Classification

In recent years, large corporations and businesses have become increasingly involved in the commodities investment business; they have many research departments that increasingly proffer statistical analysis on commodity markets. Commodity forecasts are offered by companies specialized in market intelligence. These practices have been more associated with companies that deal in stock exchange markets industries and manufacturing but less food commodities. There has been a tremendous increase in academic interest lately in the field of market classification. In this context, identifying groups of local food markets has significance in helping families mitigate the challenge of poverty. Families would be able to identify markets that are highly priced generally and those that are lowly priced for effective purchasing decisions. Classifying markets is a technical way of identifying various groups of markets according to some appropriate index. A similar classifying scheme is used in other studies for example the study of poverty among households and among individuals in a country (Adjasi & Osei, 2007; Cooke, Hague, & McKay, 2016; Kyereme & Thorbecke, 1991). Studies in identifying extreme weather also use some kind of index to categorize various occurrences as being severe, less severe or mild(de Freitas, Scott, & McBoyle, 2008; Hansen, Sato, Glascoe, & Ruedy, 1998).

Price Volatility

Commodity price volatility and macroeconomic market risk can have severe long-term impacts on economic growth and development (Ramey &Ramey, 1995; van der Ploeg & Poelhekke, 2009) in particular in countries with underdeveloped financial institutions (Aghion, Bacchetta, Rancière, & Rogoff, 2009). Extensive literature suggests that poor countries are much more volatile than rich countries and that does not engender economic growth (Deaton & Miller, 1996; Deaton, 1999; Ramey & Ramey, 1995). Infact more evidence have been established recently to confirm that high price volatility affects the growth of any country (Acemoglu, Johnson, Robinson, & Thaicharoen, 2003; Hnatkovska & Loayza, 2005; Loayza, Rancière, Servén, & Ventura, 2007; Poelhekke & Ploeg, 2007). The steep increase in the food commodity prices globally in the recent years have raised concerns for various economies as the implications for the most vulnerable low income group is serious (Trostle, 2010).

Analysts attribute the increasing volatility in commodity food prices to some underlying factors such as price speculations in future commodity markets, and changes in the price of crude oil (Esmaeili & Shokoohi, 2011; FAO, 2008). The increasing commodity food prices is also attributed to the general upsurge in the global demand for food items, mainly due to the ever expanding world economies like China. And so China is a classic case of the impact of the consumption patterns of big economies on the global demand in commodity food and on prices (Coxhead & Jayasuriya, 2010; Hernandez, Ibarra, & Trupkin, 2014). According to another researcher, price volatility among poor countries is a result of these countries specializing only in agriculture and mineral production. Ghana is a classic case. Indicating that this primary commodities experience far greater volatility than manufacturing and services oriented economies. Higher price volatility is also associated with greater potential losses for producers and poor subsistence farmers: because high volatility implies large, rapid changes in prices, and it becomes more difficult for producers to make optimal decisions on the allocation of inputs into the agricultural sector (UNCTAD, 2008). Consequently, in a time of high price volatility, producers may use fewer inputs like fertilizer and high-quality seeds in their production, and they may reduce their investments in areas that improve productivity, which could adversely affect their income and the overall availability of food in the country (Ivanic & Martin, 2008). It is therefore so important to study the extreme prices of food items.

Impact of food price volatility on households

In the short run, price increases lead to increased poverty in most countries, as many poor households are net buyers of food. More so is the consequence of such increase on child malnutrition. Child malnutrition indeed is considered to be one of the key human development indicators (Gabriele & Schettino, 2008; Pelletier & Frongillo, 2003). Smith & Haddad (2000) estimated that child malnutrition prevalence would remain high by the end of the year 2020 with about 20 percent (140 million) malnourished children under age five in developing countries. In the medium to long term, higher commodity prices may lead to higher wages due to agricultural to labour market linkages. This, in turn, would also reduce poverty for many direct consumers of food who receive wages, leading to lower poverty rates in most countries and on the global scale. Persson (1999) actually argued this reasoning; that price volatility is a disincentive to the efforts and investments generally in agriculture and that it is probably one of the causes for the distressed state of agriculture in many developing countries. Countries should therefore establish or expand safety nets to adversely affected households as a form of remedy.

In Ghana, seasonal variability in food supply and prices due to climate changes and other natural disastrous occurrences such as drought, irregular rainfall, and bush fires among others make it difficult to meet food demands all year round. A sharp increase in commodity prices of food products creates severe famine and other health related problems. As recorded in the case of food crops most especially Maize in 1983 and some other cereals and vegetables (Ofori-Sarpong, 1986). We can therefore say that large variations in prices have serious consequences on the purchase behaviours of consumers. A common purchase behaviour among consumers is the phenomenon of shifting from diets that have high micronutrients to staple foods because of down shift of income levels. As a result, Pindyck & Rotemberg (1990) discover that prices of seemingly unrelated commodities move together.

It is a known fact that despite the differing food habits, consumers switch their consumption patterns to the closest food substitute because of a change in the price of a particular food item. However, the shortage of one commodity can lead to a demand in another substitute food commodity, which will likely increase the price of that substitute commodity. For example "since wheat is a close substitute for grains it is likely that when there is a shortfall in the production of cereals, particularly grains, the pressure of adjustment in demand falls on wheat" (Sharma, Gosh, & Kumar, 2000). Cassava is one of the most common staple foods in Ghana, which is less expensive; it is a rich source of carbohydrate but has very low level of protein or other essential micronutrients (Stadlmayr et al., 2012). When households consume starchy foods in place of other nutritional foodstuffs it leads to high risk of stunting and associated nutritional health challenges and increased immune deficiencies. Suggestions have been prescribed to prevent a reduction in the volume and quality of foods (Thompson, 2010). There is evidence that macroeconomic indicators explain very little of the variation in commodity food items(Bailey & Chan, 1993; Bessembinder & Chan, 1992; Bjornson & Carter, 1997; Park, Wei, & Frecka, 1988) and so it is significant to understand the underlying reasons behind extreme prices.

Impact of food price variations on food security

According to IFAD (2011) food prices tend to have a major impact on food security, at both household and country levels. Many of the world's poorest families spend more than half their income on food. Price hikes for cereals and other staples can force them to cut back on the quantity and quality of their food.

Higher food prices have high reaching negative impact on members of a household (Meerman & Aphane, 2012). This means that at the national level this can lead to various forms of malnutrition related diseases and health challenges like stunting, and underweight, which eventually will also lead to slow human development and economic growth (Brinkman, De Pee, Sanogo, Subran, & Bloem, 2010). Besides these microeconomic considerations, there are macroeconomic effects of changes in food prices, and some studies have revealed fundamental supply and demand as well as macroeconomic factors as the main contributing factors for the significant rise in the prices of commodities in the year 2007/2008 (Boyd, Harris, & Li, 2018; Will, Prehn, Pies, & Glauben, 2013). Increasing commodity prices serves as incentives to farmers and help them deal with food security and hence achieve self-sufficiency (Orbach, 2008); however, irregular changes in food prices have a ripple effect on other sectors of the economy, which often results in the frequent increase in food .

According to (Kaldor, 1939 ; Working, 1948), there is the short run dynamics of commodity prices, which they explain using the theory of storage in the general economic theory. Food insecurity and insufficient nutrition reduce health status and human capital, affecting labour productivity and economic output (Behrman & Rosenzweig, 2004; Fogel, 1994; Gyimah-Brempong & Wilson, 2004; Weil, 2007). Prices for cereals and other major food commodities have experienced two global spikes, one within 2007 to 2008, the other within 2010 to 2011. Moreover, they have generally remained higher than the early years of the 20th century.

In many developing countries, prices have seen hikes or remained at higher levels because of higher consumption. According to a World Bank report, about 39% of the increase in consumption of food globally, between the 1996 and 2006 is merely accounted for by developing economies like Ghana (World Bank, 2018). In developing countries, the key factors behind inadequate supply are low and stagnating productivity in agriculture, a deteriorating natural resource base, and weak rural and agricultural infrastructure and markets. Between 2006 and 2008, international food prices doubled. The effects of the price surge reverberated globally, though the worst hit were low-income, food-deficit countries with meagre stocks. In total, about 100 million poor rural and urban people dropped into the ranks of the world hungry. Commodity food prices increased dramatically between late 2006 and mid-2008. And reaching high levels in later years (i.e., during 2010, early 2011, and the third quarter of 2012). Growing population and income in emerging and developing countries will add significantly to the demand for food in the coming decades and so boost the increasing global demand for food commodities by these developing economies (Carter, Rausser, & Smith, 2012; Hamilton, 2009; Kilian & Murphy, 2014; Krugman, 2008). It is projected that by the year 2050, the world's population would reached more than 9 billion people (Kochhar, 2010) and the demand for food to increase by between 70% and 100%. This alone is sufficient to exert pressure on commodity prices.

According to the latest OECD/FAO medium term outlook projections, prices of crops and most livestock products will be higher in both real and nominal terms during the decade to 2019 than they were in the decade before the 2007/08 price spikes. If the rate of growth of agricultural production does not keep pace with demand, upward pressure on prices is inevitable. A demand or supply shock in a situation where the supply-demand balance is already tight, can, for the reasons explained in the previous paragraph, result in increased volatility around the upward trend." and "Food price inflation can also be a serious issue in middle income countries, where many consumers expend as much as half of their budget on basic foods. Nevertheless, consumers in developed countries face wider choices in terms of their ability to adjust spending on different types of foods and they have safety net mechanisms that are well suited to delivering targeted assistance to the most affected. The sharp increase in food prices both in international markets and in local markets since 2006 has raised serious concerns about the food and nutrition situation of poor families in many countries (Brinkman et al., 2010). Particularly in urban areas, where people cannot grow their own food, household budgets become stifled. The rapid price increases are especially bad news for young children, as any disruption to their nutrition tends to have serious long-term implications, both in terms of stunting, and lower educational outcomes, affecting their earning potential in later life. An important motivation for food consumption is to maintain good health (Sarkar, 2013; Urala, 2005).

Poverty

Poverty has many dimensions. It is mostly characterized by low income, malnutrition, ill health, illiteracy, and insecurity, among others. In addition, the incidence of poverty has tremendous impact on the nation as a whole. In Ghana the profile of poverty based on surveys (2005/06, 2012/13, and 2016/17) shows that the country made a marginal progress in the pursuit of poverty reduction since the last round of the survey in 2012/13. In contrast to the period between 2005/06 and 2012/13 which recorded a decline in the poverty headcount of 7.7 percentage points, the decline in the poverty headcount between 2012/13 and 2016/17 was small at 0.8 percentage points (that is, from 24.2 percent to 23.4 percent). Much needs to be done if the country is to achieve the Sustainable Development Goal (SDG) on ending poverty in all its forms by 2030. Extreme poverty (people unable to meet their basic food needs) declined from 8.4 percent in 2012/13 to 8.2 percent in 2016/17. However global poverty appears to be on the increase.

Poverty Measurements Index

From as far back as 1987, the constitutionally mandated body vested with the power to conduct various forms of enumeration for policy making in

Ghana is the Ghana Statistical Service (GSS). This organic agency of state has conducted various forms of surveys; Agriculture censuses and surveys, Economic Censuses and Surveys, Ghana living standards survey, Demographic and Health surveys and Multiple cluster surveys, specifically Multiple indicator survey six (MICS 6), Integrated Business Survey (IBES), and the Ghana Living Standards Survey (GLSS 7) with the aim of measuring the living conditions and well-being of the population. A report from the seventh round of the Ghana Living Standards Survey conducted between 2016/2017 shows that 45.6 percent of Ghana's population are multi dimensionally poor. A Multi-dimensional Poverty Index (MPI) (Alkire & Santos, 2014) unlike the internationally commonly used poverty line (Gillie, 1996) makes visible the joint distribution of deprivations, starting with a profile of each person's simultaneous challenges, in order to measure non-monetary poverty. Overall, MPIs provide not only a headline figure, but also an associated information platform on national and subnational conditions across population groups and joint deprivations in different dimensions of poverty.

Globally the Oxford Poverty and Human Development Initiative (OPHI) (Alkire & Santos, 2010) at the University of Oxford and the United Nations Development Program's Human Development Report Office (HDRO) jointly compute and publish a global MPI that compares acute multidimensional poverty across more than 100 countries. However, this measure is intended for international comparability and is not adapted for the specific circumstances of a given country. Thus, many countries have developed their own national MPIs in much the same way that they use national monetary poverty lines as well as the $1.90/day measure which are already existing measurements of poverty.

National MPIs are increasingly being adopted as official permanent poverty statistics, which provide a more detailed exposition of the various dimensions of people's living standards to complement monetary poverty statistics. Updated regularly, national MPIs are used to shape and energize effective policy actions. Using the Alkire-Foster Methodology (Alkire et al., 2015; Alkire, Roche, Santos, & Seth, 2011; Santos, 2019), the Multi-dimensional poverty measurement is computed.

This methodology allows for the construction of individual and household level deprivation profiles that can then be used to identify multi-dimensionally poor people. It first identifies who is poor, by summing up the deprivations each person experiences in a weighted deprivation score, and then aggregates this information into a headline and associate information platform for a given population. This methodology for multidimensional poverty measurement has come to be widely used because of its simple, yet specific approach. The MPI can equivalently be computed as the weighted sum of censored headcount ratios; which shows the percentage of people who are identified as poor and deprived in an indicator.

The MPI is always broken down by an indicator to show the composition of multidimensional poverty. This feature of dimensional detail brings added policy relevance to the analysis. In addition, the MPI can be disaggregated by different population groups, such as, urban/rural areas, age groups, and sub-regions. The weights computed from the MPI are used as indicators to allocate or assign households whose computed weights fall within a certain cut-off called poverty cut-offs. Poverty cut-offs are indices used to assign interpretation to groups which fall into their identified index. In Ghana, the MPI cut-off is specified at one-third of the indicators; that is, a person whose deprivation constitutes at least 33 percent of the weighted indicators is identified as multi-dimensionally poor. The chosen cut-off reflects the global MPI, which suggests that a person must be deprived in at least one full dimension's worth of indicators to be considered multi-dimensionally poor. A person deprived in 20-33.3 percent of the weighted indicators is considered 'vulnerable to poverty' and a person deprived in at least 50 percent of the weighted indicators is identified as being in severe poverty (GSS, 2018).

Extreme Events

Extreme events refer to "unusual" events that do not occur regularly and such occurrences do have major negative impacts. The phenomenon of the rarity of extreme events is important, since their occurrence mostly are not normally expected and has become difficult and mostly expensive to be ready and prepared for the outcome and hence cope with them (Sarris, 2014) in (Kalkuhl, Von Braun, & Torero, 2016). This challenge is not peculiar to only individuals, firms, or public institutions (governments) but to markets that are not always prepared and able to insure against such extreme events (Jaffee & Russell, 1997). Extreme events are mostly events outside a certain quantile.

For instance, emerging extreme weather conditions due to change in climate conditions (Aikins, 2012) like global warming, conflicts and political instabilities in the Middle East and Africa, and the ongoing use of expansive monetary policy leading to low interest rates—could lead to new sudden extreme events. Markets where goods and services are exchanged and prices are decided on daily are deeply linked to food price volatility (Brunner, 1998; Deaton et al., 2003; Roll, 1984). Food markets cannot be considered in isolation.

Many markets are linked through trade; commodity, asset, and financial markets influence food markets; and these, in turn, influence trading and allocation decisions of actors that also engage in food markets.

Global Cost of Extreme Climatic Events

The Global Climate Risk Index (CRI) 2019 (Kreft, Eckstein, & Melchior, 2013) analyses to what extent countries and regions have been affected by impacts of weather-related loss events (storms, floods, heat waves, Hurricanes etc.). It indicates a level of exposure and vulnerability to extreme events, which countries should understand as warnings in order to be prepared for more frequent and/or more severe events in the future. People all over the world have to face the reality of climate change – in many parts of the world manifesting as increased volatility of extreme weather events. Between 1998 and 2017, more than 526,000 people died worldwide and losses of US$ 3.47 trillion were incurred as a direct result of more than 11,500 extreme weather events.

The UNEP Adaptation Gap Report 2016 (Asokan, Obando, Kwena, & Luwesi, 2020; Chapagain, Baarsch, Schaeffer, & D'haen, 2020) warns of increasing impacts and resulting increases in global adaptation costs by 2030 or 2050 that will likely be much higher than currently expected, the "mean net present value of the costs of damages from global warming for 1.5°C and 2°C (including costs associated with climate change-induced market and non-market impacts, impacts due to sea level rise as a result of the polar cap melting, and impacts associated with large scale discontinuities) (Cline, 1992; Marshall, Pettersen, Bode, & White, 2020). Costs resulting from residual risks or unavoidable loss and damage are not covered in these numbers. Similarly, the

Intergovernmental Panel on Climate Change (IPCC) estimates same in its recent Special Report on “Global Warming of 1.5°C”(Broome, 1992; Cline, 1992; Fankhauser, 1994; Huppmann, 2020) .

The CRI quantifies the impacts of extreme weather events on countries. The CRI examines both absolute and relative impacts to create an average ranking of countries in four indicating categories, with a stronger emphasis on the relative indicators. The countries ranking highest are the ones most impacted and should consider the CRI as a warning sign that they are at risk of either frequent events or rare, but extraordinary catastrophes.

Table 1 shows the ten countries that were most affected in 2017, with their average weighted ranking (CRI score). This is adapted from the German Watch (Eckstein et al., 2019). In all these literature, the use of indices to categorize the extremity of an event or experience has been demonstrated. My study attempts to use a similar indexing scheme for classifying market prices as being extremely high priced, extremely low priced, high priced, low priced and medium priced.

Table 1: The Climate Risk Index for 2017:10 Most Affected Countries

Ranking	Country	CRI Score	Death Toll	Deaths per 100000 Inhabitats
1	Puerto Rico	1.5	2978	90.242
2	Sri Lanka	9	246	1.147
3	Dominica	9.33	31	1686.894
4	Nepal	10.5	164	1909.982
5	Peru	10.67	147	6240.625
6	Vietnam	13.5	298	4052.312
7	Madagascar	15	89	693.043
8	Sierra Leone	15.67	500	99.102
9	Bangladesh	16	407	2826.678
10	Thailand	16.33	176	4371.16

(Adapted from the German Watch, 2017)

Statement of the Problem

Many attempts have been made at identifying extremes in prices on the Ghanaian Market by using the classical principal component analysis, (PCA) approach. These attempts are at identifying extreme prices on the Ghanaian markets using the classical PCA approach and outlier displaying component (Eyiah-Bediako, 2019; Nkansah & Gordor, 2012).

These attempts examine extremes in multivariate observations and have usually relied on classical PCA. The classical PCA has been used as both an interim technique as well as the ultimate. It is clear from literature that the classical PCA cannot be used as the only technique for detecting extreme observations. This is because it is not known to be a robust approach. The study will therefore attempt to use the General Principal Component Analysis to obtain classifications in the general price levels of the markets. This approach has wider application in areas such as image processing, big data dimensionality reduction and facial recognition(Vidal, Ma, & Sastry, 2006).Studies that have so far made use of the intended datasets have identified varied results of market classification. It appears that the results of this kind of studies are influenced by the type of technique that is employed. So far, techniques such as cluster analysis (Seglah, 2014), PCA and displaying components (Eyiah-Bediako, 2019) have been utilised for the analysis of this type of data. The results are known to differ because of the technique used. It will be relevant therefore to explore this data from a different perspective by employing a more robust approach in addition to those already used.

Objectives of the Study

The general purpose of the study is to examine the general price levels of the local markets in Ghana. In order to achieve the above objective, the study will aim at the following specific objectives;

1. Determine suitable classification index that may be used for categorising levels of prices of local food markets in Ghana.
2. Identify specific local markets that may be regarded as extreme in prices.

Significance of the Study

1. The study will lead to improved growth in personal income of both produce growers and consumers, resulting in increasing food security.
2. The study will test the robustness and the limits of the algorithm developed for computing GPCA for future researches.
3. Information of the price-based groupings of various markets will help promote specific commodities.
4. We present new insights into the structure of local markets and their classification.

Data

We obtain a secondary data for this study from The Statistical, Research and Information Directorate (SRID) of the Ministry of Food and Agriculture (MoFA). The Ministry of Food and Agriculture is a ministry of the executive arm of government. The Statistics Research and Information Directorate (SRID) was created in 1999 in response to Civil Service Law (PNDC Law 135 of 1985) that requires for a Statistics and Information Directorate within the Ministry of Food and Agriculture (MoFA, 2020). With the vision to generate relevant, reliable and timely statistical information on the agricultural sector through the establishment of an operational food and agricultural statistics system within MoFA. The dataset covers the wholesale prices (quoted in GH¢ and GHp), in non-consecutive years within the period 2008 to 2015 (Eyiah-Bediako, 2019) on nineteen (19) foodstuffs, from Ninety-one (91) markets. All the prices are the annual wholesale prices collected from the markets.

Table 2: Variables Selected for the Study

No.	Commodity	Variable	Unit Of Sale
1	Maize	Mz	100kg
2	WhYam	YmWt	250kg(100 Tubers)
3	Cassava	Cv	91kg
4	Tomatoes	Tm	52kg (Crate)
5	Garden Egg	GEg	27kg
6	Dried Pepper	PpDr	16kg
7	Red Groundnut	GnR	82kg
8	WhCowpea	CpWt	109kg
9	Palm Oil	PmOil	18Litres
10	Orange	Org	20kg(100 Singles)
11	Banana	Ban	6-8kg(Aver. Bunch)
12	Smoked Herring	HrSmk	18 Litres
13	Koobi	Kbi	100 Singles
14	Onion	On	73kg
15	Egg	Eg	1Crate (30 Singles)
16	Plaintain(Apentu)	Pltn	9-11kg
17	Gari	Gri	68kg
18	Loc Rice	RiLoc	100kg
19	ImpRice	RiImp	50kg

Source: Field Data, SRID-MoFA 2008

The prices of the 19 food items selected for the study, with the standard units of sale or measurements are as given in Table 2. According to Asche and

Guthormssen (2001), it is valid to represent the various prices of a commodity by a unified price. This means that, the average prices of the commodity food items over the five-year period can represent the individual year's prices of the food items, for the analysis as is done in some other studies on commodity futures price analysis. We however will do our analyses on year-by-year basis. The commodities have been further classified based on their type. The categories are given in the Table 3. We observe that the various items are in seven categories. It can be observed that each category has at least two food items in the table. In addition, Gari is placed in the same category as Cassava and Onion is classified as vegetable.

Table 3. Labelling of Food Commodities

No.	Label	Food Items
1	Cereals	Maize, Imported Rice, Local Rice
2	Roots and Tubers	Cassava, White Yam, Plantain, Gari
3	Vegetables	Tomatoes, Garden Eggs, Dry Pepper, Onion
4	Pulses	White Cowpea, Red Groundnut
5	Fish and dairy	Smoked Herrings, Koobi, Egg
6	Oil	Palm Oil, Groundnut Oil
7	Fruits	Orange, Banana

Source: Field Data, SRID-MoFA 2008

Table 4 shows the various markets with their regional distribution used for the study. It can be observed from the table that there are at least three markets in each region. With Ashanti region having the largest distribution of markets, followed closely by the Brong Ahafo, Central, Greater Accra and the Western regions. Upper West has the least number of markets considered for the study. Probably because the number of people in the region are far less than the number of people in the other regions. The choice of these markets is based partly on the results of some previous related work done by Seglah (2014).

