UNIVERSITY OF CAPE COAST

TECHNIQUES FOR DETERMINING CLASSIFICATIONS IN MULTIPLE MULTIVARIATE DATA: APPLICATION TO PRICES OF LOCAL FOOD ITEMS ON MARKETS IN GHANA

BY

FRANCIS EYIAH-BEDIAKO

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

JANUARY 2019

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 27/05/2019

Name: Francis Eyiah-Bediako

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 27/05/2019

Name: Dr. Bismark Kwao Nkansah

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

Name: Prof. Nicholas Nicodemus Nana Nsowah-Nuamah

ABSTRACT

The study makes use of time-dependent displaying components and structural equation modelling to determine the price levels of key local food items on several markets in Ghana over non-consecutive time periods. The application of the techniques to this multiple multivariate complex data structure identifies suitable dimensions along which to assess the major influence of the price data over a time-period and highlight possible extreme prices simultaneously. It also examines multiple sets of data-generating variables that may be put together in a single model. These variables include a 'vectorised' factor solution and some market-feature covariates. The displaying components comprise the principal component and the outlier displaying component (ODC). The study has provided necessary extensions that would make the components suitable for the study. The plots for the first five components in addition to the preliminary results give the set of suspect outlying markets. Using this set, the Modified 1-ODC is applied based on the pooled reduced sample Sum of Squares and Cross Product matrix. Markets 17 and 65 are clearly identified as the most consistently low and high priced, respectively, over the period. Using the factor solution and two covariates, a structural model is obtained for determining the price levels. Even though the factors constitute a significant model by themselves, they are not significant in the model that contains significant covariates, which are 'Region' and the 'number of days' of trading. The model shows that extreme markets, which are few, are predominantly associated with large number of market days. Equitable and increased production of cereals and spices in particular in all regions could reduce price variations across markets and enhance well-being.

KEY WORDS

Displaying Components

Markets

Outliers

Price Levels

Structural Equation Model

Vectorised Factor Solution

ACKNOWLEDGEMENTS

I am extremely grateful to my principal supervisor, Dr. B. K. Nkansah, of the Department of Statistics, University of Cape Coast (UCC). His unwavering interest, valuable contributions and suggestions, and close supervision during the planning and execution of this study were splendid. My special thanks also go to my co-supervisor, Prof N. N. N. Nsowah-Nuamah who helped to shape this research.

Thanks to Prof. D. K. Essumang, the Dean of the School of Physical Science, University of Cape Coast, who continuously showed interest in the work and always encouraged me to finish on time. To all colleague lecturers of Departments of Statistics and Mathematics, University of Cape Coast, I give sincere gratitude for your various contributions and continual encouragements. My profound appreciation also goes to Dr. (Mrs.) Henrietta Nkansah and Mr. Joseph Mensah both of the Department of Mathematics for their singular contributions to this work.

My heartfelt thanks also go to Rev. J. N. Nortey and all staff of the Statistical, Research and Information Directorate (SRID) of the Ministry of Food and Agriculture (MOFA), for giving me the secondary data used for the study. To all past students of UCC who helped me in the administration and collection of my questionnaires, I am very grateful. I am thankful to the staff of Oguaa Hall, UCC, for their support in diverse ways. Above all, I am thankful to the Almighty God for another great honour done me.

I appreciate immensely the forbearance, support and encouragement given me by my wife and the children in the pursuit of this research.

DEDICATION

To my wife, Sylvia and the children, Blessing and Matilda

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

ODC	-	Outlier Displaying Component
CFA	-	Confirmatory Factor Analysis
SEM	-	Structural Equation Modelling
MKT	-	Market
LOC	-	Location

CHAPTER ONE

INTRODUCTION

In this introductory chapter, the importance of the price of an item as a strategic marketing variable is highlighted. Common market forces or factors that influence price determination are reviewed, in addition to the effect of movements in item prices in recent past in the country and around the globe. These will be captured in the background of the study which also will include the nature of the data at hand and how it prescribes the techniques for the analysis. The background will provide the necessary motivation for the study which will be expressed in the statement of the problem, out of which the objectives of the study will be spelt out. The description of the data problem and the research design will be given adequate space in this chapter, as the problem is of a national interest and involves multiple multivariate data. The techniques that are employed in this research are originally not designed for the purpose. As a result, an attempt will be made to introduce the techniques and justify their use and possible extensions. The significance of this study and the outline of the entire thesis will be the final portions of this chapter.

Background to the Study

The market price of a commodity is the short-term equilibrium price that is decided on daily basis resulting from short-term to medium-term stable conditions prevailing in the market. Thus, price is a strategic marketing variable that influences consumer purchase behaviour as well as that of traders since they will always want to break-even. It consequently has effect on the patronage in the markets.

Other factors such as seasonal variation in food supply due to climatic changes and other occurrences such as flood, fire among others make it difficult for Ghana to meet its food demands all year round. It is obvious that poor rural road infrastructure limits the effective distribution of food and could increase market prices arbitrarily. The poor nature of roads to production centres which seriously contribute to the high cost of transporting food commodities to the market centres directly or indirectly affect food prices in our markets. Inadequate market information also leads to weak market integration between local, district and regional markets. However, the commodity's own price, price of related goods, inflation rate, income levels of consumers, cost of production, proximity to source of production, prevailing market conditions are among factors that are likely to influence price determination in Ghana (Food and Agriculture Sector Development Policy Document, FASDEP II, August 2007; Meyer & Yu, 2013).

According to Clapp (2015), food price tends to have a major impact on food security, at both household and country levels. Many of the world's poorest people spend more than half of their income on food. Price hikes for cereals and other staples can force them to cut back on the quantity or quality of their food. This may result in food insecurity and malnutrition, with tragic implications in both the short and long term. In the 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. While international food prices have declined since mid-2008, they are still substantially higher than the time prior to the price surge between 2006 and 2008, and they remain at the same levels in 2010 or higher for the next decade (IFAD, 2011).

The sharp increase in food prices both in world markets and in local markets since 2006 has raised serious concerns about the food and nutrition situation of poor families in many countries. Particularly in urban areas, where people cannot grow their own food, household budgets have been squeezed. Globally, the costs of food and fuel increased significantly between February 2005 and February 2008. On the average, food prices rose by 82 percent and oil prices increased by about 80 percent (over US$140 per barrel), thus, establishing linkage between oil prices and food prices. Higher food prices have been steep for some basic food items such as corn and wheat which doubled during the period, (Scott-Joseph, 2009). Consequently, these price movements in the global market had a dominion effect on inflation and depleted foreign exchange earnings. Undoubtedly, the impact of high food prices affects every fragment of the society: poor households, middle and upper class households, profit and non-profit organizations and governments. The most severe is the negative impact on the most vulnerable group. Scott-Joseph (2009) contested that efforts towards eradicating poverty will be significantly affected by persistent high food prices as this situation will push more people into poverty.

In Ghana, the 2007/2008 fiscal year observed a high rate of food price increases following the global food crises. For instance, the prices of cereals increased from 20% to 30% between 2007 and 2008. Also food component of the consumer price index rose from 193.9 to 246.7 indicating a 27% food inflation within the same period. (Ghana Statistical Service, 2008; Wodon, 2008; Ghana Statistical Service, 2009).

Several policy interventions have been implemented to protect consumers from rising food prices in Ghana. For instance, during the year 2008, Ghana Government was compelled to suspend import duties on rice, yellow corn and wheat in an attempt to cushion Ghanaian consumers from the severe impacts of further price increases.

In Ghana, studies concerning market prices are usually conducted periodically by national institutions and agencies such as the Ghana Statistical Service (GSS) and the Statistical, Research and Information Directorate (SRID) of the Ministry of Food and Agriculture (MoFA). They basically study prices of selected items from key markets in all regions of the country. The SRID, for example, have identified some sixty-four key markets and others from which they collect monthly food prices for monitoring. The GSS is also engaged in market surveys, and compute some periodic economic measures such as the monthly Consumer Price Index (CPI). It is conceivable that studies in this regard could be pursued from other perspectives other than those followed by the national agencies. Academic corroborations, such as what this research seeks to do, is needed to provide an exhaustive study of our market statistics. It is the view of this study that such corroborative research will unearth market information that could eventually lead to market organization and creation as a poverty-reduction strategy, since a good number of our population is engaged in trading.

The only work in this regard that comes close to the attempt made in this thesis is that of Seglah (2014) which sought to determine the level of prices in major markets of Ghana. However, that work utilises data involving only a single year of 2008. This work, which is quite restricted in coverage and methodology, already points out certain perspectives to market studies. The results of such studies could be influenced by two main issues: (1) the number and composition of the markets involved; and (2) the methodology applied in the analysis. On one hand, in as much as a large number of markets is important, the nature of the markets involved must be designated key markets from all the regions of the country. Major changes in the selection of these markets for different studies could alter the findings. The methodology, on the other hand, must be specifically designed for detecting the features of the markets as intended by the study. If the intended methods are not originally designed for the study, but has the desired properties, then adequate extensions and modifications must be made to accommodate the aim of the research. For example, the Principal Components Analysis, which are used in such studies (Sharma, 1996; Seglah, 2014) are recommended only as a preliminary technique (Barnett & Lewis, 1994). Thus, the study will take some rigorous methodological steps to enhance the credibility of the results. In this regard, it will combine both existing softwares and computationally intensive methods with codes written in MATLAB to carry out several transformations of the original data, among others. In the work of Seglah, Tepa market (a town in the Ashanti region and designated as Market 2 in this study) emerged as the lowest priced market in Ghana. It will be realised that by virtue of the methods used and slight changes in the composition of the markets in this study, Market 2 does not emerge as a conspicuously low priced in that year and across all the years considered in the study.

Among the main activities in the research is the development of necessary procedures that would enable, among others, the detection of extreme observations in the data which is a multiple multivariate datasets. By multiple multivariate dataset, we mean a replication of the usual multivariate data on several variables under different conditions, which results in categorised multivariate dataset. This type of data is frequently generated in practice in several fields. For example, performance of insurance companies may be assessed on a number of financial ratios, $R_1, R_2, \ldots, R_8$. However, these ratios may themselves be categorised into three types, for example, Capital Asset Ratio (CAR), Operating Efficiency (OE), and Profitability (P). Suppose that R_1, R_2, R_3 fall under CAR, R_4, R_5, R_6 fall under OE and R_7, R_8 fall under P. Thus, the type of financial ratio serves as the categorising variable in the multivariate dataset. In the context of this study, we may be interested, not just in a company with extreme performance on the eight ratios, but rather in the extent of simultaneous extremeness on each of the three types of ratios. The categorisation of the data that will constitute a multiple multivariate data required for our study must have the same number of the variables in each categorization. This type of data will further be described in a later section on the layout of the data problem.

Usually in the texts covering such data layout, the interest of the multivariate statistical tools (e.g., Hoteling's T test, MANOVA) has been on determining differences in means, especially in the case where the same variables are covered in each category. However, this will not be the focus of this study.

Studies on market price levels across various locations have usually made use of techniques such as the Principal Components Analysis (Sharma, 1996). This study makes use of this technique and others that also belong to the class of displaying components. This class of components could be affected by variations in the data. In the data problem for this study, this consideration is of prime concern as a result of potentially wide variations in prices across the entire nation over a number of years. This variation is seen in Table 1 which shows the total sample sum of squares and corresponding variance in the data for each of the five years under study.

Table 1: Variations in the Price Data for All Years

Year	Sample Sum of Squares	Sample Variance
1	5.2609×10^5	5845.44
2	9.4660×10^5	10517.78
3	2.0223×10^6	22470.00
4	2.9098×10^6	32331.11
5	7.5747×10^6	84163.33
Total	1.3980×10^7	30792.95

The values in the second column are obtained as $tr(\mathbf{S}_k)$, where $\mathbf{S}_k$ is the sum of squares and cross-product matrix of the data for the kth year. The total variance is the variance of the combined data, without reference to the years. There is a clear increasing trend in variation which could be assigned to either or both of two causes: (1) general increases in prices across almost all markets each year; and (2) increases in prices in one or few markets each year.

It will be observed in this study that the second is a more dominant cause of the variations.

The determination of level of prices may be driven by the methodology adopted for the study. In the view of this research, price levels may be determined from either or both of two sources: (1) the actual prices of items in the market; or (2) latent factors that influence the determination of these prices. Optimal information will undoubtedly be obtained if both sources are pursued jointly in a single study. Since two sources of 'explanatory' variables are to be brought together, it will call for a kind of some mediation models (Baron & Kenny, 1986; James & Brett, 1984; MacKinnon, 2008; Ledermann & Macho, 2015) which may be captured through structural equation modelling (SEM) technique. This technique analyses the interrelationships among the latent attributes. Compared to an ordinary regression analysis, the SEM approach has some appealing features. First, through grouping multiple indicators into a few latent attributes, the SEM reduces the model dimension significantly. Second, based on the condensed information, the SEM provides clearer and simpler model interpretation. Thus, the overall intention of the study is to obtain an optimal market and item price information for effective decision making on market creation and organisation.

Statement of the Problem

The problem of assessing extreme observations in statistical data has been studied mainly in single multivariate datasets. It means that methods that are usually used in such studies have not been designed and applied in multiple multivariate datasets. For example, the Principal Components

Analysis is not primarily designed as a technique for detecting extreme observations. However, since it is used as a means of constructing indices, it may be used as a preliminary measure for detecting extreme observations. Its weakness may be compounded therefore and highlighted, if it is applied as a time-dependent technique or in several multivariate datasets simultaneously. Similar reservations may be expressed about the applications of techniques that are actually designed for studying extreme observations, such as the Outlier Displaying Component (Nkansah & Gordor, 2012; 2013). Their use in extended multivariate data would therefore require substantial modifications.

There are obvious other reasons for the apparent lack of application of these techniques in extended multivariate data. As multivariate data varies, so their overall features may vary which could pose problems for techniques if they are not robust. In order to accommodate the effect of variations in observations on the original Outlier Displaying Component (ODC) (Gordor & Fieller, 1994), a modification of the technique has been carried out (Nkansah & Gordor, 2013). The modification enables the use of alternative measure of the mean vector in the estimation of the variance-covariance matrix and the Outlier Displaying Component (ODC) so that results are not influenced by extreme variations in the data. The applications so far have been limited to single multivariate data.

Another issue with the use of such techniques is that they are usually an end in themselves, not a means to an end. If by applying these techniques, some observations could be classified as extreme, then the others may be viewed as average or moderate. The resulting classified data may further be studied. Thus, the techniques, though advanced, could be used as interim techniques.

To handle multivariate data, the individual observations should be properly characterised. As the application in this study is a novelty, the first challenge will be a clear characterisation of an observation in a multiple multivariate dataset. In each step of the data processing, this characterisation of the general observation will be a basic problem that has to be fully described.

Apart from the usual national institutions such as the Ghana Statistical Service and the Ministry of Food and Agriculture whose mandate it is to monitor level of prices and produce periodic reports, there appears not to be any academic corroboration in this direction. What this means is that assessment of level of prices is being carried out from almost the same perspective over the years. This will certainly cause some vital developments on the subject to escape notice. There is therefore the need to develop other perspective for monitoring price levels for the benefit of the wellbeing of the citizenry. The relevance of rigorous research into market information is an important avenue for poverty reduction. This is because, price increases in basic food commodities creates insecurity and erodes the wealth and health of the populace, with the vulnerable being the hardest hit. It is the view of this study that such corroborative research will uncover market information that could eventually lead to market organization and creation as a poverty-reduction strategy, since it is assumed that a good number of our population is engaged in trading.

Objectives of the Study

The main objective of the study is to obtain a general assessment of price levels of local food items in Ghana. The specific objectives are as to:

1. identify and develop suitable techniques for assessing the price levels of markets over some non-consecutive years.
2. identify latent factors that influence price levels of food items;
3. obtain a model that may be used to determine the price level of a market in terms of the identified factors and other relevant market features.

Layout of the Data Problem and Research Design

In this study, the methodology will be applied to assess price levels, which is seen as a measure of market performance, on 19 commodities, $X_1, X_2, X_3, \ldots, X_{19}$. Thus, the commodities will serve as the variables of study. The food items are selected to include those that form the basis for the computation of the monthly Food Price Index (FPI) by the Ghana Statistical Service (GSS). These items will also include all those that consistently have data on them from all identified centres for over fifteen years. In all, nineteen food items would be studied. These food items include the following in Table 2 with their variable names.

Table 2: Items Constituting Variables for the Study

No	Commodity	Variable Name	Unit of Sale
1	Maize	Mz	100 kg
2	Imported Rice	RiImp	50 kg
3	White Yam	YmWt	250 kg (100 Tubers)
4	Cassava	Cv	91 kg
5	Tomatoes	Tm	52 kg (Crate)
6	Garden Eggs	GEg	27 kg
7	Dried Pepper	PpDr	16 kg
8	Red Groundnuts	GnR	82 kg
9	White Cowpea	CpWt	109 kg
10	Palm Oil	PmOil	18 Lit
11	Orange	Org	20 kg (100 single)
12	Banana	Ban	6-8 kg (Av. bunch)
13	Smoked Herring	HrSmk	100 singles
14	Kobi	Kobi	100 singles
15	Onion	Onn	73 kg
16	Egg	Egg	1 crate (30 singles)
17	Local Rice	RiLoc	50 kg
18	Plantain	Pltn	Apentu (Av. bunch)
19	Gari	Gr	68 kg

The items were selected to cover the eight main categorisations of local foodstuff, namely: Cereals, Roots and Tubers, Vegetables, Pulses, Fish, Spices, Oil and Fruits. The categorisations are as seen in Table 3.

Table 3: Categorisation of Foodstuff

SN	Category of Foodstuff	Example of Food Items
1	Cereals	Maize, Imported and Local Rice
2	Roots and Tubers	Cassava, Yam, Gari, Plantain
3	Vegetables	Tomatoes, Garden Eggs
4	Pulses	Cowpea, Red Groundnut
5	Fish	Smoked Herring, Kobi, Egg
6	Spices	Dried Pepper, Onion
7	Oil	Palm Oil
8	Fruits	Orange, Banana

Out of the many markets in Ghana, some are identified by the Statistical, Research and Information Directorate (SRID) of the Ministry of Food and Agriculture (MoFA) as key or leading venues whose description are based on their strategic location in the various regions, reliability of data generation procedures, their size, and relevant results of previous related studies of Seglah (2014).

The markets, which are given by the names of the towns and villages in which the markets are located and the region, are given in Table 4. The spatial distribution of the markets is also given in Figure 1.

Table 4: List of Markets

Mkt ID	Name of Market	Region	Mkt ID	Name of Market	Region
1	Agona	Ashanti	51	Mallam	Greater Accra
2	Tepa		52	Tema	
3	Adugyama		53	Madina	
4	Bekwai		54	Bole	Northern
5	Obogu		55	Salaga	
6	Agogo		56	Nalerigu	
7	Juaben		57	Gushiegu	
8	Ejura		58	Bimbila	
9	Kumasi		59	Tamale	
10	Obuasi		60	Damongo	
11	Mampong		61	Yendi	
12	Nsuta		62	Bawku	Upper East
13	Goaso	Brong Ahafo	63	Zebilla	
14	Kukuom		64	Bolgatanga	
15	Atebubu		65	Fumbusi	
16	Brekum		66	Garu	
17	D/Ahenkro		67	Navrongo	
18	Kintampo		68	Bugubelle	Upper West
19	Yeji		69	Tumu	
20	Sunyani		70	Wa	
21	D/Nkwanta		71	Akatsi	Volta
22	Techiman		72	Ho	
23	Nsawkaw		73	Hohoe	
24	Dunkwa	Central	74	Logba Alakpeti	
25	Ajumako		75	Abotoase	
26	Assin Praso		76	Kute	
27	F/Nyankomase		77	Denu	
28	Kasoa		78	Adidome	
29	Bawjiase		79	Mafi-Kumase	
30	Cape Coast		80	Kpeve	
31	Elmina		81	A/Nkwanta	Western
32	Mankessim		82	Bibiani	
33	Swedru		83	Sefwi Bekwai	
34	Winneba		84	Tikobo	
35	Nsawam	Eastern	85	Juaboso	
36	Anyinam		86	Bogoso	
37	Akoase		87	Asawinso	
38	New Tafo		88	Sefwi Dwinase	
39	Ahoman		89	Sekondi	
40	Mpraeso		90	Takoradi	
41	Agormanya		91	Tarkwa	
42	Koforidua				
43	Suhum				
44	Asamankese				
45	Agbogbloshie	Greater Accra			
46	Kaneshie				
47	Makola				
48	Ashaiman				
49	Kaseh				
50	Dome				

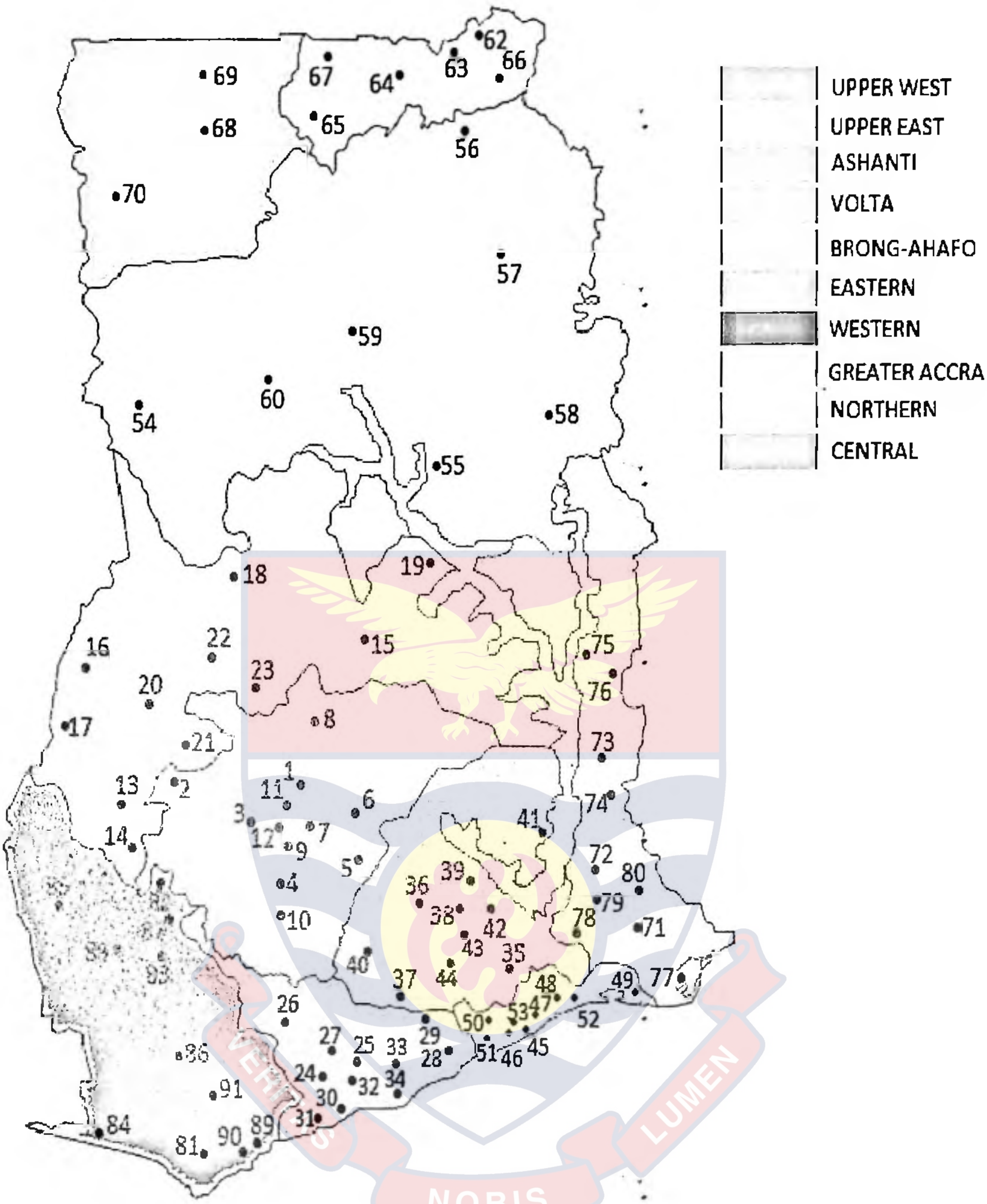

Figure 1: Map of Ghana Showing the Study Markets in Dots.

Now, since the prices are obtained across a number of years, the 'Year' could be used as a categorisation variable. The extent of extremeness of price levels could then be assessed simultaneously for all years. In the previous section, we have explained why it may not be expedient to determine extremeness of price level for the entire data without consideration for particular years. The layout of the data problem is displayed in Table 5.

As pointed out in the problem statement, it is crucial in such studies to identify a description of the general observation in this layout. Data on each year constitutes a single multivariate dataset. An observation in each year therefore is a vector of dimension p.

Table 5: Layout of the Data

			Year 1								Year 2		
Mkt	X_1	X_2	...	X_j	...	X_p	Mkt	X_1	X_2	...	X_j	...	X_p
1							1						
2							2						
⋮							⋮						
s				X_{sj1}			s				X_{sj2}		
⋮							⋮						
n_1							n_2						
			Year k								Year r		
Mkt	X_1	X_2	...	X_j	...	X_p	Mkt	X_1	X_2	...	X_j	...	X_p
1							1						
2							2						
⋮							⋮						
s				X_{sjk}			s				X_{sjr}		
⋮							⋮						
n_k							n_r						

To constitute a multiple multivariate data in a manner required for the applications of the techniques used in the study, it is important that the order of the observations is maintained for all years. It means that, for example, the first market in Year 1 is the same as that in the second year and in all other years; and Market s is the same for all years; the actual observations on them would obviously vary due to yearly differences in prices even in the same market and the same item. It therefore requires that in Table 5, $n_1 = n_2 = \cdots = n_k = \cdots = n_r$. Let the equal data points in each year be denoted by n. So the same set of markets is studied each year, and on the same items. Thus, in the layout, the general market, s, is repeated in a total of nr data points as follows:

s	in	Year 1
$s+n$	in	Year 2
$s+2n$	in	Year 3
$\vdots$	$\vdots$	$\vdots$
$s+(t-1)n$	in	Year t
$\vdots$	$\vdots$	$\vdots$
$s+(r-1)n$	in	Year r

The general observation X_{ijt} $(i = 1, 2, \ldots, nr)$ in Table **5** is therefore characterised as

$$X_{s+(t-1)n,j,t}; \quad s = 1, 2, \ldots, n; \; t = 1, 2, \ldots, r; \; j = 1, 2, \ldots, p. \qquad (1.1)$$

This characterisation is important as, in the data processing in MATLAB used in most parts of the work, this specification needs to be updated each time there is a change in the data points n. By varying the values of j and t, we can deduce that a single observation π_s on Market s in Table 5 therefore is a matrix with Equation (1.1) as the general element. It is given as

$$\pi_s' = \begin{pmatrix} X_{s11} & X_{s+n,1,2} & \cdots & X_{s+(t-1)n,1,t} & \cdots & X_{s+(r-1)n,1,t} \\ X_{s21} & X_{s+n,2,2} & & X_{s+(t-1)n,2,t} & & X_{s+(r-1)n,2,t} \\ X_{s31} & X_{s+n,3,2} & & X_{s+(t-1)n,3,t} & & X_{s+(r-1)n,3,t} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ X_{sj1} & X_{s+n,j,2} & \cdots & X_{s+(t-1)n,j,t} & \cdots & X_{s+(r-1)n,j,r} \\ \vdots & \vdots & & \vdots & \vdots & \vdots \\ X_{sp1} & X_{s+n,p,2} & \cdots & X_{s+(t-1)n,p,t} & \cdots & X_{s+(r-1)n,p,r} \end{pmatrix}. \tag{1.2}$$

Therefore, an observation in the data is a $p \times r$ matrix. Apart from this general observation, in Chapter Four, it will be found repeatedly important to denote the data in the kth year as $\mathbf{X}_k = (\mathbf{X}_{1k}, \mathbf{X}_{2k}, \ldots, \mathbf{X}_{pk})$ on p variables, which is of dimension $n \times p$.

The perspective taken in the study of such data problem has the benefit of extended studies on observations covered in the data. In most multivariate data that are designed for outlier detection (e.g., the Iris Setosa data in Johnson and Wichern, 2007), there is no presence of a variable that may be regarded as the 'response' variable. In the data problem under study, the 'level of prices' as a variable does not originally exist. The study will determine the extent of extremeness of the observations using the prices on the 19 items, and based on their projections on the time-dependent displaying components, at least for some key observations. This will enable a classification of these observations into some k classes; for example, 'high-priced', 'moderate-priced' and 'low-priced' markets. These classifications of observations into a price level class will constitute the 'new' response variable, Y. Thus, the values of Y has $k\ (=3)$ levels.

In Table 6 (item No. 3), we intend to determine latent factors which, in the opinion of traders, influence price levels of items. These factors would be determined from a set of structured questionnaire that would be administered to traders in selected markets from all categorisations of price levels that will be determined from the actual prices of items from the markets. We would examine the significance of these factors in the determination of price levels. This would also serve as a validation mechanism for the result from procedures that make use of data on the other observable variables in the table. The indicators of the factors cover twelve broad areas. Some of these relate to market features that could serve

Table 6: Variables Considered to Influence Price of Commodities

No.	Variable	Type	Level	Description
1	Type of market	Qualitative	2	Rural/Urban
2	Location of market	Qualitative		Name of Town/Village
3	Factors of choice of market	Quantitative	< 5	To be determined from analysis
4	Market days	Quantitative		Number of days of trading in the week by trader
5	Region	Qualitative	10	All regions of Ghana

as separate variables in Table 6. Since such market features could be many, it is more convenient to use them as indicators of main latent factors to be determined. Others also cover the usual economic considerations for fixing prices. The coverage of the indicators are as listed.

1 Proximity to source of production

2 Competition within market

3 Cost of other supporting market services (e.g. rent)

4 Profit margin

5 Sales orientation

6 Market mix (e.g. distribution/promotion strategies)

7 Price regulation by market authorities

8 Cost of procuring items from source

9 Customer consideration

10 Market share drive

11 Prevailing market conditions

12 Level of perishableness of item

As can be seen in Table 16 (in Chapter Four), fifty-three indicators have been created from these twelve broad areas guided by obvious prevailing market forces and other standard conditions. Responses on these indicators will generate the data on five-point Likert scale for determining the latent factors that are influential in fixing prices in the markets. The factors will then serve as one set of variables for determining the price levels. In Chapter Four, we will determine further transformation of these factors that will be appropriate to serve as a variable for the development of the intended model.

Model in Terms of the Two Sets of Variables

The design described above shows that there are two sources of data on two separate sets of variables that are required for the study: (1) the actual multiple multivariate data on prices of items covering some selected years; and (2) the single multivariate data generated on Likert scale on indicators of latent factors that influence the actual prices. The two divergent sources of data shows that the methods applied in each case must also be different in applications. However, in a single study, these methods must be brought together in the end. This is the ultimate task of the study. In the first case, the methods that will be employed will enable us to identify categorisations in the markets as high priced, moderate priced and low priced; the multivariate data is thus reduced to a univariate data with three categories. In the second case, the factors identified will be reduction of high dimensional dataset to a smaller dimension of factors. The two sets of variables may be then be reconciled by allowing the factors to explain the categorised observations of price levels. In factor extraction, a parsimonious factor model is usually the guiding principle. However, since these factors are intended to explain some 'created' dependent variable, we will rather be interested in a significant factor model, not necessarily parsimonious. This step will actually satisfy a condition for such a model. It is a model that brings together the result of the examination of the structural forms of two different datasets. In many studies (Li, Feng, Lu, & Song, 2015; DiStefano, Leighton, Ene, & Monrad, 2015; Ledermann, & Macho, 2015), extraction of factors is usually considered as structural modelling. This model is then linked to another structural form derived as the price levels of the markets.

In this study, the structural equation modeling (SEM) technique is used to overcome the above mentioned problems. In the first stage, a measurement model relates observed indicators to latent attributes. The second stage assumes a structural model to explore the effects of latent attributes (firm-specific factors) on three debt ratios. Specifically, the measurement model is defined as follows:

$$\mathbf{y} = \mathbf{\Lambda F} + \boldsymbol{\varepsilon}, \qquad (1.3)$$

where $\mathbf{y}$ is $q \times 1$ vector of observable indicators; $\mathbf{F}$ is $m \times 1$ vector of latent market forces, which is assumed to follow a multivariate normal distribution $N(\mathbf{0}, \mathbf{\Phi})$; $\mathbf{\Lambda}$ is a $q \times m$ matrix of factor loadings; and $\boldsymbol{\varepsilon}$ is $q \times 1$ vector of measurement errors, which is independent of $\mathbf{F}$, and distributed as $N(\mathbf{0}, \mathbf{\Psi})$ with a diagonal covariance matrix, $\mathbf{\Psi}$. As shown earlier (see page 21), we have identified twelve item pricing factors (attributes) with 53 indicators. Thus, $\mathbf{y}$ is known to be 53×1 vector of indicators. The matrix $\mathbf{\Lambda}$ and the vector $\mathbf{F}$ are yet to be determined in Chapter Four. Through the measurement model in Equation (1.3), SEM simultaneously accommodates highly correlated explanatory indicator variables without encountering multicollinearity, and measures latent attributes through indicators with different weights, reflecting different contributions of indicators in measuring the latent market forces. This is an attempt to reduce to the barest minimum several indicators by incorporating different market features and other characteristics, along with their importance, thereby reflecting the forces that determine prices more accurately and completely. Basically, the structural model is defined as

$$\eta = \gamma \mathbf{M} + \Gamma \mathbf{F} + \delta \qquad (1.4)$$

where η is a $p\times 1$ vector of endogenous variables; $\mathbf{M}$ is $n\times 1$ vector of covariates; γ and Γ are matrices of regression coefficients; and δ is a vector of error terms, which is independent of $\mathbf{F}$, and distributed as $N(\mathbf{0}, \mathbf{\Psi}_{\delta})$ with a diagonal covariance matrix, $\mathbf{\Psi}_{\delta}$.

The data problem suggests that the intended model would rather be an adaptation of Equation (1.4). In Chapter Three, we will further explain how the adaptation of Equation (1.4) will be carried out. Then in Chapter Four, further description of the design for obtaining the data for the CFA will be given. In that chapter, it will be realised in particular that the layout of the data from which the matrix of factor solution obtained is not of the same dimension as the price data, especially for the rows that represent the markets.

Displaying Components

In this study, we will refer to the main techniques of Principal Components Analysis and Outlier Displaying Components (Nkansah & Gordor, 2013; Gordor & Fieller, 1994) as a class of displaying components. This is as a result of the manner in which they are utilised in the study of the problem. These techniques are originally designed for single multivariate data in which the observation in Equation (1.2) is just a column vector. We intend to extend these techniques to accommodate multiple multivariate data problem.

The original Principal Component (PC) is given by

$$y_i = \sum_{j=1}^{p} a_{ij} x_j ; \quad i = 1, 2, \ldots, p. \qquad (1.5)$$

Traditionally, this transformation of original variables X_j, $j = 1, 2, \ldots, p$, is used as a measure of some indices that are orthogonal to each other. It may also be used as a dimensionality-reduction technique. We will benefit from these two properties of y_i. As the components are determined by the weights, a_{ij}, which are influenced by the standard deviations of X_j, it will be necessary to investigate the type of weights that make the components meaningful in the context of the study. Secondly, since the data is time-dependent, we will use the component in Equation (1.5) as time-dependent component of the form

$$y_i(t) = \sum_{j=1}^{p} a_{itj} x_{jt} ; \quad i = 1, 2, \ldots, p \qquad (1.6)$$

for all $t = 1, 2, \ldots, r$, for each component i, $i = 1, 2, \ldots, m$ $(m < p)$. Such a treatment will enable a time-dependent display of the values of the indices given by the components. The outlying values of the indices will provide indication of suspect market with extreme price levels across the various years.

The results of the PC extraction in addition to descriptive statistics of the data will provide initial candidates for assessment of level of prices. Based on these results, the Outlier Displaying Component (ODC) will be used to determine the actual extent of outlyingness of the observations. The original ODC (Gordor & Fieller, 1994) which is given by

$$\boldsymbol{\beta}_{\pi_v} = \mathbf{S}^{-1}(\mathbf{x}_{\pi_v} - \overline{\mathbf{x}}) \tag{1.7}$$

is usually criticised for being affected by the very extreme observation it seeks to highlight. This is because the overall mean, $\overline{\mathbf{x}}$, and the variance-covariance matrix, $\mathbf{S}$, are both influenced by the presence of the suspect extreme observation, $\mathbf{x}_{\pi_j}$. Thankfully, Nkansah and Gordor (2013) have provided a modification of this projection vector of the form

$$\boldsymbol{\beta}_{(\pi_v)} = \mathbf{S}^{-1}_{(\pi_v)}(\mathbf{x}_{\pi_v} - \overline{\mathbf{x}}_{(\pi_v)}) \tag{1.8}$$

which excludes the suspect outlying observation in the computation of $\overline{\mathbf{x}}$ and $\mathbf{S}$.

The use of this modified ODC (M-1ODC) may not help much in obtaining accurate results as the data covers different years with inherent large variations. This is evident from Table 1. To derive the full benefit of the modification in Equation (1.8), we will explore the use of a pooled variance-covariance matrix, which will be explained in the methodology. As is the case of the principal component, the second challenge with the displaying component and its modification is that it is designed for single multivariate data. As has been pointed out in Equation (1.2), the single observation is now a matrix. Thus, the projection vector $\boldsymbol{\beta}_{(\pi_v)}$ now becomes a matrix of dimension $p \times r$. Projections based on this vector will enable us obtain a simultaneous display of data for all years in order to determine the exact year in which an observation (or market) is extreme in price levels. This will be a novelty application of the ODC and its modification.

Significance of the Study

The standard of living of any people is based on the prices of commodities and services that they consume on regular basis. It will therefore be relevant in this study to determine the important items of consumption that should make up the food basket. By this, we would also be able to identify sets of items that may be regarded as homogeneous with respect to what constitute the food baskets of most consumers. We would also determine where these items abound and at relatively cheap prices and vice versa. This attempt would inform the basis for advocating for the creation of markets that may be designated for specific items. It is anticipated that such an attempt would lead to reduced prices and become a poverty-reduction strategy.

It is anticipated that the result of this project would serve as an academic corroboration to the work of national agencies that have the mandate to provide statistics on price levels of local items in Ghana.

The techniques that will be employed in the study are not typically known for such data problem. The study will therefore serve as an important demonstration for the extended use of the techniques. It will hopefully highlight some further features of the techniques for future studies.

Organisation of the Study

The study is organised into five chapters. Chapter One is the introductory chapter. It discusses the background, statement of the problem, the objectives, and the layout of the data and research design.

In Chapter Two, we consider the relevant literature on studies on prices of food items around the world, and applications of the intended techniques on the subject of market prices.

Chapter Three reviews the relevant theory of the techniques of displaying components and structural equation modelling and adaptions to this study. Displaying Components that have been discussed in that chapter are the Principal Components Analysis and the Outlier Displaying Component and their extensions as time-dependent components.

Chapter Four contains the applications of the techniques to the data problem for the study. It demonstrates the analysis of the data and identifies the main results.

The last chapter is Chapter Five. It presents the general summary of the entire work, and presents the conclusions and relevant recommendations.

Chapter Summary

In this chapter, we have presented the background of the study in which four main issues have been brought to the fore regarding food prices. These are common factors that influence price determination, effect of movements in item prices in recent past in the country and around the globe, the nature of the data at hand and how it prescribes the techniques for the analysis. Among the prevailing factors that influence prices are the poor nature of roads to production centres which seriously contribute to the high cost of transporting food commodities. Others include the proximity to production centres and prevailing market conditions. It is observed that item prices have basically been on the rise over the few years prior to 2008. The consequences of the price rise have been felt by almost every class of the society as individuals and national incomes have been seriously eroded. The most vulnerable group is the hardest hit of the negative impact. The situation shows that a persistent high food prices deepens poverty. The background also identifies effort at price monitoring by government agencies and points out that academic corroboration in this regard has been scanty. It also highlights the need for the right techniques to be employed in the analysis of market statistics as the results could be affected by the technique if it not specifically designed for the study. It also identifies the composition of the markets in such studies as crucial in arriving at the kind of results obtained.

Consequently, the statement of the problems specifies three key area of motivation for the study. These are the appropriateness of the techniques that have been used in this area of study and the lack of application of suitable methods for the study of the subject. It also points out the complex nature of the data problem usually involved in this type of studies. The last concern is the need for academic corroboration of studies into market statistics as it has the potential of generating market information that could lead to efficient market organisation as a national strategy for poverty-reduction. Subsequently, three main objectives have been specified that broadly assess the price levels of local food items in Ghana.

Considerable space has been given to the description of the data problem and the research design in the chapter. It explains the nature of the data problem with a layout of a multiple multivariate dataset. As the techniques that are employed in this research are originally not designed for the purpose, an attempt has been made to introduce the techniques and justify their use and possible extensions. These are techniques that are of the class of time-dependent displaying components and structural equation modelling, as the main problem covers a period of time and may be explained from two different sets of generating variables. The significance of this study has clearly been pointed out and the outline of the entire thesis has been presented.

CHAPTER TWO

LITERATURE REVIEW

Introduction

This chapter reviews works related to dimension reduction techniques applied to multi-dimensional datasets. This is necessary because an aspect of this study relies on data reduction techniques. The chapter also examines literature on market information and price variations. Global changes and its effects on food sustainability are also discussed. This has direct impact on prices of food commodities and the total effect on food markets all over the world. Finally, the chapter reviews applications of displaying components and structural equation modelling (SEM), which are the two broad techniques that have been employed in this study.