Table 4: Regional Distribution of the various commodity markets

Region	Markets	Total
Ashanti	Adugyama, Agogo, Agona, Bekwai, Ejura, Juaboso, Kumasi(Kejetia), Mampong, Nsuta, Obogu, Obuasi,Tepa	12
BrongAhafo	Atebubu, Berekum, D/Ahenkro,D/Nkwanta, Goaso, Kintampo, Kukuom, Nkoranza, Sunyani, Techiman,Yeji	11
Central	Ajumako, AssinPraso, Bawjiase, Cape Coast, Dunkwa, Elmina, F/Nyankomase, Kasoa, Mankessim, Swedru, Winneba	11
Eastern	Agormanya, Ahomam, Akoase, Anyinam, Asamankese, Koforidua, Mpraeso, New Tafo, Suhum	09
Greater Accra	Agbogbloshie, Ashaiman, Dome, kaneshie, Kasseh, Madina, Makola, Mallam, Nalerigu, Nsawam,Tema	11
Northern	Bimbilla, Bole, Damongo, Gushiegu, Salaga, Tamale, Yendi	07
Upper East	Bawku, Bolgatanga, Fumbisi, Garu, Navrongo, Zebilla	06
Upper West	Bugubelle, Tumu, Wa	03
Volta	Abotoase, Adidome, Akatsi, Denu, Ho, Hohoe, Kpeve, Kute, Logba-Alakpeti, Mafi Kumasi	10
Western	AgonaNkwanta, Asawinso, Bibiani, Bogoso, Dwinase, Juaben, S/Bekwai, Sekondi, Takoradi, Tarkwa, Tikobo	11

Source: Field Data, SRID-MoFA 2008; The SRID Market Enumerator's Manual, June 2000

Delimitations of the Study

The study focused on the commodity prices of food items in the Ghanaian markets only, even though the literature covered prices of commodities internationally. And only major markets in the districts are considered for the study. No variable from the secondary datasets used for the study were excluded. All the variables in the datasets featured in the analysis of the work.

Limitations of the Study

The limitations of a study can be defined as the external factors that could limit the scope of application of the study beyond a certain boundary. Since the study is time dependent, a time series analysis of the prices of the commodities would reveal the trend in the data, which would not be done in this study. Also the prices of the food items considered cannot be generalised for those food items which cannot be quantified with the same measurements used for the commodities under study. The study is counting on the accuracy of the secondary data collected from the Ministry of Food and Agriculture (MOFA). Any inconsistencies in the data could affect the quality of the output of the analysis done in this study and so thorough pre-processing of the data is done. The researcher encountered some limitations in the course of this study.

Time limitations: There was not adequate time, as the study has a time line to follow. The use of thesis timetable help mitigate the challenge. Every research has the financial constraint, which is also a limitation in itself, as the researcher needs to be online to update application packages to be used for the analysis of data, to research materials, to print stationeries and other necessary expenditures. Financial support from friends and family greatly help alleviate this financial burden. Matlab was used to manipulate large datasets. In quantitative analyses of large datasets, reliability is of important. Therefore, the researcher was rigorous in the performance of the analysis.

Organisation of Thesis

The work is organized into five chapters, namely Chapter one, which comprises of the introduction, background, statement of problem, objectives of the study and then the nature of the data used for the study. Chapter two reviews relevant theories and methods that were utilised to achieve the purpose of the study. Discussions covers reviews of some multiple multivariate methods mainly Principal Component Factor Analysis and the Generalized Principal Component Analysis. Chapter three discuss the methodology used for Generalized Principal Component Analysis and the preliminary exploratory analysis of the commodity prices data. It also covers exploratory analysis such as descriptive statistics discussing the relationships among the variables in the data. The theoretical framework of the Generalized Principal Component Analysis was explored. Chapter four discuss results from the analysis of the data using the principal component factor analysis, by showing results and their interpretation, it discuss the indices that are used to achieve the objectives of the study. The results of the analysis from the application of the theorems of Generalized Principal Component Analysis applied are also displayed in this chapter. The final Chapter five, summarises the discussion so far made in the work, the findings established, conclusion and the recommendations. Then the References that were cited throughout the work.

Chapter Summary

Chapter One has generally highlighted the implication of levels of market prices in their variation on general economic wellbeing. It was observed that food prices are generally on the increase around the world, and that the phenomenon is likely to heighten poverty levels. The chapter has also revealed the relevance of general market classification and noted that this practice is particularly associated with sectors other than the local food markets. The importance of extending market classification to the area of local food markets has been highlighted. In particular, it has the potential of enhancing purchase intentions of families in order to improve their choices for their wellbeing. This serves as the main motivation that has been captured in the statement of the problem. The problem identifies a number of multivariate statistical methods that have already been used for the study of similar data problem, and justifies the need to explore further robust techniques for the same purpose. The chapter has described in detail the source and nature of the data and explains that this data constitutes a multiple multivariate dataset and therefore requires combination of computational techniques to maximize information that could be derived from it.

CHAPTER TWO

LITERATURE REVIEW

Introduction

This chapter delves into some works done by other researchers and authors using some multivariate techniques. It involves reviewing Principal Component Analysis and General Principal Component Analysis, which are the main research methods used in the further analysis of the data in the study. Studies reveal many factors significantly influence the movement of prices in the commodity market. Some establish that macroeconomic factors are involved in the co-movement of price of Agriculture goods (Bailey & Chan, 1993; Carter, Rausser, & Smith, 2011). A similar study considers commodity specific factors that account for the formation of market price bubbles (extremes) for important commodity markets. That study used a right side rolling window ADF (Augmented Dickey-Fuller) test with a bootstrap procedure. The price extremes are first accurately identified in Corn and Soybeans markets, after which the study applied adopted penalized maximum likelihood estimation of a multinomial logistic modelling technique to explore the potential factors that contribute to these seeming price extremes, for the commodities considered (Mao, Yanjun, & Loy, 2020). Using pooled data from different commodity prices in various markets, another study attempted to determine changes in the price levels of these commodities when attempting to estimate the common potential influencing factors (Etienne, Irwin, & Garcia, 2015, 2017; Li, Li, & Chavas, 2017). Other studies attempted to understand the dynamics of the prices of the commodities in the markets locally for a particular year (Seglah, 2014). A study on techniques of classification of commodity of prices in the local markets have also been attempted (Eyiah-Bediako, 2019). Many other such studies have been carried out on the prices of commodities in the Ghanaian market. Our studies however seeks to identify classification of price levels of commodity prices in the local market for a number of commodities using a technique called generalized principal component Analysis (Caussinus & Ruiz-Gazen, 2007).

Multivariate techniques

Multivariate techniques are the statistical procedures used in multivariate statistics for the simultaneous observation and analysis of more than one variable. According to Sharma, Kleinbaum, & Kupper (1978), Statisticians generally use the term multivariate analysis to describe a method whose theoretical framework allows for the simultaneous considerations of several dependent variables. They further explained that researchers in the biomedical and health sciences who it appears are not statisticians view this term as describing any statistical technique involving several variables, even if only one dependent variable is considered at a time, prefer the use of the term multivariable analysis. Therefore multivariable analysis are used to analyse data involving varieties of variables that require reducing the number of the variables but without losing useful and sufficient enough information for decision making. There are many such techniques for the multivariable analysis of multivariate data. The following are some examples of multivariate techniques used in multivariate analysis of data:

1. Hoteling's T^2: Is a generalization of Student's t statistic that is used in multivariate hypothesis testing (Hotelling, 1936; Tabachnick & Fidell, 2012).

2. Multivariate Analysis of Variance (Covariance) (MANOVA or MANCOVA) methods extend ANOVA methods to cover cases where there is more than one dependent variable and where the dependent variables cannot simply be combined (Johnson & Wichern, 2007).
3. Multiple Regression Analysis: attempts to determine a linear formula that can describe how some dependent variables respond to changes in others. Regression analyses are based on forms of the general linear model (Kleinbaum & Kupper, 1978; Sharma, 1996).
4. Discriminant Analysis attempts to establish whether a set of variables can be used to distinguish between two or more groups (Anderson, 2003).
5. Principal Components Analysis: attempts to determine a smaller set of new variables that could explain the variation in the original set (Jolliffe, 2002b).
6. Clustering Analysis: assigning different objects into some designated groups (called clusters) so that objects from the same cluster (Homogenous in nature) are more similar to each other than objects from different clusters (Sharma, 1996).
7. Factor Analysis: this method uses several variables to define one or more new composite variables called factors (Sharma, 1996).
8. Multidimensional Scaling covers various algorithms to determine a set of synthetic variables that best represent the pairwise distances between records (Govaert, 2003).

9. Canonical Correlation Analysis tries to establish whether there are linear relationships between two sets of variables covariates and response (Mardia & Kent, 1979).

These are some of the notable multivariate analysis techniques that are widely used among others and treated in standard texts like Johnson & Wichern (2007). The choice of which techniques to use or apply depends largely on the objectives of the study and the characteristics of the variables under study. However these techniques can be classified generally under two categories; Analysis of covariance structure group comprising of Principal Component Analysis, Correspondence Analysis, Factor Analysis, and Canonical Correlation Analysis, and the Classification and Grouping Techniques group which comprises of Cluster Analysis, and Discrimination and Classification.

According to Johnson & Wichern (2007), the objectives of the multivariate methods just like most scientific investigations include the following;

1. Data reduction or Simplification of structure. The study is concerned with the reduction of dimension of the data as much as possible without loss of valuable information. The attempt is to make interpretation easier.
2. Sorting and grouping. "Similar" Groups of objects or variables are created, with consideration for measured characteristics. Alternatively, rules for classifying objects into well-defined groups may be required.
3. Investigation of the dependence among variables. The interest of the study is in the nature of the relationship between the variables being studied. The questions considered are whether all the variables are

mutually independent or whether there are one or more variables dependent on the others. If so, how?

4. Prediction. Relationships between variables must be determined for forecasting the values of one or more variables based on observations on the other variables.
5. Hypothesis construction and testing. Testing of Specific statistical hypotheses, which are formulated in terms of the parameters of multivariate populations, which may be done to validate assumptions or to reinforce prior convictions.

Challenges of multivariate data

One major challenge in analysing real time data by researchers is the visualization of a high dimension data in multivariate statistics with many variables, "Curse of dimensionality" (Schelter & Winterhalder, 2006). Most statistical packages are able to analyse and display the results of the relationships between two to three variables. However when the variables are more than three, it becomes quite difficult to depict such relationship between them.

Most variables have some underlying relationships among themselves, such that they move together. This could be because more than one feature could be used to measure a certain characteristic of the system under study. In recent years where measurement developments have far advanced, characteristics of variables can be measured many times and in many ways, and this is what happens in real life situations. This leads to a pool of data, which provides a near accurate measurement of variables. However, in other to solve problems of such systems easily, a great many variables can be replaced by a single variable with just as much information as those many variables it can replace. The statistical technique that is used to do such quantitative simplification of data is the principal component. Therefore, the new sets of variables generated is called the principal components. Now the principal components formed are orthogonal to each other but are linear combination of the original variables. The principal components then becomes the orthogonal basis for the entire data space. Principal components can then be defined as a linear dimensionality reduction technique that is used to extract useful information from a high dimension space, to a low dimensional space by preserving the essential parts with more variation in the data and take away the parts with fewer variations. This essentially leads to the loss of some supposedly less useful information

A literature review of Principal Component Analysis

Preisendorfer & Mobley (1988) noted that mathematicians Beltrami (1873) and Jordan (1874) independently derived the singular value decomposition (SVD), an important mathematical technique that underlies Principal Component Analysis. Some other literature state that Fisher & Mackenzie (1923) used the singular value decomposition (SVD) technique in an Agricultural trial saying it is “more suitable than analysis of variance for the modelling of response data”, they developed their own algorithm. However many researchers generally agreed that the Principal Component Analysis or (PCA) as it is known today (Jolliffe, 2002), is a product of the earlier works done by Pearson (1901) and Hotelling (1933) . Even though the approaches used by the two authors were different, both were researching using the same technique. Whiles Hotelling used standard algebra derivations, Pearson used the geometric deductions to arrive at the PCA that is finding “lines and planes of closest fit to systems of points in space”. Even though computers were not commonly available then to aid in complex computations, their methods were accurate even when smaller number of variables were used in their calculations. Bryant & Atchley (1975) has done extensive and exhaustive literature review of the genealogy of the principal component analysis. There were also some other works done on Principal component Analysis by some researchers, Rao (1964) and Thurstone (1931). I would mention some important available literature that had significance in the expansion and growth of this statistical technique.

Anderson (1963) which is widely cited in theoretical works discussed the asymptotic sampling distributions of the coefficients and variances of the sample PCs which according to Jollife is an improvement on the work of Girshick (1939). Rao(1964) offers some study on PCA offering some important geometric insights. Jeffers (1967) gave an impetus to the practical application of the PCA by discussing case studies in a manner that shows that PCA can be used in ways other than just dimensional reduction tool. So much research is currently being done in the general area of PCA, and it is very widely used in most areas in research. In Chemistry, PCA was said to be introduced by Malinowski (Malinowski & Howery, 1980) under the name principal factor analysis, and after that, a large number of chemical applications have been published. PCA is also used in geology (Davis, 1986; Jareskog, Klovan, & Reyment, 1976). The technique has also been used in the study of the weather (Barnett & Preisendorfer, 1987).

Related works and methods that use PCA for market price studies

Principal component analysis (PCA) investigates the differences among genders on the Adult Nowicki-Strickland Internal-External Control Scale. Random samples of 86 male and 108 female were taken from among the university students. Sixteen factors with eigenvalues greater than one were extracted. It was found out that 72.5% of the variance accounted for males and 69.6% of the variance for females. Orthogonal Varimax rotation was then used and items loading over .40 were retained. Items with loadings on more than one factor were kept with the factor on which they loaded the highest. Five factors which were identified and given the names powerlessness, helplessness, hard work verses passivity, futility and luck. Males and females shared many common items on the first two factors of powerlessness, peer relations and helplessness at home. Differences became apparent between genders on the factor three with males concerned with hard work verses passivity while females focused on futility. Factor four, belief in luck and factor five, a sense of helplessness were relevant only in males. The study suggests that females unlike males emphasize feelings of helplessness and hopelessness experience from their interactions with life (Kearney & Kearney, 1983).

Interpretation of Principal Components

The simple correlation between the original and the new variables is known as the loadings. These loadings give an indication of the influence of the original variables in the formation of the new variables. The higher the loading, the more influential the variable is in forming the principal components scores. Most often, if the ideal loading is greater or equal to 0.5, then it is concluded that the variable is influential in the formation of the Principal Component scores. However, the cut off, is considered based on the purposes of the researcher.

Assumptions underlying Principal Component Analysis

The reliability of data should be ensured to reduce the standard error of measurement by carefully examining any instrument, for example responses from questionnaires. Principal components analysis includes in its extraction common, specific and error variance. Thus specific and error variance caused by random variables are not recognized by principal components analysis and contaminates the outcomes. Such contamination makes it difficult or impossible to name and understand the components retained. Validity of items or a measurement of the degree to which variables represent the construct being measured may be tested by inserting random results for a few dummy items.

These dummy items should not load on any dimension shared by the questionnaire variables. If they do load on these dimensions, the analysis is trapping error variance. Data should be based on a sound theory and sample all aspects of that theory. Harman (1960) warns that dimensions may not be fundamental if the necessary measures have been omitted from the questionnaire. For instance, if variables related to an effective school omit student time on task, this variable will not appear as a component. A pitfall of the process is the lack of internal criteria by which to judge the interpretation of Principal components. A Principal components analysis on sub groups may be used to inform the researcher of the relative standings of each individual in the sample in relationship to each dimension. The higher the score, the higher the ranking of an individual on a dimension. Validity of the construct may be tested by making predictions based on the loadings and then observing the actual behaviour of people who score highly or lowly on a component to determine the degree of congruence. Behaviour in accordance with prediction suggests a valid measurement of a dimension. The idea of independence of the variables is an important

Dimension Reduction

Dimension reduction methods have in common the goal of using the correlation structure (or the variance-covariance structure) among the predictor variables to accomplish the following:

1. To reduce the number of predictor components without any substantial loss of useful information (Fodor, 2002).

2. To help ensure that these components are independent.

3. To provide a framework for interpretability of results (Jollife & Cadima, 2016).

Data Visualization

Principal Component analysis is used for data visualization for big data. There is no gain saying the fact that, in today's world, the sheer volume of data churned out daily is gigantic (Hammer, Kostroch, & Quiros, 2017). The burden however is how to view the essential information in this data to make decisions that can impact individuals and organizations (Jin, Wah, Cheng, & Wang, 2015). Exploration of data in order to understand the distribution of the variables or to see how the variables are correlated is necessary to solve the problems using data. However, the challenge is when the dimension of the data is high; data exploration becomes difficult and almost impossible in some cases (Kshetri, 2014). This challenge of visualizing large volumes of data and higher dimensions can be solved by the use of Principal Component Analysis (Acharjya & Ahmed, 2016). The data is projected onto a lower dimension, which allows for easier visualization (Siluvairajah, 2011).

Machine Learning

Machine learning is a new branch of data science, which involves the training of machines to independently learn by identifying patterns and modelling data to develop correct solutions (Fodor, 2002). According to (Ben-david, 2014; Shalev-Shwartz & Ben-David, 2013) machine learning is the "automated detection of meaningful patterns in data". In PCA reduction and recovery is performed by linear transformations. PCA finds a transformation that is linear, for which the least squared differences between the new and the original vectors are minimum and orthogonal. In machine learning, the following are performed;

1. Finding new patterns, extracting and summarizing data (Wu, 2019)
2. Forecasting based on the analysis from the data (Wu, 2019)
3. Computing the probabilities for the results
4. Autonomously adapting to new discoveries
5. Then optimizing the processes based on identified patterns (ICMLC, 2019)

High dimension data needs to be trained to the system, however when PCA is used, it is reduced to a lower dimension. This makes it easy for the system to be trained faster, as the algorithms are executed at half the time. Because if the dimension is higher the learning algorithm will be slow. Therefore, PCA speeds up the training and testing of the system. As this is an emerging field of data analytics, PCA is at the core of its algorithms to optimize its operation.

Generalized Principal Component Analysis

Engineers or scientists use pre-processed data for solving scientific problems for their analysis. Most data collected contain fewer amounts of noise and adequately for the specific task or problem of interest. The assumptions underlying the data analysis methods are that the data collected by these professionals have fewer effective dimension than that considered in the general space. For example, the number of pixels in an image can be rather large, yet image-processing models used only a few parameters to describe the appearance, geometry, and dynamics of a scene. This assumption motivated the development of a number of techniques for identifying low-dimensional structures in high-dimensional data, a problem that is important not only for understanding the data, but also for many practical purposes such as data compression and transmission.

A popular technique for discovering low-dimensional structure in data is principal component analysis (PCA), which assumes that the data are drawn from a single low-dimensional affine subspace of a high-dimensional space (Jolliffe, 1986, 2002a). This is the simplest and most popular dimensionality reduction tool, and it has found widespread applications in many fields such as computer vision (Turk & Pentland, 1991). One of the biggest endeavours of the 21st century is the ability to extract meaningful information from large volumes of data; big data, the concept of big data is simply a pool of data made up of billions of variables. Analysing big data requires using a more robust technique since existing classical methodologies have not been designed for this regime.

Such large volume of data, comes with much noise or error, therefore there is the need to use a more robust technique to analyse the data whiles minimizing the error. We call the generalization of the classical principal component analysis in order to make them robust to nuisance in the data, generalized principal component Analysis. Not only is it robust for detecting and minimizing error but also it can robustly find the correct subspace structure of the data by identifying clusters.

Another challenge that arises in the new regime is that we can no longer assume that the data lie on a single low-dimensional subspace or sub-manifold. This is because many modern data sets are not task specific. Instead, the data is collected, and the task emerges only afterward. For example, Facebook data, made up of several user statistics. Hence a data set can be mixed with multiple classes of data of different natures, and the intrinsic structure of the data set may be inhomogeneous or hybrid. In this case, the data set may be better represented or approximated by not one, but multiple low-dimensional subspaces or manifolds

Given a set of data points from a mixture of affine subspaces, how does one automatically learn or infer those subspaces from the data? A solution to this problem requires one to cluster or segment the data into multiple groups, each belonging to one subspace, and then identify the parameters of each subspace. To model data with such mixed subspace structures, we generalize the classical PCA method, so that it can simultaneously identify multiple subspaces from the data. This leads to the so-called subspace-clustering problem, which has received great attention in the last decade and has found widespread applications in computer vision, image processing, pattern recognition, and system identification (Vidal, Ma, & Sastry, 2016).

Generalized Principal Component Analysis (GPCA) has clear advantage over the classical PCA in the following manner:

1. GPCA aids in visual inspection of the data, and this helps to give a fair idea of the number of possible homogenous classes and their major characteristics.
2. When we cluster from the first principal components of GPCA rather than the raw data, this improves efficiency of estimation. The sequence of eigenvalues provides a good guideline for choosing the suitable number of components; incidentally, using the results of GPCA gets round the problem of possible linear transformations of the data, in particular standardization, since GPCA is invariant under any affine transformation (Caussinus & Ruiz, 1990).
3. The robust nature of GPCA gets rid of discording observations that are likely to spoil the displays as well as the clustering (in fact, non-robust GPCA is useful to detect outliers, but another generalization of PCA is simpler and more efficient for that (Caussinus, Fekri, Hakam, & Ruiz-Gazen, 2003).

Principal components contain most significant information, but they can be subjected to clustering algorithm, this usually expel any unnecessary nuisance and help reveal any cluster. Many researchers have investigated this approach by subjecting the PCA to clustering algorithms (Chae & Warde, 2006). PCA display data by using a criterion of maximum variance, which some researchers say is not necessarily the best way to visualize clusters or any salient feature of the data. However projection pursuit techniques (Ruiz-Gazen, Marie-Sainte, & Berro, 2010) aim to display the data by maximizing a criterion of heterogeneity, which is more closely related to the search of a partition.

Definition of Some Key terminologies

Extremely High Price or Extremely Low Price

Extreme values can defined simply as characteristic values or measures that are the smallest (Minimum) or the largest (Maximum). They describe categories that range from values existing in very high degrees to existing in very low degrees. These two terms are situated at the farthest end of the range of values. In other words, they are the Maximum and the minimum respectively for a number of statistics. Extreme High price, also known as Maximum price is often seen as exorbitant as it is far from the normal price charged for commodities generally in a market (Zafeiriou, Arabatzis, Karanikola, Tampakis, & Tsiantikoudis, 2018).

Whereas we might consider excessive expenditure as immoderate, extremely high prices seem to be what marketers sought to achieve for their wares so as to maximize their profit. This however does not help a buyer in the market, as is popularly said, ‘one man’s meat is another man’s poison’. The relevance of these definitions in measuring the poverty or otherwise of nations cannot be underrated, in an online article on extreme poverty, Hoy (2015) stated that the World Bank has made an announcement about arriving on a new definition of ‘extreme poverty’. According to the article, the world body has made ending extreme poverty a strategic goal as tabled in the sustainable development goals. Their goal is to end extreme poverty by the year 2030. Such extreme value measures of characteristics have been adopted over the years in dealing with some of the world's biggest challenges for example extreme hunger among others.

Calculating the number of people in extreme poverty across various countries depends heavily on data collected on the prices of goods sold in the countries. Now, as goods and services become more expensive, the less people can purchase with their limited amount of money. Measuring the difference in the prices across countries is known as purchasing power parity (PPP) (McNown & Wallace, 1989).However performing comparison of a standard consumption basket across countries is bedevilled with challenges. Some of which are inadequate market information, and inconsistencies in the information available. Having said that, Economists embark on this challenge every 5 to 10 years – most recently in 2011 – and estimate changes in PPP between countries.

This new definition shows a significant departure from the previous definitions. In the past, the extreme poverty line (Goedhart, Halberstadt, Kapteyn, & Van Praag, 1977) was developed by finding the mean value of the national poverty lines of the world's poorest countries. Because every country has its own defined poverty lines using their own accepted parameters, which are not very different from what was accepted worldwide. The previous $1 a day measure (using 1985 PPP) was based upon an average of the 8 poorest countries in the world that had data available, while the $1.25 a day line was based on the average of the poorest 15 countries. This approach however has a problem that such that it has not been able to address the issue of eradicating extreme poverty.

The problem is that as long as some of the poorest countries in the world set their national poverty line (Hagenaars & Van Praag, 1985), that identifies some of their population as poor then there will always be people in extreme poverty. So the World Bank has attempted to counter this challenge by coming up with a new definition by increasing the $1.25 (2005 PPP) poverty line to $1.90 (2011 PPP), to reflect a general rise in the price of goods around the world (Handbury & Weinstein, 2015). This approach is independent of the changes in the national poverty lines of countries over time. It implicitly concedes that the $1.25 (2005 PPP) line is the most appropriate extreme poverty line and it only needs to be updated to reflect those changes in price levels. Even though some industry players see this approach as a phenomenon of shifting the goal post, the World body believes that this approach is more effective definition to help it achieve its goal of lifting millions from extreme poverty by 2030.