Studies on Food Prices in Ghana

In Ghana food production across the country is not uniform. For instance, Northern Ghana is relatively poor, isolated and dry when compared to the rapidly developing and urbanizing south. This has a lot of effects on food production and hence prices of food commodities across the market centres all over the country. A specific study has been conducted on the production of maize, rice and soyabean in Ghana (Gage et al., 2012). Each commodity market was assessed independently in the country to evaluate current market demand; current supply and likely future market demand; and potential supply to meet the shifting market. They reported that the total quantity of maize produced and marketed annually in Ghana reported by maize market stakeholders is about one million tons. They also reported that a significant volume of maize produced remains within the producing households as their primary staple food. However, industrial maize buyers estimate that processing and utilisation of maize represents about 20 to 25 percent of the total maize marketed. They also revealed that most of the maize produced in Ghana is produced in the northern part of the country. White maize is the most common variety of maize produced in Ghana. Gage et al. (2012), also reported that maize and soyabeans are used significantly in the poultry feed industry. However, it was discovered that Ghanaian broiler production is about 50% more expensive than imported ready-to-cook chicken. This, the local poultry producers attribute to the high cost of additives to the maize and the soyabean for proper feeding of the fowls. In addition, they observed that the market for rice in Ghana as in other countries is highly segmented. One of the most important opportunities identified as viable is therefore the substitution for imported non-aromatic rice, approaching it in phases and targeting specific markets segments. They also contended that there is an expected high yield from Southern Volta area which will satisfy this specific import substitution opportunity, hence as Ghanaian incomes rise and more consumers shift into this market segment, demand will grow, but production around urban centres in the south is likely to keep pace.

According to Cudjoe, Breisinger and Diao (2008), Ghana is largely self-sufficient in the major staples except for rice and wheat. Changes in the prices of these staples therefore may primarily be driven by domestic market forces. However, substitution in consumption between rice and local staples such as maize, cassava, and yam may cause domestic prices for non-tradable products to rise when the world prices for imported food products increase. They argued that prices for all staple products have generally moved together in recent history, with a short exceptional period in 2006, particularly for maize. Furthermore, after the decline in price of maize in 2006, a year in which maize and rice prices moved in opposite directions, the maize price started to increase again in 2007 and 2008, moving in the same direction as rice prices. Prices for cassava and sorghum follow a similar trend to that of rice prices in late 2007 and 2008, while yam prices started to increase only in 2008.

Perhaps, a study that comes closest to the attempt made in this work is that of Seglah (2014) that simultaneously examines the price levels of major local food items in almost all major markets in Ghana. It made use of only a single year of 2008 and involved one hundred markets across all regions of the country. It identified Tepa market, in the Ashanti Region, as the lowest-priced market along a certain component that was determined in the study. However, markets that are predominantly located in the Northern, Upper East and Upper West Regions generally were found to be high priced. The interest in these markets in this study is therefore very high. In that study by Seglah, two important dimensions were obtained for determining the levels of prices using Principal Components and Cluster Analysis. The first is a ‘‘weighted sum of all the food items’’ and the second is what was referred to as the ‘‘weighted sum of only food items that are considered as the main constituent of a typical local diet’’. That study is based on fifteen food items. Clearly, this study is quite restricted in coverage and methodology. This is what makes this study relevant in order to corroborate the findings of the study to provide more generalised conclusions.

Sustainability of Food Commodities

There are several categories of food items, and each food item has a period of time that it could stand a test of time. Traders seem to be careful when dealing with food items that are easily perishable. The easily perishable once are not purchased from a long distance for onward sales. Traders of perishable commodities are likely to price such commodities quite different from the other commodities in other to avoid huge losses. It is also perceived that when such commodities are not sold quickly, they lose their freshness and therefore affect its pricing. A study conducted by Konuk (2015) considered the effect of price consciousness and sale proneness on purchase intention towards Expiration Date-Based Priced (EDBP) perishable foods. EDBP is defined as a pricing tactic in which a retailer charges different prices for the same perishable products, according to their respective expiration dates. Structural equation modelling was used in order to test the proposed hypothesis. They sought to examine the relationships that exist between price consciousness, sale proneness and purchase intentions in the context of EDBP. The results of a structural model reveal positive relationship between price consciousness and sale proneness. The finding also confirms the effect of price consciousness on purchase intentions toward EDBP perishable foods. Similar studies broadly analysed the effects of promotions, for example, price discounts, on consumer behaviour,. These are frequently used forms of sale promotion in order to increase store traffic and enhance purchase (Chen, Monroe & Lou, 1998; Gilbert & Jackaria, 2002; Aggarwal & Vaidynathan, 2003; Chung & Li, 2013). Retailers often consider EDBP an effective revenue management tool that encourage the purchase and reduce waste by adjusting the sale price of perishable foods as the expiration date approaches (Chun, 2003; Warde, 1997).

Global Changes and Market Impacts

Demand and supply are likely to be affected by global changes and this in effect is likely to affect prices of food commodities. For instance, if commodities are in short supply, it is likely that prices of food commodities will shoot up all things been equal. Large lags between the production decision, completion and sale output, any uncertainty during production such as bad weather, disease or financial crisis can affect the prices within the food market systematically. Huge spikes in world food prices in 2006 – 2008, and recent food crises have triggered a lot of research in food production and prices. Many institutes such as FAO (2008), OECD (2008), and the World Bank (Mitchell, 2008) have published numerous research papers in which an attempt to provide explanation to the problem of price explosion is made. On the other hand, scholars discussed supply events, which could have been responsible, such as weather effects, reduced stocks or changes in impute prices (Konuk, 2015). Other factors such as rapid urbanization and fast income growth in transition countries, especially China and India could have increased the demand for agricultural products (Von Braun, 2007). A study by Meyer and Yu (2013) focused on world's wheat and corn prices. They found that agricultural production involves a lot of uncertainties, comprising of natural risks and market risks, resulting from time lags between planning, realisation and sale of outputs. It was also discovered that uncertainty factors of wheat have significant impacts on both wheat and corn prices, while that of corn is not significant either for wheat or corn prices. Their study also suggested a policy implication that farmers should be organised by themselves or by the government to coordinate production so as to stabilise the market price.

According to Alem and Soedermbom (2012), consumption pattern is one of the most important drivers of the development pattern of industrialised world. One of the main factors explaining food consumption pattern is the level of disposable income. However, consumption pattern changes not only measure in the amount of calories consumed with the rising income, but also the share of animal products in overall diets. Recently, soaring food prices have become a major concern among policy makers. In the instance, the prices for cereals, cooking oils and sugar increased most, whiles increase of meat prices was more moderate (Alem & Soedermbom, 2012). Global supply and demand imbalances in agricultural commodity markets appear to have been a main driving factor for this recent increase. Unfavourable weather conditions in important producing countries and growing world population are main drive of food prices. Other factors driving food prices are high energy prices and the expansion of bio-fuel production (OECD, 2012). Imad, Abdul and Safar (2014) in their paper on 'Effects of high food on consumption pattern of Saudi consumers: A case study of AL Riyady City', found that there is a trend of world food price increase due to drought in major producing areas and changing of consumption pattern especially in emerging developing economies. This has impact on consumers' demand, choices and welfare. The study of their result also showed that the consumption quantities of major food commodities decrease due to high prices, however, expenditure increases lead to erosion of some consumers' savings. They concluded by suggesting governments' intervention through food policy to mitigate the effects of food price volatility.

The impact of higher food prices affects every fragment of the society: poor households, middle and upper class households, profit and non-profit organizations and governments. The most severe is the negative impact on the most vulnerable group – poor households, women and children. Efforts towards eradicating poverty will be significantly affected by persistent high food prices as this situation will push more people into poverty (Scott-Joseph, 2009; Belemmare, 2015).

Differential vulnerabilities exist on the scales of social, economic and environmental bases. These spatial differentials will lead to variations in climatic conditions per region. Adger (1999) argues that the impacts of extreme climate events are the principal climate phenomena which enhance vulnerability. The consensus of scientific opinion is that countries in temperate and polar regions will enjoy increased agricultural production, while countries in tropical and sub-tropical regions are likely to suffer agricultural losses. The FAO projects that the impact of climate changes on global crop production will be slight up to 2030. After that year, however, widespread decline in the extent and potential productivity of cropland could occur, with some of the severest impacts likely to be felt in the currently food-insecure areas of sub-Saharan Africa, which has the least ability to adapt to climate change or to compensate through greater food imports (Fischer, Shah, Tubiello & Velhuizen, 2005; Timmer, 1989).

Competition Policy

One of the main indicators of the level of competition in an industry or market is the presence or absence of dominant firms. It is believed that price control always reduces, unbalances, distorts, and discoordinates production. Price control becomes progressively harmful with every passage of time. Even a fixed price or price relationship that may be right or reasonable on the day it is set can become increasingly unreasonable. According to Vickers and Hay (1987), market dominance defines the power a single firm or group has over the supply of goods and services in one or more markets. This analysis shows how a profit-maximising firm with market power will restrict the amount of output produced in order to be able to charge a price above marginal cost. As a result of this there is a distortion in the resource allocation and welfare loss which is usually measured in terms of consumer and producer surplus (Vickers & Hay, 1987; Ferris, 1999).

Firms can only exercise market power once they have acquired it. It is therefore important to look at how this market power is achieved in the first place. There are a variety of ways to acquire market power. For instance, market power can be granted by a public authority, as is common for firms in the utility industries or those with natural monopolies. Firms such as these are often in public ownership and generally operate under economic, financial and other criteria which are not as stringent as that of private firms. Market power can also be obtained through collusion, looked at as multi-firm dominance. This dominance can be gained through the explicit or implicit co-operation between firms and more specifically through the co-ordination of their strategies and the free flow of information between these firms. Co-operation between firms is generally deemed to be acceptable if it is for the sole purpose of research and development, although firms often use issues of risk reduction and forward planning as other acceptable motives for collusion. Another way of obtaining market power is by predatory behaviour. Predatory behaviour is often carried out by firms already in the market with the aim of driving existing competitors from the market, but more importantly of deterring potential rivals from entering the market (Vickers & Hay, 1987). Apart from market dominance, the extent to which there is concentration in the market is significant for competition analysis. They will either co-ordinate their activities closely so as to form a virtual monopoly, compete fiercely or fluctuate somewhere in the middle. It is often argued that depending on the behaviour of oligopolists (a specific marketing practice in which the individual relates to each other in different ways), their combined market share can simply become a diluted version of the dominance that a monopolist exerts. Although the calculation of concentration has had limited success in determining actual profit ratios, it still remains useful as a means of conveying the main shape of an industry (Shepherd, 1997; Chabane, 2002; Elis, 1992).

Overly high food prices for poor populations are often the result of artificial price distortion. Traders influence competition and therefore price by controlling the supply entering the markets and the number of traders allowed to sell. This control varies from market to market. This means that traders have to control access to the market. Exerting such control is possible if the authorities and local government recognize the power of trader associations to control market spaces and if non-members are not able to sell in other places. For this reason, associations actively lobby traditional authorities, some of whom own the land in many of the markets and act as intermediaries in disputes (Lyon, 2003).

Market Information

From the studies of Robbins and Ferris (2000) in Uganda, market information services (MIS) should be designed to benefit farmers, traders and consumers. The services are tested to seek gains in farmer sales prices and improved prices for collectively sold produce. The gains have been achieved through the local MIS. Informal survey data from farmers in Rakai district, claim to have received between 5 to 15 percent higher returns on their sales when they are able to negotiate on known market prices, compared with farmers who simply accept prices they are offered by traders. Similarly, farmers associations in Jinja use the local marketing agent as a link to markets and has proved successful for farmers in bulking for higher value sales to larger traders. The local agribusiness centre forecast crop sales prices based on the market data, and Non-Government Organizations in northern Uganda also use the trend data to support a credit and storage scheme. However, beneficiaries are mainly targeted at the small-scale producers, as these groups are believed to be most vulnerable to situations where market information is unavailable or asymmetric.

Robbins and Ferris (2000) note that for a very long time, farmers in Africa have had to make decisions on what crops to plant, when to plant, where they will sell their produce and at what price. During the 1960-70s production was supported by Governments, who operated commodity marketing boards to purchase major export and staple food crops. When commodity boards were in operation, the focus for market information services in African countries was to:

- Advise government on marketing policy
- Set intervention prices
- Organize marketing training for marketing authorities and cooperatives
- Document market prices as part of the Government policy analysis process.

This highly interventionist system was good for farmers, as risk was shared by the Government and farmers were able to plan production based on a known buyer price. Similar agricultural support programmes are still practiced in the major OECD nations, who currently provide approximately one billion dollars a day to support the agricultural sector. These subsidies are considered an effective use of resources, as it allows the greater part of the workforce to be employed in more remunerative industrial and service activities. Unfortunately for farmers in Africa, the commodity boards were unable to adapt to changing times and industrialization failed to occur.

Market information is also advocated by Shepherd (1997, 2001) who recommends formats for the information packaging and presentation to farmers. Among these are the use of local FM radio stations in local languages at times convenient to most farmers. The aim of the intervention is to improve the marketing skills of farmers for better prices.

According to Robbins, Ferris, and Muganga, (2000), public provision of market information aims at avoiding asymmetry of information in the marketplace towards one group of actors. The rationale is that more equal access to market information will lead to greater uniformity in prices of a given commodity at the same time across the country and this transparency will enhance trading efficiency and assist in reducing transaction costs. In imperfect markets, traders make exorbitant profits by buying at a low price on one market and simultaneously sell on another market at a high price. This encourages arbitrage, a process of exchange of commodities with the objective of taking advantage of price differences that exceed transaction costs. Arbitrage is typical of three main situations: (1) markets information is not accessible; (2) markets are highly colluded; or (3) high transaction costs are caused by problems in the supply chain; as a result of too many traders in the marketing chain, or specific traders that can take advantage of market inefficiencies.

Robbins and Ferris (1999) report further market characteristics in Uganda. These are long chains of transactions between farm gate and consumers, lack of competitiveness between traders, collusion at all levels of trading and poor access to appropriate market information. As a result of these predicament, it is reported (Ferris, Legg, Bua, Agona, & Whyte, 2000) that for small-scale farmers, the provision of market information is the second highest priority in their efforts to gain access to better prices and markets after roads.

Price Variations

Prices of similar food commodities vary from market to market in Ghana and the world in general. This is not limited to only few commodities but among almost all category of food commodities. Notable among these food commodities are cereals, roots and tubers, vegetables, pulses, fish, fruits and oil. Apart from variation in food prices in various markets, the rate of increase in prices of these commodities also varies. According to Claro and Monteiro

(2010), fruit and vegetable participation in total food purchases increased as the price of these foods decreased, or as income increased. A 1% decrease in the price of fruit and vegetables would increase their participation by 0.79%, whereas a 1% increase in family income would increase participation by 0.27%. The effect of income tended to be smaller among higher income strata. Interestingly, Zimmerman (1999) asserted that as income increases, the percentage spent on food commodities and housing decreases, the percentages on clothing and household expenses remain about constant and the percentage on education, health and recreation increase.

Roache (2010) in his study on 'What Explains the Rise in Food Price Volatility?' elucidated that 'the macroeconomic effects of large food price swings can be broad and far-reaching, including the balance of payments of importers and exporters, budgets, inflation, and poverty. For market participants and policymakers, managing low frequency volatility may be more challenging as uncertainty regarding its persistence is likely to be higher. It finds that low frequency volatility is positively correlated across different commodities, suggesting an important role for common factors. It was pointed out that the timing of price collection was not crucial in Austria although it might play a very important role in the Consumer Price Index compilation in countries with high inflation. Other issues he discussed included the training of price collectors, the use of broad specifications and the greater importance of product variety over using a larger number of outlets and covering more regions. Finally, Roache (2010) affirmed that composition of diets can have implications for the magnitude and distribution of rising staple food prices. Households in countries where the diet is largely composed of non-tradable food staples tend to be less affected, to the extent that the prices of non-tradables do not trail the prices of tradables. For instance, in their simulations, Ghanaian households appear to be relatively insulated from swings in international food markets, because a large share of their diet is based on non-tradable staples such as cassava and sorghum. The price of these non-tradables increase, as demand for them also increases, given sharp rise to food prices. The fact that the poor are hit the hardest by rising food prices in both urban and rural areas is clearly a cause for concern. The erosion of real income in poor households not only harms their current ability to cover basic needs but has the potential to do so for some time to come, thus, diminishing their prospects of escaping poverty. Poor households may be forced to cope with the added stress of high food prices by depleting their asset base, reducing the number or variety of meals they consume.

Inflation, tariffs and other market factors can also be determinants of price variations in food commodities. Prices of food commodities on world markets, adjusted for inflation, declined substantially from the early 1960s to the early 2000s, when they reached a historic low. They increased slowly from 2003 to 2006 and then surged upwards from 2006 to the middle of 2008 before declining in the second half of that year. The sudden increase took many by surprise, and led to increase concern over ability of the world food economy to adequately feed billions of people, now and in the future. Although various observers attach deferring degrees of importance to assorted factors, there is a relatively strong consensus that multiple factors had a role in the price increases that began in 2003 (Gilbert & Morgan, 2010).

Gilbert and Morgan (2010) asserted that it is possible for average prices to change without any change in variability of domestic prices. One simple way this might happen would be if food-imported country were to impose a constant tariffs on import; the tariff would make food expensive, but most often it would have no effect on the variability of domestic prices. Prices exhibit variability for many reasons, but some price change may be largely predictable. The classic example of predictable changes in food prices is seasonality, whereby prices are lowest during and soon after harvest and highest immediately before harvest. They stressed that the simplest way to measure price volatility is the coefficient of variation. This makes it easy to compare, for example, domestic price volatility measured in different countries. They also indicated that the economic impacts of commodity prices are important because they affect the level of per capita income, which ultimately is a key determinant of living standards for individuals and families. High international prices of food commodities benefit countries that export those products, while low prices benefit importing countries.

Barnett (2008) reported that the world price of wheat rose to $400 a ton, the highest level on record in the world. That's twice the inflation-adjusted average price of wheat for the past 25 years, and twice as high as it was in May, 2008. In 2007, the price of corn also hit a record of $175, a ton more than 50 percent above the average for the previous year. Other staples, such as rice, have also hit records, ricocheting off the price of other staples as farmers switch land to high-paying commodities from other uses. Interestingly, prices reached record high during a time of equally record abundance: cereal crop yields were higher than ever (2007), an outcome the economist attributes to two trends: Growing demand for meat in (increasingly wealthy) China and India (livestock production requires more crops for feed); and skyrocketing demand for corn-based ethanol. As farmers have shifted croplands to feed America's growing demand for ethanol, the cost of other crops went up, and stockpiles went down. In affluent countries like the US, the ready abundance of cheap, highly processed carbohydrates have pushed obesity and diabetes to epidemic proportions'.

The standard of living varies from country to country. This paves way for a high variation in prices of food commodities, especially, prices of cereals, since cereal constitutes about the highest food consumption in the whole world wide. The study on the topic 'Understanding price variation in Agricultural commodities in India: Minimum support Price (MSP), Government Procurement and Agriculture Markets' was a research conducted by Chatterjee and Kapur (2016). The study specifically focused among others how regulation and physical location of wholesale agriculture commodity markets affects price variation across space. They examined the relative contributions of different factors in explaining this price variations. The major commodities studied were rice and wheat which together accounted for about three-fourths food grain output in India. They found large variances in prices of agricultural food commodities across the country. Real wholesale prices across wholesale markets have an average standard deviation of 0.18, much higher than United State (US) and also many developing countries. They were of the opinion that the large variation in prices is important to understand because it implies not only that consumers pay different prices at different locations for the same product but producers get different prices at different locations as well.

Price variation of food commodities is also highly associated with Ghanaian markets. The 2007/2008 fiscal year in Ghana observed a high rate of food price increases following the global food crises. For instance, Wodon (2008) indicated that the prices of cereals increased from 20 percent to 30 percent between 2007 and 2008, food component of the consumer price index also rose from 193.9 to 246.7 indicating a 27% food inflation within the same period. (See also GSS, 2009). The food crises put extra burden on consumers by reducing their real income of household expenditure on food in urban and rural Ghana (GSS, 2008).

Dimensionality Reduction Techniques

Multivariate data naturally is a complex one, making it difficult to extract information straight away from it. It is therefore found useful in the literature to use data reduction techniques to obtain information from this type of data. According to Shyam, Shanmugapriya and Kumaran (2016), principal component can be used for data mining or data reduction for low dimensional datasets. However, there can be breakdown of the usage of the technique if there is a very high-dimensional datasets. Generally, high-dimensional data has two main implications: Finding the relative contrast between similar and dissimilar points is difficult as the dimensionality of the data grows. Also grouping of different datasets with each other becomes extremely difficult. Similar study by Caprihan, Pearlson and Calhoun (2009), used principal component analysis on biological data to modify the principal component analysis to a discriminant principal component analysis. This was applied to diffusion tensor based fractional anisotropy images to distinguish age matched schizophrenia subject from health controls.

Applications of PCA in the literature (Armeanu & Lache, 2008; Gulumbe, Dikko & Suleiman, 2014) are quite overwhelming. It is applied from Economics to Medicine using data from Insurance Market to that of cholesterol levels of the human body.