Low, High and Medium

As is expected low and high are the less extreme limits of a range covering a range of measure of a characteristic. High measures, the maximum price in a less extreme degree. High measures the upper limit of a measure of a characteristic whiles a low measure the lower limit of a measure of the same characteristic from the median of population or sample. High prices are usually not unexpected in a commodity price (Dewbre et al., 2008). They are common because of sharp drop in the quantity of food crops during the harvest season. On the other hand, for an imported food item like imported rice, the import duties when increased can cause its price in the local market to appreciate (Headey & Fan, 2008). Even though there might be other causes of the high pricing of commodity by marketers daily (Prices, 2008). Similarly, a reduction in the price of a commodity to low could be due to a number of reasons, notable among them is that there was a bumper harvest in the previous harvest season. This can cause the prices of commodities to be low. Medium price is the price that is considered reasonable by most marketers. Moderate prices appear to be a compromise between the sellers and the buyers of a food product. It seems only proper as it serves the interest of sellers and buyers too. It can be said that high price favours the seller but is not what the buyer would want to settle for. On the other hand, low price favours the buyer (Solaymani, 2017) and does not in any way help the seller except in very bizarre situation where the seller is 'trading off' the wares because it is getting old or rotten or to make room for new stock.

Chapter Summary

Chapter two covers areas such as the introduction of the chapter with the studies done in the area of food commodity using similar data. It considers the multiple multivariate techniques by looking at some of the most commonly used multivariate techniques and how they are applied. in addition it reviews some challenges of multivariate data with a brief review on the classical principal component analysis PCA and other related works. The assumptions that underline the use of the PCA were mentioned and the interpretations assigned. The chapter reviews dimension reduction, data visualizations with machine learning. A thorough review of the generalized principal component was done followed by the definition of some key terminologies.

It has generally reviewed the literature on the multivariate techniques employed in this work and other related studies that use this same kind of data. Some commonly used techniques of multiple multivariate analysis were briefly reviewed. It was noted that classical principal component analysis is extensively used in studies that employ this kind of dataset. However, there appears to be a challenge as to how robust the classical principal component analysis is. The need to employ a more robust technique in analysing such multivariate data was made. This technique is introduced in the chapter as Generalized Principal Component Analysis. The motivation for the use of this technique is to be able to address the challenges identified in the problem statement, which identifies a number of multiple multivariate statistical methods that were already used for the study of similar data problem, hence justifying the need to use Generalized Principal Component Analysis (GPCA). The chapter has reviewed in detail, the literature around the multivariate technique GPCA. The reviews also covers areas in which this technique is known to be popularly used, for instance in facial recognition, visual data inspection, image processing and modelling of subspaces of multiple multivariate data structures. The chapter further review some key terminologies that throws light on the objectives of the study and the need to explore the subject further.

CHAPTER THREE

RESEARCH METHODS

Introduction

In this chapter, we discuss the basic concepts and definitions of Generalized Principal Component Analysis. We further discuss the generalized principal Component technique. We formulate and demonstrate the algorithm for the generalized principal component. We develop an appropriate index for classifying data. We perform exploratory analysis to reveal the association between the variables. We use the descriptive statistics, Scatter plots to reveal the linear relationship among the variables.

Organization of Multivariate Data

Multivariate data arise whenever an investigator, seeking to understand a social or physical phenomenon selects a number $P \geq 1$ of variables to record (Johnson & Wichern, 2007). The observations are all recorded for each distinct item, individual, or experimental unit. We will use the notation x_{jk} to indicate the particular value of the k^{th} variable that is observed on the j^{th} item. That is

x_{jk} = measurement of the k^{th} variable on the j^{th} item.

Consequently, *m* measurements on *p* variables can be displayed as follows:

$$
\begin{array}{lcccccc}
 & \text{Variable 1} & \text{Variable 2} & \ldots & \text{Variable k} & \ldots & \text{Variable p} \\
\text{Item 1:} & x_{11} & x_{12} & \ldots & x_{1k} & \ldots & x_{1p} \\
\text{Item 2:} & x_{21} & x_{22} & \ldots & x_{2k} & \ldots & x_{2p} \\
\vdots & \vdots & \vdots & \ldots & \vdots & \ldots & \vdots \\
\text{Item } j\text{:} & x_{j1} & x_{j2} & \ldots & x_{jk} & \ldots & x_{jp} \\
\vdots & \vdots & \vdots & \ldots & \vdots & \ldots & \vdots \\
\text{Item } m\text{:} & x_{m1} & x_{m2} & \ldots & x_{mk} & \ldots & x_{mp}
\end{array}
$$

These data can be displayed as rectangular array called X, of m rows and p columns.

$$X_{(m\times p)} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2k} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{j1} & x_{j2} & \cdots & x_{jk} & \cdots & x_{jp} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mk} & \cdots & x_{mp} \end{bmatrix}$$

The array **X,** then, contains the data consisting of all of .the observations on all of the variables.

Characteristics of Principal Components

First Principal Components

The first principal component is the linear combination $a_1'X$ that maximizes $\text{var}\left(a_1'X\right)$ subject to $a_1'a_1 = 1$

Second Principal Components

The second PCA is the linear combination $a_2'X$ that maximises $\text{var}\left(a_2'X\right)$ subject to $a_2'a_2 = 1$, $\text{cov}\left(a_1'X, a_2'X\right) = 0$.

Third Principal Components

The third principal component is a linear combination $a_3'X$ that maximises $\text{var}\left(a_3'X\right)$ subject to $a_3'a_3 = 1$, $\text{cov}\left(a_1'X, a_3'X\right) = 0$ and $\text{cov}\left(a_2'X, a_3'X\right) = 0$.

i^th Principal Components

The i^{th} principal component is the linear combination $a_i'X$ that maximizes $\mathrm{var}(a_i'X)$ subject to $a_i'a_i=1$, and $\mathrm{cov}(a_i'X, a_k'X)=0$, $\forall i<k$

Suppose the $\sum$ is the covariance matrix associated with the random vector $X'=[X_1, X_2, \ldots, X_p]$. Let $\sum$ have the eigenvalue-eigenvector pairs (λ_1, e_1), $(\lambda_2, e_2), \ldots, (\lambda_p, e_p)$, where $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots \geq \lambda_p \geq 0$. Thus, the ith Principal component is given by $Y_i = e_i'X = e_{i1}X_1 + e_{i2}X_2 + \ldots + e_{ip}X_p$ where $i = 1, 2, \ldots, p$.

$\mathrm{var}[Y_i] = e_i' \sum e_i = \lambda_i$, $i = 1, 2, \ldots, p$ and $\mathrm{cov}(Y_i, Y_k) = e_i' \sum e_k = 0$, $i \neq k$

Note that, the PCs are uncorrelated and their variances are equal to the eigenvalues of $\sum$.

Generalized Principal Component Analysis

Generalized Principal Components Analysis provides a way to draw out a structure from a noise. Most often, projection pursuit techniques deal with the search for a 'global' structure, in particular, the search for clusters (Govaert, 2003);

Suppose that the expectation

$$\mathrm{E}[x_i] = 0$$

We expressed the theoretical variance covariance matrix of x_i as

$$\sum = \sum_W + \sum_B , \ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

where $\sum_B$ is variance of the structure part of the model (1)

and $\sum_W$ is the variance of the error part (noise) and it is assumed non-singular.

We consider a linear combination $a'x_i$ that maximises the structural variance verses the error variance by maximising the ratio

$$\frac{a'\sum_B a}{a'\sum_W a}$$

According to the property of eigenvectors and eigenvalues, this maximum is attained when **a** is eigenvector of $\sum^{-1}{}_W \sum_B$ associated with the largest eigenvalue. If **u** is the $\sum^{-1}{}_W$ normed eigenvector of $\sum_B \sum_{Ww}^{-1}$ associated with the largest eigenvalue $(a'x_i)u = \left(u'\sum_W^{-1} x_i\right)u$ then $\sum_W^{-1}$ is orthogonal projection of x_i on the subspace spanned by **u** (Govaert, 2003). For a positive definite matrix M, the M-norm of a vector $\mathbf{x}$ is $x'Mx$ and the vectors $\underset{\sim}{x}$ and $\underset{\sim}{y}$ are M orthogonal if and only if $x'Mx = 0$. This gives the best projection in one dimension and this process can be extended to obtain projections on k-dimensional subspaces.

The best k-dimensional projection of the units x_i is the $\sum_W^{-1}$ orthogonal projection on the subspace spanned by the k-eigenvectors associated with the k largest eigenvalues of $\sum_B \sum_W^{-1}$. This is known as the generalization of principal components with metric $\sum_W^{-1}$. At this point, the dimension k is not known or rather it is arbitrary. Also if the rank $rank(B) = q$, that is all significant structural features lie in a q-dimensional subspace, this subspace is spanned by the eigenvectors associated with the q strictly positive eigenvalues of $\sum_B \sum_W^{-1}$ (q eigenvalues of $\sum \sum_W^{-1}$ strictly larger than 1) (Caussinus & Ruiz-Gazen, 2010). In practice however neither $\sum_B$, $\sum_W$ nor even the reality of the model (1) is certain. Therefore, if the units are divided into two known groups with different means and variance, then it becomes a discriminant factor analysis problem. If the true structure of the model is unknown, it is necessary to use a robust estimator for $\sum$ which is S. The robust estimate of the covariance matrix is an estimate of $\sum_W$.

The robust estimate of $\sum$ is the S_n whiles V_n is the robust estimator of $\sum_W$ (Caussinus & Ruiz- Gazen, 1990). We implement a projection pursuit with the S_n or V_n in place of $\sum$ and $\sum_W$ (Yenyukov, 1988).

Suppose S_n is an estimate of $\sum$ in the form of a local variance;

$$S_n = \frac{1}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \delta(X, i, j)(x_i - x_j)(x_i - x_j)',$$

Where δ is dependent on the whole data set and the unit is (i, j). In practice δ is a kind of cut off in order to ensure that S is in the same units. The choice of δ is not fixed but some researchers agree that δ is equal to 1 or 0 according to literature, depending on whether

$\|x_i - x_j\|_{S^{-1}{}_w} = (x_i - x_j)' S^{-1}{}_w (x_i - x_j)$ is the least of the Mahalanobis distance.

But Caussinus & Ruiz, (1990) suggested that

$\delta = k\left(\|x_i - x_j\|^2_{S^{-1}{}_w}\right)$, where k = kernel function.

We consider

$$S_n = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x}_n)(x_i - \bar{x}_n)'$$

$$V_n = \frac{\sum_{i=1}^{n} \exp\left(\frac{-\beta}{2}\left\|x_i - \bar{x}_n\right\|^2_{S^{-1}_n}\right)(x_i - \bar{x}_n)(x_i - \bar{x}_n)'}{\sum_{i=1}^{n} \exp\left(\frac{-\beta}{2}\left\|x_i - \bar{x}_n\right\|^2_{S^{-1}_n}\right)}$$

where $\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$, $\left\|X\right\|^2_M = x'Mx$ and β is the parameter.

We must indicate that the matrices depend on the tuning parameter. The choice of a suitable β is vital and this choice is different depending on the matrix involved. The GPCA is formulated based on this matrix.

Now exploratory data analysis achieves the purpose of visualizing and classifying data. The data mostly used consist of an n number of objects times p variables real matrix, when *n* and *p* are large, the aim is to synthesize the huge quantity of information into an easy and understandable form. Principal Component Analysis (PCA) and related techniques (e.g. correspondence analysis) are the most popular tools of visualization and are often used to complement partition-type clustering techniques (Caussinus & Ruiz-Gazen, 2007).

Let X be a matrix of n x p dimensions. Let X_i denote the transpose of the i^{th} row of the matrix X. The empirical mean of $\bar{X} = 0$ is also denoted by i^{th} , and the empirical variance covariance matrix be represented by the non-singular matrix V. For a $p \times 1$ column vector x, we define a norm $\left\|x\right\|_{V^{-1}} = x'V^{-1}x$ where x' is the transpose of the vector x.

We set $T(\beta)$ a variance covariance as a function of a parameter β

$$T(\beta) = \frac{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} w_{ij}(\beta)\left(X_i - X_j\right)\left(X_i - X_j\right)^T}{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} w_{ij}(\beta)}$$

where

$$w_{ij}(\beta) = \exp^{\left(-\frac{\beta}{2}\|X_i - X_j\|^2_{V^{-1}}\right)}$$

and β is a tuning parameter according to literature. (Caussinus & Ruiz-Gazen, 2007).

The Generalized Principal Component Analysis (GPCA) involves projecting the matrix X_i onto the subspace spanned by the m eigenvectors of $VT^{-1}(\beta)$ associated with the m largest eigenvalues. According to Caussinus & Ruiz-Gazen (2007), one property of these projections, which is very important, is that it is invariant under any affine transformation of the data rows. In particular, the raw data or the standardized data provides the same display. Another useful property of the projections is about the sequence of the eigenvalues. That is the dimensions associated with the theoretical eigenvalues lower than a β value of +0.5 contains noise using a probabilistic model. Which gives a great information about how many principal components should be retained.

The cut off value of β to be +0.5 is cautiously used due to variability of sampling technique used, and the possibility of inadequacy of the models built. Some researchers have developed testing procedures (Caussinus, Fekri, Hakam, & Ruiz-Gazen, 2003).

The results of the GPCA are sometimes interpreted by using biplots or scatter (Gabriel, 1971). Our study used scatter plots to display the GPCA. In implementation, however the presence of extreme values may pull the GPCA towards their detection rather than towards the detection of clusters. Outliers are also seen as clusters themselves. In order to avoid this problem, we use a sample

Variance Covariance S(α),instead of the covariance matrix V, as S(α) is a more robust estimator of V (Ruiz-Gazen, 1996) to develop the GPCA.

Generalized Principal Component Analysis

We compute the GPCA based on the sample variance-covariance matrix, S.

$$S = \tfrac{1}{n}\sum\left(X_i - \bar{X}\right)\left(X_j - \bar{X}\right)'$$

$$S = \frac{1}{n} X_i X_i' \text{, if } \bar{X} = 0$$

Mahalanobis distance (Mahalanobis, 1936) is given as

$$D_M\left(X_i, X_j\right)^2 = \left(X_i - X_j\right)' S^{-1}\left(X_i - X_j\right)$$

We define a variance covariance matrix S^*, given by the equation

$$S^* = \frac{\sum_{i=1}^{n} k\left(\partial_m\right)\left(X_i - \widehat{X}_m\right)\left(X_i - \widehat{X}_m\right)'}{\sum_{i=1}^{n} k\left(\partial_m\right)},$$

Where,

$$\partial_m = \left\| X_i - \widehat{X}_m \right\|_{S^{-1}}^2$$

$$\left\| X \right\|_M^2 = X'MX$$

$$k(u) = \exp(-hu).$$

In the literature, h is usually assigned a value of 0.1 (Caussinus & Ruiz, 1990).

$S \times S^*$ is a matrix cross product whose Eigenvectors associated with the Eigenvalues is the GPCA. We obtain a projection by multiplying the Eigen vectors by the standardization of the data.

Codes for Computing GPCA

We write the algorithm for computing the GPCA in Matlab, shown in Appendix E; and we will further explain the algorithm used for the computation of the GPCA in Chapter four. We named the variable assigned to the data in the Matlab program FOODCOMB, a 455 by 19 numeric matrix. We designed the algorithm for each year specifically, and for all five years combined. We computed the GPCA for all the years individually and for the years combined as recommended in some studies elsewhere (Etienne et al., 2015, 2017).

Some matrix concatenation functions

cat function: This matrices concatenation function enables two matrices to be arrayed into one matrix. $C = cat(\dim, A, B)$ Concatenates matrix B to the end of Matrix A along dimension dim when matrices A and B have the same sizes (the lengths of the dimensions are similar except for the operating dimension dim). The same function can be written as $C = cat(\dim, A_1, A_2, ..., A_n)$ concatenates $A_1, A_2, ..., A_n$ along dimension $\dim$.

Horzcat function: This matrices concatenation function enables two matrices A and B to be arrayed into matrix $C = horzcat(A, B)$. $C = horzcat(A, B)$ Concatenates the matrix B horizontally to the end of matrix A when A and B have same sizes. $C = horzcat(A_1, A_2, ...A_n)$ concatenates A1, A2, … , An horizontally. Horzcat is equivalent to using square brackets for horizontally concatenating arrays. For example, [A,B] or [A B] is equal to horzcat (A,B) when A and B are compatible arrays. When the Concatenation method parameter is set to *horizontal*, the block concatenates the input matrices along *rows*. For horizontal concatenation, inputs must all have the same row dimension, M, but can have different column dimensions.

Vertcat function: This matrices concatenation function enables two matrices to be vertically arrayed into a single matrix. $C = vertcat(A, B)$ Concatenates B vertically to the end of A when A and B have compatible sizes (the lengths of the dimensions match except in the first dimension). $C = vertcat(A_1, A_2, \ldots A_n)$ Concatenates $A_1, A_2, \ldots A_n$ vertically. *vertcat* is equivalent to using square brackets for vertically concatenating arrays. For example, $[A; B]$ is equal to *vertcat* (A,B) when A and B are compatible arrays.

repmat function: This matrices concatenation function enables a matrix to be copied into the new matrix a number of times. $B = repmat(A, n)$ Returns an array containing n copies of A in the row and column dimensions. The size of B is size $(A) * n$ when A is a matrix.

blkdiag function: $B = blkdiag(A_1, A_2, \ldots, A_n)$ Returns the block diagonal matrix created by aligning the input matrices $A_1, A_2, \ldots, A_n$ along the diagonal of B.

Factor Analysis

Is a technique for modelling observed variables and their covariance structure in terms of a relatively small number of unobservable (latent) factors, to represent relationships among variables that are interrelated (Rencher, 2002). These factors are thought of as underlying constructs that cannot be measured by a single variable. The focus of this technique is to identify the underlying factors that describes the variability in a data set (Kleinbaum &Kupper, 1978). This method of data analysis has been used extensively in better part of the 2000s. It has been used for data reduction, instrument development, identifying patterns, hypothesis testing. Factor analysis is applied for two main reasons: To detect structure in the relationship between variables and for dimensional reduction.

We denote X_i to be measurements of observables i .the vector of observation is given by the expression

$$X = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_p \end{pmatrix}$$

In addition, X is a random vector, with population mean μ . If we assume that X is drawn from a population with population mean vector μ

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \vdots \\ \mu_p \end{pmatrix}$$

$\mathrm{E}(X_i) = \mu_i$, denote the population mean of the variable i

We consider m unobservable common factors $f_1, f_2, f_3, \ldots, f_m$. The ith common factor is denoted by f_i. But $m <<< p$

The vector of common factors are given by the vector **f**

$$f = \begin{pmatrix} f_1 \\ f_2 \\ f_3 \\ \vdots \\ f_m \end{pmatrix}$$

We compute the covariance of any two observable variables, Y_i and Y_j

$$Y_i = \beta_{i0} + \beta_{i1}F_1 + \beta_{i2}F_2 + (1)\varepsilon_i + (0)\varepsilon_j$$

$$Y_j = \beta_{j0} + \beta_{j1}F_1 + \beta_{j2}F_2 + (0)\varepsilon_i + (1)\varepsilon_j$$

$$\operatorname{cov}(Y_i, Y_j) = \beta_{i1}\beta_{j1}\operatorname{var}(F_1) + \beta_{i2}\beta_{j2}\operatorname{var}(F_2) + (1)(0)\operatorname{var}(\varepsilon_i) + (0)(1)\operatorname{var}(\varepsilon_j)$$

$$\operatorname{cov}(Y_i, Y_j) = \beta_{i1}\beta_{j1} + \beta_{i2}\beta_{j2}$$

We consider the factor model as a series of multiple regressions, predicting each of the observable variables from the values of the underlying common factors.

$$X_1 = \mu_1 + l_{11}f_1 + l_{12}f_2 + \ldots + l_{1m}f_m + \varepsilon_1$$

$$X_2 = \mu_2 + l_{21}f_1 + l_{22}f_2 + \ldots + l_{2m}f_m + \varepsilon_2$$

$$X_3 = \mu_3 + l_{31}f_1 + l_{32}f_2 + \ldots + l_{3m}f_m + \varepsilon_3$$

$$\vdots$$

$$X_p = \mu_p + l_{p1}f_1 + l_{p2}f_2 + \ldots + l_{pm}f_m + \varepsilon_m$$

We consider μ_1, variable means through μ_p as the intercept of the regression equations. The regression coefficients l_{ij} for this regression is the factor loadings. l_{ij} =loadings of the ith variables on the jth factor. We call these matrices of these l_{ij} as matrix of factor loadings.

$$L = \begin{pmatrix} l_{11} & l_{12} & \cdots & l_{1m} \\ l_{21} & l_{22} & \cdots & l_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ l_{p1} & l_{p2} & \cdots & l_{pm} \end{pmatrix}$$ = matrix of factor Loadings

The errors ε_i , are called specific factors for variable i . So that vectors of specific factors.

$$\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_p \end{pmatrix}$$ =vector of unique factors.

We assumed the basic model as a regression model. Each of the response variables X_i is predicted as a linear function of the unobserved common factors $f_1, f_2, ..., f_m$ also known as explanatory variables. Implying that m observed factors explained the variability in the data. In the simplest model, we assume that the factors affect the variable X linearly so that we consider a linear model of the form:

$$X = \mu + Lf + \varepsilon$$

ε is the error term. The ε are the specific or unique factors. We assume that f, and ε have zero means and are uncorrelated.

Model Assumptions

The assumptions of factor analysis are as follows;

$\mathrm{E}(\varepsilon_i) = 0$, $\mathrm{E}(f_1) = 0$, $\mathrm{E}(X_i) = 0$, Given $i = 1, 2, ..., p$

Variance:

$$\mathrm{var}(f_i) = 1$$

$\mathrm{var}(\varepsilon_i) = \psi_i$ Where $i = 1, 2, ..., p$

ψ_i is the specific variance

$\text{cov}\left(f_i, f_j\right) = 0, \ \forall i \neq j$

$\text{cov}\left(\varepsilon_i, \varepsilon_j\right) = 0 \ \forall i \neq j$

$$\text{cov}\left(\varepsilon\right) = \psi = \begin{pmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_p \end{pmatrix}$$

$\text{cov}\left(\varepsilon_i, f_i\right) = 0 \ \forall i \neq j$

The covariance $\text{cov}\left(f, \varepsilon\right)$ indicates a rectangular matrix containing the covariance of f and ε

$$\text{cov}\left(f, \varepsilon\right) = \begin{pmatrix} \sigma_{f_1\varepsilon_1} & \sigma_{f_1\varepsilon_2} & \cdots & \sigma_{f_1\varepsilon_p} \\ \sigma_{f_2\varepsilon_1} & \sigma_{f_2\varepsilon_2} & \cdots & \sigma_{f_2\varepsilon_p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{f_m\varepsilon_1} & \sigma_{f_m\varepsilon_2} & \cdots & \sigma_{f_m\varepsilon_p} \end{pmatrix}$$

Now, variance for the i^{th} observed variable is equal to the sum of the squared loadings for that variable and specific variance

$\sigma_i^2 = \text{var}\left(X_i\right) = \sum_{j=1}^{m} l_{ij}^2 + \psi_i$, where l_{ij}^2 is the communality and ψ_i is the specific or unique variance.

Factor Loadings

The loadings associated with the factor solution is represented by a matrix display where the number of the correlation of a specific factor with the original variable is known, as the factor loading is a correlation of a variable with a factor.

$corr\left(X_i, f_j\right) = \lambda_{ij}$

Principal Component Model

We consider each of the p possible principal components (y_i) expressed as a linear combination of the original variables (x_i) as follows;

$$
\begin{aligned}
Y_1 &= a_1^{'}X = a_{11}X_1 + a_{12}X_2 + \ldots + a_{1p}X_p \\
Y_2 &= a_2^{'}X = a_{21}X_1 + a_{22}X_2 + \ldots + a_{2p}X_p \qquad (1) \\
&\vdots \\
Y_p &= a_p^{'}X = a_{p1}X_1 + a_{p2}X_2 + \ldots + a_{pp}X_p
\end{aligned}
$$

$$y_i = \sum_{j=1}^{p} a_{ij}x_j .....(2)$$

$$j = 1, 2, 3, \ldots, \Lambda, \ldots, p$$

$$y_i = a_i^{'}x$$

$$\text{var}[y_i] = \text{var}[a_i^{'}x] = a_i^2\,\text{var}[x] = a_i^{'}\Sigma a_i,$$

where a_i are the loadings

$$\text{eigen vectors} = \begin{bmatrix} a_{11} \\ a_{12} \\ \vdots \\ a_{1p} \end{bmatrix},$$

Principal Component estimation of the factor loadings

We assume X_i be a vector of observations for the i^{th} variables,

$$X_i = \begin{pmatrix} X_{i1} \\ X_{i2} \\ \vdots \\ X_{ip} \end{pmatrix}$$

And we denote S as the variance covariance matrix, and we express it as

$$S = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(X_i - \bar{X}\right)'$$

The p eigen-values and corresponding Eigen vectors of the variance covariance matrix as

$\hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3, \ldots, \hat{\lambda}_p$

$\hat{e}_1, \hat{e}_2, \hat{e}_3, \ldots, \hat{e}_p$

$$\Sigma = \sum_{i=1}^{p} \lambda_i e_i e_i' \cong \sum_{i=1}^{m} \lambda_i e_i e_i' = \left(\sqrt{\lambda_1} e_1 \; \sqrt{\lambda_2} e_2 \quad \cdots \quad \sqrt{\lambda_m} e_m \right) \begin{pmatrix} \sqrt{\lambda_1} e_1' \\ \sqrt{\lambda_2} e_2' \\ \vdots \\ \sqrt{\lambda_m} e_m' \end{pmatrix} = LL'$$

The estimator for the factor loadings is given as $\hat{l}_{ij} = \hat{e}_{ji} \sqrt{\hat{\lambda}_j}$

The specific variances are computed using the expression

$$\hat{\psi}_i = s_i^2 - \sum_{j=1}^{m} \lambda_j \hat{e}_{ji}^2$$

Where s is the variance covariance matrix and $\sum_{j=1}^{m} \lambda_j \hat{e}_{ji}^2$ is the communality.

Discussion of Principal Components in Market Prices

Use of principal component analysis should be based, however, on grouping the analysed variables according to some hypothesis or design, literature suggests "that such studies are likely to be more productive of definite results" than analysing unrelated variables in the hope of discovering the underlying factors or groupings within the data. Discussion of the computation of component scores is meaningful in relation to the data under examination. A generalization of the discussion may run into difficulties. Thus, we review this section with specific reference to the market price data.

Y_i is the ith principal component, written as $Y_i = \sum_{j=1}^{p} a_{ij} X_j$. Clearly, the value of this expression will be influenced by the observed values of the variables X_j. The variability in the values of these variables will then be reflected in the component scores. This can distort interpretation of the score. Variation in measurement scales is dealt with by standardizing the data; otherwise, we can also use the correlation matrix. We first standardize the data. Thus, the component score corresponding to Y_i , is

$$C_i = \sum_{j=1}^{p} a_{ij} \frac{X_j - \mu_j}{s_j}$$

The standardization process determines the magnitude and the sign of the j^{th} . Assuming that all the weights a_{ij} , are positive, we consider three typical scenarios C_i , could be a high positive value, a high negative value or close to zero. A high positive score is obtained if the values of most of the items X_j , are higher than the average values of the respective items. Thus, in relation to the market price data used for this study, a high positive score C_i , indicates that the market's prices in all the commodities, X_j are consistently much higher than the average price for all the commodities. Such a market is a typically high-priced market. A high negative score is obtained if the values of most of the items X_j are lower than the average values of the respective items. Thus, in relation to the market price data, a high negative score C_i , indicates that the market prices in most of the commodities are consistently much lower than the average price for all commodities. Such a market is a typically high-priced market.

A very small score (close to zero) is obtained if the values of most of the items are just about the same as the average values of the respective items. Thus, in relation to the market price data, a small value of C_i, indicates that the market consistently has prices of all the commodities that are about the same value as the average price of all the commodities. Such a market is a typical average (moderate) priced market. The above discussion is on the effect of the standardization of the data on the sign and size of the component score assuming that the loadings are all positive. However, the magnitude of the score is determined largely by the size of the loading a_{ij}, on the variable X_j. In the case where the loadings are almost equal, the principal component, Y_i, is usually referred to as a *weighted sum* of the original variables. In this case, all the variables have about the same influence in the formation of the component. In relation to the market price data, a high component score means that the market has commodities with high prices, and hence, a high-priced market. If the component score is small (close to zero) then it implies that the two commodities have average (moderate) prices in that market.

Standard Error of the mean and the coefficient of variation

Two basic statistics will be important to consider, these are the standard error of the mean and the coefficient of variation. The standard error of the mean of a variable X_i is given by the expression: $SE_{(\bar{x})} = \frac{S_{X_i}}{\sqrt{n_i}}$. Where n_i is the sample size on the variable X_i whose means is $\bar{X}$. The measure of the spreads of the observations from the variable about the mean. Considering the nature of commodity prices, the smaller the $SE_{(\bar{X})}$ indicates that the prices are stable. And the higher the $SE_{(\bar{X})}$ indicates that the variation in the prices of the food item are wide from market to market. We ensure the number of observations are the same for all the variables so we can correctly imply the meaning of the $SE_{(\bar{X})}$.

The coefficient of variation of the variable X_i is given by the ratio of the standard deviation to the mean of the variable expressed in percentage. For a sample of variable

$$CV(X_i) = \frac{S_{X_i}}{\bar{X}} \times 100$$

Descriptive Statistics

Table 5 displays the basic descriptive statistics of the nineteen commodities. The first column is the names of the various variables, which are the commodities. The second column is the mean price of each of the commodities from the various markets. The third column, which is the standard error of the mean (*SE* Mean) is the ratio of the standard deviation of the commodity food prices to 21.331 (the square root of 455). Using the variance covariance matrices, we obtain the standard deviation of the prices. There appears to be a direct relation between Standard deviation and the standard error increase in standard error recorded a corresponding increase in the standard deviation. Standard deviation shows how widely the various prices are spread out from their respective average prices. There is also the Coefficient of variation, CoefVal. The statistic measures the relative variability in a data around the mean. It compares the relative variability in from one food item to the other. We obtain the value of the coefficient by dividing the Standard deviation found in the price of the commodity by its mean and multiply the result by 100.This statistic variation, which makes it possible to check accuracy of the measurements of the respective averages of the prices of the food items. This is particular important since the pricing of the various commodities to a large extent depends on the how each food item is packaged as there measuring unit for sale for each differs from one item to the other. For example, whiles palm oil is measured in Litres, Cassava is measured in the weight of the tubers.

The range of the prices of the commodities is the values, which are highest and lowest, denote the prices, which are not close to the general price value. The skewness of the distribution of the commodity prices is the distortion in the data. The extent to which a data varies from the normal distribution. All the values in the table are positive indicating positively skewed commodity price data.

Tables 5: Descriptive Statistics of commodities for the year 2008

Var	Mean	**SE Mean**	**StDev**	**CoefVar**	**Minimum**	**Maximum**	**Skewness**
Mz	52.59	1.21	11.56	21.98	29.88	92.00	0.78
YmWt	96.90	3.35	31.94	32.96	30.00	205.00	0.71
Cv	17.25	1.08	10.26	59.51	3.64	45.33	1.10
Tm	49.44	1.84	17.55	35.50	15.20	91.67	0.17
GEg	16.336	0.89	8.476	51.89	6.00	46.250	1.81
PpDr	58.06	2.06	19.62	33.79	18.00	108.29	0.39
GnR	113.41	3.02	28.86	25.44	59.82	185.00	0.65
CpWt	115.06	2.99	28.51	24.78	57.50	194.22	0.15
PmOil	24.551	0.81	7.685	31.30	14.00	46.750	0.73
Org	5.452	0.43	4.125	75.66	1.40	23.333	2.67
Ban	3.145	0.25	2.359	75.00	0.80	14.125	1.75
HrSm	29.17	1.66	15.85	54.32	10.00	66.34	1.04
Kbi	45.66	2.09	19.96	43.72	17.00	130.00	1.28
On	97.08	2.96	28.20	29.04	47.78	160.67	0.27
Eg	4.4219	0.06	0.584	13.21	3.50	7.67	2.30
Pltn	4.432	0.21	1.977	44.62	1.80	9.29	0.92
Gri	40.298	0.89	8.446	20.96	20.14	65.96	0.39
RiLoc	100.95	2.28	21.79	21.59	40.65	144.00	-0.31
RiImp	62.69	1.05	9.98	15.92	40.67	119.20	1.90

Source: Field Data, SRID-MoFA 2008

Relationship between the Prices of the commodities

We may not be able to do comparison between the prices of the commodities since they are not all priced with the same metric. However, we can do comparison of other basic statistics like the standard deviation, Mean and the coefficient of variation. The average unit price of Egg is in the market over the five year period is 8.22 with a coefficient of variation of 44.62 and StDev of 3.95 whereas the unit price of Banana is 6.622 with a coefficient of variation of 80.85 and StDev of 5.354. Clearly, the dispersion in the prices of both commodities are not extreme. Unlike red groundnut, which have very large standard deviation of 117, indicating a large dispersion in the prices of the commodity. Another characteristic to consider is whether the movement of another commodity causes one commodity. There is a linear association between the prices of commodities as seen by the positive correlation coefficients in the Appendix A. Many scholars adopted the idea of causality to explain certain behaviours between the commodity prices (Geweke, 1984; Cartwright, Kamerschen& Huang, 1989; Price, 1979; Uri, Howell, & Rifkin, 1985). One reason probably for this is the correlation does not imply causality. To explain the co-movement of prices among food prices, researchers use the granger causality tests to understand the causal relationship between them. According to Cartwright et al., (1989)

> "Significant instantaneous causality indicates that price movements in one area (product) are temporally correlated with price movements in the other area (product)" and "Statistically significant unidirectional causality indicates that current prices

in one area (product) are influenced by price movements in the other area (product)”

Variance-Covariance Matrix

We display the variance-covariance matrix given in Appendix C. From the variance-covariance matrix results, the covariance between pairs of the variables is high and this can be seen as translated into high correlations among the food items. The correlation matrix is given in Appendix L. The diagonals represented the variances of the food items. The high covariance between pairs of variables indicates that the nineteen food items adequately indicate the large variability in the prices of food commodities in the various markets selected for the study. The entries displayed in the diagonal show the variations present in the prices of the food items. We observe that Egg has the smallest variation. This is followed by Banana and Plantain. This means the prices of the three food items do not adequately capture the total variability in the prices of these food items in the markets considered for the study. However the other food items have high variations with Red Groundnut having the highest variation followed by White Yam, Tomatoes, Dry Pepper and then the White cowpea followed closely by the rest. These commodities adequately explain the variability in the prices of these items in the commodity prices.

Chapter Summary

Chapter three introduces the multivariate organisation of the data, and then considers some characteristics of classical principal component analysis (PCA). The methodology for the GPCA was formulated and developed. The algorithm for the full application of the methodology of the GPCA was developed and implemented. The study employed an important matrix manipulation function known as concatenation. Principal component factor analysis was introduced to aid in the labelling of the food commodities. The models of the principal and factor analysis were developed. A discussion of the principal factor analysis of the market prices were done. Some important basic descriptive statistics were shown to highlight the relationship among various food items. In all, the chapter three is about the methods and techniques engaged in the study; GPCA. This technique is envisaged to address the issues raised in the statement of the problem.

The GPCA was developed mathematically using very robust estimators of the variance covariance models (Caussinus & Ruiz, 1990). It was obvious from the method developed that the GPCA is more computationally tractable and hence more robust. This makes it a more suitable technique for carrying out the objectives of this study outlined in chapter one of the study. Indeed the results from the implementation of this technique using the algorithm formulated appears to be similar to those results from similar studies using different technique. It seems the results differ slightly based on the technique used even within the same study (Eyiah-Bediako, 2019). The similarities between the two studies is not in doubt, as the GPCA is effectively a kind of extension on the classical PCA. The difference is in the core of the algorithm developed and used throughout the studies. Whereas the classical PCA uses the mean in the variance covariance, the GPCA uses the median as the location parameter. The median is a more robust location parameter than the mean. This makes the GPCA a lot more robust against outliers. This study therefore makes use of a different methodology that combines GPCA and Principal Component Factor analysis.

CHAPTER FOUR

RESULTS AND DISCUSSION

Introduction

The study examines the general price levels of the local markets in Ghana by identifying the classification of markets by groups based on these classes: Extremely High, Extremely Low, High, Low, or Medium. The generalized principal component analysis is used to compute the Eigenvectors associated with the Eigenvalues. The data used is non-consecutive five-year secondary data with nineteen variables made of certain foodstuffs sampled from ninety-one (91) different markets across ten (10) regions of Ghana, obtained from the Statistics, Research and Information Directorate of the Ministry of Agriculture.

For a particular year, the generalized principal components analysis (GPCA) is computed and it helps identify the various groupings underlying the market prices for that year showing that the classification is regionally based. That is the discrimination identified with each group is based on region. A similar generalized principal component is computed for all the other four years and the results reported. Finally, the GPCA is computed for all the years pooled together and the results reported and interpreted.

Statistical packages

The following computer programs were used for the analysis of the data at various stages of the study; they are also the most common statistical packages used for data analysis (IBM Corporation, 2020; MATLAB, 2020; Minitab, 2020). These applications are such that users can execute preprogrammed codes using tabs. Until recently, most researchers relied heavily on these kinds of applications for the analysis of data for various research works. Recent developments in script-based data analysis tools have revolutionized the way and manner in which analysis of research data is conducted. The development of R, has enjoyed worldwide patronage in the analysis of research data. Following the successful role out of the windows based program R, RStudio (2020) has also been developed. R, R-studio and Python are the most commonly used script based programs. Script based applications have the advantage that, the underlining codes behind the execution of the program can be seen and shared. These codes can be peer reviewed and optimized, adapted for use in related fields of study. It also allows others to follow up on how the codes are written and to be shared in the scientific research community.

Algorithm and its Explanation

The algorithm for the k^{th} year is presented at the appendix E

The algorithm uses the approach of redefining the variance covariance matrix in terms of a new location parameter, the median. The eigen vector derived from the matrix product of the covariance matrices gives the GPCA. Multiplying the eigenvectors (re-ordered according to decreasing order of the eigen values) by the standardization of the raw scores is known as the projection of the standard scores onto the eigenvectors. Some literature term this as the projection pursuit approach (Caussinus & Ruiz-Gazen, 2010).

Analysis of Year 2008 Data

In this section, we present the implementation of the codes to data in year 2008. It is demonstrated how the GPCA is obtained and how the indices are subsequently generated. Following the generation of the indices, the markets are then categorized according to the prescribed rule.

Results of Computation of GPCA for the year 2008

We compute the covariance matrix for the year 2008 data, and we computed a median centred Mahalanobis distance, and the sum of squares cross product for each observation with respect to the median. We computed the eigen-value and the associated eigenvectors. The following observations are made about the matrix of the eigenvectors: we observe that all the values of the GPCA are less than +0.50 with a few exceptions; GPCA8 has 0.5239, GPCA12 has 0.514, GPCA13 has 0.5322, GPCA14 has 0.7685, GPCA15 has 0.8757, GPCA19 has 0.5045 and 0.5542.

In order to obtain the indices to serve as the basis for classification of markets the standardized data for Year 2008, is projected onto the GPCA of that year. The resulting scores are found to range between -3 and 6. Figure 1, shows a plot of the resulting scores for the markets.

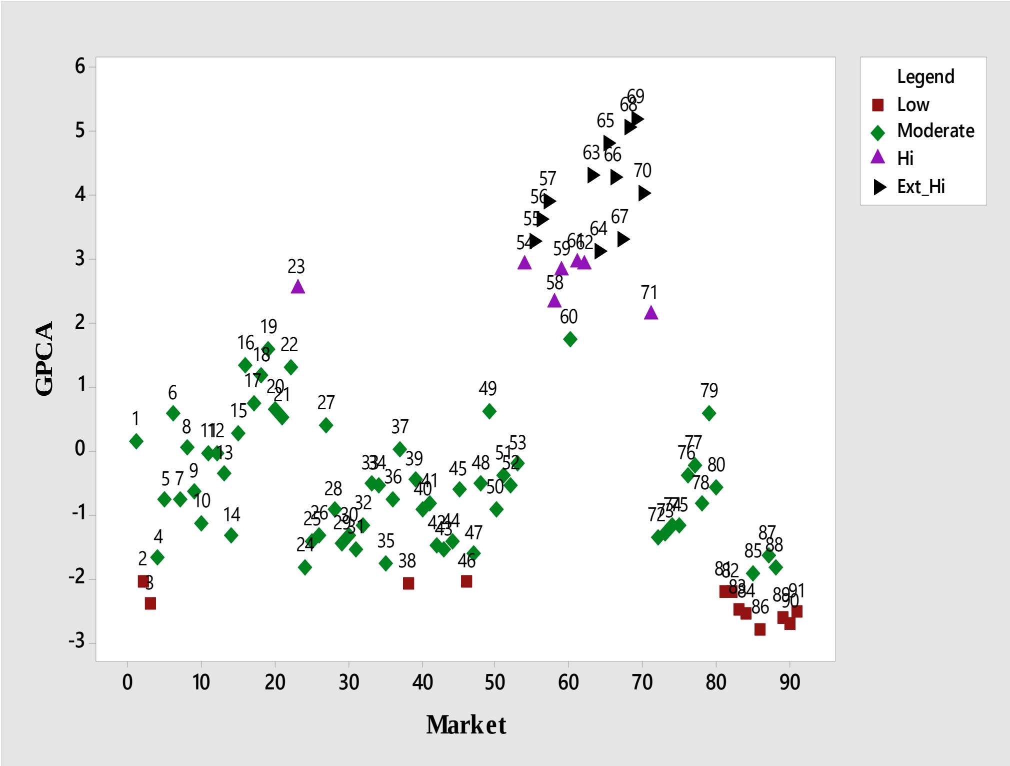

Figure1: Classification of price levels according to markets, 2008

Figure 1 shows the summary plot of the General Principal Component for the year 2008. The plot shows that there's no market with a score less than -3. This means that, according to our classification rule, no market is found to be extremely low priced that year. However, there are a number of markets that have scores greater than 3, indicating that in that year, several markets are found to be extremely high priced. There are several markets that have scores between -2 and 2. Thus, majority of the markets may be regarded as moderately priced. Table 6, presents a summary of categorization of the markets based on our adopted rule.

Table 6, shows a summary of the categorization of prices of food commodities in the market for the year 2008.

Table 6: Categorization of the prize levels for the year 2008

Limits of Index	Representation	Markets	Percentage
$Z \geq 3$	Extremely High	11	12.1
$2 \leq Z < 3$	High	7	7.7
$-2 < Z < 2$	Moderate	61	67.0
$-3 < Z \leq -2$	Low	12	13.2
$Z \leq -3$	Extremely Low	0	0.0

Source: Field Data, SRID-MoFA 2008

Extremely High Priced Markets in the year 2008

The values of the GPCA that falls in the range of 3.0 to 10.0, we refer to as Extremely High. The markets with corresponding GPCA falling within this range are extremely high priced. Eleven markets are seen in the extremely high priced category. From Figure 1, there are identified groups of markets. Table 6 shows that the most extremely priced market is market number 69, which is Tumu in the Upper West Region. Another extremely priced market is market 68 which is Bugubelle. It appears that, for the year 2008, the most expensive markets are in the northernmost part of Ghana.

High Priced Markets for the year 2008

We identified seven markets that are high priced in the year 2008. Notable among these markets are markets 61 and 62. Which are both located in the northern part of the country. From Table 6, we can identify that the rest of the markets are from the Volta and Brong Ahafo regions.

Moderate Priced Market for the year 2008

From table 6, we realized that about 61% of the markets across the country are medium-priced in the year 2008. The majority of the markets appear to come from the southern part of the country. Whereas a few others can be seen in the northern sector, we identified only one market from the northern part of the country as a medium-priced market; market 60 namely Damongo.

Low Priced Markets for the year 2008

The values of the GPCA that falls in the range -3.0 to -2.0 are referred to as low. The markets with corresponding GPCA falling within this range are referred to as extremely low prices. 12 Markets are low priced accounting for about 13% of the total number of markets sampled for the study. As can be seen from Table 6, there are no markets from the northernmost part of the country identified as a low priced market. Some notable markets are markets 2, 3 and 86, which are Tepa, Adugyaman, and Bogoso respectively. The lowest-priced market is market number 86, which is Bogoso. In the year 2008, the lowest priced market is Bogoso, in the Western region. It can be observed that the groupings are identified on a regional basis. Whereas some markets stand-alone, they turn to form groups with their neighbouring markets. The grouping formed by New Tafo and Kaneshie market appears to follow this line of reasoning. Another such example is the market 86 which appears to stand alone, however forming a group with the nearest neighbours markets 89, 90, 91, 83, 84, 81, and 82.

Extremely Low Priced Markets for the year 2008

Using the projection of the transformed data on the Eigenvectors associated with the Eigenvalues for the year 2008, we obtained the categorization for the five prize levels. The values of the GPCA that falls in the range -10 to -3 are referred to as extremely low. The markets with corresponding GPCA falling within this range are referred to as extremely low prices. From figure 1, there are no extremely low-priced markets. This implies that in the year 2008, no commodity market is identified as extremely low-priced. There could be some reasons for this; it could be a result of global commodity prices being high and as a result, the prices of the local commodities markets were impacted. Therefore, there are no identifiable lowly priced markets. This in summary implies that in the year 2008, no market recorded an extremely low price for any commodity.

Table 7: Market Categorizations for the year 2008

Market Category	Market Number	Region
Extremely High	54,55,56,57, 63,64,65,66,67,68,69,	Northern, Upper West & Upper East,
High	23,54,58,59,61,62,71	BrongAhafo,Northern , Upper East, Volta
Moderate	1,4,5,6,7,8,9,10,11,12,13,14,15,16, 17,18,19,20,21,22,24,25,26,27,28,29,30, 31,32,33,34,35,36,37,39,40,41,42,43,44, 45,47,48,49,50,51,52,53,60,72,73,74,75, 75,76,77,78,79,80,85,87,88	Ashanti, Brong Ahafo ,Central,Eastern, Greater,Accra, Northern,Volta, Western
Low	2,3,38,46,81,82,83,84,86,89,90,91	Ashanti,Eastern,GreaterAccra,Western
Extremely Low	-	N/A

Source: Field Data, SRID-MoFA 2008

Analysis of Year 2009 Data

In this section, we present the implementation of the codes to data in year 2009. It is demonstrated how the GPCA is obtained and how the indices are subsequently generated. Following the generation of the indices, the markets are then categorized according to the prescribed rule.

Results of Computation of GPCA for the year 2009

We compute the covariance matrix for the year 2009 data, and then we calculate the median centred Mahalanobis distance, and the sum of squares cross product for each observation with respect to the median. We compute the Eigenvalue and the associated Eigenvectors. The Eigenvectors associated with the Eigenvalues for the year 2009. Table 8 is the summary of the range of values of GPCA within which a market belongs in other to be referred to as belonging to that particular range. Z is the GPCA value.

Table 8: Categorization of the prize levels for the year 2009

Limits of Index	Representation	Markets
$Z \geq 3$	Extremely High	2
$2 \leq Z < 3$	High	6
$-2 < Z < 2$	Moderate	81
$-3 < Z \leq -2$	Low	2
$Z \leq -3$	Extremely Low	-

Source: Field Data, SRID-MoFA 2008

Figure 2 is the scatter plot of the first GPCA for the year 2009 which gives the classifications.