As indicated in the introduction, methods such as the Principal Components Analysis are recommended as the preliminary methods in detecting price levels of markets (Barnett & Lewis, 1994). The use of the technique helps to observe sample values of the projection of observations on to the principal components of different order. In outlier detection in particular, the distinction in the relative utility of the first few, and the last few, components is basic to the methods in the literature on outlier detection. It has been remarked (Gnanadesikan & Kettenring, 1972; Gnanadesikan, 1977) that the first few principal components are sensitive to outliers inflating variance or covariances or correlations, whilst the last few are sensitive to outliers adding spurious dimensions to the data or obscuring singularities.

Schreiber, Nora, Stage, Barlow, and King (2006) reported on the results of structural equation modelling (SEM) and confirmatory factor analysis (CFA). They stressed that both SEM and factor analysis (both exploratory and confirmatory) are statistical techniques that can be used to reduce the number of observed variables into smaller number of latent variables by examining the covariation among the observed variables. It was reported that SEM in comparison with confirmatory factor analysis (CFA), extends the possibility of the relationships among the latent variables and encompasses two components: a measurement model and a structural model. The measurement model of SEM is the CFA which depicts the pattern of observed variables for those latent constructs in the hypothesized model. In addition, it is also used to examine the extent of interrelationships and covariation among the latent construct. As part of the process, factor loadings, unique variances, and modification indexes are estimated for one to derive the best indicators of latent variables prior to testing a structural model. They also observed that the use of single indicator to fully capture the complexities of such a construct as required in path analysis is impractical in the case of the SEM. In both SEM and CFA, sample size is an important issue because it relates to the stability of the estimates. Just like the PCA, SEM has also been applied widely in various studies. Other studies that use the concept of SEM include studies on the determinants of capital structure choice for Chinese listed companies (Li, Feng, Lu, & Song, 2015); predictors and outcomes related to school climate (DiStefano, Leighton, Ene & Monrad, 2015); and mediation models (Ledermann, & Macho, 2015).

The nature of the factor component extracted from a given dataset is of importance to this study as it anticipates some difficulty in component extraction using the type of data for our implementation. The use of a single indicator to capture a latent factor construct has been identified (Benyi, 2018) as potentially problematic. In a one-indicator factor, the factor is influenced in its formation by only one indicator. It is demonstrated that a one-indicator factor can only be acceptable if that indicator is almost independent of the other set of indicators. For consistency with the rationale of factor extraction, the one-indicator factor is only plausible if one can observe in the correlation matrix that the influencing indicator has generally low (or negative) correlations with all other indicator variables. It is further illustrated that the incidence of one-indicator factor usually arises as a result of an attempt to resolve the incidence of a contrasting factor by factor rotation.

Another class of problematic factors is found to be contrasting factors (Benyi, 2018) or bipolar factor (Russel, 2002). For this type of factors, they are identified by items with strong positive loadings whiles some others have strong negative loadings. A direction of influence of an indicator on the factor is specified by the common sign of the set of variables that have high loadings on the factor. Benyi has demonstrated that contrast factors could be problematic with respect to their plausibility and further analysis. It should therefore be avoided if rotation cannot be used to resolve it. It is cautioned that contrasting factors should be anticipated in datasets with prevalent negative correlations. In such datasets, rotation cannot resolve the contrast, and it is in the context of negative correlations that contrasting factors should be entertained. It has been demonstrated that the existence of contrast factors that constitute the main dimensions in the data does not give the data a good measure of factor-suitability. It is therefore admonished that contrast factors should require a good knowledge of the dimensionality of the dataset as these factors may not be plausible. It is demonstrated that the maximum likelihood method could be a preferred approach to obtaining a plausible solution that also resolves the incidence of contrasting factors.

The identification of such factors could only be based on a cut-off loading value which could be much smaller than 0.3. The use of a chosen cut-off value in factor identification is demonstrated to vary depending on the nature of the data, and hence, the correlation matrix. Usually, the cut-off value could range between 0.3 to 0.5 (Nkansah, 2018). Types of factors have the focus of quite a number of other studies (Fabrigar, Wegener, MacCallum, & Strahan, 1999; Comrey & Lee, 1992; Gorsuch, 1988).

Chapter Summary

The literature review contains dimensions along which prices are fixed for food items. These include the perishableness or sale proneness of the item, and promotions. Prices are also influenced by prevailing conditions of global changes and market impacts. Predominantly, these impacts are due to uncertainties during production and spikes in world food prices in the period of 2006 – 2008, rapid urbanisation and fast income growth with associated changes in consumption patterns, high energy prices and expansion of bio-fuel production.

The literature also contains ways in which prices have been controlled. These may be by the exercise of some authority gained by single or group of individual persons or companies, and by organised information systems that seek to create transparency about prevailing prices. This control system is found in both the industrial sector as well as the agrarian business. Artificial price distortions, particularly among the poor population, have been noted. The distortions are as a result of perturbed competition that controls supply and entry of traders into the market. It is found that for most of such control avenues, the effect is generally harmful. For a few other measures, it has helped to improve prices to the benefit of members of associations in some parts of the world.

Variations in prices of various types of items have been reported with cereals appearing to be the most affected, which has also gained considerable attention for research. It is also observed that the consumption of fruits and vegetables do change with changes in income. It stresses the need for the inclusion of a variety of products as a key component in market price assessment rather than the use of a large number of outlets and covering more regions. Movement in prices could also be influenced by the composition of diet, and prices are likely to be stable in communities whose share of diet is largely based on non-tradable staple foods, for example, cassava and sorghum. An important observation is that changes in average prices may not be connected to changes in variability of domestic prices. This may be as a result of imposition of constant tarrifs on some imported items that generally affects domestic prices. Variations have also been assigned to seasonality of the items. What is ironical in the literature is that prices have reached record high over the period of 2007/2008 during a time of equally record abundance. This suggests that variations are not likely to be attributable to shortage in production. This phenomenon is attributable to growing demand for meat in increasing-wealthy nations of China and India, and demand for corn-based ethanol in America.

It is further observed that swings in food prices have broad and far-reaching macro-economic effects. It is a common concern in the literature that rising food prices could be a potential for deepening poverty, as it is a key component for the determination of standard of living.

The review shows that Principal Component Analysis has enjoyed widespread applications on the subject. It has been applied from Economics to Medicine. The technique, however, is identified in various authority texts as being suitable for preliminary studies, rather than the main technique. Another major technique found in the studies is structural equation modelling (SEM). The studies that use SEM shows that the technique encompasses the use of Confirmatory Factor Analysis, Multiple Regression, Latent Class Cluster and Mediation Models. The literature shows that the use of factors could encounter difficulties depending on the type of the factors extracted. Two main types of factors that have been identified as problematic are the one-indicator factor and contrasting factors.

The review shows that a few studies have been conducted on food prices in Ghana, and have targeted few specific commodities. There is perhaps a single study in the entire literature that has taken the perspective of market price research that is similar to the attempt made in this research. However, this study in the literature is quite restricted in coverage and methodology.

CHAPTER THREE

RESEARCH METHODS

Introduction

In this chapter, we will examine the main methods employed in this study. The notation of the multiple multivariate data is reviewed. It would be relevant to also review the general orthogonal factor model and the basic conditions and assumptions regarding the use of the model. The method of confirmatory factor extraction will be examined. Component scores will also be discussed under conditions for generating these scores. This discussion will enable us determine time-dependent series of scores that can highlight extreme observations. Two broad methods underlie the implementations in the next chapter. These are: the time-dependent displaying components and structural equation modelling. The class of displaying components in this study includes the principal component and outlier displaying component. Since outlying observations in this study could be many, our use of displaying components is to obtain suspect outlying observations. As a result, only the technique of one-outlier displaying component will be discussed, since by highlighting a single extreme observation, other potentially extreme observations would lie in its neighbourhood. The review of these methods will also examine the development of extensions to the techniques that would enable the implementations in this study.

Notation

In the introductory chapter, we have already presented a notation of the multiple multivariate observation as used in this study. A further description of the entire data is provided in this section.

The general observation X_{ijs} $(i = 1, 2, \ldots, nr)$ that represent the price in Market s on the jth commodity in the tth year was characterised in Chapter One as

$$X_{s+(t-1)n,j,t}; \qquad s = 1, 2, \ldots, n;\ t = 1, 2, \ldots, r;\ j = 1, 2, \ldots, p. \tag{3.1}$$

The representation means that each observation is repeated r times on the same set of p items. By varying the values of j and t, the single observation π_s on Market s is a matrix with Equation (3.1) as the general element. It is given as

$$\pi_s' = \begin{pmatrix} X_{s11} & X_{s+n,1,2} & \cdots & X_{s+(t-1)n,1,t} & \cdots & X_{s+(r-1)n,1,t} \\ X_{s21} & X_{s+n,2,2} & & X_{s+(t-1)n,2,t} & & X_{s+(r-1)n,2,t} \\ X_{s31} & X_{s+n,3,2} & & X_{s+(t-1)n,3,t} & & X_{s+(r-1)n,3,t} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ X_{sj1} & X_{s+n,j,2} & \cdots & X_{s+(t-1)n,j,t} & \cdots & X_{s+(r-1)n,j,r} \\ \vdots & \vdots & & \vdots & \vdots & \vdots \\ X_{sp1} & X_{s+n,p,2} & \cdots & X_{s+(t-1)n,p,t} & \cdots & X_{s+(r-1)n,p,r} \end{pmatrix}. \tag{3.2}$$

Therefore, an observation in the data is a $p \times r$ matrix. In this matrix, the columns represent the n observations or markets (or individuals) each with p measurements for a fixed year. Alternatively, it may be seen as p variables (items) on which measurements are taken for n individuals. The matrix $\mathbf{X}_{s+(t-1)n,j,t}$ can be written simply as the column vector

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_{11} & \mathbf{X}_{21} & \cdots & \mathbf{X}_{s1} & \cdots & \mathbf{X}_{n1} \\ \hline \mathbf{X}_{12} & \mathbf{X}_{22} & \cdots & \mathbf{X}_{s2} & \cdots & \mathbf{X}_{n2} \\ \hline \vdots & \vdots & & \vdots & & \vdots \\ \hline \mathbf{X}_{1k} & \mathbf{X}_{2k} & \cdots & \mathbf{X}_{sk} & \cdots & \mathbf{X}_{nk} \\ \hline \vdots & \vdots & & \vdots & & \vdots \\ \hline \mathbf{X}_{1r} & \mathbf{X}_{2r} & \cdots & \mathbf{X}_{sr} & \cdots & \mathbf{X}_{nr} \end{pmatrix}'.$$

The above treatment considers the data as r points each having n observations and each on p dimensions. Geometrically, the representation may be constructed as data on n points in p-dimensional space with r replicates. In the above vector representation, there are r column stacked under each other in the form of partitioned sets of columns. The kth set of column $\mathbf{X}_k = (\mathbf{X}_{1k}, \mathbf{X}_{2k}, \ldots, \mathbf{X}_{sk}, \ldots, \mathbf{X}_{nk})'$ is of dimension $n \times p$ such that each of the n observations is assessed on p variables. Computations based on this representation is carried out by first defining a vector of n ones, $\mathbf{1}_n = (1, 1, \ldots, 1)'$. The components of the mean vector for the kth year is then obtained as

$$\bar{\mathbf{x}}_k = \frac{1}{n}\mathbf{X}'_k \mathbf{1}_n.$$

The general (or total) sample mean vector is thus obtained as

$$\bar{\mathbf{x}} = \frac{1}{nr}\mathbf{X}'\mathbf{1}_{nr}.$$

These quantities and notations will be used to compute a number of useful results including the sample mean vector, sample variance among others.

The Orthogonal Factor Model

Suppose that we have the observable random variable vector $\mathbf{X}=\left(X_1, X_2, \ldots, X_p\right)$ of p components such that $E(\mathbf{X})=\boldsymbol{\mu}$ and $\operatorname{cov}(\mathbf{X})=\boldsymbol{\Sigma}$. The factor model assumes that $\mathbf{X}$ is linearly dependent on a few $(m<p)$ unobservable dimensions represented by the random variables, $f_1, f_2, \ldots, f_m$ which are the common factors, and p additional sources of variation $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p$ which are the specific factors. The factor model is thus given as

$$x_i - \mu_i = \sum_{j=1}^{m} l_{ij} f_j + \varepsilon_i; \qquad i = 1, 2, \ldots, p \tag{3.3}$$

or

$$\mathbf{X} - \boldsymbol{\mu} = \mathbf{LF} + \boldsymbol{\varepsilon},$$

where l_{ij} is the loading of the ith variable on the jth factor. Thus, $\mathbf{L}$ is a $p \times m$ matrix of factor loadings.

Now by considering the covariance structure

$$(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})' = (\mathbf{LF}+\boldsymbol{\varepsilon})(\mathbf{LF}+\boldsymbol{\varepsilon})'$$
$$= \mathbf{LFF'L'} + \mathbf{LF}\boldsymbol{\varepsilon}' + \boldsymbol{\varepsilon}(\mathbf{LF})' + \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'$$

Taking expectation of both sides, we have

$$\operatorname{cov}(\mathbf{X}) = \mathbf{L}E(\mathbf{FF}')\mathbf{L}' + \mathbf{L}E(\mathbf{F}\boldsymbol{\varepsilon}') + E(\boldsymbol{\varepsilon}\mathbf{F}')\mathbf{L}' + E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'). \tag{3.4}$$

In order to obtain an orthogonal model, we assume that

$E(\mathbf{F}) = \mathbf{0}$, $E(\boldsymbol{\varepsilon}) = \mathbf{0}$, and $\operatorname{cov}(\boldsymbol{\varepsilon}, \mathbf{F}) = \mathbf{0}$. Thus, $E(\boldsymbol{\varepsilon}\mathbf{F}') = \mathbf{0}$.

Further, $\text{cov}(\boldsymbol{\varepsilon}) = E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \text{diag}(\psi_1, \psi_2, \ldots, \psi_p)$. We shall denote $\text{diag}(\psi_1, \psi_2, \ldots, \psi_p) = \boldsymbol{\Psi}$.

In particular,

$$\text{cov}(\mathbf{F}) = E(\mathbf{F}\mathbf{F}') = \underset{m\times m}{\mathbf{I}}$$

The assumptions mean that the factors are independent among themselves with $\text{var}(f_i) = 1$, and that common factors and specific factors are also independent.

Equation (3.4) then becomes

$$\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi} \qquad (3.5)$$

Based on the above assumptions, it can then be inferred that

$$\begin{aligned} E(\mathbf{x} - \boldsymbol{\mu})\mathbf{F}' &= E(\mathbf{L}\mathbf{F} + \boldsymbol{\varepsilon})\mathbf{F}' \\ &= \mathbf{L}E(\mathbf{F}\mathbf{F}') + E(\boldsymbol{\varepsilon}\mathbf{F}') \\ &= \mathbf{L} \end{aligned}$$

is the covariance between $\mathbf{X}$ and the factors. That is, $\text{cov}(X_i, f_j) = l_{ij}$.

Equation (3.3) further means that

$$\text{var}(X_i) = \mathbf{L}_i\mathbf{L}_i' + \psi_i, \qquad (3.6)$$

which is the inner product of the ith row of $\mathbf{L}$ and itself plus the variance specific to X_i. That is, $\text{var}(X_i) = \sum_{j=1}^{m} l_{ij}^2 + \psi_i$. Thus, in line with Equation (3.5), the variation in X_i is in two parts: the sum of the variation accounted for by the m factors, called the communality of the variable with the factors; and the other is the variation that is specific to the variable, and unaccounted for by the factors. Further, the covariance between X_i and X_k is

$$\text{cov}(X_i, X_k) = \mathbf{L}_i\mathbf{L}'_k = \sum_{j=1}^{m} l_{ij} l_{kj},$$

which is the sum of the cross-product of the *i*th row and the *k*th column of **L**. If factors are extracted from the correlation matrix, then $\text{var}(X_i) = 1, \forall i$, and **L** become a matrix of correlation coefficients between the variables and the factors. In this case, $\psi_i = 1 - \sum_{j=1}^{m} l_{ij}^2$.

If the *m* factors adequately explain the relationship among the variables, then we expect $\psi_i, \forall i$ to be small. In this case the main aim of factor analysis model is to reproduce $\text{cov}(\mathbf{X})$ or $\mathbf{R} = corr(\mathbf{X})$, the correlation matrix by $\mathbf{LL}'$. The usefulness of factor model therefore is derived when *m* is small relative to *p*. However, for many datasets, $\mathbf{R}$ cannot be factored as in Equation (3.5) even when *m* is close to *p*.

Factor Scores

The generalised expression for a factor is

$$f_i = \sum_{j=1}^{p} l_{ij} x_j, \tag{3.7}$$

where $m \geq 1$ and l_{ij} are the factor loadings, such that the measure

$$v_i = \sum_{j=1}^{p} l_{ij}^2 \tag{3.8}$$

is the eigenvalue of factor *i*. This represents the amount of variation in the data accounted for by the factor. There is an ordering principle in factor extraction. By this principle, we expect the relation $v_i \geq v_k, \quad i < k$. Thus, the first factor will have the largest eigen-value. Since factors are extracted from the correlation matrix, it follows that

$$\sum_{i=1}^{p}\sum_{j=1}^{p} l_{ij}^2 = p$$

That is, we expect all p factors to explain the entire variation in the dataset. It is expected therefore that for $m < p$, $\sum_{i=1}^{m}\sum_{j=1}^{p} l_{ij}^2 < p$. Thus, any extracted factor solution with reduced dimensions will have some associated loss of information.

It is easy to interpret the factors if the coefficients have almost the same signs. There is a problem, however, when there is a mixture of positive and negative signs in the construction of the factor. Even though a rotation of the solution should eliminate mixed signs, for some portions of the data, the sign problem remains even after rotation. This constitutes one of the problems studied by Benyi (2018).

Construction of Time-Dependent Principal Components

Denote the variables of study by $X_1, X_2, \ldots, X_p, p = 19$. Denote also the ith PC of the kth year by y_{ik} and define as

$$y_{ik} = \sum_{j=1}^{p} a_{ikj} x_j. \tag{3.9}$$

For purposes of plausible interpretation, it is necessary to identify the influential indicators in the formation of the y_{ik} by a large weight a_{ij}, usually greater than 0.5 (Nkansah, 2018, Frempong, Nkansah & Nkansah, 2017). However, in some cases as will be encountered in the study (see Table 9 and

Appendix C), the nature of the data may not facilitate this. This is because, the real importance of x_j can be influenced by its variation or may be masked by a high variation in another variable. In such a case, the problem may be resolved by expressing y_{ik} rather in terms of the loadings and as a factor component given by

$$F_{ik} = \sum_{j=1}^{p} l_{ikj} x_j . \tag{3.10}$$

Values l_{ikj} are the loadings with $\sum_{j=1}^{p} l_{ikj}^2$ equal to the eigenvalue λ_i of $\mathbf{f}_{ik}$ and represents the variation in the data explained by the component, and l_{ij} is the correlation coefficient between the ith component and the jth variable. The weight in Equations (3.9) and (3.10) are connected by

$$a_{ij} = \frac{l_{ij} s_{x_j}}{\sqrt{\lambda_i}}. \tag{3.11}$$

where s_{x_j} is the standard deviation of the indicator variable x_j.

Proof

We can write the component in Equation (3.9) as

$$Y_i = \mathbf{a}_i' \mathbf{X}$$

where $\mathbf{X} = (X_1, X_2, \ldots, X_p)'$ is the p-dimensional variable vector. If we represent the p-dimensional vector $\mathbf{a}_j'$ by $\mathbf{a}_j' = (0,0,\cdots,0,1,0,0,\cdots 0)$, with $a_{jj} = 1$, then we can write the component variable X_j as

$$X_j = \mathbf{a}_j' \mathbf{X}$$

Given the variance-covariance matrix of $\mathbf{X}$ as then

$$\text{Var}(Y_i) = \mathbf{a}_i'\Sigma\mathbf{a}_i.$$

But since $\mathbf{a}_i'\Sigma\mathbf{a}_i = \lambda_i$, it implies that $\lambda_i a_i = \Sigma a_i$. Now given $\text{Var}(X_j) = s_j^2$ we have

$$\text{Cov}(Y_i, X_j) = \text{Cov}(\mathbf{a}_i'\mathbf{X}, \mathbf{a}_j'\mathbf{X}) = \mathbf{a}_j'\Sigma\mathbf{a}_i$$

$$= a_j'(\lambda_i a_i)$$

$$= \lambda_i a_j' a_i$$

$$= \lambda_i a_{ij}$$

The correlation coefficient between the i^{th} principal component, Y_i, and the j^{th} variable, X_j, which is given by the loading l_{ij} is generally given by the expression

$$l_{ij} = \frac{Cov(Y_i, X_j)}{\sqrt{Var(Y_i)}\sqrt{Var(X_j)}}.$$

Making appropriate substitutions, we obtain an expression for the correlation coefficient, $\rho_{Y_i X_j}$, between Y_i, and the jth variable, X_j, in terms of the loading, λ_i, and the weights, a_{ij}, as

$$l_{ij} = \frac{a_{ij}\sqrt{\lambda_i}}{s_j} \qquad (3.12)$$

which may also be expressed as Equation (3.9).

End of proof.