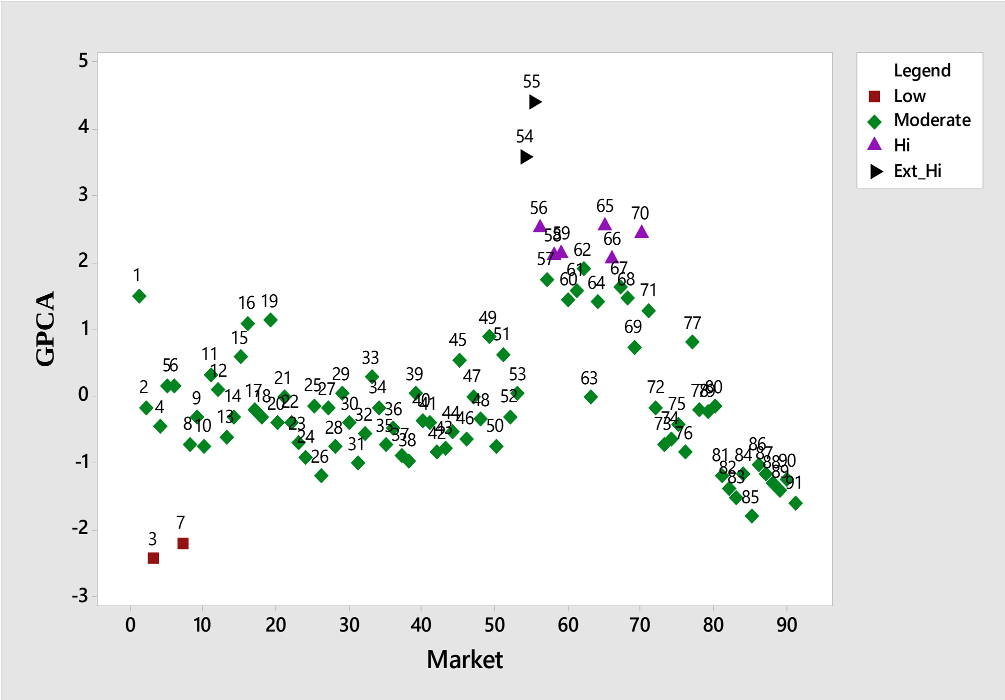

Figure 2: A scatter plot of the GPCA for the year 2009

Figure 2 shows the summary plot of the General Principal Component for the year 2009. The plot shows that there's no market with a score less than -3. This means that, according to our classification rule, no market is found to be extremely low priced that year. However, there are a number of markets that have scores greater than 3, indicating that in that year, several markets are found to be extremely high priced. There are several markets that have scores between -2 and 2. Thus, majority of the markets may be regarded as moderately priced. Table 8, presents a summary of categorization of the markets based on our adopted rule.

Extremely High Priced Market for the year 2009

Table 9 shows the group of markets that are extremely high priced. We identified only two markets as extremely high priced. Both of these Markets are based in the Northern region. We identified these two markets as extremely high in the previous year 2008, markets 55 and 54, namely Salaga and Bole respectively, both of which are markets situated in the Northern Region. This implies that in the year 2009, these two markets 54 and 55, respectively Bole and Salaga are the most expensive to purchase commodity food items from.

Highly Priced Markets for the year 2009

Highly-priced markets are those that are higher than average. We identified six markets as highly-priced. Table 8 shows the classification for markets that are high-priced. As can be seen from the table, the three Northern regions have high-priced markets. But we take note of market 56 which is Nalerugu and 66 which Garu, which in particular are high priced in 2009 but extremely high priced in the previous year, 2008.

Medium priced Markets for the year 2009

Table 9 shows the various identified groupings of moderately priced markets for the year 2009. The majority of the markets are medium-priced, as seen in the previous year 2008. But there are about 89% of the total markets sampled for the study that are identified as medium priced. In 2008 the majority of the markets across the country fall in the class of moderately priced markets. This is important because it shows that commodities in the majority of the markets across the country are not expensive. About 61% of the markets in Ghana fall within the class of moderately priced markets in Ghana in 2008, indicating an increase from the previous year of about 28%. So that we can say that fairly most commodities sold in the markets across Ghana are priced moderately for the year 2009.

Low Priced Markets for the year 2009

We identified only two markets 3 and 7 which are Adugyaman and Juaben respectively, both in the Ashanti Region as low-priced. As indicated earlier the two markets appear to be grouped and they belong in the same region. In the previous year 2008, market 3 was identified as low priced market and market 7 medium priced.

Extremely Low Priced for the year 2009

From Figure 2, the scatter plot of the first general principal component for the year 2009 reveals that there are no extremely low priced markets. A look at table 9, also indicates that we did not identify any extremely low-priced markets. The previous year 2008 also we did not identify any extremely low priced markets.

Table 9: Market Categorizations for the year 2009

Market Category	Market Number	Region
Extremely High	54,55	Northern
High	56,58,59,65,66,70	Northern,UpperEast,Upper West
Moderate	1,2,4,5,6,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,57,60,61,62,63,64,67,68,69,71,72,73,74,75,76,77,78,79,80,81,82,83,84,85,86,87,88,89,90,91	Ashanti,BrongAhafo, Central, Eastern,GreaterAccra, Northern,UpperEast,UpperWest,Volta, Western
Low	3,7	Ashanti,
Extremely Low	-	N/A

Source: Field Data, SRID-MoFA 2008

Analysis of Year 2012 Data

In this section, we present the implementation of the codes to data in year 2012. It is demonstrated how the GPCA is obtained and how the indices are subsequently generated. Following the generation of the indices, the markets are then categorized according to the prescribed rule.

Results of Computation of GPCA for the year 2012

We calculate the covariance matrix for the year 2012 data, we compute a median cantered Mahalanobis distance, the sum of squares cross product for each observation with respect to the median. Appendix I is the matrix of the GPCA for the year 2009 which is the Eigen Vectors associated with the Eigenvalues computed using Matlab. Figure 3 is the scatter plot of the first GPCA for year 2012. Table 10 shows the various categorizations of the markets for the year 2012.We see from figure three that there are a few extremely high markets in the year 2012. And majority of the scatter points are moderately priced. With low priced market thinly dotted below the pool of moderate scatter points.

Table 10: Categorization of the prize levels for the year 2012

Limits of Index	Representation	Markets
$Z \geq 3$	Extremely High	3
$2 \leq Z < 3$	High	8
$-2 < Z < 2$	Moderate	74
$-3 < Z \leq -2$	Low	6
$Z \leq -3$	Extremely Low	-

Source: Field Data, SRID-MoFA 2008

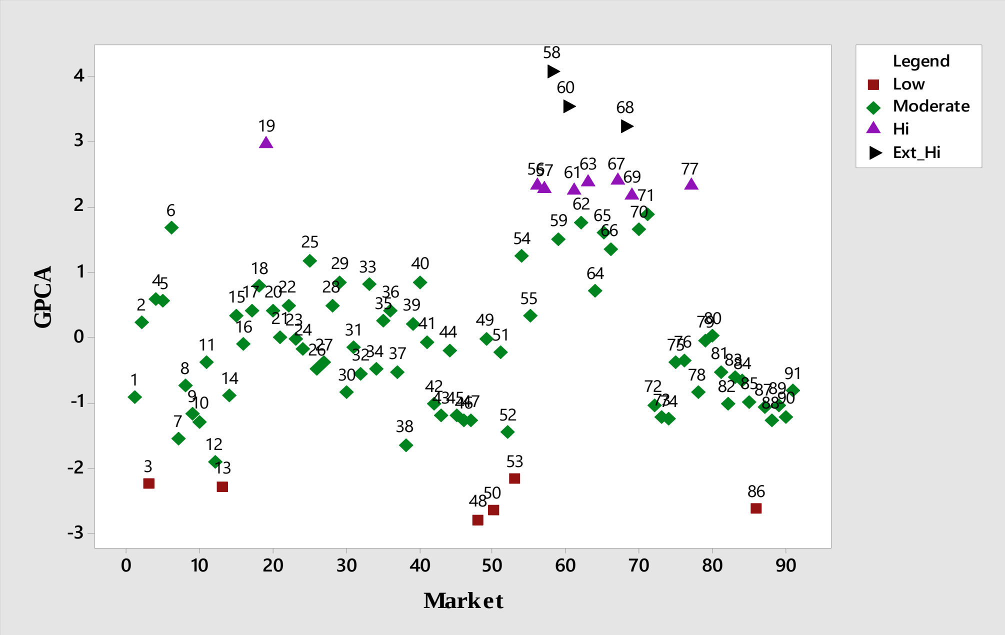

Figure 3: A scatter plot of the GPCA for the year 2012

Figure 3 shows the summary plot of the General Principal Component for the year 2008. The plot shows that there are a number of markets that have scores greater than 3, indicating that in that year, several markets are found to be extremely high priced. There are several markets that have scores between -2 and 2. Thus, majority of the markets may be regarded as moderately priced. Table 10, presents a summary of categorization of the markets based on our adopted rule.

Extremely High Priced Market for the year 2012

Table 11 shows that Market 58, Bimbilla is the most extremely priced in the year 2012. The market number 60 known as Damongo in the Northern Region is also an extremely high priced market so also is Markets 68 known as Bugubelle of the Upper West region. These three markets, accounting for 3.3% of markets in Ghana are extremely priced markets in the year 2012, which implies that very few markets across the country in the year 2012 are extremely priced. So, if we consider the total outlook of the country in terms of how expensive the country is, we can say that Ghana is not an expensive country indeed in the year 2012, very few markets are extremely high priced.

Highly Priced Markets for the year 2012

From Table 11, Market 19, Yeji in the Brong Ahafo Region stands alone as a high priced market. Markets 56 and 57 known as Nalerugu and Gushiegu respectively of the Northern region are both identified as been grouped. Market 61 and 63 Yendi and Zebilla also appear to group together. Again the number of high priced markets is few, less than 10% of the total markets sampled belong to the high priced market group. Worthy of note is market numbers 56 and 57 which are Nalerugu and Gushiegu respectively which were high price in the previous year and also high priced in the year 2012. This could mean that nothing fundamentally different has changed over the past three years for which the markets remain in the same High priced category. Indeed customers will benefit more if the market is less expensive as it reduces the amount of money they spend on purchasing a unit quantity of foodstuff. This allows more space on their expenditure to invest or spend toward other important family needs, thereby improving the livelihood of the customers and the nation at large. Hitherto, Yeji, Zebilla, Navrongo were medium-priced markets in year 2009. For some reason, these markets have moved into the high priced category. Might be that the farmers who supply foodstuff to these regions did not have a good harvest, which fundamentally, is usually the case when food items become high priced. Of course, some food items are imported into the local markets, implying that the source of supply is from an international market and so might not be subject to change in the price as a result of the occurrence or no occurrence of a bumper harvest.

Medium Priced Markets for the year 2012

Table 11 shows that majority of the markets are in the medium-priced category. About 87% of all markets in the country fall into this medium priced category. However, some notable markets among them are market 65, Fumbisie which was high priced in the 2009 year analysis. Market 66 which is Garu also was high priced in the 2009 year review. These two markets are in the Upper East region of Ghana. Whatever might have happened to reduce the two to be less expensive than the previous year's remain to be found out. We could only suggest that it might be due to the early rains of the years preceding the 2012 harvest season. Generally, it appears there has been an increase in the number of markets that belong to the medium-priced category. A 20% increase from the previous years. This could mean that the prices of markets generally across the country in the 2012 season were medium. This has significant economic benefits for the country as more people can afford to buy food items which are less expensive. This will also go a long way to alleviate the poverty of the people. Also, it seems groupings are on a regional basis.

Low Priced Markets for the year 2012

Table 11 shows the categories of prices of commodities in the markets. The market that is lowest in this group is the market 48 which is Ashaiman in the Greater Accra Region of Ghana. The rest are market number 50 which is Dome also in the Greater Accra Region, market number 53 which is the Madina market. There is also market number 86 known as Bogoso in the western region of Ghana. Market 3 is Adugyaman in the Ashanti region and market 13 is Goaso in the Brong Ahafo region. Market 3 was low in 2009 and is a low priced market in 2012. Only six markets out of the total markets sampled are identified as low priced. It is to be expected that these markets are located in local food crop growing communities.

Extremely Low Priced Markets for the year 2012

There appears to be no extremely low priced market for the year 2012. This could be due to some reasons. That the 2011 crop season has not been a very high yield one and so affected the volume of food crops harvested and sold in the subsequent year. It can also be due to a very low bumper harvest. In many cases where there happen to be bumper harvests, the general price of the commodities is cheaper or extremely low as farmers tend to sell much food products at very low prices. Further studies should be conducted if there is none done to link agriculture productivity to the extremity in the prices of the markets.

Table 11: Market Categorizations for the year 2012

Market Category	Market Number	Region
Extremely High	58,60,68	Northern,Upper West
High	19,56,57,61,63,67,69,71,77	Northern,UpperEast ,Upper West Volta, Brong Ahafo
Moderate	1,2,4,5,6,7,8,9,10,11,12,14,15,16, 17,18,20,21,22,23,24,25,26,27,28 ,29,30,31,32,33,34,35,36,37,38,3 9,40,41,42,43,44,45,46,47,49,51, 52,54,55,59,62,64,65,66,70,72,73 ,74,75,76,78,79,80,81,82,83,84,8 5,87,88,89,90,91	Ashanti,BrongAhaf o, Central, Eastern,GreaterAcc ra,Northern,UpperE ast,UpperWest,Volt a, Western
Low	3,13,48,50,53,86	Ashanti, Brong Ahafo, Greater Accra,Western
Extremely Low	-	N/A

Source: Field Data, SRID-MoFA 2008

Analysis of Year 2013 Data

In this section, we present the implementation of the codes to data in year 2013. It is demonstrated how the GPCA is obtained and how the indices are subsequently generated. Following the generation of the indices, the markets are then categorized according to the prescribed rule.

Results of Computation of GPCA for the year 2013

We calculate the covariance matrix for the year 2013 data, we computed a median cantered Mahalanobis distance, the sum of squares cross product for each observation with respect to the median. Figure 4 is the scatter plot of the first GPCA for year 2013 showing the categories dotted across the plot. Table 12 shows the various categorizations of the markets with the number of markets in each category for the year 2013.

Table 12: Categorization of the prize levels for the year 2013

Limits of Index	Representation	Markets
$Z \geq 3$	Extremely High	2
$2 \leq Z < 3$	High	5
$-2 < Z < 2$	Moderate	70
$-3 < Z \leq -2$	Low	12
$Z \leq -3$	Extremely Low	2

Source: Field Data, SRID-MoFA 2008

From Figure 4, only two extremely low priced commodities are seen. However the medium priced category is seen to span across the plot. With High priced markets thinly dotted on top of the medium priced markets. The extremely high price markets are seen at the far right on top of the plot.

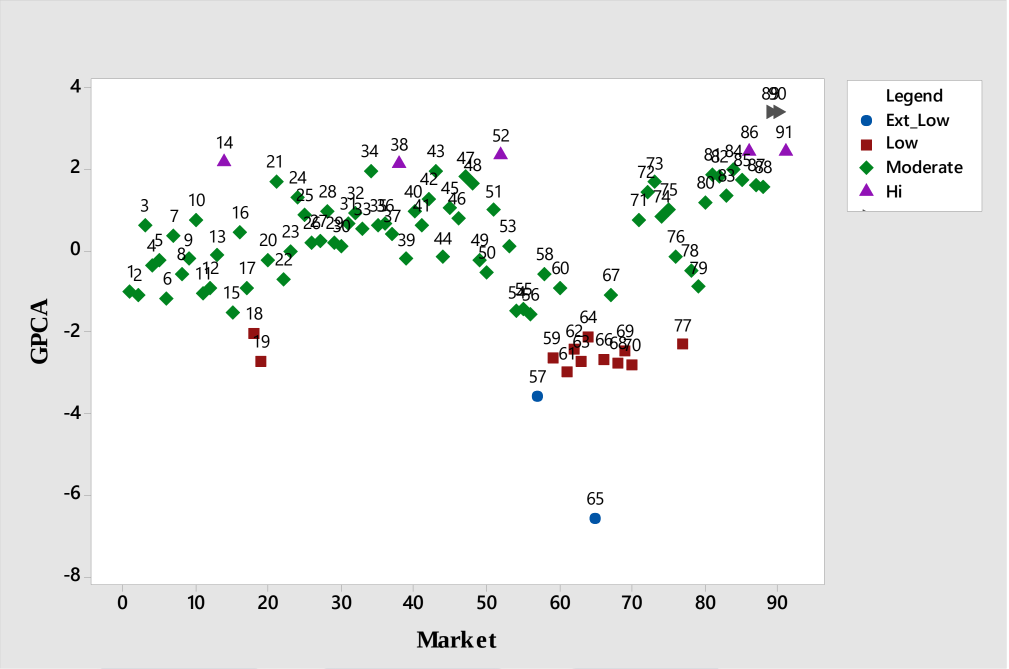

Figure 4: A scatter plot of the GPCA for the Year 2013

Extremely High Priced Market for the year 2013

Table 13, shows that there are only two markets that are extremely high priced in the year 2013. These two markets 89, 90 which are Sekondi and Takoradi respectively. This is to say that the most exorbitant markets in the country for the year 2013 are the markets in the western region, Sekondi and Takoradi. Until 2013, almost all the extremely High priced markets are all identified in the Northern regions in the previous years. These two markets were moderately priced in the previous year 2012. Again we see that very few markets are extremely high priced across the selected markets in the country.

High Priced Markets for the year 2013

From Table 13, in the year 2013, there were five highly priced market. These are market numbers 14, 38, 52, 86, 91 which are Kukuom in the Brong

Ahafo Region, New Tafo in the Eastern region, Tema in the greater Accra region, Bogoso in the western region and Tarkwah also in the Western region. Market 86 and 91 appear to belong to a group. It also goes to explain that fewer markets are highly priced across the nation.

Medium Priced Markets for the year 2013

Table 13 shows that majority of the markets are in the medium-priced category. About 77% of all markets in the country fall into this category. Market 67 which is Navrongo was high priced in the year 2009 but medium priced in 2013. As is usually expected the majority of the markets are medium-priced this appears to be the trend over the past few years from the data so far analysed. And indeed this appears to be based on a regional basis. From the previous year we see that the percentage of markets that were medium priced is about 87%, indicating a slight reduction in the number of markets by 10% that are medium priced in the year 2013. We can say that the markets generally in the country are moderately priced. This is a fairly good outlook for the country.

Low Priced Markets for the year 2013

Table 13 shows the markets that are low priced. The market that is lowest in this group is the market 19, which is Yeji, and Kintampo which is market number 18 both are in the Brong Ahafo Region. The low priced market constitute 13% of the total markets sampled for the study. And again we can say that generally Ghana's markets are not low priced, as a large percentage of the markets are medium priced.

Extremely Low Priced Markets for the year 2013

From Table 13, only two markets 57 and 65 known as Gushiegu and Fumbisie respectively are the extremely low priced market. Fumbisie is the extremely low priced market in the country in Ghana for the year 2013. Again from the data, 2013 is the first to record any lowest priced market. This means that in the year 2013, some markets became very cheap, perhaps due to rains and availability of food crops. In the previous year 2012, Fumbisie was a medium priced market. To ascribe reason to the change from a medium priced market to an extremely low priced market, this could be due to a good harvest season, that is high yields were recorded in the previous harvest season and so this markets tend to be extremely low priced. In many cases where there happen to be bumper harvests, the general price of commodities are cheaper or extremely low as farmers tend to sell much food products at very low prices. The percentage of extremely low markets in the country is only 2.2%.

Table 13: Market Categorizations for the year 2013

Market Category	Market Number	Region
Extremely High	89,90	Western
High	14,38,52,86,91	Brong Ahafo, Eastern, Greater Accra, Western
Moderate	1,2,3,4,5,6,7,8,9,10,11,12,13,15,16,17,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37,39,40,41,42,43,44,45,46,47,48,49,50,51,53,54,55,56,58,60,64,67,71,72,73,74,75,76,78,79,80,81,82,83,84,85,87,88,	Ashanti,BrongAhafo, Central, Eastern,GreaterAccra,Northern,UpperEast,UpperWest,Volta, Western
Low	18,19,59,61,62,63,66,68,69,70,77	Brong Ahafo, Volta, Northern, Upper East, Upper West
Extremely Low	57,65	Northern, Upper East

Source: Field Data, SRID-MoFA 2008

Analysis of Year 2015 Data

In this section, we present the implementation of the codes to data in year 2015. It is demonstrated how the GPCA is obtained and how the indices are subsequently generated. Following the generation of the indices, the markets are then categorized according to the prescribed rule.

Results of Computation of GPCA for the year 2015

We calculate the covariance matrix for the year 2015 data, we computed a median cantered Mahalanobis distance, the sum of squares cross product for each observation with respect to the median. Figure 5 is the scatter plot of the first GPCA for year 2015 showing the categories dotted across the plot. Table 14 shows the various categorizations of the markets with the number of markets in each category for the year 2015.

Table 14: Categorization of the prize levels for the year 2015

Limit of index	Representations	Markets
$Z \geq 3$	Extremely High	3
$2 \leq Z < 3$	High	1
$-2 < Z < 2$	Moderate	85
$-3 < Z \leq -2$	Low	2
$Z \leq -3$	Extremely Low	-

Source: Field Data, SRID-MoFA 2008

Figure 5, shows that only two low priced food commodity markets can be seen. However, the medium priced category is seen to span across the plot. With one High priced market spotted on top of the medium priced markets. The extremely high price markets are seen at the far top right corner of the plot. From the Figure 5, there are no extremely priced market identified.

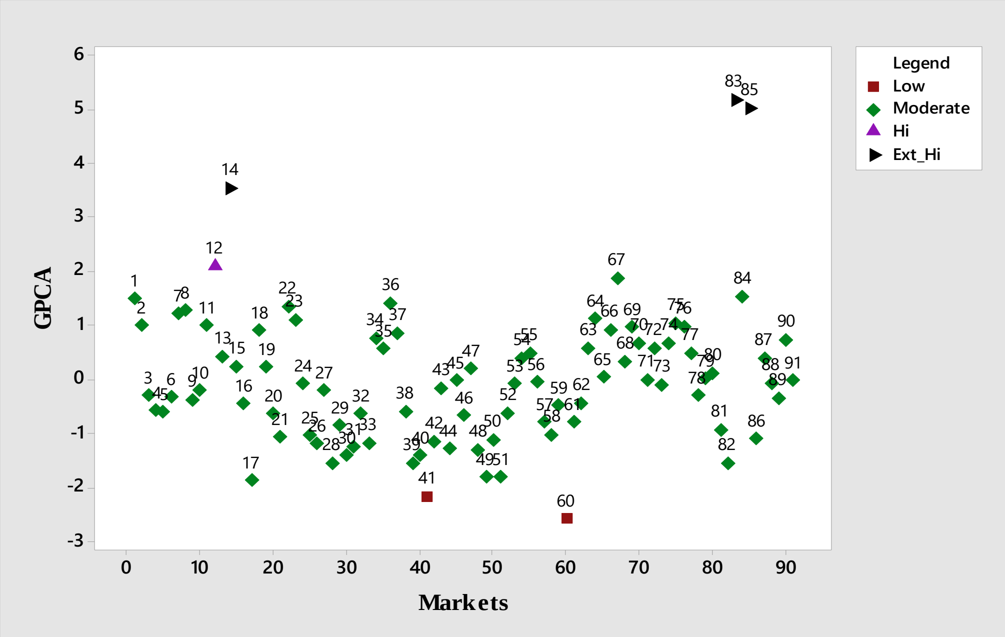

Figure 5: A scatter plot of the GPCA for the Year 2015

Extremely High Priced Markets for the year 2015

Three markets are identified as the most expensive commodity markets. Market 14, Kukuom in the Brong Ahafo Region, was very expensive. Markets 83 and 85 which are SefwiBekwai and Juabeso respectively all in the western region are identified as the most expensive markets in the country in the year 2015.

High Priced Market for the year 2015

Table 15 shows that only one market is high priced in the year 2015. That is market number 12, Nsuta by name, in the Ashanti region. This market have been moderately priced for the past years.