Equation (3.12) suggests that the size of the weights a_{ij} that constitute the components is influenced by the variation in the observations on the variable. If the values are dispersed, the weight will be large and the variable will dominant the formation of the component. On the other hand, if there is little spread in the values of the variable, the weight will be small and the variable will not be influential in the formation of the component.

In this study, the prices are likely to vary from market to market. Thus, it will be necessary to ensure that this variability does not affect the determination of the component. It may be necessary to standardize the prices in the case where wide variations exist in the prices. In this case, we have

$$\rho_{Y_i X_j} = a_{ij}\sqrt{\lambda_i}$$

Thus, the formation of the components would not be influenced by variables with high variations.

Let f_{isk} be the *i*th component score for Market *s* in year *k*. The vector of factor scores is given by

$$\mathbf{f} = \mathbf{z}_{1\times p}\mathbf{R}^{-1}_{p\times p}\mathbf{L}_{p\times m}, \tag{3.13}$$

where the vector/matrices have their indicated dimensions and $\mathbf{R}^{-1}\mathbf{L}$, which we denote by $\mathbf{C}$ is the factor score coefficient matrix and $\mathbf{R}$ and $\mathbf{L}$ are the correlation and loading matrix, respectively. The vector $\mathbf{z}$ gives the standardised values of the random variables. Suppose we rewrite the value of f_{isk} from Equation (3.11) as

$$f_{isk} = \sum_{j=1}^{p} c_{ij} z_j, \tag{3.14}$$

where $z_j = \frac{x_{ij} - \bar{x}_j}{s_{x_j}}$ and $\bar{x}_j$ is the mean of item *j*. If $c_{ij} > 0$ and large and $x_{ij} > \bar{x}_j$, and that x_{ij}, $j \in \mathbf{h} = \{h_1, h_2, \ldots, h_m\}$ are much more in number in

Market s, then f_{isk} for that market could be high and positive. Even in such a market, there could be some commodities $x_{st}, t \in \mathbf{X} \backslash \mathbf{h}$, that are rather much fewer than x_{sj}s. Thus, we can split Equation (3.14) into two components as

$$f_{isk} = \sum_{x_j \in \mathbf{h}} c_{ij} z_j + \sum_{x_t \in \mathbf{X} \backslash \mathbf{h}} c_{it} z_t \quad (3.15)$$

In the second summand in Equation (3.15), $c_{it} z_t < 0$, since $c_{it} > 0$ and $x_{it} - \bar{x}_t < 0$ for some t. Therefore, a high positive score reflects a market that has high prices. Similarly, a low positive score reflects a market that has low prices with reference to the variables x_{ij}, $j \in \mathbf{h}$.

Now, suppose that we generate the PCs f_{ik}, $i = 1, 2, \ldots, m$; $k = 1, 2, \ldots, r$ that constitute an m-factor solution for each of the r years. Then on the m PCs, we obtain

$$\mathbf{f} = [\mathbf{f}_{1k}, \mathbf{f}_{12}, \ldots, \mathrm{sign}_{k \in I} \mathbf{f}_{ik}, \ldots, \mathbf{f}_{mk}] \quad (3.16)$$

which is $nr \times m$ matrix of component scores, and $I = \{i_1, i_2, \ldots, i_q\}, q \le r$, are the years on which there are sign inconsistencies. The sign of the score helps to identify the correct label for the suspect extreme market. For the purpose of generating time-dependent component graphs, the direction of interpretation must be the same for all PCs. However, this is generally not the case in time-dependent component scores. For example, in 2008 (i.e., $k = 1$), the direction of f_{2k} and f_{3k} are negative as the influential loadings are negative, whilst that of f_{1k} is in the positive direction (see Table D1 of Appendix D). This means that a high positive score on PC1 would mean that the market is high priced The same cannot be said of PC 2 and PC 3. On PC 2 for example, a high priced market is one that has a high negative score. To resolve this inconsistency, we simply multiply $\text{sign}_{k\in l}\mathbf{f}_{ik}$ by -1. We therefore define the sign of $\mathbf{f}_{ik}$ in year k as follows:

$$\text{sign}_{k\in l}\mathbf{f}_{ik} = \begin{cases} -\sum_{j=1}^{p} l_{ikj}z_j, & l_{ikj} < 0 \quad \forall |l_{ikj}| > \tau \\ \sum_{j=1}^{p} l_{ikj}z_j, & l_{ikj} > 0 \quad \forall |l_{ikj}| > \tau \end{cases} \tag{3.17}$$

The value of τ is a reasonably chosen loading which serves as a cut-off for determining the those x_j that influence the formation of $\mathbf{f}_{ik}$. Usually $\tau = 0.5$ (Frempong et al., 2017; Nkansah, 2018). By this treatment, the resulting score signs are in line with those of other years.

The rule specified in Equation (3.17) is not applicable to certain years. For those years (Component 4 of $k=1$ in Table D4, Appendix D) we have contrasting components (Benyi, 2018) on which some groups of indicators have high positive loadings whilst others have high negative loadings. This means that high component score represents a market that is high priced on items that have positive loadings but low priced on those with negative loadings. However, high negative score would mean that the market is high priced on items with negative loadings but low priced on items with positive loadings. This exceptional situation must be monitored in the behavior of the time-dependent component.

Confirmatory Factor Extraction

Suppose that data are obtained on the variables $\mathbf{X}=(X_1, X_2, \ldots, X_p)$. In factor extraction, we seek to determine why the variables relate among themselves the way they do. That is, we determine the salient dimensions that explain (or reproduce) the correlation matrix of $\mathbf{X}$. Thus, primarily, there should be high correlations among the variables to justify factor analysis of the data. In this case, the population correlation matrix, $\boldsymbol{\rho}$, or the population variance-covariance matrix, $\boldsymbol{\Sigma}$, deviates significantly from being diagonal. The problem of extracting factors is essentially a determination of two parameter matrices: the matrices $\mathbf{L}$ and $\boldsymbol{\Psi}$. We consider two popular methods of estimating these parameters: the principal components method and the maximum likelihood method. Solutions from either of these methods may be rotated to obtain more interpretable factors.

Maximum likelihood estimation

If we assume that the factor matrix $\mathbf{F}$ and the specific factors $\boldsymbol{\varepsilon}$ are normally distributed, then it is possible to obtain the maximum likelihood estimates of the factor loadings and specific variances. When $\mathbf{F}_i$ and $\boldsymbol{\varepsilon}_i$ are jointly normal, the observation $\mathbf{x}_j-\boldsymbol{\mu}=\mathbf{LF}_j+\boldsymbol{\varepsilon}_j$ are then normally distributed. The likelihood function is then given by

$$L(\boldsymbol{\mu},\boldsymbol{\Sigma}) = (2\pi)^{-\frac{np}{2}}|\boldsymbol{\Sigma}|^{-\frac{n}{2}}\exp\left[-\tfrac{1}{2}tr\left\{\boldsymbol{\Sigma}^{-1}\left(\sum_{j=1}^{n}(\mathbf{x}_j-\mathbf{x})(\mathbf{x}_j-\mathbf{x})'\right)+n(\overline{\mathbf{x}}-\boldsymbol{\mu})(\overline{\mathbf{x}}-\boldsymbol{\mu})'\right\}\right]$$

$$= (2\pi)^{-\frac{(n-1)p}{2}}|\boldsymbol{\Sigma}|^{-\frac{n-1}{2}}\exp\left[-\tfrac{1}{2}tr\left\{\boldsymbol{\Sigma}^{-1}\left(\sum_{j=1}^{n}(\mathbf{x}_j-\mathbf{x})(\mathbf{x}_j-\mathbf{x})'\right)\right\}\right]$$

$$\times(2\pi)^{-\frac{p}{2}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}}\exp\left[-\tfrac{n}{2}(\overline{\mathbf{x}}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{x}}-\boldsymbol{\mu})\right] \qquad (3.18)$$

The expression above depends on $\mathbf{L}$ and $\boldsymbol{\Psi}$ since $\boldsymbol{\Sigma}=\mathbf{LL}'+\boldsymbol{\Psi}$. It is noted under the orthogonal factor model that there are multiplicity of choices for **L**. This makes the factorisation of $\boldsymbol{\Sigma}$ undefined. To ensure a well-defined **L**, we impose a computationally convenient uniqueness condition

$$\mathbf{L}'\boldsymbol{\Psi}\mathbf{L}=\boldsymbol{\Delta}$$

which is a diagonal matrix. The maximum likelihood estimates of $\mathbf{L}$ and $\boldsymbol{\Psi}$ are obtained by a maximisation of Equation (3.16). Thus, we should obtain $\hat{\boldsymbol{\mu}}=\overline{\mathbf{x}}$ and $tr(\boldsymbol{\Sigma}^{-1}\mathbf{S}_n)=p$ by noting that $tr(\boldsymbol{\Sigma}^{-1}\mathbf{S}_n)\geq p$. By considering the log-likelihood of $L(\boldsymbol{\mu},\Sigma)$ the maximisation process also leads to a maximisation of $-\frac{n}{2}\left[\ln|\boldsymbol{\Sigma}|+tr(\boldsymbol{\Sigma}^{-1}\mathbf{S}_n)\right]$. Now, since variables are standardised so that

$$\mathbf{Z}=\mathbf{V}^{-\frac{1}{2}}(\mathbf{x}-\boldsymbol{\mu}), \qquad (3.19)$$

then

$$\text{cov}(\mathbf{z})=\mathbf{V}^{-\frac{1}{2}}\text{cov}(\mathbf{x})\mathbf{V}^{-\frac{1}{2}}$$

$$=\mathbf{V}^{-\frac{1}{2}}\boldsymbol{\Sigma}\mathbf{V}^{-\frac{1}{2}}$$

$$=\mathbf{V}^{-\frac{1}{2}}(\mathbf{LL}'+\boldsymbol{\Psi})\mathbf{V}^{-\frac{1}{2}}$$

$$=\mathbf{V}^{-\frac{1}{2}}\mathbf{LL}'\mathbf{V}^{-\frac{1}{2}}+\mathbf{V}^{-\frac{1}{2}}\boldsymbol{\Psi}\mathbf{V}^{-\frac{1}{2}}$$

$$=(\mathbf{V}^{-\frac{1}{2}}\mathbf{L})(\mathbf{V}^{-\frac{1}{2}}\mathbf{L})'+\mathbf{V}^{-\frac{1}{2}}\boldsymbol{\Psi}\mathbf{V}^{-\frac{1}{2}}$$

Thus, if $\Sigma = LL' + \Psi$, then $\rho (= \text{cov}(x))$ has analogous factorisation with loadings $L_z = V^{-\frac{1}{2}}L$ and specific variance $\Psi_z = V^{-\frac{1}{2}}\Psi V^{-\frac{1}{2}}$. Thus, if $\hat{V}^{-\frac{1}{2}}$ and $\hat{L}$ are maximum likelihood estimates of L and Ψ, respectively, then

$$\rho = (V^{-\frac{1}{2}}L)(V^{-\frac{1}{2}}L)' + V^{-\frac{1}{2}}\Psi V^{-\frac{1}{2}}$$
$$= L_z L_z' + \Psi_z'$$

Principal components estimation method

By the spectral decomposition of symmetric matrices, the correlation matrix $\mathbf{R}$ could be expressed as

$$\mathbf{R} = \sum_{i=1}^{p} \lambda_i \mathbf{e}_i \mathbf{e}_i' = \mathbf{P\Lambda P'} \qquad (3.20)$$

The vector $\mathbf{e}_i$ is the ith $p\times 1$ eigen-vector of $\mathbf{R}$ with corresponding eigen-value λ_i. The matrix $\mathbf{P}$ is such that $\mathbf{PP'} = \mathbf{P'P} = \mathbf{I}$. The matrix $\mathbf{\Lambda}$ is diagonal given as

$$\mathbf{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_p \end{pmatrix}$$

Equation (3.10) may be written as

$$\mathbf{R} = (\mathbf{P\Lambda}^{\frac{1}{2}})(\mathbf{P\Lambda}^{\frac{1}{2}})', \qquad (3.21)$$

fits the prescribed structure of the factor analysis technique having as many factors as variables (i.e., $m=p$) and specific variance, $\psi_i = 0, \ \forall i$. Thus, we can write

$$\mathbf{R} = \mathbf{LL}', \tag{3.22}$$

where $\mathbf{L} = \mathbf{P\Lambda}^{\frac{1}{2}}$, $p \times p$ matrix of scaled columns of $\mathbf{P}$. To allow for variation in the specific factors, $\boldsymbol{\varepsilon}$, we prefer models that explain $\mathbf{R}$ in terms of few factors (i.e., $m < p$). We could therefore discard the last $(p-m)$ eigenvalues and obtain an approximation to Equation (3.12) as

$$\mathbf{R} = (\mathbf{P\Lambda}^{\frac{1}{2}})(\mathbf{P\Lambda}^{\frac{1}{2}})' + \mathbf{\Psi} \tag{3.23}$$

where $\mathbf{L} = \mathbf{P\Lambda}^{\frac{1}{2}}$ is now a $p \times m$ matrix of loadings and

$$\mathbf{\Psi} = \begin{pmatrix} \psi_1 & 0 & 0 & \cdots & 0 \\ 0 & \psi_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \psi_p \end{pmatrix}$$

and $\psi_i = 1 - \sum_{j=1}^{m} l_{ij}^2; \quad i = 1, 2, \ldots, p$ or the diagonal elements of the matrix of $\mathbf{\Psi} = \mathbf{R} - (\mathbf{P\Lambda}^{\frac{1}{2}})(\mathbf{P\Lambda}^{\frac{1}{2}})'$. The representation in Equation (3.22) is the principal components solution since the factor loadings are the scaled coefficients of first m sample principal components.

For any factor model, that the principal component factor model explains a higher variation in the data than the likelihood method. The variance-maximisation property would naturally favour the principal components extraction as it is specifically obtained by a maximisation of the ratio $\frac{\mathbf{L}_i'\mathbf{R}\mathbf{L}_i}{\mathbf{L}_i'\mathbf{L}_i}$, where $\text{var}(Y_i)=\mathbf{L}_i'\mathbf{R}\mathbf{L}_i$ is the variance explained by the *i*th principal component.

The maximum likelihood factor model, on the other hand, appears to have much smaller elements of the residual matrix

$$\mathbf{R}_{res} = \mathbf{R} - \mathbf{L}\mathbf{L}' - \mathbf{\Psi} \qquad (3.24)$$

Obviously, for a good factor model, the residual correlations should be as small as possible. This will translate into small average squared values of the off-diagonal elements of $\mathbf{R}_{res}$. If the square root of this quantity, known as the Root Mean Square Residual (RMSR) is small, then we could obtain a good factor structure. The RMSR of $\mathbf{R}_{res}$ is given by

$$RMSR = \sqrt{\frac{\sum_{i<j}(\mathbf{R}_{res})_{ij}^2}{\frac{1}{2}p(p-1)}} \qquad (3.25)$$

As a consequence of the properties, either of these two methods may be preferred depending on the nature of the study. For example, since a small RMSR is an indication of a good factor structure, the maximum likelihood solution may be preferred in confirmatory factor analysis which is concerned with the goodness of the factor model and is focused on the structure of the correlation (or the variance-covariance) matrix. The test of the number of extracted factors considered in the next section will therefore make use of the maximum likelihood method. If the focus is on the amount of variation accounted for by the solution, then one may consider the principal components method.

Test for Extracted Number of Common Factors

Suppose that an m common factors underlie the sample correlation matrix **R** generated on the variables $X_1, X_2, \ldots, X_p$. The test of adequacy of the m-factor model is equivalent to the test of the hypothesis

$$H_o: \underset{p\times p}{\boldsymbol{\rho}} = \underset{p\times m}{\mathbf{L}}\,\underset{m\times p}{\mathbf{L}'} + \underset{p\times p}{\boldsymbol{\Psi}} \tag{3.26}$$

against $H_a: \underset{p\times p}{\boldsymbol{\rho}} \neq \underset{p\times m}{\mathbf{L}}\,\underset{m\times p}{\mathbf{L}'} + \underset{p\times p}{\boldsymbol{\Psi}}$

where

$\mathbf{LL}'$ is the reproduced correlation matrix based on m factors,

$\boldsymbol{\Psi}$ is a diagonal matrix of specific variances, whose elements are given by $\psi_i = 1 - \sum_{j=1}^{m} l_{ij}^2$, $i = 1, 2, \ldots, p$.

The alternative hypothesis means that $\boldsymbol{\rho}$ is any other positive definite matrix that cannot be factorised as under H_o. Under H_o, the maximum of the likelihood function, with $\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$ and $\hat{\boldsymbol{\Sigma}} = \hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\Psi}}$, where $\hat{\mathbf{L}}$ and $\hat{\boldsymbol{\Psi}}$ are the maximum likelihood estimates of $\mathbf{L}$ and $\boldsymbol{\Psi}$, is proportional to

$$\hat{\Sigma}^{-\frac{n}{2}} \exp\left[-\tfrac{1}{2} tr\left\{\hat{\Sigma}^{-1}\left(\sum_{j=1}^{n}(\mathbf{x}_j - \mathbf{x})(\mathbf{x}_j - \mathbf{x})'\right)\right\}\right]$$

$$= \left|\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\Psi}}\right|^{-\frac{n}{2}} \exp\left[-\tfrac{1}{2} ntr\left\{(\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\Psi}})^{-1}\mathbf{S}_n\right\}\right]$$

where

$$\mathbf{S}_n = \frac{1}{n}\sum_{j=1}^{n}(\mathbf{x}_j - \mathbf{x})(\mathbf{x}_j - \mathbf{x})'$$

By the general likelihood method,

$$-2\ln\Lambda = -2\ln\left(\frac{\max_{\theta\in\Theta_o} \mathbf{L}(\boldsymbol{\theta})}{\max_{\theta\in\Theta} \mathbf{L}(\boldsymbol{\theta})}\right)^{-\frac{n}{2}}$$

$$= -2\ln\left(\frac{|\hat{\boldsymbol{\Sigma}}|}{|\mathbf{S}_n|}\right)^{-\frac{n}{2}} + n\left[tr(\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{S}_n) - p\right]$$

with degrees of freedom

$$\upsilon - \upsilon_o = \tfrac{1}{2}\left[(p-m)^2 - p - m\right] \tag{3.27}$$

Under a maximum likelihood estimate of the parameters in H_o, $tr(\hat{\mathbf{R}}^{-1}\mathbf{R}_n) - p = 0$. Thus, the statistic becomes

$$-2\ln\Lambda = n\ln\left(\frac{|\hat{\boldsymbol{\Sigma}}|}{|\mathbf{S}_n|}\right)^{-\frac{n}{2}} \tag{3.28}$$

It is shown (Bartlett, 1954) that the chi-square approximation to the sampling distribution of $-2\ln\Lambda$ can be improved by replacing n in Equation (3.28) with the multiplicative factor $\left[n-1-\tfrac{1}{6}(2p+4m+5)\right]$. Using the Bartlett's correction, we reject H_o at α level of significance if

$$\left[n-1-\tfrac{1}{6}(2p+4m+5)\right]\ln\frac{|\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\psi}}|}{|\mathbf{S}_n|} > \chi^2_{\frac{1}{2}[(p-m)^2-p-m]}(\alpha) \tag{3.29}$$

provided n and $n-p$ are large.

Remark 3.1

Since $\tfrac{1}{2}\left[(p-m)^2 - p - m\right] > 0$, we should have

$$m < \tfrac{1}{2}\left(2p+1-\sqrt{8p+1}\right) \tag{3.30}$$

The degrees of freedom provides a constraint on the number of factors that could be extracted to obtain a solution with reasonable adequacy.

Remark 3.2

If n is large and m is small relative to p, H_o will usually be rejected. This will lead to retaining more common factors, defeating the purposing of achieving a parsimonious solution, which is a key goal in factor extraction. It is pointed out (Johnson & Wichern, 2007) that $\hat{\mathbf{\Sigma}} = \hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\mathbf{\Psi}}$ may be close to $\mathbf{S}_n$ so that addition of more factors does not provide additional information, though they ‘‘are significant’’, in the sense that it does not help to factorise $\mathbf{\Sigma}$ under H_o. It should further be pointed out (as will be seen in the study) that in certain situations, adding more factors beyond a certain number does provide additional information, though they are ‘‘not significant’’, in the sense that it does help to factorise $\mathbf{\Sigma}$ under H_o. The question then arises, in this case, regarding the particular value m for which the factor solution may be regarded as just adequate. It is widely admonished that some judgement must therefore be exercised in the choice of m (e.g., Johnson & Wichern, 2007). A guide to the exercise of this judgement would be very necessary.

As indicated in the introductory part of this section, factor analysis is essentially based on the correlation matrix, $\mathbf{R}$. It is therefore expedient to state the test statistic in terms of $\boldsymbol{\rho}$. Let $\mathbf{V}^{-\frac{1}{2}} = diag\left(\frac{1}{\sigma_1}, \frac{1}{\sigma_2}, \ldots, \frac{1}{\sigma_p}\right)$, a matrix with diagonal elements being the standard deviations of the variables. Then, we should have, $\mathbf{V}^{-\frac{1}{2}}\mathbf{S}_n\mathbf{V}^{-\frac{1}{2}} = \mathbf{R}$. By properties of determinants,

$$\left|\mathbf{V}^{-\frac{1}{2}}\right|\left|\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\mathbf{\Psi}}\right|\left|\mathbf{V}^{-\frac{1}{2}}\right| = \left|\mathbf{V}^{-\frac{1}{2}}\hat{\mathbf{L}}\hat{\mathbf{L}}'\mathbf{V}^{-\frac{1}{2}} + \mathbf{V}^{-\frac{1}{2}}\hat{\mathbf{\Psi}}\mathbf{V}^{-\frac{1}{2}}\right|$$

and

$$\left|\mathbf{V}^{-\frac{1}{2}}\right|\left|\mathbf{S}_n\right|\left|\mathbf{V}^{-\frac{1}{2}}\right| = \left|\mathbf{V}^{-\frac{1}{2}}\mathbf{S}_n\mathbf{V}^{-\frac{1}{2}}\right|$$

Consequently,

$$\begin{aligned}\frac{|\hat{\boldsymbol{\Sigma}}|}{|\mathbf{S}_n|} &= \frac{\left|\mathbf{V}^{-\frac{1}{2}}\right|}{\left|\mathbf{V}^{-\frac{1}{2}}\right|}\frac{\left|\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\Psi}}\right|}{|\mathbf{S}_n|}\frac{\left|\mathbf{V}^{-\frac{1}{2}}\right|}{\left|\mathbf{V}^{-\frac{1}{2}}\right|} \\ &= \frac{\left|\mathbf{V}^{-\frac{1}{2}}\hat{\mathbf{L}}\hat{\mathbf{L}}'\mathbf{V}^{-\frac{1}{2}} + \mathbf{V}^{-\frac{1}{2}}\hat{\boldsymbol{\Psi}}\mathbf{V}^{-\frac{1}{2}}\right|}{\left|\mathbf{V}^{-\frac{1}{2}}\mathbf{S}_n\mathbf{V}^{-\frac{1}{2}}\right|} \\ &= \frac{\left|\hat{\mathbf{L}}_z\hat{\mathbf{L}}_z' + \hat{\boldsymbol{\Psi}}_z\right|}{|\mathbf{R}|}\end{aligned} \tag{3.31}$$

The statistic in Equation (3.29) is then based on the ratio in Equation (3.31) extracted from standardised variables.

Remark 3.3

In this remark, we point out two peculiar features of the factor model that could arise in this study as a result of the nature of the correlation structure, and its consequence on confirmatory factor analysis and hence, structural equation modelling. Suppose that

$$H_o^{(m)}:\ \underset{p\times p}{\boldsymbol{\rho}} = \underset{p\times m}{\mathbf{L}}\,\underset{m\times p}{\mathbf{L}'} + \underset{p\times p}{\boldsymbol{\Psi}} \quad \text{and} \quad H_o^{(r)}:\ \underset{p\times p}{\boldsymbol{\rho}} = \underset{p\times r}{\mathbf{L}}\,\underset{r\times p}{\mathbf{L}'} + \underset{p\times p}{\boldsymbol{\Psi}} \tag{3.32}$$

are two hypotheses on two factor models based on m and r factors (suppose that $m \leq r$) assumed to underlie the correlation matrix. Generally, the model under $H_o^{(r)}$ does not necessarily contain all the factors under $H_o^{(m)}$. Now, three peculiar features could emerge: (1) The first m factors under $H_o^{(r)}$ the factors under $H_o^{(m)}$; (2) For almost all models under $H_o^{(r+i)}$, $i = 1, 2, \ldots, p-r$, the factors are one-indicator factors beyond a certain point $\nu > r$; (3) The factors in the interval $(m, r]$ under $H_o^{(r)}$ are have very low number of influential indicators, usually less than 3.

Figure 2 is a typical scree plot of a data whose factor solution exhibits the two features described.

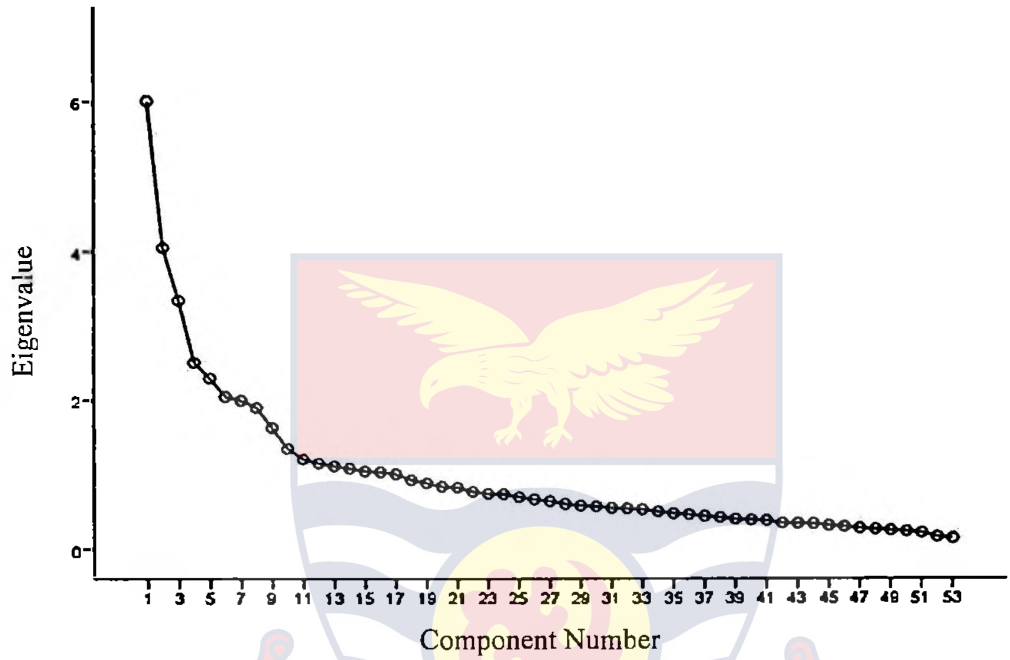

Figure 2: Scree Plot Indicating Possible Incidence of Repeating Factor Solutions and One-Indicator Factors.

This is the scree for the data generated on the indicator variables described in Chapter One. In the graph, the scree tends to level off after component 11.

Now, if the test is significant for both $H_o^{(m)}$ and $H_o^{(r)}$, then for some i, $(r < i < p - r)$, the test would be non-significant. Let this point be $i = s$ $(0 < s < p - r)$. Then the ratio in Equation (3.31) based on $H_o^{(r+i)}$, $i = 0, 1, 2, \ldots, s-1$, may be reduced to

$$\frac{|\hat{\Sigma}|}{|\mathbf{S}_n|} = \frac{|\mathbf{V}^{-\frac{1}{2}}|}{|\mathbf{V}^{-\frac{1}{2}}|} \frac{|\hat{\mathbf{L}}^{(r+i)}\hat{\mathbf{L}}'^{(r+i)} + \hat{\Psi}^{(r+i)}|}{|\mathbf{S}_n|} \frac{|\mathbf{V}^{-\frac{1}{2}}|}{|\mathbf{V}^{-\frac{1}{2}}|}$$

$$= \frac{|\mathbf{V}^{-\frac{1}{2}}\hat{\mathbf{L}}^{(r+i)}\hat{\mathbf{L}}'^{(r+i)}\mathbf{V}^{-\frac{1}{2}} + \mathbf{V}^{-\frac{1}{2}}\hat{\Psi}^{(r+i)}\mathbf{V}^{-\frac{1}{2}}|}{|\mathbf{V}^{-\frac{1}{2}}\mathbf{S}_n\mathbf{V}^{-\frac{1}{2}}|}$$

$$\approx \frac{|\hat{\mathbf{L}}_z^{(m)}\hat{\mathbf{L}}_z'^{(m)} + \hat{\Psi}_z^{(m)}|}{|\mathbf{R}|} \qquad (3.33)$$

which is based on the m-factor solution. It should be noted that in some datasets, the point of change, s, from significance to non-significance of the test of $H_o^{(r)}$ in Equation (3.32) could be difficult to determine.

Remark 3.4

Remark 3.3 makes it necessary for us to identify, in this study, two main considerations for determining factor solutions that may be relevant for various aspects of the study. We will identify: (1) for the purpose of interpretation, factors under $H_o^{(m)}$ in which all the factors are plausible, but may not constitute a non-significant solution in the sense that the factor-solution does not fit the correlation matrix; (2) factors under $H_o^{(r+s)}$, are all not plausible, but constitutes a non-significant solution in the sense that it fits the correlation matrix.

The Pooled Variance-Covariance Matrix

The nature of the dataset used for the study presents two ways of computing the variance-covariance or the sum of squares cross-product matrix, **S**. It can be measured as the total SSCP matrix based on the entire $nr \times p$ dataset. However, this can unduly enhance the projection of the suspect outlier in a year that has a very high variation in the x_is. A fair projection could be obtained by using the pooled SSCP matrix. Thus, consider the years as constituting r groups and denote the year by a categorical variable, T. Let $T(\omega)$ be the year of observation ω. The frequency of all year classes is n. The multivariate data is thus obtained in the layout $\mathbf{X}_{nr \times p} = \left(X_{1k}, X_{2k}, \ldots, X_{pk}\right);\ k = 1, 2, \ldots, r,$ measured from r years. The within-year SSCP matrix for the kth year is given as

$$\mathbf{W}_k = \sum_{\omega:T(\omega)=k}^{n} \left(\mathbf{X}'_{i,k}(\omega) - \bar{\mathbf{x}}_{i,k}\right)\left(\mathbf{X}'_{i,k}(\omega) - \bar{\mathbf{x}}_{i,k}\right)'; \quad i = 1, 2, \ldots, p;\ k = 1, 2, \ldots, r.$$

Dropping the index for the variable, we have

$$\mathbf{W}_k = \left(\mathbf{X}'_k(\omega) - \bar{\mathbf{x}}_k(\omega)\mathbf{1}'\right)\left(\mathbf{X}'_k(\omega) - \bar{\mathbf{x}}_k(\omega)\mathbf{1}'\right)'$$

The pooled within-year SSCP is then given by

$$\mathbf{S}_{pooled} = \sum_{k=1}^{r} \left(\mathbf{X}'_k(\omega) - \bar{\mathbf{x}}_k(\omega)\mathbf{1}'\right)\left(\mathbf{X}'_k(\omega) - \bar{\mathbf{x}}_k(\omega)\mathbf{1}'\right)' \qquad (3.34)$$

In the above two equations, we note that $\mathbf{X}_k(\omega)$ is of dimension $n \times p$. But $\bar{\mathbf{x}}_k$ is of dimension $p \times 1$. We therefore introduce a vector of ones, defined as $\mathbf{1} = \text{ones}\ (n, 1)$ so that the product $\bar{\mathbf{x}}_k(\omega)\mathbf{1}'$ is of dimension $p \times n$. This way, we ensure that the same average price value of a particular commodity is subtracted from the value of that commodity in each of the n markets of that year.

Outlier Displaying Component

The Outlier Displaying Component is one of the techniques that form part of the one what has been described as the class of displaying components in this study. The outlier displaying component is generally a graphical technique used for highlighting multivariate observations that are known or suspected to be outliers. We will provide the theoretical background of this technique. As explained earlier, the focus will be on the one-outlier displaying component.

The one-outlier displaying component (1-ODC) is designed for highlighting a single known outlier, $\mathbf{x}_{\varepsilon}$, in a multivariate dataset $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p]'$ in one-dimensional plane. In this plot, the projected outlier is expected to stick out most clearly. It is assumed that the sample is drawn from common p-dimensional normal distribution with a $p \times 1$ mean vector μ and $p \times p$ variance-covariance matrix, Σ. However, derivations of the technique are formally algebraic and geometric and do not actually require the use of the distributional assumptions.

In displaying the projected multivariate data in one-dimensions, we have our eyes on the labelled outlier, $\mathbf{x}_{\varepsilon}$. Using a projection vector, $\boldsymbol{\beta}$, one can convert a p-dimensional observation $\mathbf{x}_j$ into a corresponding univariate observation y_j given by

$$y_j = \boldsymbol{\beta}'\mathbf{x}_j. \tag{3.35}$$

The variance of these projected values is given as

$$T = \boldsymbol{\beta}'\mathbf{S}\boldsymbol{\beta}.$$

Now, the studentised distance of y_j from the mean $\bar{y}$ is given by

$$U_j(\boldsymbol{\beta}) = (y_j - \bar{y})'T^{-1}(y_j - \bar{y}).$$

Making substitution for T and y_j, we have

$$\begin{aligned} U_j(\boldsymbol{\beta}) &= [\boldsymbol{\beta}'(\mathbf{x}_j - \bar{\mathbf{x}})]'(\boldsymbol{\beta}'\mathbf{S}\boldsymbol{\beta})^{-1}[\boldsymbol{\beta}'(\mathbf{x}_j - \bar{\mathbf{x}})] \\ &= \frac{(\mathbf{x}_j - \bar{\mathbf{x}})'\boldsymbol{\beta}\boldsymbol{\beta}'(\mathbf{x}_j - \bar{\mathbf{x}})}{\boldsymbol{\beta}'\mathbf{S}\boldsymbol{\beta}} \\ &= \frac{\boldsymbol{\beta}'(\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})'\boldsymbol{\beta}}{\boldsymbol{\beta}'\mathbf{S}\boldsymbol{\beta}}. \end{aligned} \tag{3.36}$$

This result shows that for any scalar $\lambda \neq 0$, $U_j(\boldsymbol{\beta}) = U_j(\lambda\boldsymbol{\beta})$. That is, the projection given in Equation (3.35) is said to be invariant. From Equation (3.36), the studentised distance of the suspected outlier $\mathbf{x}_\varepsilon$ is given as

$$U_\varepsilon(\boldsymbol{\beta}) = \frac{\boldsymbol{\beta}'(\mathbf{x}_\varepsilon - \bar{\mathbf{x}})(\mathbf{x}_\varepsilon - \bar{\mathbf{x}})'\boldsymbol{\beta}}{\boldsymbol{\beta}'\mathbf{S}\boldsymbol{\beta}}.$$

Now, every choice of $\boldsymbol{\beta}$ results in a different set of y-values, and hence, different value of $U_\varepsilon(\boldsymbol{\beta})$. The best choice of $\boldsymbol{\beta}$ is one that maximises $U_\varepsilon(\boldsymbol{\beta})$ subject to the constraint $\boldsymbol{\beta}'\mathbf{S}\boldsymbol{\beta} = c$, where c is an arbitrary constant. The constant is introduced to remove the ambiguity in scale of $\boldsymbol{\beta}$ which is as a result of the invariant property of the projection. The maximization problem is solved by introducing the Lagrange multiplier λ as

$$\Phi = \boldsymbol{\beta}'(\mathbf{x}_\varepsilon - \bar{\mathbf{x}})(\mathbf{x}_\varepsilon - \bar{\mathbf{x}})'\boldsymbol{\beta} - \lambda(\boldsymbol{\beta}'\mathbf{S}\boldsymbol{\beta} - c).$$

The maximization process shows that

$$\lambda = \boldsymbol{\beta}'(\mathbf{x}_\varepsilon - \bar{\mathbf{x}})(\mathbf{x}_\varepsilon - \bar{\mathbf{x}})'\boldsymbol{\beta} = U_\varepsilon(\boldsymbol{\beta}).$$

This means that for a maximum value of $U_\varepsilon(\boldsymbol{\beta})$, λ must be chosen as the largest eigen-value of the $p \times p$ matrix $\mathbf{S}^{-1}(\mathbf{x}_\varepsilon - \overline{\mathbf{x}})(\mathbf{x}_\varepsilon - \overline{\mathbf{x}})'$. Now, this matrix is of rank 1, and hence, λ corresponds to the only non-zero eigenvalue of the matrix, which is given by

$$\lambda_{\max} = (\mathbf{x}_\varepsilon - \overline{\mathbf{x}})'\mathbf{S}^{-1}(\mathbf{x}_\varepsilon - \overline{\mathbf{x}}).$$

The associated eigenvector can be determined to be

$$\boldsymbol{\beta}_\varepsilon = \mathbf{S}^{-1}(\mathbf{x}_\varepsilon - \overline{\mathbf{x}}). \tag{3.37}$$

This eigenvector is what is referred as the One-Outlier Displaying Component, 1-ODC (Gordor & Fieller, 1994).

Remark 3.5

The Outlier Displaying Component (ODC) has a number of properties. One important property which will be exploited in this study is stated as follows:

> All information on the discordancy of the *p*-dimensional outlier is held in the single dimension of the 1-ODC.

The modified 1-ODC

Remark 3.5 means that the outlyingness of the suspect multivariate outlier observation is preserved in the corresponding projected univariate observation. In spite of this important property, the vector is usually criticised for being affected by the very extreme observation it seeks to highlight, as the overall mean, $\overline{\mathbf{x}}$, and the variance-covariance matrix, $\mathbf{S}$, are both influenced by the presence of the suspect extreme observation, $\mathbf{x}_\varepsilon$. Thankfully, Nkansah and Gordor (2012b) have proved that the difference $\mathbf{x}_j - \bar{\mathbf{x}}$ may be expressed as

$$\mathbf{x}_j - \bar{\mathbf{x}} = \frac{n-k}{n}\left(\mathbf{x}_j - \bar{\mathbf{x}}_{(I_k)}\right) + \frac{k}{n}\left(\mathbf{x}_j - \bar{\mathbf{x}}_{I_k}\right),$$

which is a partitioning of the difference into a weighted sum of two components in the presence of k outliers. These components are: (1) the difference between $\mathbf{x}_j$ and the mean, $\bar{\mathbf{x}}_{(I_k)}$, of the remaining $(n-k)$ observations which excludes the set of k outliers; and (2) the difference between $\mathbf{x}_j$ and the mean, $\bar{\mathbf{x}}_{I_k}$, of the set of outliers. In the case of the single outlier, the difference reduces to

$$\mathbf{x}_\varepsilon - \bar{\mathbf{x}} = \frac{n-1}{n}\left(\mathbf{x}_\varepsilon - \bar{\mathbf{x}}_{(\varepsilon)}\right),$$

as $\bar{\mathbf{x}}_{I_k} = \mathbf{x}_\varepsilon$. Now, following the approach in the initial part, the projection of $\mathbf{x}_j$ onto a corresponding univariate observation y_j such that for some $\boldsymbol{\beta}_\varepsilon$, $y_\varepsilon = \boldsymbol{\beta}'_\varepsilon \mathbf{x}_\varepsilon$. In terms of our notation, the distance of y_{π_v} from the remaining $(n-1)$ observations is

$$U(y_{\pi_v}; \bar{\mathbf{x}}_{(\pi_v)}, \mathbf{S}_{(\pi_v)}) = (y_{\pi_v} - \bar{y})\mathbf{S}_y^{-1}(y_{\pi_v} - \bar{y})$$
$$= \frac{\boldsymbol{\beta}'_{\pi_v}(\mathbf{x}_{\pi_v} - \bar{\mathbf{x}}_{(\pi_v)})'(\mathbf{x}_{\pi_v} - \bar{\mathbf{x}}_{(\pi_v)})'\boldsymbol{\beta}_{\pi_v}}{\boldsymbol{\beta}'_{\pi_v}\mathbf{S}_{(\pi_v)}\boldsymbol{\beta}_{\pi_v}}.$$

Maximising this expression subject to the constraint $\boldsymbol{\beta}'_{\pi_v}\mathbf{S}_{(\pi_v)}\boldsymbol{\beta}_{\pi_v} = c$, a constant, gives a modification of Equation (3.35) of the form

$$\boldsymbol{\beta}_{(\pi_v)} = \mathbf{S}^{-1}_{(\pi_v)}(\mathbf{x}_{\pi_v} - \bar{\mathbf{x}}_{(\pi_v)}). \qquad (3.38)$$

This modification excludes the suspect outlying observation in the computation of $\bar{\mathbf{x}}$ and $\mathbf{S}$. Equation (3.38) is what is referred to as the Modified One-Outlier Displaying Component (M1-ODC). In the M1-ODC, the matrix $\mathbf{S}_{(\pi_v)}$ and the vector $\bar{\mathbf{x}}_{(\pi_v)}$ are the corresponding matrix/vector in Equation (3.37) computed with the suspected outlier $\mathbf{x}_{\pi_v}$ deleted from the data. The extent of *outlyingness* of an observation is usually obtained by a projection of the data $\mathbf{X}$ onto $\boldsymbol{\beta}$ given by $\mathbf{X}\boldsymbol{\beta}$. In fact, we should rather have the projection $(\mathbf{X}' - \bar{\mathbf{x}}\mathbf{1}')'\boldsymbol{\beta}$ of the mean-corrected data. However, it will be apparent in this study that it is more expedient to obtain a projection of the raw data.

As has been pointed out in Equation (1.2), the single observation in the study data is now a matrix. Thus, the projection vector $\boldsymbol{\beta}_{(\pi_v)}$ now becomes a matrix of dimension $p \times r$. Projections based on this matrix will enable us to obtain a simultaneous display of data for all years in order to determine the exact year in which an observation (or market) is extreme in price levels.