Medium Priced Market in the year 2015

Table 15, reveals that majority of the markets are in the category of medium priced markets. About 93% of all markets sampled for the study are identified in the medium priced category. This could be interpreted to mean that generally markets in Ghana are moderate priced in the year 2015.

Low priced Markets for the year 2015

Table 15 shows that only two markets 41 and 60, Agormanya in the Eastern region and Damongo in the Northern region were the lowest priced market across Ghana in the year 2015.

Extremely Low priced market for the year 2015

Table 15, shows that there were no markets that are extremely low priced in the year 2015. The markets which are in that category in the previous year 2013; markets 57 and 65, respectively Gushiegu and Fumbisie have all moved into the category of moderate priced markets in 2015.

Table 15: Market Categorizations for the year 2015

Market Category	Market Number	Region
Extremely High	14,83,85	BrongAhafo,Western
High	12	Brong Ahafo, Eastern, Greater Accra, Western
Moderate	1,2,3,4,5,6,7,8,9,10,11,13,15,16, 17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,37, 38,39,40,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56,57,58,59, 61,62,63,64,65,66,67,68,69,70,71,72,73,74,75,76,77,78,79,80,81, 82,84,86,87,88,89,90,91	Ashanti,BrongAhafo, Central, Eastern,GreaterAccra,Northern,UpperEast,UpperWest,Volta, Western
Low	41,60	Eastern, Northern,
Extremely Low	-	

Source: Field Data, SRID-MoFA 2008

Analysis of all the five years together

In this section, we present the implementation of the codes to all the five years pooled together. It is demonstrated how the GPCA is obtained and how the indices are subsequently generated. Following the generation of the indices, the markets are then categorized according to the prescribed rule.

In application of the algorithm for the entire five non-consecutive years, the variance covariance matrices for the five years are pooled. The sum of squares cross product for each year is also computed and the results of all five years are pooled. The product of both pooled matrices are calculated and their eigenvalues computed. The standardization of the data for each year is computed and then concatenated together into one broad matrix. The results from this standardization are projected onto the re-arranged eigenvectors from the product. Then we recode the values into the categories and then plot the scatter graph.

Explanation of the Codes for the five years pooled together

The codes for the implementation of the algorithm for the five years pooled together can be seen in Appendix F.

We calculate the pooled covariance matrix for the years together; we then computed a median cantered Mahalanobis distance, the sum of squares cross product for each observation with respect to the median. Appendix C is the matrix of the GPCA for the years together which is the Eigen Vectors associated with the Eigenvalues computed using Matlab. Figure 6 is the scatter plot of the first GPCA for all the five years together showing the categories across the plot. Table 17 shows the various categorizations of the markets with the number of markets in each category for the years together.

From Figure 6, we can see the low priced commodity markets dotted beneath the plots. However, the medium-priced category is seen to be densely spread across the plot.

Table 16: Categorization of the prize levels for all five years together

Limit of index	Representations	Markets
$Z \geq 3$	Extremely High	23
$2 \leq Z < 3$	High	46
$-2 < Z < 2$	Moderate	336
$-3 < Z \leq -2$	Low	38
$Z \leq -3$	Extremely Low	12

Source: Field Data, SRID-MoFA 2008

With one High priced market sandwiched between the heavily dense medium priced and the few extremely high priced markets on top. The extremely high price markets can be seen on top of the plot.

Extremely High Priced Market for all five years together

We identify that about 5% of markets across the country over the five years together. The extremely high priced markets are few but market 65 which is Fumbisie appears to stand out. Fumbisie is a market in the upper East region of Ghana. And this market is the most exorbitant in the country when we combine all the years.

High Priced Market for all five years together

Sandwich between the extremely high priced and the largely shown moderate market, the high price markets can be seen in Figure 6 thinly. There are a few high priced markets. About 10% of the total markets sampled across the country are high priced for the five years combined.

Medium priced market for five years combined

About 74% of the markets are medium priced for the five years together in the whole country. This category accounts for the majority of the markets across the country. We can say that during the entire five non-consecutive years considered in this study, the country is generally a medium priced country. By that we mean that prices of commodity food items are basically not expensive and not cheap too.

Low priced Market for five years combined

We identified that only 8% of the markets in the country are low priced. There are less low priced market than High priced markets in Ghana. Indeed majority of the low priced markets are in the food growing communities, where it is expected that prices of some commodity food items should be low priced.

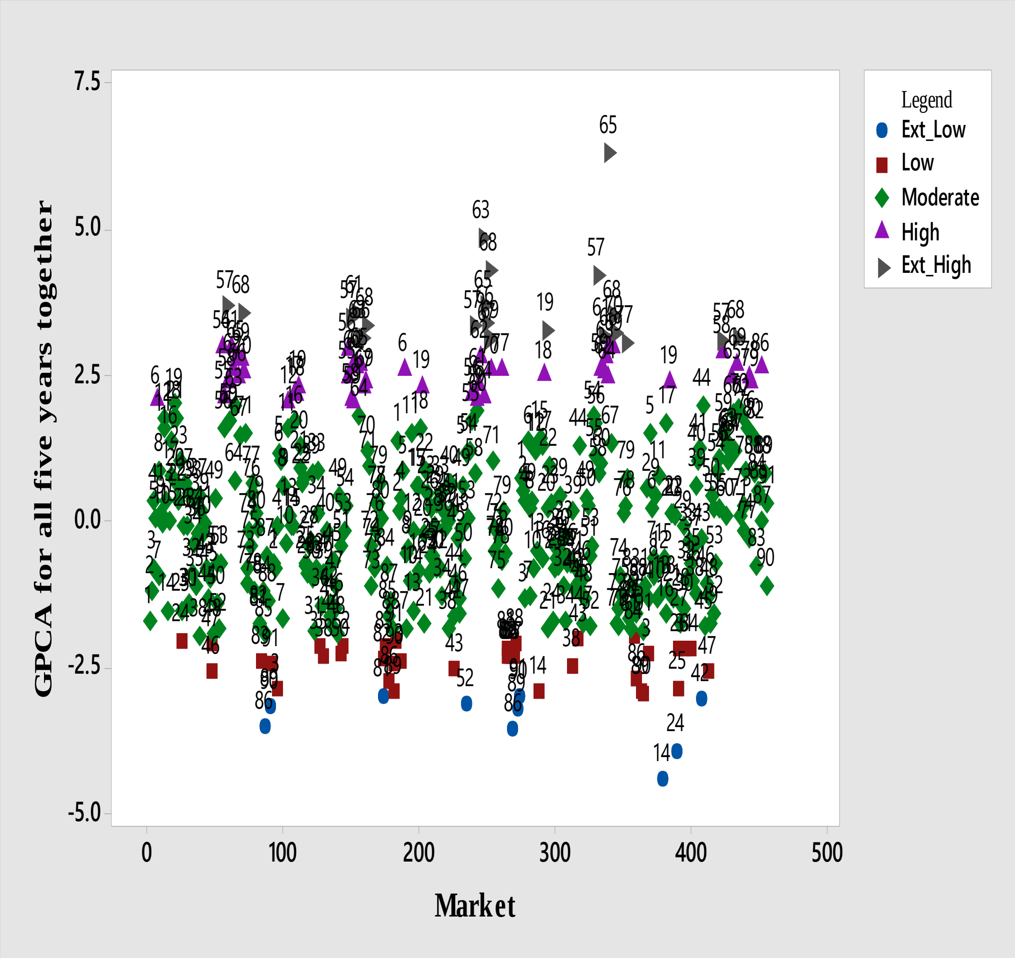

Figure 6: A scatter plot of the GPCA for all five years together

Extremely Low priced market for all five years together

When we combine the covariance matrices and compute the GPCA for the five years combined. We identify the following markets as extremely low priced: Bogoso, Sekondi, Takoradi, Agona Nkwanta, Tema, Kukuom, Dunkwa, and Koforidua. This shows that very few markets, about 3% are extremely low priced. However Market 14 which is Kukuom in the Brong Ahafo Region is the lowest of the extremely low priced markets.

Details of the classification of the cross tabulation for the categories of price levels can be seen in the appendix B.

Table 17: Summary of the market category cross tabulation

Classification	Nos of Markets	Percentage
Extremely Low	12	2.64
Low	38	8.35
Moderate	336	73.85
High	46	10.10
Extremely High	23	5.06
Total	455	100

Source: Field Data, SRID-MoFA 2008

Table 17 shows the summary of the markets category cross tabulation for the pooled data. It reveals that 74% of the total markets across the country are moderately priced. We observed from Table 17, that there appears to be more extremely high priced markets than extremely low priced markets across the country. 10% of the markets across the country are high priced.

Factor Loadings

In order to label the groupings of food items, we perform a rotation of the components. We can see from the Table 18 that the matrix of rotated components for PC1 loaded high on the following food items: Cassava (Roots and Tubers), Garden Eggs, Orange and Banana which are collectively called Fruits. Smoked Herring, Koobi and Egg which are also collectively called Fishes.

Table 18: Rotated Factor Loadings for the year 2008

No	Variable	PC1	PC2	PC3	PC4	PC5
1	Maize	-0.072	0.776	0.034	-0.047	-0.043
2	White Yam	-0.001	0.573	-0.185	-0.259	0.347
3	Cassava	0.863	0.067	0.069	0.017	0.073
4	Tomato	0.049	0.682	0.459	-0.069	0.104
5	Garden Egg	0.721	0.100	-0.065	-0.130	0.046
6	Dry Pepper	0.058	0.150	0.718	0.060	0.111
7	Red Groundnut	-0.305	0.806	0.018	-0.190	-0.010
8	White Cowpea	-0.313	0.828	0.150	0.132	0.005
9	Palm Oil	0.404	0.023	0.178	0.706	0.205
10	Orange	0.765	-0.281	-0.219	-0.038	-0.108
11	Banana	0.880	-0.234	-0.069	0.025	-0.070
12	Smoked Herring	0.729	-0.409	-0.235	0.177	-0.099
13	Koobi	0.534	-0.407	0.089	0.260	-0.046
14	Onion	-0.335	0.003	0.766	-0.138	0.048
15	Egg	0.644	-0.013	0.329	0.330	-0.092
16	Plantain	0.890	-0.032	-0.092	0.115	-0.032
17	Gari	0.123	0.176	0.205	-0.828	0.074
18	Local Rice	0.248	0.492	0.096	0.066	-0.298
19	Imported Rice	-0.046	-0.013	0.210	0.073	0.890

Source: Field Data, SRID-MoFA 2008

So then we conclude that Roots and Tubers, Garden Eggs, Fruits and Fishes loads highly on the first PC in the Year 2008. Then the following food items also loads highly on PC2: Maize which is classified as cereals, Tomatoes, White Yam, Red Groundnut and White Groundnut, both of which we classify as pulses. This commodity food items load high on PC2 in the Year 2008.

The following food items load high on the third PC: Red Pepper and Onion which are both classified as Spices. Gari and Oil both loads high on PC4. They are contrasting as a high price in one will see a reduction in price of the other in specific markets. Imported rice appears to be the only commodity food item that loaded high on the PC 5.

The rotated factor loadings for each year have been computed separately and shown in the appendices 1A, 1B, 1C, 1D and 1E, starting from the year 2009. The summary of the results of the labels of the food commodity items of the years combined are displayed in Table 20. A contrast is indicated by the '/' when the Principal Component factor of a labelled food item is negative. That is the component score is negative. From table 20, we consider PC1 for year 3, we observe the following items Maize, White Yam, Vegetable but the rotated PC scores is negative for fruit and fishes. We interpret this to mean that there is a contrast between the following items, white yam, and maize, vegetable on one hand and fruits and fishes on the other hand. This simply means that if these commodities in this group are highly priced in a market, then Fruits and Fishes will be low priced in the same market. That is to say that, high priced markets identified by the first PC in the third year are those in the set of food items; Maize, Yam and Vegetable but low priced on the Fruits and Fishes.

Table 19 shows the first five principal components for the five years. Roots and Tubers appears to be dominant on the PCs for most of the years. Fruits and Fishes also appears to dominate for the first four PCs for the five years. For both PC1 and PC2, Roots and Tubers, Fishes and Fruits featured across the years. This can be explained away as Root and tubers are the staple food of most Ghanaian communities, and so they appear to stand out as the principal food item dominating in all five years. From the Table 19, we see that the last principal components have fewer labels, mostly one food items are identified across all the year groups.

We identify PC1 to have more than two food groups for each of the five years. PC1 for year one for example has Roots and Tubers, Garden Eggs, Fruits and Fishes as the food groups that are high on PC1. Similarly for year two, Roots and Tubers, Fruits, Fishes, Oil are high on PC1 for that year.

Table 19: Labels for the food items

Years	PC1	PC2	PC3	PC4	PC5
Year 1	Roots and Tubers, Garden Eggs, Fruits and Fishes,	Cereals, Yam, Tomato, Pulses	Spices	Gari, Oil	Imported Rice
Year 2	Roots and Tubers, Fruits, Fishes, Oil	Maize, Tomato Yam, Gari, Spices, Pulse	Pulse, Local Rice	Dry Pepper, Imported Rice	Garden Eggs
Year 3	Maize, Yam, Vegetable, /Fruits, Fishes	Spices, Pulses	Roots and Tubers, Vegetables, Egg	Pulse, Oil, Koobi, Local Rice	Cereal, Gari
Year 4	Maize, Yam, Spices, Vegetables, Pulses	Oil, Egg	Root and tubers, Garden Eggs	Onion, Gari, Local Rice	Fishes
Year 5	Roots and Tubers, Fruits, Fishes	Tomato, Onion	White Cowpea, Fishes, Cereals	Dry Pepper, Gari, Pulses	Oil, Egg

Source: Field Data, SRID-MoFA 2008

Chapter Summary

Chapter Four discusses the results from the implementation of the codes. We started the chapter with revision of the objectives of the study as outlined in chapter One. Secondary data was obtained from the SRID for five non-consecutive years. The statistical packages used for the computation were also briefly outlined in this chapter. The codes written for execution of the technique are explained line by line. It was obvious from the results of the computation of the GPCA for the five years pooled together that majority of the markets are moderately priced as about 74 % of all the markets used for the study reveal that they belong to the moderately priced category. A number of scatter was plotted which shows the distribution of the markets in their categories. In all we see that the markets are mostly seen as moderately priced. With a few extremely high and very few extremely low priced markets. We also observe that there are more extremely priced markets than extremely low priced markets. About 5% of markets are extremely high priced whiles 2.1% of the markets are extremely low priced. The results of the factor loadings for the five years pooled together reveal that roots and tubers is the most consumed food item for the period under considerations.

CHAPTER FIVE

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

Overview

In this chapter, a summary of the entire study is presented. It covers the main techniques used for the study, and a justification for their choice. It also includes the data, for the study, a summary of the literature is also presented. It will be apparent from the literature that even though a number of multivariate techniques have already been employed for this kind of study, there is still the need to explore other techniques since results appear to be influenced by the technique adopted. A summary of the technique used in this study, which is the Generalized Principal Component Analysis, is therefore provided. The procedure for the use of this technique for market categorization is presented. In the end, a conclusion is given and appropriate recommendations are also provided.

Summary

The study seeks to examine the level of prices of local food items over five non-consecutive years within the period 2008 to 2015, in order to determine the category of the market based on the price levels. Secondary data from the SRID of MoFA is obtained for the purpose. The principal component was reviewed thoroughly and the generalized principal component was the technique used to arrive at the results.

The study begins with a discussion of the definition of the prices of commodities, of a market as an equilibrium price as a result of daily interaction of market forces. It was seen that prices of commodities are not static but subject to change as a result of either internal market forces or external influence. Natural occurrences were also identified as a cause of changes in the prices of commodities. It was observed that prices of food items have seen increases over the years, in fact for the last part of the 20th century. It was evidence from literature, that the growth and economic development of a country can be greatly affected by the price volatility.

It was seen also that Ghana is heavily reliant on home-grown foods items for the provision of its nutritional needs. And so any economic phenomenon that affects the production of such food items or causes an increase in the prices of such food items, will adversely affect the nutritional health of the members of the households who cannot meet their needs. An index to help classify the extremities in the prices of the food items in the market were adopted. An index similar to those used to calibrate extreme poverty among household or used to study extreme weather phenomenon. The problem statement was made, which was an attempt to identify the extreme price markets of selected markets across the country using the technique of generalized principal component analysis. Other studies from literature use the outlier displaying component and the classical principal component analysis to identify such extreme prices among commodities in the local markets (Eyiah-Bediako, 2019). The main objective of the study was to use the technique to examine the general price levels of the local market in the country by identifying the marked prices levels in the local markets and also to identify such markets that are regarded as extremely high priced and those that are extremely low priced.

A number of multivariate techniques in literature were reviewed; the history of principal component analysis was also reviewed. The literature on some principal components studies was reviewed. The definitions of Principal component analysis was appraised and the objectives underlining the use of the PCA were reviewed. The study assesses the derivative of the Principal Components. The literature reveals that a number of procedures are used to determine the retention of the principal components in any study. The chapter reviewed dimensionality reduction as a multivariate approach to data manipulation. Other important data modelling techniques, like machine learning, was reviewed. Matlab, Minitab, SPSS, R-studio were identified as some common statistical packages used to aid in the analysis of the data. The sampling and inferences concerns with the use of Principal components were highlighted in the study. The generalized principal component was reviewed as the main technique to be used to achieve the objectives of the study. The contrast between the classical PCA and the GPCA was made in the chapter under review. The chapter ended with a summary of the whole chapter.

The study proceeded to use a robust estimate of the variance-covariance matrix; the eigenvectors were computed as the matrix cross product of the defined variance-covariance matrix. The GPCA was found by projecting the standardized data onto the eigenvector. The study developed an algorithm in Matlab for the computation of the GPCA for each year for all five years. The data used for the study was a secondary data.

A scatter plot was used to display the groupings of the markets. It was also observed that the most predominant category of the market was the medium-priced markets. Medium priced markets accounted for more than 70% of the market classifications across the country for each year. The results also show that there were extremely high priced markets located at the northern regions. Notable were the markets at Bugubelle and Gushiegu. The percentage of markets that were extremely high in the year 2008 was about 12% of all markets across the country, whereas there were no extremely low markets in the same year. The results indicated that food prices are moderately priced in that year. In the year 2009 however, only two markets were identified as extremely high priced with no markets identified as extremely low priced. However, 88% of the markets across the country were seen in the medium-priced category. In the year 2012, again, there were no extremely-low priced markets. However, only three markets were identified as extremely high priced. Again 81% of the markets are identified as moderately priced. This implies that the outlook of the markets in the year 2012 is moderately priced across the country.

In the year 2013, only two markets were seen as extremely high priced whereas the two markets were identified as extremely low priced. 77% of the rest of the markets were identified in the medium-priced category. Finally, it was observed that in the year 2015, only three markets belong to the extremely high priced category with no market identified as extremely low priced. Again in this year 2015, about 93% of the markets are medium-priced. And so it can be said that the outlook of the markets generally in the year 2015 is moderate. Various scatter diagrams were drawn to show the groups of markets to be seen in each year.

Conclusion

Some importance of the study was identified, notable among them is that it will improve the personal livelihood of members of the household. Generally, we conclude that markets across the country are moderately priced. As we can see from the individual year analysis and the pooled five years' analysis. We have also been able to obtain unique classifications in the general price levels of the local markets in Ghana. Therefore, in conclusion, we have been able to identify extreme price levels using an appropriate index. We also obtain specific markets that we regarded and highly-priced and those that we recognise as extremely low price.

Recommendation

In the course of my study, I found out that according to Rencher (2002), the plot could also reveal groupings within the data. We recommend that a researcher takes it up and find out the theoretical feasibility and practicality of Rencher's claim. We recommend that efforts should be concentrated on the production of cereal, roots, and tubers food crops, further analysis should be carried out to find what variables might be responsible for the categorization that appear to underline the way the markets are grouped together. Finally, we would recommend for further studies, a modelling of the changes in prices of food commodity items.

REFERENCES

Acemoglu, D., Johnson, S., Robinson, J., & Thaicharoen, Y. (2003). Institutional causes, macroeconomic symptoms: Volatility, crises and growth. *Journal of Monetary Economics*.

Acharjya, P. D., & Ahmed, K. (2016). A Survey on Big Data Analytics: Challenges, Open Research Issues and Tools. *International Journal of Advanced Computer Science and Applications*, *7*(2).

Adjasi, C. K. D., & Osei, K. A. (2007). Poverty profile and correlates of poverty in Ghana. *International Journal of Social Economics*, *34*(7), 449–471.

Aghion, P., Bacchetta, P., Rancière, R., & Rogoff, K. (2009). Exchange rate volatility and productivity growth: The role of financial development. *Journal of Monetary Economics*.

Aikins, E. K. W. (2012). Evidence of Climate Change (Global Warming) and Temperature Increases in Arctic Areas. *International Scholarly and Scientific Research & Innovation Journal*, *6*(12), 1453–1458.

Alkire, S., Foster, J. E., Seth, S., Santos, M. E., Roche, J., & Ballon, P. (2015). Multidimensional poverty measurement and analysis: Chapter 5--the Alkire-Foster counting methodology.

Alkire, S., Roche, J. M., Santos, M. E., & Seth, S. (2011). Multidimensional poverty index 2011: brief methodological note.

Alkire, S., & Santos, M. E. (2010). *Multidimensional Poverty Index 2010: research briefing*. Oxford Poverty & Human Development Initiative (OPHI).

Alkire, S., & Santos, M. E. (2014). Measuring acute poverty in the developing world: Robustness and scope of the multidimensional poverty index.

World Development, *59*, 251–274.

Anderson, T. W. (1963). Asymptotic Theory for Principal Component Analysis. *The Annals of Mathematical Statistics*.

Anderson, T. W. (2003). *An introduction to Multivariate Statistical Analysis* (Third). New Jersey: John Wiley & Sons, Inc.

Asokan, S. M., Obando, J., Kwena, B. F., & Luwesi, C. N. (2020). Climate Change Adaptation Through Sustainable Water Resources Management in Kenya: Challenges and Opportunities. *African Handbook of Climate Change Adaptation*, 1–11.

Bailey, W., & Chan, K. C. (1993). Macroeconomic Influences and the Variability of the Commodity Futures Basis. *Journal of Finance*, (48), 555– 574.

Barnett, T. P., & Preisendorfer, R. (1987). Origins and levels of monthly and seasonal forecast skill for United States surface air temperatures determined by canonical correlation analysis. *Monthly Weather Review*, *115*(9), 1825.

Behrman, J. R., & Rosenzweig, M. R. (2004). Returns to birthweight. *Review of Economics and Statistics*.

Beltrami, E. (1873). Sulle funzioni bilineari. *Giornale Di Mathematiche Di Battaglini*, (11), 98–106.

Ben-david, S. (2014). *Understanding Machine Learning : From Theory to Algorithms*.

Bessembinder, H., & Chan, K. (1992). Time-varying risk premia and forecastable returns in futures markets. *Journal of Financial Economics*.

Bjornson, B., & Carter, C. A. (1997). New Evidence on Agricultural Commodity Return Performance under Time-Varying Risk. *American Journal of Agricultural Economics*.

Boyd, N. E., Harris, J. H., & Li, B. (2018). An update on speculation and financialization in commodity markets. *Journal of Commodity Markets, Financialization of Commodities*, *10*, 91–104.

Brinkman, H.-J., De Pee, S., Sanogo, I., Subran, L., & Bloem, M. W. (2010). High food prices and the global financial crisis have reduced access to nutritious food and worsened nutritional status and health. *The Journal of Nutrition*, *140*(1), 153S--161S.

Broome, J. (1992). Counting the cost of global warming.

Brunner, A. D. (1998). *El Nino and World Primary Commodity Prices: Warm Water or Hot Air* (No. 608).

Bryant, E. H., & Atchley, W. R. (1975). *Multivariate Statistical Methods: Within Group Covariation*. Stroudsberg: Stroudsberg: Halsted Press.

Carter, C. A., Rausser, G. C., & Smith, A. (2011). Commodity booms and busts. *Annual Review of Resource Economics*, *3*(1), 87–118.