In the study, the categorisation of the extreme markets will be based on the pooled SSCP only. One reason is that as estimate based on the total SSCP is likely to be influenced by the extreme. The (likelihood-based) statistic for testing the extreme p-dimensional observation $\mathbf{x}_{\pi_v}$ is the same as that in the single dimension provided by the outlier component. The statistic is given by

$$D_{(\pi_v)} = (n-1)\mathbf{S}_{pooled}^{-1}(\mathbf{x}_{\pi_{v.}} - \bar{\mathbf{x}}_{\pi_{v.}}) \qquad (3.39)$$

based on the pooled SSCP. Observation $\mathbf{x}_{\pi_v}$ is an outlier for large values of $D_{(\pi_v)}$ higher than tabulated outlier critical values, (see Table XXXII, in

Barnett & Lewis, 1994) for $p \leq 5$. This tabulated values however will not be relevant in this study as the number of variables are much larger than 5. It has been shown (Nkansah & Gordor, 2013) that based on $\mathbf{S}_{(\pi_\nu)}^{-1}$ and $\overline{\mathbf{x}}_{(\pi_\nu)}$, with $\mathbf{x}_{\pi_\nu}$ deleted, the distance

$$U(\mathbf{x}_{\pi_\nu};\overline{\mathbf{x}}, \mathbf{S}) = (\mathbf{x}_{\pi_\nu} - \overline{\mathbf{x}})\mathbf{S}^{-1}(\mathbf{x}_{\pi_\nu} - \overline{\mathbf{x}})$$

is related to $U(\mathbf{x}_{(\pi_\nu)};\overline{\mathbf{x}}_{(\pi_\nu)}, \mathbf{S}_{(\pi_\nu)})$ by the relation

$$\frac{U(\mathbf{x}_{\pi_\nu};\overline{\mathbf{x}}, \mathbf{S})}{U(\mathbf{x}_{(\pi_\nu)};\overline{\mathbf{x}}_{(\pi_\nu)}, \mathbf{S}_{(\pi_\nu)})} = \left(\frac{n-1}{n}\right)^2 \left[1 - \lambda U(\mathbf{x}_{(\pi_\nu)};\overline{\mathbf{x}}_{(\pi_\nu)}, \mathbf{S}_{(\pi_\nu)})\right], \qquad (3.40)$$

where

$$\lambda = \frac{n-1}{n} \cdot \frac{1}{1 + tr\mathbf{A}_I \mathbf{S}_{(\pi_\nu)}^{-1}},$$

and

$$\mathbf{A}_I = \frac{n}{n-1}(\mathbf{x}_{\pi_\nu} - \overline{\mathbf{x}})(\mathbf{x}_{\pi_\nu} - \overline{\mathbf{x}})'$$ (Nkansah & Gordor, 2012b, 2013; Barnett & Lewis, 1994).

Equation (3.40) gives the amount by which the distance of $\mathbf{x}_{\pi_\nu}$ from $\overline{\mathbf{x}}_{(\pi_\nu)}$ exceeds its distance from $\overline{\mathbf{x}}$. It therefore measures the relative efficiency (RE) of the modified method over the original. Clearly, $U(\mathbf{x}_{(\pi_\nu)};\overline{\mathbf{x}}_{(\pi_\nu)}, \mathbf{S}_{(\pi_\nu)}) > U(\mathbf{x}_{\pi_\nu};\overline{\mathbf{x}}, \mathbf{S})$. It means that if an observation is found to be significantly extreme on the pooled reduced SSCP, $\mathbf{S}_{(\pi_\nu)}$, then it is necessarily significantly extreme on the pooled $\mathbf{S}$. The efficiency of the modified method over the original has been examined (Nkansah & Gordor, 2012b) in three commonly used datasets and presented in Table 7.

Table 7: Relative Efficiency of the Modified 1-ODC in Detecting Single Outliers in Three Datasets

Dataset	n	Outlier	$U(\mathbf{x}_{(\pi_v)};\bar{\mathbf{x}}_{(\pi_v)}, \mathbf{S}_{(\pi_v)})$	$tr\mathbf{A}_l\mathbf{S}^{-1}_{(\pi_v)}$	RE
Iris Setosa	50	42	0.3524	0.3454	1.4:1
Milk Transport Cost	36	9	1.1075	1.0768	2.2:1
USA Food Price	23	10	1.3110	1.2540	2.5:1

Source: (Nkansah & Gordor, 2012b)

The results so far show that the projection obtained by the M1-ODC could be two and half times more distinct than that obtained by the original. This particular result is obtained using the USA Food Prices in Sharma (1996). The other two datasets, are well studied in standard text in multivariate statistical techniques (Johnson & Wichern, 2007; Anderson, 2003). The graphical presentation of the results in Table 7 is given in Figures 3 to 5 obtained from Nkansah and Gordor, (2012b). In the graphs, the projection given by the original ODC is marked as A and that given by the modified method is marked as B.

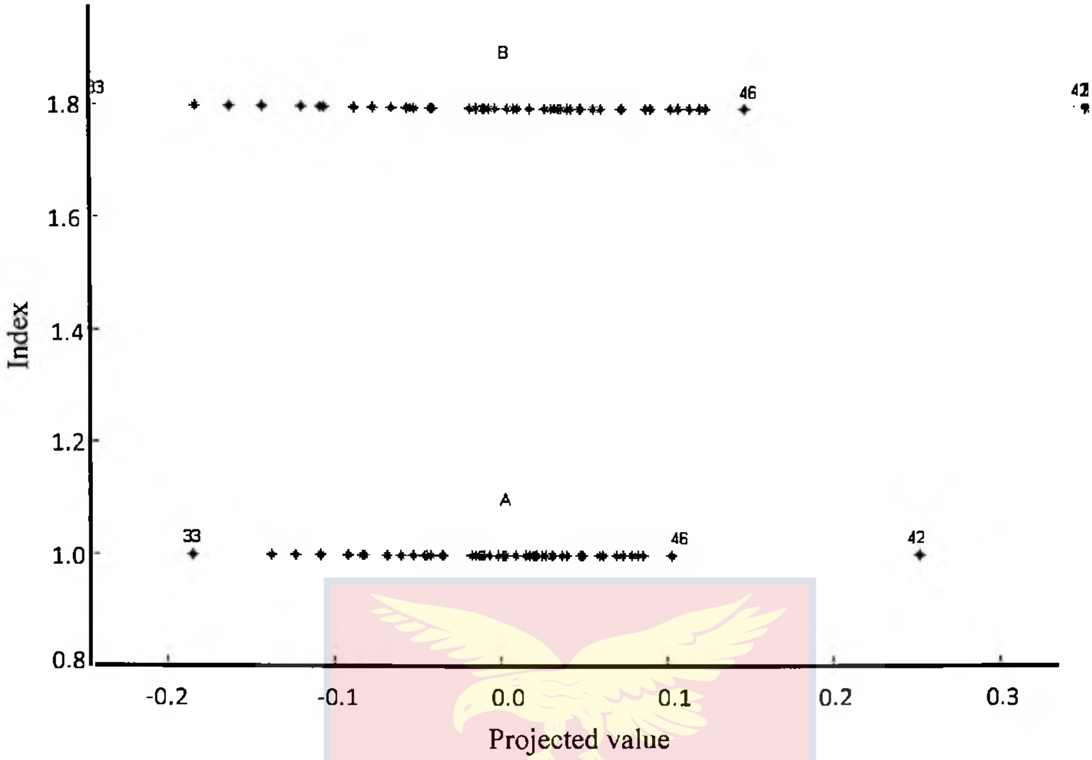

Figure 3: Projection of the Iris Setosa Data onto the Original and Modified ODCs.

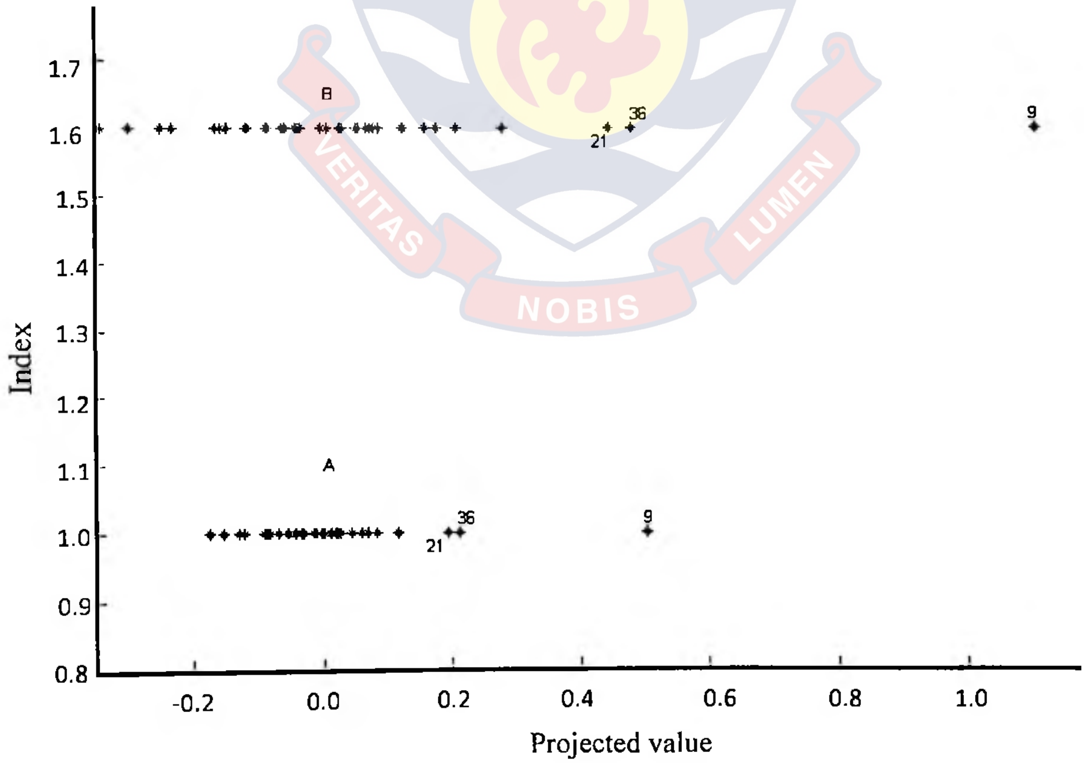

Figure 4: Projection of the Milk Transportation Cost onto the Original and Modified ODCs.

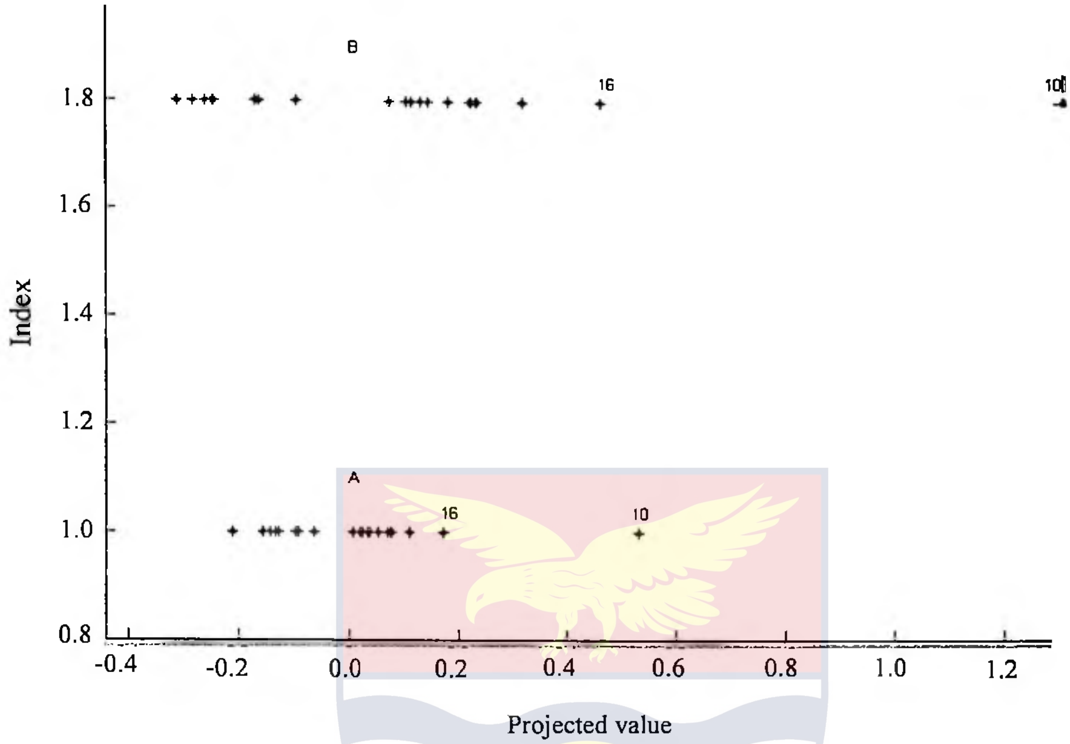

Figure 5: Projection of the US Food Data onto the Original and Modified ODCs.

The graphs in Figures 3 to 5 show other features of the modified method of the Outlier Displaying Component (ODC). The graphs show that in addition to providing a better projection of the suspect outlier, the modified method also gives a better spread in the projected data. Again, a study of the pair of outliers (Nkansah & Gordor, 2012a) shows that the projection of the single outlier reveals other potential outliers within its neighbourhood. These features will be beneficial to our study, in that we will also be able to assess the actual spread in the data based on the projection with reference to any particular suspect outlying market observation, and be able to identify almost all possible suspect outliers by means of methods that involve the 1-ODC.

Vectorisation of a Matrix

A property of the original Outlier Displaying Component (ODC) is that "all information on the discordancy of the p-dimensional outlier is held in the single dimension of the 1-ODC". We will apply this property of 'information retention' ability of the 1-ODC to determine a vectorisation of a matrix in a situation where the matrix as a whole will be suitable for further processing of our intended result, but cannot be used in the form of a matrix.

In this thesis, we will consider each observation as potentially an outlier and determine its discordancy from the centre of the response data. By this property, the single discordancy value of the m-dimensional observation may be represented by a single value. Thus, we can obtain a 'discordancy vector' to represent an entire $n \times m$ data matrix. This representation may be possible by considering any observation $\mathbf{v}_i$, $i = 1, 2, \ldots, n$ with its 1-ODC, $\boldsymbol{\beta}_i$. It should be noted that the observation vector $\mathbf{v}_i$ may itself be an element of a transformation $\mathbf{V} = \mathbf{TF}$ of some original dataset, $\mathbf{T}$, by a transformation matrix, $\mathbf{F}_{n \times m}$. The projection $\{y_i = \mathbf{v}_i'\boldsymbol{\beta}_i;\ i = 1, 2, \ldots, n\}$ constitutes the discordancy sample. The likelihood ratio statistic for this sample is given as

$$U_{y(n)} = \max_{i=1,2,\ldots,n} \left(\frac{y_i - \bar{y}}{s} \right)^2,$$

where $s^2 = \operatorname{var}(y_i) = \boldsymbol{\beta}_i' \Delta \boldsymbol{\beta}_i$, and $\Delta = C(\mathbf{V}\boldsymbol{\beta})$, the variance-covariance matrix of $\mathbf{V}$. Now, for any i, $U_{y(n)}$ becomes the discordancy value D_{y_i} given by

$$D_{y_i} = \frac{\boldsymbol{\beta}_i'(\mathbf{v}_i - \overline{\mathbf{v}})'(\mathbf{v}_i - \overline{\mathbf{v}})\boldsymbol{\beta}_i}{\boldsymbol{\beta}_i'\Delta\boldsymbol{\beta}_i}$$

$$= \frac{(\mathbf{v}_i - \overline{\mathbf{v}})'\Delta^{-1}(\mathbf{v}_i - \overline{\mathbf{v}})(\mathbf{v}_i - \overline{\mathbf{v}})'\Delta^{-1}(\mathbf{v}_i - \overline{\mathbf{v}})}{(\mathbf{v}_i - \overline{\mathbf{v}})'\Delta^{-1}\Delta\Delta^{-1}(\mathbf{v}_i - \overline{\mathbf{v}})'}$$

$$= (\mathbf{v}_i - \overline{\mathbf{v}})'\Delta^{-1}(\mathbf{v}_i - \overline{\mathbf{v}})$$

$$= U_{(n)}$$

Since $U_{(n)}$ is the statistic for testing the extremeness of the m-dimensional observation $\mathbf{v}_i$, in the sample, the result shows that the discordancy of the observation in the sample may be represented by a single value along a single dimension provided by what we will refer to in this thesis as 'an observation displaying component'. Since the observations $\mathbf{v}_i$ may be transformation by some matrix **F**, the effect of the transformation could therefore be captured in the discordancy value, D_{y_i}. In terms of extremeness of the observations, we can therefore represent the matrix, $\mathbf{F}_{n\times m}$ by the single vector $\mathbf{d} = (D_{y_i})'$, $i = 1, 2, \ldots, n$.

Remark 3.6

Even though the modified 1-ODC is shown to enhance the discordancy of the outlier, the original 1-ODC is preferred in the vectorisation procedure proposed above. This is because, as shown in this section, the discordancy on the modified 1-ODC is not exactly equal to the test statistic. Thus, information in the original data is not preserved along this vector.

Structural Equation Modelling

Structural equation modelling (SEM), also called causal modelling, is a multivariate statistical technique that is used to analyse relationships. The technique is a combination of factor analysis and multiple regression analysis, and it is used to analyse the structural relationship between measured variables and latent constructs. The method is most often preferred to other multivariate methods because it is able to estimate the multiple and interrelated dependence in a single analysis. SEM technique makes use of two variables, the exogenous (independent) and endogenous (dependent) variables. There are two types of SEM. These are measurement model which represent the theory that specifies how much variables come together to represent the theory and structural model which represents the theory that shows how constructs are related to other constructs.

There are several assumptions made when using the SEM technique. These include multivariate normal distribution; linear relationship between the endogenous and the exogenous variables; data is free of outliers; sequence of cause and effect relationship: that is, a cause must occur before the event; non-spurious relationship: that is, the observed covariance must be true; model identification: indicating that equations must be greater than the estimated parameters, also models should be over identified or be exact identified; the sample size should be preferably between 200 and 400 with 10 to 15 indicator variables; error terms are assumed to be uncorrelated with other variable error terms; and finally data should be interval scale. For detailed discussions on structural equation modelling see Bollen (1989), Kline, (2005), Raykov and Marcoulides, (2006), and Thompson, (2000).

The introductory chapter has provided detailed description of the data. In that chapter, we have identified two different sources of data generated on two sets of variables. The first data on actual prices cover nineteen variables of food items. In this multivariate data, there are no variables, out of the p original variables, that may be designated as dependents variables from the perspective of our study. To make use of SEM, however, there is the need to create such dependent variables. In the first stage, a measurement model should relate observed indicators to latent attributes. The second stage assumes a structural model to explore the effects of latent attributes (price-fixing factors) on the p variables. Specifically, the measurement model is defined as follows:

$$\mathbf{y} = \mathbf{\Lambda F} + \varepsilon , \qquad (3.41)$$

where $\mathbf{y}$ is $q \times 1$ vector of observable indicators; $\mathbf{F}$ is $m \times 1$ vector of latent market forces, which is assumed to follow a multivariate normal distribution $N(0, \mathbf{\Phi})$; $\mathbf{\Lambda}$ is a $q \times m$ matrix of factor loadings; and $\boldsymbol{\varepsilon}$ is $q \times 1$ vector of measurement errors, which is independent of $\mathbf{F}$, and distributed as $N(0, \mathbf{\Psi})$ with a diagonal covariance matrix, $\mathbf{\Psi}$. As shown earlier, we have identified twelve item pricing factors (attributes) with 53 indicators. Thus, $\mathbf{y}$ is known to be 53×1 vector of indicators. However, the actual number, m, of factors will emerge from the confirmatory factor-extraction of the correlations among the 53 indicators. Thus, the matrix $\mathbf{\Lambda}$ and the vector $\mathbf{F}$ are yet to be determined in Chapter Four. Through the measurement model in Equation (3.41), SEM simultaneously accommodates highly correlated explanatory indicator variables without encountering multicollinearity, and measures latent attributes through indicators with different weights, reflecting different contributions of indicators in measuring the latent market forces. This is an attempt to reduce to the barest minimum several indicators by incorporating different market features and other characteristics, along with their importance, thereby reflecting the forces that determine prices more accurately and completely.

Basically, the structural model is defined as

$$\boldsymbol{\eta} = \boldsymbol{\gamma}\mathbf{M} + \boldsymbol{\Gamma}\mathbf{F} + \delta \tag{3.42}$$

with all vectors and matrices having indicated dimensions. The vector $\boldsymbol{\eta}$ is a reduced endogenous variables created from a source of the original $p \times 1$ vector of item variables; $\mathbf{M}$ is $n_{MC} \times 1$ vector of covariates, in this case, of market characteristics; $\boldsymbol{\gamma}$ and $\boldsymbol{\Gamma}$ are matrices of regression coefficients; and δ is a vector of error terms, which is independent of $\mathbf{F}$, and distributed as $N(\mathbf{0}, \boldsymbol{\Psi}_{\delta})$ with a diagonal covariance matrix, $\boldsymbol{\Psi}_{\delta}$. In this study, the dependent vector, $\boldsymbol{\eta}$, may be derived through a link function which will be specified in Chapter Four when the actual form of the factor vector $\mathbf{F}$ is determined.