Carter, C., Rausser, G., & Smith, A. Y. (2012). *The Effect of the US Ethanol Mandate on Corn Prices*. UC Davis.

Cartwright, P. A., Kamerschen, D. R., & Huang, M. Y. (1989). Price correlation and granger causality tests for market definition. Review of Industrial Organization, 4(2), 79–98.

Caussinus, H., Fekri, M., Hakam, S., & Ruiz-Gazen, A. (2003). A monitoring display of multivariate outliers. *Computational Statistics & Data Analysis*, *44*(1–2), 237–252.

Caussinus, Henri, & Ruiz-Gazen, A. (2007). Classification and Generalized Principal Component Analysis (pp. 539–548).

Caussinus, Henri, & Ruiz-Gazen, A. (2010). Exploratory Projection Pursuit. In *Data Analysis*.

Caussinus, Henri, & Ruiz, A. (1990). Interesting projections of multidimensional data by means of generalized principal component analyses. In *Compstat* (pp. 121–126). Springer International Publishing.

Chae, S. S., & Warde, W. D. (2006). Effect of using principal coordinates and principal components on retrieval of clusters. *Computational Statistics and Data Analysis*, (50), 1407–1417.

Chapagain, D., Baarsch, F., Schaeffer, M., & D'haen, S. (2020). Climate change adaptation costs in developing countries: insights from existing estimates. *Climate and Development*, 1–9.

Chen, S.-T., Kuo, H.-I., & Chen, C.-C. (2010). Modeling the relationship between the oil price and global food prices. *Applied Energy*, *87*(8), 2517–2525.

Chen, Z. M., Wang, L., Zhang, X. B., & Zheng, X. (2019). The co-movement and asymmetry between energy and grain prices: Evidence from the crude oil and corn markets. *Energies*, *12*(7), 1–17.

Cline, W. R. (1992). *The economics of global warming.* Institute for International Economics.

Cooke, E., Hague, S., & McKay, A. (2016). The Ghana poverty and inequality report: Using the 6th Ghana living standards survey. *University of Sussex.*

Coxhead, I., & Jayasuriya, S. (2010). China, India and thecommodity boom: economic and environmental implications for low-income countries. *The World Economy*, *33*(4), 525–551.

Davis, J. (1986). *Statistics and Data Analysis in Geology*. New York: John Wiley & Sons, Inc.

de Freitas, C. R., Scott, D., & McBoyle, G. (2008). A second generation climate index for tourism (CIT): specification and verification. *International Journal of Biometeorology*, *52*(5), 399–407.

Deaton, A., & Miller, R. (1996). International commodity prices, macroeconomic performance and politics in sub-Saharan Africa. *Journal of African Economies*.

Deaton, Angus. (1999). Commodity prices and growth in Africa. *Journal of Economic Perspectives*.

Deaton, Angus, & Laroque, G. (2003). A model of commodity prices after Sir Arthur Lewis. *Journal of Development Economics*.

Dewbre, J., Giner, C., Thompson, W. and Von Lampe, M. (2008), High food commodity prices: will they stay? who will pay?. Agricultural Economics, 39: 393-403.

Eckstein, D., Künzel, V., Schäfer, L., & Winges, M. (2019). Global climate risk index 2020. Bonn: Germanwatch

Esmaeili, A., & Shokoohi, Z. (2011). Assessing the effect of oil price on world food prices: Application of principal component analysis. *Energy Policy*, *39*(2), 1022–1025.

Etienne, X. L., Irwin, S. H., & Garcia, P. (2015). Price explosiveness, speculation, and grain futures prices. *American Journal of Agricultural*

Economics, *97*(1), 65–87.

Etienne, X. L., Irwin, S. H., & Garcia, P. (2017). New evidence that index traders did not drive bubbles in grain futures markets. *Journal of Agricultural and Resource Economics*, *42*, 23.

Eyiah-Bediako, F. (2019). *Techniques for determining Classifications in multiple multivariate data: Application to prices of local food items on markets in Ghana*. University of Cape Coast.

Fankhauser, S. (1994). The economic costs of global warming damage: a survey. *Global Environmental Change*, *4*(4), 301–309.

FAO. (2008). *Food Outlook Report*.

Fisher, R. A., & Mackenzie, W. A. (1923). Studies in crop variation: II. The manurial response of different potato varieties. *The Journal of Agricultural Science*.

Fodor, I. K. (2002). *A Survey of Dimension Reduction Techniques*. Livermore, CA.

Fogel, R. W. (1994). Economic growth, population theory, and physiology: the bearing of long-term processes on the making of economic policy. *American Economic Review*.

Gabriel, K. R. (1971). The biplot graphic display of matrices with application to principal component analysis. *Biometrika*.

Gabriele, A., & Schettino, F. (2008). Child malnutrition and mortality in developing countries: Evidence from a cross-country analysis. *Analyses of Social Issues and Public Policy*.

Geweke,John (1984).Inference and causality in economic time series models,Handbook of Econometrics,Elsevier,Volume 2,Pages 1101-

1144.

Gillie, A. (1996). The origin of the poverty line. *Economic History Review*, 715–730.

Girshick, M. A. (1939). On the Sampling Theory of Roots of Determinantal Equations. *The Annals of Mathematical Statistics*.

Goedhart, T., Halberstadt, V., Kapteyn, A., & Van Praag, B. (1977). The poverty line: concept and measurement. *Journal of Human Resources*, 503–520.

Govaert, G. (2003). *Data Analysis*. (G. Govaert, Ed.). New Jersey: John Wiley &Sons, Inc.

GSS. (2018). Ghana Living Standards Survey (GLSS7): Poverty trends in Ghana; 2005-2017. *Ghana Statistical Service*, 14–21.

Gyimah-Brempong, K., & Wilson, M. (2004). Health human capital and economic growth in Sub-Saharan African and OECD countries. *Quarterly Review of Economics and Finance*.

Hagenaars, A. J. M., & Van Praag, B. M. S. (1985). A synthesis of poverty line definitions. *Review of Income and Wealth*, *31*(2), 139–154.

Hamilton, J. D. (2009). *Causes and Consequences of the Oil Shock of 2007–2008* (No. 15002). Cambridge, MA.

Hammer, C., Kostroch, M. D. C., & Quiros, M. G. (2017). *Big data: Potential, challenges and statistical implications*. New York.

Handbury, J., & Weinstein, D. E. (2015). Goods Prices and Availability in Cities. *The Review of Economic Studies*, *82*(1), 258–296.

Hansen, J., Sato, M., Glascoe, J., & Ruedy, R. (1998). A common-sense climate index: Is climate changing noticeably? *Proceedings of the National*

Academy of Sciences, *95*(8), 4113–4120.

Harman, H. H. (1960). *Modern factor analysis*. Chicago: University of Chicago Press.

Headey, D. and Fan, S. (2008), Anatomy of a crisis: the causes and consequences of surging food prices. Agricultural Economics, 39: 375-391.

Hernandez, M. A., Ibarra, R., & Trupkin, D. R. (2014). How far do shocks move across borders? examining volatility transmission in major agricultural futures markets. *European Review of Agricultural Economics*, *41*(2), 301–325.

Hnatkovska, V., & Loayza, N. (2005). Volatility and Growth. In *Managing Economic Volatility and Crises: A Practitioner's Guide*.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, *24*(6), 417–441.

Hotelling, Harold. (1936). Relations Between Two Sets of Variates. *Biometrika*, *28*(3/4), 321.

Hoy, C. (2015). The definition of extreme poverty has just changed – here's what you need to know.

Huppmann, D. (2020). *Climate change mitigation & sustainable development-Qualitative and quantitative analysis in the IPCC's "Special Report on Global Warming of 1.5° C."*

IBM Corporation. (2020). IBM SPSS Statistics for Windows, Version 24.0. New York: IBM.

ICMLC '19: Proceedings of the 2019 11th International Conference on Machine Learning and Computing. (2019). New York, NY, USA: Association for

Computing Machinery.

IFAD. (2011). *Price Volatility in Food and Agricultural Markets: Policy Responses*. Retrieved from www.ruralpovertyportal.org.

Ivanic, M., & Martin, W. (2008). *Implications of higher global food prices for poverty in low-income countries*. The World Bank.

Jaffee, D. M., & Russell, T. (1997). Catastrophe insurance, capital markets, and uninsurable risks. *Journal of Risk Insurance*, *64*, 205–230.

Jareskog, K., Klovan, J., & Reyment, R. (1976). Geological Factor Analysis. *Elsevier, Amsterdam*.

Jeffers, J. N. R. (1967). Two Case Studies in the Application of Principal Component Analysis. *Applied Statistics*.

Jin, X., Wah, B. W., Cheng, X., & Wang, Y. (2015). Significance and Challenges of Big Data Research. *Big Data Research*, *2*(2), 59–64.

Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis* (Six). Pearson Prentice Hall.

Jollife, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2065).

Jolliffe, I. . (1986). Principal Component Analysis and Factor Analysis. In *Principal component analysis* (pp. 115–128). New York: Springer International Publishing.

Jolliffe, I. T. (2002a). *Principal Component Analysis, Second Edition, New York: Springer Science & Business Media*. *New York: Springer Science & Business Media*.

Jolliffe, I. T. (2002b). *Principal components analysis* (Second). New York: Springer International Publishing.

Jordan, M. C. (1874). M´emoire sur les Formes Bilin´eaires. *J. Math. Pures Appl.*, (19), 35–54.

Kaldor, N. (1939). Speculation and economic stability. *Review of Economic Studies*.

Kalkuhl, M., Von Braun, J., & Torero, M. (2016). Volatile and Extreme Food Prices, Food Security, and Policy: An Overview. In *Food Price Volatility and Its Implications for Food Security and Policy* (pp. 3–31). Springer International Publishing.

Kearney, M. J., & Kearney, J. F. (1983). Investigation of sex differences and consistency of factor pattern on the adult Nowicki-Strickland internal-external control scale. *Journal of Clinical Psychology*.

Kilian, L., & Murphy, D. P. (2014). The role of inventories and speculative trading in the global market for crude oil. *Journal of Applied Econometrics*, *29*(3), 454–478.

Kleinbaum, D. G., & Kupper, L. L. (1978). *Applied Regression Analysis And Other Multivariate Methods*. California: Duxbury Press.

Kochhar, R. (2010). projections for the global population in 2050. *Pew Research Center*.

Kreft, S., Eckstein, D., & Melchior, I. (2013). Global climate risk index 2014. *Who Suffers Most from Extreme Weather Events*, *1*.

Krugman, P. (2008). More on oil and speculation.

Kshetri, N. (2014). The emerging role of Big Data in key development issues: Opportunities, challenges, and concerns. *Big Data & Society*, *1*(2), 205395171456422.

Kyereme, S. S., & Thorbecke, E. (1991). Factors affecting food poverty in Ghana. *The Journal of Development Studies*, *28*(1), 39–52.

Li, J., Li, C., & Chavas, J. (2017). Food price bubbles and government intervention: is China different? *Canadian Journal of Agricultural Economics*, *65*(1), 135–157.

Loayza, N. V., Rancière, R., Servén, L., & Ventura, J. (2007). Macroeconomic volatility and welfare in developing countries: An introduction. *World Bank Economic Review*.

Mahalanobis, P. C. (1936). On the generalized distance in statistics (pp. 50–54).

Malinowski, F., & Howery, D. (1980). *Factor Analysis in Chemistry*. New York: John Wiley & Sons, Inc.

Mao, Q., Yanjun, R., & Loy, J. (2020). Price bubbles in agricultural commodity markets and contributing factors : evidence for corn and soybeans in China. *China Agricultural Economic Review*.

Mardia, K. V, & Kent, T. (1979). *MULTIVARIATE ANALYSIS*. New York: Academic Press.

Marshall, D. J., Pettersen, A. K., Bode, M., & White, C. R. (2020). Developmental cost theory predicts thermal environment and vulnerability to global warming. *Nature Ecology & Evolution*, *4*(3), 406–411.

MATLAB. (2020). MATLAB version 9.4.0.8.13684 (R2018a). Natick, Massachusetts.

McNown, R., & Wallace, M. S. (1989). National price levels, purchasing power parity, and cointegration: a test of four high inflation economies. *Journal* *of International Money and Finance*, *8*(4), 533–545.

Meerman, J., & Aphane, J. (2012). Impact of high food prices on nutrition. Retrieved from http://www.fao.org

Minitab. (2020). Minitab version 18. Pennsylvania State University. Retrieved from http://www.minitab.com/

MoFA. (2020). Ministry of Food and Agriculture. Retrieved September 5, 2020, from www.mofa.gov.gh.

Nkansah, B. K., & Gordor, B. K. (2012). On the One-Outlier Displaying Component. *Journal of Informatics and Mathematical Sciences*, *4*(2), 229–239.

Obadi, S. M., & Korcek, M. (2014). Are food prices affected by crude oil price: Causality investigation. *Review of Integrative Business and Economics Research*, *3*(1), 411–427.

Ofori-Sarpong, E. (1986). The 1981–1983 Drought in Ghana. *Singapore Journal of Tropical Geography*.

Orbach, B. Y. (2008). Antitrust vertical myopia: The allure of high prices. *ARIz. L. REv.*, *50*, 261.

Park, H. Y., Wei, K. C. J., & Frecka, T. J. (1988). A Further Investigation of the Risk Return Relationship for Commodity Futures. *Advances in Futures and Options Research*, *3*, 357–77.

Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *2*(11), 559–572.

Pelletier, D. L., & Frongillo, E. A. (2003). Changes in child survival are strongly associated with changes in malnutrition in developing countries. *Journal* *of Nutrition*.

Persson, K. G. (1999). *Grain Markets in Europe, 1500–1900*. *Grain Markets in Europe, 1500–1900*.

Pindyck, R. S., & Rotemberg, J. J. (1990). The Excess Co-Movement of Commodity Prices. *The Economic Journal*. https://doi. org/10.2307/2233966

Poelhekke, S., & Ploeg, V. der. (2007). *Volatility, financial development and the National Resource Curse* (No. 6513).

Preisendorfer, R. W., & Mobley, C. D. (1988). Principal Component Analysis in Meteorology and Oceanography. *Elsevier*.

Price,M. J. (1979). The characterization of instantaneous causality. A correction. Journal of Econometrics.

Prices, S. F. (2008, June). Facts, Perspectives, Impacts and Actions Required. In Proceedings of High-level Conference on World Food Security: The Challenges of Climate Change and Bioenergy.

Rao,R. C. (1964). The use and interpretation of principal component analysis in applied research. *Sankhya - Series A*.

Ramey, G., & Ramey, V. a. (1995). Cross-Country Evidence on the Link Between Volatility and Growth. *American Economic Review*.

Rencher, A. C. (2002). Methods of Multivariate Analysis, 727.

Roll, R. (1984). Orange juice and weather. *American Economic Review*.

RStudio Team. (2020). RStudio: Integrated Development Environment for R. Boston, MA. Retrieved from http://www.rstudio.com/

Ruiz-Gazen, A. (1996). A very simple robust estimator of a dispersion matrix. *Computational Statistics and Data Analysis*, (21), 149–162.

Ruiz-Gazen, A., Marie-Sainte, S. L., & Berro, A. (2010). Detecting multivariate outliers using projection pursuit with particle swarm optimization. In *Proceedings of COMPSTAT'2010* (pp. 89–98). Springer.

Sarkar, S. (2013). Potential of probiotics as pharmaceutical agent: A review. *British Food Journal*.

Sarris, A. (2014). *Options for developing countries to deal with global food commodity market volatility* (No. 84).

Schelter, B., & Winterhalder, M. (2006). *Handbook of Time Series Analysis*. *Handbook of Time Series Analysis*. https://doi.org/10.1002/9783527609970

Seglah, F. (2014). Multivariate Analysis of food prices in Ghana, Mphil Thesis. *University of cape coast*.

Shalev-Shwartz, S., & Ben-David, S. (2013). *Understanding machine learning: From theory to algorithms*. *Understanding Machine Learning: From Theory to Algorithms*.

Sharma, A. K., Gosh, N., & Kumar, P. (2000). *A Policy Model for Open Market Operations of Wheat*. New Delhi.

Sharma, S. (1996). *Applied multivariate tecnhiques*. *John Willey & Sons: Inc.* John Wiley & Sons, Inc.

Sharma, S., Kleinbaum, D. G., & Kupper, L. L. (1978). Applied Regression Analysis and Other Multivariate Methods. *Journal of Marketing Research*.

Siluvairajah, M. (2011). Supply and demand vs. price driven graphs financial data visualization. *International Journal on Computer Science & …*, *3*(8), 2935–2941.

Smith, L. C., & Haddad, L. (2000). Overcoming Child Malnutrition in Developing Countries - Past Achievements and Future Choices. 2020 Vision for Food, Agriculture, and the Environment. Discussion Paper 30 (IFPRI, 2000, 73 p.): *International Food Policy Research Institute*.

Solaymani, S. (2017) Agriculture and Poverty Responses to High Agricultural Commodity Prices. Agric Res 6, 195–206.

Stadlmayr, B., Charrondiere, U. R., Enujiugha, V. N., Bayili, R. G., Fagbohoun, E. G., Samb, B., … Burlingame, B. (2012). *West African Food Composition Table*. *Organization*.

Tabachnick, B. G., & Fidell, L. S. (2012). *Using multivariate statistics (6th ed.)*. *New York: Harper and Row*.

Thompson, B. (2010). Impact of the Financial and Economic Crisis on Nutrition – Policy and Programme Responses. *Nutrition and Consumer Protection Division*.

Thurstone, L. L. (1931). Multiple factor analysis. *Psychological Review*.

Trostle, R. (2010). *Global agricultural supply and demand: factors contributing to the recent increase in food commodity prices (rev*. Diane Publishing.

Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *J. of Cognitive Neuroscience*, *3*(1), 71–86.

Tynan, A. C., & Drayton, J. (1987). Market segmentation. *Journal of Marketing Management*, *2*(3), 301–335.

UNCTAD. (2008). *19 Export Performance Following Trade Liberalisation (Report 2008) - Some patterns and policy perspectives*. *United Nations Conference on Trade and Development*.

Urala, N. (2005). Functional foods in Finland. Consumer's views, attitudes and willingness to use. *VTT Publications*.

Uri, N. D., Howell, J., & Rifkin, E. J. (1985). On defining geographic markets. Applied Economics.

Van der Ploeg, F., & Poelhekke, S. (2009). Volatility and the natural resource curse. In *Oxford Economic Papers*.

Vidal, Rene, Ma, Y., & Sastry, S. S. (2006). *Generalized Principal Component Analysis Estimation & Segmentation of Hybrid Models*.

Vidal, René, Ma, Y., & Sastry, S. S. (2016). *Generalized Principal Component Analysis*. *Springer*. New York: Springer International Publishing.

Weil, D. (2007). Accounting for the effect of health on economic growth. *Q J Econ*, *122*(3), 1265– 1306.

Will, M., Prehn, S., Pies, I., & Glauben, T. (2013). *Is Financial Speculation with Agricultural Commodities Harmful or Helpful? – A Literature Review of Current Empirical Research* (No. 2333076). Rochester, NY.

Working, H. (1948). Theory of the Inverse Carrying Charge in Futures Markets. *Journal of Farm Economics*.

World Bank. (2018). Global Economic Prospects 2018: The Role of Major Emerging Markets in Global Commodity Demand.

Wu, Q. (2019). A Review of Methods Used in Machine Learning and Data Analysis. In *Proceedings of the 2019 11th International Conference on Machine Learning and Computing - ICMLC '19* (pp. 43–51). New York, New York, USA: ACM Press.

Yenyukov, I. (1988). Detecting structures by means of Projection Pursuit. In *COMPSTAT* (pp. 47–58). Heidelberg: Physica-Verlag HD.

Zafeiriou, E., Arabatzis, G., Karanikola, P., Tampakis, S., & Tsiantikoudis, S.

(2018). Agricultural commodities and crude oil prices: An empirical investigation of their relationship. *Sustainability*, *10*(4), 1199.

Zhang, Q., & Reed, M. R. (2008). *Examining the impact of the world crude oil price on China's agricultural commodity prices: the case of corn, soybean, and pork.*

APPENDICES

APPENDIX A

ROTATED FACTOR SOLUTION FOR EACH YEAR

AI Factor Solution for Year 2008

No	Variables	1	2	3	4	5
1	Maize	-0.072	0.776	0.034	-0.047	-0.043
2	White Yam	-0.001	0.573	-0.185	-0.259	0.347
3	Cassava	0.863	0.067	0.069	0.017	0.073
4	Tomato	0.049	0.682	0.459	-0.069	0.104
5	Garden Egg	0.721	0.100	-0.065	-0.130	0.046
6	Dry Pepper	0.058	0.150	0.718	0.060	0.111
7	Red Groundnut	-0.305	0.806	0.018	-0.190	-0.010
8	White Cowpea	-0.313	0.828	0.150	0.132	0.005
9	Palm Oil	0.404	0.023	0.178	0.706	0.205
10	Orange	0.765	-0.281	-0.219	-0.038	-0.108
11	Banana	0.880	-0.234	-0.069	0.025	-0.070
12	Smoked Herring	0.729	-0.409	-0.235	0.177	-0.099
13	Koobi	0.534	-0.407	0.089	0.260	-0.046
14	Onion	-0.335	0.003	0.766	-0.138	0.048
15	Egg	0.644	-0.013	0.329	0.330	-0.092
16	Plantain	0.890	-0.032	-0.092	0.115	-0.032
17	Gari	0.123	0.176	0.205	-0.828	0.074
18	Local Rice	0.248	0.492	0.096	0.066	-0.298
19	Imported Rice	-0.046	-0.013	0.210	0.073	0.890

AII Factor Solution for Year 2009

No	Variable	1	2	3	4	5
1	Maize	-0.359	0.537	0.364	-0.298	0.181
2	White Yam	-0.060	0.407	0.633	-0.325	0.056
3	Cassava	0.839	0.058	-0.092	-0.044	0.061
4	Tomato	0.001	0.748	0.322	0.161	0.021
5	Garden Egg	0.210	0.108	0.141	0.006	0.857
6	Dry Pepper	0.231	0.245	0.064	0.737	0.169
7	Red Groundnut	-0.499	0.495	0.494	0.016	0.070
8	White Cowpea	-0.260	0.433	0.738	-0.067	-0.135
9	Palm Oil	0.615	-0.292	0.309	0.236	-0.345
10	Orange	0.778	-0.273	-0.241	-0.001	0.267
11	Banana	0.859	-0.161	-0.164	-0.006	0.216
12	Smoked Herring	0.678	-0.410	-0.247	0.024	0.217
13	Koobi	0.721	-0.257	-0.238	0.063	-0.107
14	Onion	-0.130	0.767	0.106	0.144	0.044
15	Egg	0.705	0.351	0.101	0.109	-0.280
16	Plantain	0.884	-0.055	-0.004	0.006	0.136
17	Gari	-0.126	0.636	-0.092	-0.484	0.066
18	Local Rice	-0.086	-0.024	0.813	0.207	0.148
19	Imported Rice	-0.142	-0.049	-0.039	0.684	-0.102

AIII Factor Solution for Year 2012

No	Variable	1	2	3	4	5
1	Maize	0.824	0.236	0.063	-0.047	0.104
2	White Yam	0.573	0.501	0.004	0.096	0.006
3	Cassava	-0.239	0.244	0.733	-0.095	-0.187
4	Tomato	0.570	0.352	0.369	-0.165	-0.066
5	Garden Egg	0.292	-0.067	0.756	-0.150	0.120
6	Dry Pepper	0.284	0.749	-0.020	0.100	-0.050
7	Red Groundnut	0.254	0.639	0.097	0.330	0.379
8	White Cowpea	0.219	0.511	0.031	0.555	0.278
9	Palm Oil	-0.223	-0.084	0.333	0.572	-0.188
10	Orange	-0.818	0.028	0.292	-0.059	-0.079
11	Banana	-0.655	-0.351	0.420	0.020	0.005
12	Smoked Herring	-0.755	-0.119	0.085	0.050	-0.165
13	Koobi	0.009	0.102	-0.070	0.630	-0.079
14	Onion	0.072	0.819	-0.109	-0.079	0.231
15	Egg	-0.222	0.047	0.602	0.358	-0.039
16	Plantain	-0.145	-0.210	0.656	0.208	0.046
17	Gari	0.171	0.379	0.167	-0.267	0.609
18	Local Rice	0.042	-0.026	0.105	0.559	0.547
19	Imported Rice	0.053	0.098	-0.145	-0.032	0.693

AIV Factor Solution for Year 2013

No	Variable	1	2	3	4	5
1	Maize	0.792	-0.132	0.018	0.084	-0.163
2	White Yam	0.588	-0.260	0.110	0.141	0.166
3	Cassava	0.046	0.087	0.860	-0.124	-0.057
4	Tomato	0.664	0.073	0.277	0.218	-0.258
5	Garden Egg	0.388	-0.123	0.697	0.074	0.211
6	Dry Pepper	0.692	0.191	0.312	-0.102	-0.041
7	Red Groundnut	0.770	0.085	-0.131	0.391	0.100
8	White Cowpea	0.717	0.053	0.112	0.155	0.282
9	Palm Oil	0.033	0.676	-0.168	-0.091	0.121
10	Orange	-0.390	0.515	0.102	-0.237	0.372
11	Banana	-0.691	0.448	0.033	0.130	0.029
12	Smoked Herring	-0.369	0.216	-0.300	0.158	0.255
13	Koobi	0.087	-0.051	0.046	0.097	0.822
14	Onion	0.154	0.103	0.080	0.660	-0.425
15	Egg	0.051	0.732	0.058	0.015	-0.229
16	Plantain	-0.353	0.490	0.349	0.111	-0.044
17	Gari	0.168	-0.349	0.345	0.578	0.054
18	Local Rice	0.080	0.103	-0.091	0.562	0.077
19	Imported Rice	0.004	-0.204	-0.097	0.379	0.112

AV Factor Solution for Year 2015

No	Variable	1	2	3	4	5
1	Maize	-0.459	0.422	0.278	0.268	-0.065
2	White Yam	0.003	0.233	0.292	0.346	-0.422
3	Cassava	0.716	0.165	-0.054	0.006	0.158
4	Tomato	-0.104	0.830	-0.130	0.063	-0.027
5	Garden Egg	0.152	0.296	-0.220	0.461	-0.260
6	Dry Pepper	-0.160	0.050	-0.088	0.723	0.297
7	Red Groundnut	-0.108	0.287	0.129	0.622	0.064
8	White Cowpea	-0.243	0.402	0.550	0.199	0.200
9	Palm Oil	0.167	-0.131	-0.024	-0.016	0.728
10	Orange	0.397	-0.293	-0.043	-0.162	0.278
11	Banana	0.823	-0.055	-0.006	-0.029	-0.088
12	Smoked Herring	0.516	-0.496	0.318	0.105	0.096
13	Koobi	-0.041	-0.332	0.573	0.077	-0.079
14	Onion	0.055	0.701	0.157	0.176	-0.079
15	Egg	0.168	0.086	0.106	0.076	0.710
16	Plantain	0.858	-0.087	0.013	0.033	0.220
17	Gari	0.117	-0.163	0.184	0.634	-0.249
18	Local Rice	0.015	-0.047	0.765	0.036	0.128
19	Imported Rice	0.076	0.348	0.626	-0.105	-0.244

APPENDIX B

DETAILED MARKET CATEGORIZATION FOR POOLED DATA

Market No.	Ext_Low	Low	Moderate	High	ExtHigh	Total
1	0	0	5	0	0	5
2	0	0	5	0	0	5
3	0	3	2	0	0	5
4	0	0	5	0	0	5
5	0	0	5	0	0	5
6	0	0	3	2	0	5
7	0	0	5	0	0	5
8	0	0	5	0	0	5
9	0	0	5	0	0	5
10	0	0	5	0	0	5
11	0	0	5	0	0	5
12	0	0	4	1	0	5
13	0	0	5	0	0	5
14	1	1	3	0	0	5
15	0	0	5	0	0	5
16	0	0	5	0	0	5
17	0	0	4	1	0	5
18	0	0	3	2	0	5
19	0	0	1	3	1	5
20	0	0	5	0	0	5
21	0	0	5	0	0	5
22	0	0	5	0	0	5

23	0	0	5	0	0	5
24	1	1	3	0	0	5
25	0	1	4	0	0	5
26	0	1	4	0	0	5
27	0	0	5	0	0	5
28	0	0	5	0	0	5
29	0	0	5	0	0	5
30	0	0	5	0	0	5
31	0	1	4	0	0	5
32	0	0	5	0	0	5
33	0	0	5	0	0	5
34	0	1	4	0	0	5
35	0	1	4	0	0	5
36	0	0	5	0	0	5
37	0	0	5	0	0	5
38	0	2	3	0	0	5
39	0	0	5	0	0	5
40	0	0	5	0	0	5
41	0	0	5	0	0	5
42	1	0	4	0	0	5
43	0	2	3	0	0	5
44	0	0	5	0	0	5
45	0	0	5	0	0	5
46	0	1	4	0	0	5
47	0	2	3	0	0	5

48	0	0	5	0	0	5
49	0	0	5	0	0	5
50	0	1	4	0	0	5
51	0	0	5	0	0	5
52	1	1	3	0	0	5
53	0	0	5	0	0	5
54	0	0	4	1	0	5
55	0	0	3	2	0	5
56	0	0	3	2	0	5
57	0	0	0	0	5	5
58	0	0	2	3	0	5
59	0	0	2	3	0	5
60	0	0	4	1	0	5
61	0	0	1	2	2	5
62	0	0	1	4	0	5
63	0	0	2	1	2	5
64	0	0	3	2	0	5
65	0	0	0	3	2	5
66	0	0	1	2	2	5
67	0	0	3	1	1	5
68	0	0	0	0	5	5
69	0	0	0	4	1	5
70	0	0	2	2	1	5
71	0	0	5	0	0	5
72	0	0	5	0	0	5

73	0	0	5	0	0	5
74	0	0	5	0	0	5
75	0	0	5	0	0	5
76	0	0	5	0	0	5
77	0	0	3	1	1	5
78	0	0	4	1	0	5
79	0	0	4	1	0	5
80	0	0	5	0	0	5
81	1	1	3	0	0	5
82	0	2	3	0	0	5
83	0	3	2	0	0	5
84	0	2	3	0	0	5
85	0	1	4	0	0	5
86	2	2	0	1	0	5
87	0	1	4	0	0	5
88	0	1	4	0	0	5
89	2	2	1	0	0	5
90	2	2	1	0	0	5
91	1	2	2	0	0	5
Total	12	38	336	46	23	455

APPENDIX C

ORDERED EIGEN VECTORS OF THE POOLED GPCA

GPC1	GPC2	GPC3	GPC4	GPC5	GPC6	GPC7	GPC8	GPC9	GPC10	GPC11	GPC12	GPC13	GPC14	GPC15	GPC16	GPC17	GPC18	GPC19
1.206	0.015	-0.006	-0.046	0.124	0.010	-0.004	0.043	-0.100	-0.015	-0.500	-0.016	0.024	-0.077	2.980	-0.900	-0.016	0.060	0.050
-0.058	1.539	0.298	-0.137	-0.171	-0.110	-0.013	0.141	0.035	-0.170	0.870	0.025	0.190	-0.161	4.800	-3.900	-0.190	0.160	0.200
-0.040	-0.002	1.416	-0.016	-0.124	0.013	-0.024	-0.037	0.120	-0.150	0.110	-0.064	0.020	0.020	-0.960	0.410	-0.024	0.040	0.020
-0.181	-0.090	0.001	1.151	0.010	0.060	0.008	-0.157	-0.063	0.194	2.880	-0.190	-0.100	0.090	0.461	-3.340	0.158	0.140	-0.110
0.06	-0.031	-0.140	-0.019	1.410	0.015	-0.034	0.035	-0.090	-0.040	0.654	0.024	-0.035	0.004	0.817	-0.460	0.043	0.004	-0.044
0.383	-0.170	0.317	-0.103	0.300	1.180	0.073	-0.001	-0.240	0.342	0.892	-0.263	-0.080	-0.068	7.040	-2.030	0.072	0.170	-0.264
-0.084	-0.005	-0.149	-0.040	-0.180	-0.030	1.731	-0.043	-0.450	0.640	0.804	0.122	-0.140	-0.211	11.370	-2.740	0.260	-0.040	-0.043
0.309	0.083	0.032	-0.194	0.341	-0.070	-0.113	1.580	-0.354	-0.583	-2.650	0.410	-0.012	-0.183	8.900	-2.150	0.058	0.025	0.163
-0.004	0.007	0.086	0.012	-0.090	0.001	-0.009	-0.028	1.570	-0.370	-0.113	0.015	-0.028	0.008	-1.840	0.660	-0.030	0.026	-0.035
-0.021	-0.006	-0.017	-0.005	-0.040	0.010	0.035	-0.035	-0.090	1.540	0.341	-0.050	-0.014	0.037	0.300	0.060	0.007	-0.006	-0.019
-0.022	0.001	-0.046	0.010	-0.010	0.010	0.007	-0.007	0.020	-0.015	2.103	-0.013	-0.006	0.016	-0.550	-0.002	-0.008	-0.002	-0.002
-0.060	0.096	-0.596	0.196	0.260	-0.121	0.029	0.268	-0.095	-1.490	-3.320	2.520	-0.111	-0.114	5.290	0.990	-0.250	-0.076	0.104
0.408	0.246	0.707	-0.140	-0.280	-0.100	-0.270	0.150	-0.293	-1.060	-2.770	0.250	1.630	-0.083	3.984	-1.430	-0.240	0.100	0.192
-0.300	-0.039	-0.218	0.058	0.060	0.023	0.013	-0.015	0.071	0.590	4.014	-0.173	-0.032	1.310	-3.300	-1.573	-0.017	-0.070	-0.074
0.002	-0.001	0.001	-0.001	0.001	0.000	0.003	-0.001	-0.006	0.000	-0.014	0.004	-0.002	-0.002	1.540	-0.008	0.000	0.002	0.002
0.034	-0.005	-0.162	-0.003	-0.010	0.004	-0.003	-0.012	0.013	-0.160	-0.760	-0.002	-0.010	0.011	-2.120	3.610	-0.071	-0.020	0.010
-0.112	-0.035	0.009	-0.003	0.080	-0.019	0.073	0.035	-0.150	0.200	0.300	-0.023	-0.032	-0.075	2.460	-1.060	1.384	0.020	0.080
0.306	0.116	0.288	0.030	-0.180	-0.017	-0.103	0.037	0.106	-0.700	-1.460	0.180	0.038	-0.230	5.350	0.370	0.031	1.480	0.200
0.216	0.050	0.222	-0.030	-0.130	-0.049	-0.082	0.055	-0.133	-0.250	0.410	0.054	0.015	-0.089	3.790	-0.875	0.124	0.052	1.550

APPENDIX D

VARIANCE-COVARIANCE MATRIX FOR POOLED DATA

Var	Mz	YmWt	Cv	Tm	GEg	PpDr	GnR	CpWt	PmOil	Org	Ban	HrSmk	Kbi	On	Eg	Pltn	Gri	RiLoc	RiImp
Mz	379.1																		
YmWt	551.8	3984.6																	
Cv	-43.4	27.2	257.9																
Tm	523.1	1123.9	122.8	3775.8															
GEg	45.3	180.3	74	313.6	270.9														
PpDr	343.3	938.4	134.3	1150.3	273.7	5414													
GnR	406.8	1020	-60.9	950.6	181.7	1218	2444												
CpWt	387.3	973.1	-25.3	1074.9	70.7	1099	1093	2518											
PmOil	-49	-201.9	33.6	-79.6	-18.9	71.9	-43.4	-41.6	217										
Org	-64.9	-117.3	35.2	-119.7	-14.6	-55	-94.5	-84.5	35.3	91.9									
Ban	-30.9	-47.4	24.1	-51.9	5	-63.6	-47.2	-63	9.3	15.9	17.6								
HrSmk	-311	-455.3	140.9	-1207	-31.4	-474	-358.3	-303.8	135	136	97.8	2391							
Kbi	2.1	314.8	53.9	-574.1	14.9	67.1	11.5	502.3	-4.4	30.6	8	735.2	3458						
On	255.7	745.4	-11.9	1167	76.6	644.8	858	694	-82.5	-131	-14	-419	-428	2478					
Eg	-0.9	-8.8	3.8	3.4	-0.4	9.6	1.2	4.1	5.3	1.5	1	8.1	1.8	1.8	1				
Pltn	-31	-58.4	40.7	-47.4	2.6	-22.8	-41.4	-42.9	22.1	17.7	11.7	76.6	-1	-19.5	1.6	36			
Gri	97.8	388.9	4.1	91.9	69.8	252.3	311	204.9	-47.5	-20.6	-5.1	-10.4	60.3	286	-2	0.7	561.9		
RiLoc	126.1	429.1	-38.1	-272.1	-37.2	109	531	590.3	90.5	-2.6	-5	276.3	424	226	2.2	24.6	87.2	1899.2	
RiImp	27.8	232.2	-37.1	244.5	25.2	58.1	165.4	323.3	-48.1	-41.2	-2.8	-55	203.4	280	-2	-11	108.4	369.5	869.6

APPENDIX E

ALGORITHM

The algorithm for the k^{th} year, given i=1,2,3,…,n

1 $Y_k = FD\left[(k-1)n+1:k*n,1:p\right]$

2 $S_k = Cov(Y_k);$

3 $m_k = median(Y_k);$

4 $I_k = ones(n,1);$

5 $MC_k = Y_k - \left(m_k' \times I_k'\right)';$

6 $D_k = MC_k \times (inv(S_k)) \times MC_k';$

7 $d_k = diag(D_k);$

8 $p_k = \exp(-h \times d_k);$

9 $s_k = sum(p_k);$

10 $X_i = MC_k(i,:)' \times MC_k(i,:);$

11 $T_i = p_k(i,:) \times X_i;\ i = 1,2,3,\ldots,n$

12 $F_{Y_k} = cat(size(T_1,1), T_1, T_2, \ldots, T_n);$

13 $sum(F_{Y_k}, size(T_1,1));$

14 $SS_k = \left(\frac{1}{s_k}\right) \times \left(sum\left(F_{Y_k}, size(T_1,1)\right)\right);$

15 $GM_k = s_k \times (inv(SS_k));$

16 $[V,D] = eig(GM_k);$

17 $GP_k = V(:,[re-arranged]);$

18 $Z_k = zscore(Y_k);$

19 $ZGP_k = Z_k \times GP_k;$

$end;$

Explanation of algorithm

Line 1: This line of code introduces the algorithm that is scripted in Matlab. This code determines the array of data to slice off from the whole data set read into the memory of the Matlab application, after reading the original data from a source on the local drive into the Matlab application. The data read into the application is given the variable name FD, a 455 by 91 numeric matrix. The K stands for the number of years. $K = 1, 2, 3, 4, 5$ Since the data spans five years. n is the number of markets, such that $n = 1, 2, 3, ..., 91$. And p is the number of variables (commodity food items) such that $p = 1, 2, 3, ..., 19$. So from this line of code, for the first year, the algorithm fetches data from FD and slices off, $FD(1:91,1:19)$, and assigns the output to a variable named Y_1. The code repeats the same for each year and slices off the appropriate data to use for computation.

Line 2& 3: This line of code computes the variance-covariance matrix for the (91×19), Y_1 data from line 1. The result is assigned the variable name S_k .line 3 computes the median of the data Y_1, a (1×19) row vector.

Line 4& 5: This computes a (91×1) column vector of ones. The product of the vectors of line 3 and line 4 gives a (91×19) matrix. Which is suitable for the computation. Line 5 median centre the data, Y_1. By subtracting the median obtained in line 4 from the raw data in line 1.

Line 6& 7: This code computes the ratio of the median centred data to the variance-covariance matrix computed in line 2. Line 7 computes the diagonal of the resulting matrix.

Line 8& 9: This line of code is used to calculate the constant p_k. Using a value of $h = 0.1$ (Caussinus& Ruiz, 1990). We find the exponent of the product of the diagonal of the matrix for line 7 and h. The results are a (91×1) column vector. Line 9 computes the sum of the individual entries. The result is assigned the variable name S_k.

Line 10: The variable X_i is assigned to the cross product of the matrix MC_k from line 5. A product of the transpose of the $(1\times19)'$ column vector by (1×19) row vector to yield a (91×19) matrix.

Line 11: This line of code is the product of the ith entry of p_k from line 8, and the X_i. This gives an output of a (1×19) row vector assigned the variable name T_i. Line 10 and 11 are iterated 91 times to generate a series of X_i and T_i row vectors respectively.

Line 12: This line of code is known as the concatenate function. It allows two or more matrices to be joined into one new matrix. The function is assigned a variable named F_{Y_k}. It puts together all the T_i matrices, arranged one on top of the other, into one new matrix.

Line 13, 14 & 15: It adds all ith row entries of F_{Y_k} for each k, whose dimension is the same as that of $(T_i, 1)$. Line 14 assigns the variable name SS_k. The line executes the ratio of the sum from line 13 with the sum from line 9. It is a (1×19) row vector. This line of code executes the ratio of the results from line 9 and the result from line 14. Line 15 assigns the variable name GM_k, a (19×19) matrix.

Line 16& 17: This line of code computes the Eigenvector and the corresponding Eigenvalues of the (19×19) matrix obtained from line 15. This function $[V, D] = eig(GM_k)$ is the standard code for calculating the eigenvector and the corresponding eigenvalues of any definite matrix. Where V the eigenvector and D is the corresponding eigenvalues respectively. Line 17 computes the eigen-values and order the eigenvectors in descending order. Before the code is executed, the eigenvalues are re-ordered using the sort function in Matlab. The results is stored in the variable GP_k.

Line 18& 19: Line 18 standardizes the data using the zscore. The result is stored in the variable named Z_k. Line 19 finds the product of the standardized data in line 8 and the re-arranged eigenvector in line 17.

APPENDIX F

CODE FOR THE FIVE YEARS POOLED TOGETHER

1 $Y_k = FD\left[(k-1)n+1 : k*n, 1 : p\right]$

2 $S_k = Cov(Y_k);$

3 $m_k = median(Y_k);$

4 $I_k = ones(n,1);$

5 $MC_k = Y_k - \left(m_k' \times I_k'\right)';$

6 $D_k = MC_k \times \left(inv(S_k)\right) \times MC_k';$

7 $d_k = diag(D_k);$

8 $p_k = \exp(-h \times d_k);$

9 $s_k = sum(p_k);$

10 $X_i = MC_k(i,:)' \times MC_k(i,:);$

11 $T_i = p_k(i,:) \times X_i;\ i = 1,2,3,\ldots,n$

12 $F_{Y_k} = cat\left(size(T_1,1), T_1, T_2, \ldots, T_n\right);$

13 $sum\left(F_{Y_k}, size(T_1,1)\right);$

14 $SS_k = \left(\frac{1}{s_k}\right) \times \left(sum\left(F_{Y_k}, size(T_1,1)\right)\right);$

15 Repeat step 1 to step 14 for $k = 1,2,3,4,5$

16 Obtain $SS_1, SS_2, SS_3, SS_4, SS_5$

17 From step 2, obtain S_1, S_2, S_3, S_4, S_5

18 $S_T = \left(\frac{1}{5}\right)(S_1 + S_2 + S_3 + S_4 + S_5);$

19 $SS_T = \left(\frac{1}{5}\right)(SS_1 + SS_2 + SS_3 + SS_4 + SS_5);$

20 $GM_T = S_T \times (inv(SS_T));$

21 $[V, D] = eig(GM_T);$

22 $GP_T = V(:, [re - arranged]);$

23 $Z_k = zscore(Y_k);$

24 Repeat Z_k for $k = 1, 2, 3, 4, 5$;

25 $Z_T = [Z_1; Z_2; Z_3; Z_4; Z_5];$

26 $GZP_T = Z_T \times GP_T;$

Line 1: Introduces the algorithm that is scripted in Matlab. This code determines the array of data to slice off from the whole data set read into the memory of the Matlab application. The data is given the variable name FD. FD is a (455×91) numeric matrix. The K stands for the number of years. $k = 1, 2, 3, 4, 5$ Since the data spans five years. n is the number of markets, such that $n = 1, 2, 3, \ldots, 91$. And p is the number of variables (commodity food items) such that $p = 1, 2, 3, \ldots, 19$. From this line of code, for the first year, the algorithm fetches data from FD and slices off, $FD[1:91, 1:19]$ and assigns the output to a variable named for the first year. We repeat the same procedure for each year and slices off the appropriate data to use for the computation.

Line 2 & 3: This line of code computes the variance-covariance matrix for the (91×19) matrix from line 1. The result is assigned the variable name S_k. We repeat this step five times for the five non-consecutive years. Line 3 computes the median of the data Y_k, a (1×19) row vector.

Line 4 & 5: This computes a (19×1) column vector of ones. The product of the vectors of line 3 and line 4 gives a (91×19) matrix, which is suitable for the computation. Code line 5 median-centre the data. By subtracting the median obtained in step 4 from the raw data Y_k.

Line 6 & 7: This code computes the ratio of the median-centred data to the variance-covariance matrix computed in line 2. This is used to do a matrix multiplication of the median-centred data and the inverse of the covariance matrix. Step 7 computes the diagonal of the resulting matrix from the execution of code line 6.

Line 8 & 9: This line of code calculates the constant P_k. Using a value of $h = 0.1$ (Caussinus& Ruiz, 1990). This line finds the exponent of the product of the diagonal of the matrix for line 7 and the h. the results is a (91×1) column vector. Line 9 computes the sum of the individual entries. The result is assigned the variable name S_k.

Line 10: The variable X_i is assigned to the cross product of the matrix MC_k from line 5. A product of the transpose of the $(1 \times 19)'$ column vector by a (1×19) row vector to yield a (91×19) matrix.

Line 11: This line of code is the product of the ith entry of p_k from line 8, and the X_i. This gives an output of a (1×19) row vector assigned the variable name T_i,

Line 10 and line 11 are iterated 91 times to generate a series of X_i and T_i row vectors respectively.

Line 12 & 13: This line of code is known as the concatenate function. It allows two or more matrices to be join into one new matrix. The function is assigned a variable named F_{Y_k} . It puts together all the T_i matrices, arranged one on top of the other, into one new matrix. Step 13 adds all ith row entries of Y_k for each k, whose dimension is the same as that of $(T_i,1)$

Line 14, 15 &16 : This line of code is assigned the variable name SS_k. The line executes the ratio of the sum from line 13 with the sum from line 9. It is a (1×19) row vector. The steps 1 to 14 is repeated five times for each of the five years; 2008, 2009, 2012, 2013, 2015.We obtain $SS_1+SS_2+SS_3+SS_4+SS_5$ respectively for each year.

Line 17, 18 & 19: We repeat line 2, for each year and assign them variable names S_1,S_2,S_3,S_4,S_5. We compute the pooled variance covariance of the five years, by adding the individual covariance and dividing it by 5. And we assign the variable name S_T .We repeat the procedure of line 16, for all five years. We add them, and divide by 5 and assign the total a variable name SS_T .

Line 20 & 21: This line of code executes the ratio of the results from step 18 and the result from step 19. It is assigned the variable name GM_T, a (19×19) matrix. Then we compute the Eigenvector and the corresponding Eigenvalues of the (19×19) matrix obtained from line 20. This function $[V, D] = eig(GM_T)$ is the standard Matlab code for calculating the eigenvector and the corresponding eigenvalues of any definite matrix. Where V is the eigenvector and D is the corresponding eigenvalues respectively.

Line 22 & 23: This next line re-writes the eigenvectors in a way that is ordered in descending order. The results is stored in the variable GP_T. Line 23 computes the standardization of the data. This is done by using the pre-defined function zscore of the data Y_k. The result is stored in the variable named Z_k.

Line 24 & 25: The step 23 is repeated, for the rest of the five years. This line of code, organises the individual matrices obtained in line 24 into one matrix named Z_T, one stacked on top of the other.

Line 26: This line of code in Matlab finds the product of the standardization of the raw data in line 25 and the re-arranged eigenvector in line 22.