Equation (3.42) looks like a Mediation model. Mediation is said to occur when the effect of an initial variable on an outcome variable is transmitted through one or more third variables, called mediator or intervening variables (Baron & Kenny, 1986; James & Brett, 1984; MacKinnon, 2008). In this study, the initial set of variables is the item variables. These will give rise to another outcome variable, the price levels with three categories. Then, we determine the price levels in terms of suitable explanatory variables, one of which is given by the factor model given by Equation (3.41).

Chapter Summary

This chapter has examined the main methods employed in the study. It involves three main techniques. These are broadly classified as displaying components techniques and structural equations modelling. The displaying components comprise the principal component and the outlier displaying component. Since the original techniques are based on single multivariate data, the review has made the necessary extensions that would make them suitable for multiple multivariate data employed in this study. Consequently, the chapter has redefined these techniques to be used as time-dependent displaying components as the data problem covers some time periods. The extensions will enable the techniques to highlight extreme observations over the time periods simultaneously. In the development of extensions to the techniques, component scores have been discussed and provision made for conditions for generating these scores that could potentially be problematic for interpretations. These conditions are informed by the literature (Benyi, 2018; Shyam et al., 2016; Schreiber et al., 2006), as well as the wide variations that exist in the data. The review on the outlier displaying component focused on only the 1-ODC since by highlighting a single extreme observation, other potentially extreme observations would lie in its neighbourhood.

The review also covered techniques of general factor analysis and confirmatory factor analysis (CFA). This technique would be useful in identifying exogenous variables that may be used to explain the endogenous variables created from the original item variables. The structural equation modelling is identified as an appropriate technique that could bring together the dependent variable of price levels and the factor model, in addition to other covariates of market characteristics. In the structural model that is envisaged in the study, it becomes necessary to obtain a combined factor effect in the model. This means that the matrix of factors has to be reduced to a single factor vector. A vectorisation procedure has been developed that would convert the factor matrix to a single column vector that retains all the properties of the original matrix. The effect of this single factor vector will therefore reflect the effect of the original factor solution in the structural model.

CHAPTER FOUR

RESULTS AND DISCUSSION

Introduction

In this chapter, we will examine some preliminary features of the dataset used in this study. Then, the methodology outlined in Chapter Three will be applied to study the data in more detail. The techniques of time-dependent principal components and outlier displaying components will be employed to determine the price levels of the markets using the secondary data that cover the actual prices of the food items obtained over the five non-consecutive years.

In the presentation, it will be noticed that factors and principal components are used interchangeably in most cases. As explained earlier, the components are obtained using the loadings as weights as used in factor extraction. This does not mean that we intend to extract factors, since the nature of the data problem does not require a representation based on factors. The presentation will demonstrate why it is not reasonable to use the usual weights of principal components even though the data problem requires the use of this technique.

Having identified the price levels of the markets, we will then determine latent factors that underlie the price levels. These factors will be determined from a confirmatory factor analysis of a primary data set obtained from respondents from selected markets that represent all the price levels. In the last part of the chapter, we will obtain a structural model for determining the price levels in terms of the factor model and other covariates. Since the data could contain outliers, the model will not be used for prediction. Rather, the rationale for the model is to enable us assess the effect of the various variables, including the factors, on price levels.

Identification of Extreme Markets based on Descriptive Statistics

The descriptive statistics of the prices of food items are summarised in Appendix A. It basically shows the mean, standard deviation and coefficient of variation as well as the minimum and maximum prices of all the various food items and the markets where they are obtained over all the five years under consideration. Portions of the results in Appendix A are presented in Table 8. In those descriptive statistics, some important observations could be made regarding markets that are frequently associated with the minimum and/or maximum prices of the respective food items.

Noticeably, Market 68 remains the only least priced market for Maize throughout the period. It is also among the least priced in two other items. Market 17 is also among the least priced in five items. It will therefore not be surprising to identify these two markets among the least priced in the application of the techniques in subsequent sections. On the other hand, Market 65 is among the highest priced in five commodities. Market 69 is also among the highest priced in four commodities over the period. Other markets that are frequently high priced are 53, 14, 55, 70 and 89. There are also markets that are among the least priced in some items but also among the highest priced on some other items. Typical of these are Market 63, 13, and 65.

The observations show that a market cannot be labeled as extremely priced merely on the basis of the descriptive statistics. It will therefore require more advanced techniques to identify which really constitutes an extreme priced market. Similarly, there could be markets that are not identified by the basic statistics that could emerge as important extreme priced markets.

Table 8: Extreme Priced Markets Based on Descriptive Statistics

Commodities	Minimum Priced Market	Maximum Priced Market
Root and Tubers		
Yam white	29, 44, 65, 63, 60	1, 3, 52, 14
Cassava	8,17, 19, 16	49, 65, 51, 89
Plantain (Apentu)	17, 16, 34, 14	59, 71, 46, 65
Gari	16, 44,19, 77, 4	9, 21,86, 10
Vegetables		
Tomato	12,17, 83, 11	48, 60, 70
Garden egg	87, 58, 13, 63	65, 55, 51, 48
Cereal		
Local Rice	6, 17, 19	87, 53, 1, 7
Imported Rice	77, 34, 57	41, 21
Maize	68	77, 10, 35, 53
Oil		
Palm oil	87, 29, 21, 26	63, 79, 40
Fruit		
Orange	39, 3, 63, 6	65, 69
Banana	39, 17, 10, 20	70, 69, 56
Fishes		
Smoked herring	29, 50, 76	69, 55, 67, 71
Koobi	82, 60, 12	23, 22, 78, 7
Egg	19, 40, 1, 6	65, 61, 55
Spices		
Dried pepper	12, 19, 59, 60	43, 69, 89
Onion	68, 86, 63, 21	46, 41, 11, 56
Pulses		
Groundnut red	68, 2, 19	83, 2, 3, 53
Cowpea white	57, 41, 6, 68, 57	78, 21, 13,14

Determination of Level of Prices Using Time-Dependent Principal Components

As pointed out in the methodology, the determination of extreme observations by the Principal Components method could be influenced by the variations in the data. We have already pointed out in the introductory chapter that variations in the prices are widening by the year and therefore, methods that make use of measures that are influenced by variations ought to make provisions for this effect. Now, since the Principal Components are highly influenced by the variations in each of the indicator items, it is rather important to make use of alternatives that reduce the effect of the variations. In Table 9, for example, we have the first two principal components extracted from the variance-covariance matrix of the data on Year 2. It can be noticed that the first principal component (PC1) is influenced by only one variable, Dried Pepper (DPep). This shows that in that year, the variation in the data is most influenced by DPep. Appendix C shows the variance-covariance matrix of the data in that year. In that matrix, we notice that the variation in DPep is not just the highest, but also extremely high. The second component (PC2) is not influenced by any indicator. This means that we cannot extract more than two components for this data. It further means that if this set of PCs is used to determine extreme markets, we will not obtain reasonable results. A possible unreasonable result is that in that year, the market with the highest price in DPep will emerge as outstandingly expensive market predominantly influenced by just one item. Usually, a reasonable component solution that reflects the importance of a single item would appear not as the first component.

Table 9: Principal Components Based on Variance-Covariance Matrix of Year 2 Data

Item	PC1	PC2
Maize	–0.011	0.121
WhYam	0.048	0.474
Cassava	0.027	–0.076
Tomato	0.144	0.317
Gegg	0.020	0.007
DPep	0.969	–0.169
RdGrnt	0.055	0.315
WhCowpea	0.045	0.427
PalmOil	0.018	–0.045
Orange	0.003	–0.035
Banana	0.006	–0.026
SmkHerr	–0.010	–0.265
Koobi	0.024	0.270
Onion	0.143	0.374
Egg	0.002	–0.001
Plantain	0.004	–0.016
Gari	–0.026	0.132
LocRice	0.091	0.197
ImpRice	0.031	–0.016

These results show that we may have to rely on Principal Component Factor Analysis method so that the weights of the items are given by the loadings. In Table 10, we observe the composition of the solution in the form of principal components factor solution up to five components.

Table 10: Components Solution for Year 2 Data

Item	F1	F2	F3	F4	F5
Maize	-0.408	0.613	0.120	0.276	0.195
WhYam	-0.137	0.594	0.435	0.339	0.059
Cassava	0.842	0.085	-0.047	0.053	0.019
Tomato	-0.054	0.802	0.084	-0.189	0.013
Gegg	0.232	0.171	0.109	-0.003	0.845
DPep	0.236	0.239	0.067	-0.741	0.152
RdGrnt	-0.557	0.590	0.276	-0.027	0.091
WhCowpea	-0.351	0.625	0.535	0.083	-0.123
PalmOil	0.583	-0.152	0.452	-0.176	-0.369
Orange	0.819	-0.279	-0.086	0.020	0.231
Banana	0.885	-0.144	-0.043	0.024	0.175
SmkHerr	0.723	-0.418	-0.053	0.002	0.187
Koobi	0.744	-0.273	-0.084	-0.044	-0.140
Onion	-0.165	0.744	-0.137	-0.193	0.040
Egg	0.666	0.403	0.049	-0.108	-0.318
Plantain	0.887	0.008	0.078	0.018	0.094
Gari	-0.148	0.586	-0.340	0.427	0.066
LocRice	-0.148	0.220	0.785	-0.149	0.157
ImpRice	-0.127	-0.098	0.031	-0.683	-0.099

In this solution, the first component, for example, is influenced by the Fishes, Oil, Cassava and Plantain, Fruits in contrast to Red Groundnuts (RdGrnt). Thus, the identification of the level of prices in a market based on Component would be based on these food groups. Dried Pepper, which is the most important food item in terms of variation is recognised along the fourth most important component. The component solutions for all the five years are presented in Appendix D. In that Appendix, we extract the main constituents of the dimensions given by the principal components for all the five years. This is given in Table 11.

Table 11: Dimensions of the Extracted Time-Dependent Principal Components

	Year				
PC	1	2	3	4	5
1	Root and tubers, Gegg, Fruits, Fishes	Root and Tubers, Fruits, Oil, Fishes,	Maize, Yam,Veg/ Fruits,Fishes	Maize, Yam,Veg, Spices, Pulses	Roots-Tubs, Fruits, Fishes
2	Cereal, Yam, Tomato, Pulses	Maize, Yam, Tom, Gari, Spices, Pulses	Spices, Pulses	Oil, Egg	Tom, Onion,
3	Spices	Pulses, LocRice	Root-Tubs, Veg, Egg	Root-Tubs, Gegg	Cowp, Fishes, Rc
4	Oil/Gari	DPep, ImpRc	Pulses, Oil, Koobi, LocRc	Onion, Gari, locRc	DPep, Pulses, Gari
5	ImpRc	Gegg	Rice, Gari	Fishes	Oil, Egg

It can be seen from Table 11 that the first PC for almost all the years is dominated by Fruits and Fishes. Roots and Tubers is another class of items that feature on the first PC. If we consider the first and the second PCs together, then Fruits, Fishes, Roots and Tubers and Cereals constitute the main dimensions of food prices. It is worth noting that on higher order PCs, they are mostly influenced by few or single items. It is therefore not practically useful to extract components beyond order five. In the table, the '/' has been used to denote a contrast. For example, the first PC in Year 3 is determined by the items Maize, Yam, Veg/ Fruits, Fishes. This means that there is a contrast between the set of items {Maize, Yam, Veg} on one hand, and the set {Fruits, Fishes}. This further means that high priced markets identified by the first PC in that year are those that are expensive in the set {Maize, Yam, Veg} but low priced on the set {Fruits, Fishes}, and the vice versa.

In Chapter Three, we have constructed the time-dependent principal components. These components are f_{ik}, $i = 1, 2, \ldots, m$; $k = 1, 2, \ldots, r$, and constitute an m-factor solution for each of the r years. Then on the m PCs, we obtain the $nr \times m$ matrix of scores

$$\mathbf{f} = [\mathbf{f}_{1k,}\ \mathbf{f}_{2k,}\ \ldots,\ sign_{k\in I}\mathbf{f}_{ik,}\ \ldots,\ \mathbf{f}_{mk}]$$

as defined in Chapter Three. Now, we note that beyond $m = 3$ component solution, each of the remaining components are influenced by few items; in the case of $k = 1$, the fourth component is influenced by only two items, whilst the fifth is influenced by only one (as seen in Table 10). Since this type of components do not help in the general solution, as seen in the literature, the matrix of component scores will be generated based on component solution containing not more than five components.

Figures 6, 7 and 8 are the plots of the time-dependent component scores for the first, second and third components, respectively. In each case, it reveals the potentially extreme as well as the moderate markets in each year based on the items that constitute the components. In Figure 6, we see that generally, prices are not noticeably spread out. We can also identify Market 65 in Year 1 and Year 4 as a suspect outlier. It is worth noting that this market is expensive in one year but least priced in another year. Although in each year, there are extreme markets, they are not as noticeable as Market 65 in the two identified years. For example, in Year 1, the next highest priced markets are 68 and 69, which are also the highest priced in Year 2.

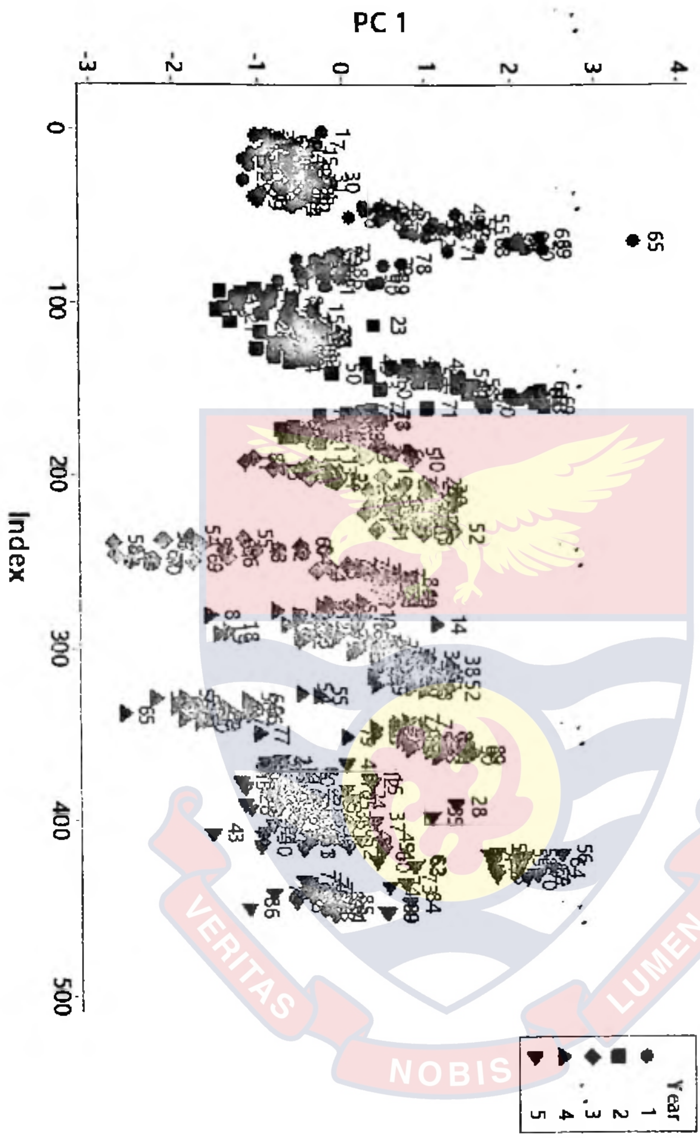

Figure 6: Plot of Scores of Component 1 for All Years.

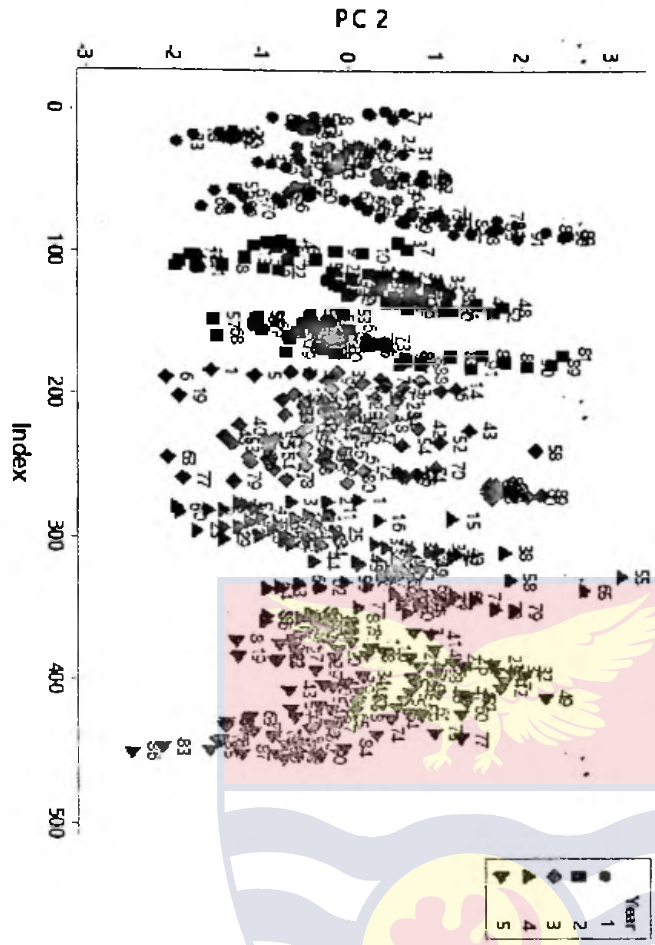

Figure 7: Plot of Scores of Component 2 for All Years.

In Figure 7, we see that generally, prices are more varied in each of the years on the second component. Extreme observations are therefore easier to identify. It can be seen that Markets 90 and 89 are extreme in Year 1 and Year 2 in addition to 81. Markets 55 and 65 are visibly the most high priced in Year 4.

In Figure 8, we see that generally, prices are more visible in the extremes. It can be seen that Market 14 is clearly high priced in Year 1 and Year 5. In Year 5, Market 83, is also high priced. Market 54 is high priced in Year 2, 46 in Year 3. Markets 65 and 48 also emerge as the lowest priced in Year 4.

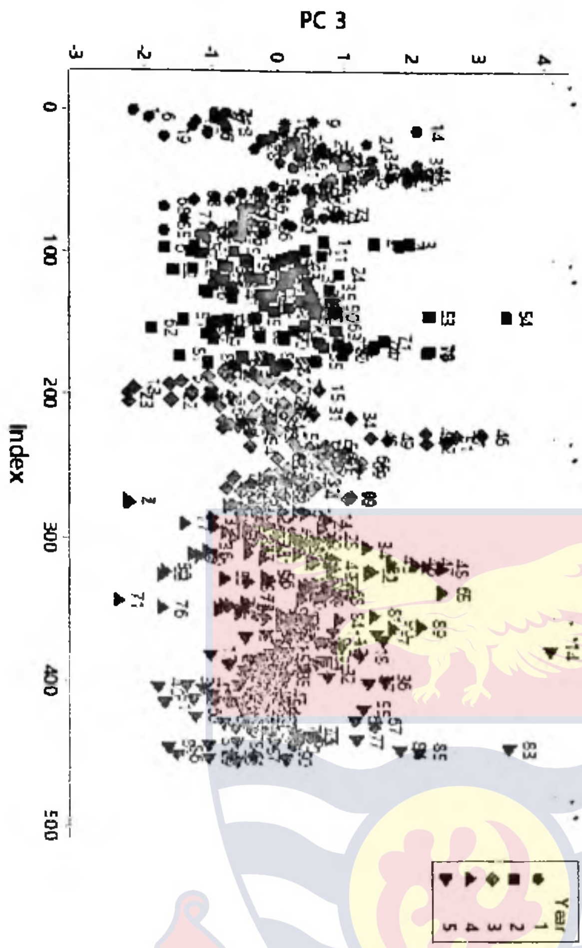

Figure 8: Plot of Scores of Component 3 for all Years.

The results in the plots show that only a few markets in the country may be classified as extreme among several markets examined. All the remaining markets are thus classified as being moderately priced. Examination of all component plots for the first five components in addition to our observations in the descriptive analysis in Table 8 gives the set of suspect outlying markets as $L = \{65, 69, 13\ 68, 54, 89, 90, 17, 56, 46, 14, 83, 55, 47\}$.

It is noticed that some markets could be high priced and low priced even in the same year on different dimensions of food items. Typically, Market 65 in Year 4 exhibits such feature. It remains to determine how such a market would be generally classified, as low priced or high priced. In the next section, we employ the technique of Outlier Displaying Components which hopefully would resolve the incidence of multiple price-level tag of certain markets.

Detection of Extreme Priced Markets Using Outlier Displaying Components

In this approach to determining extreme markets, we will make use of the three different procedures, with each procedure distinguished by the nature of the variance-covariance matrix used in the approach. These are attempt at highlighting as much as possible all possible suspect extreme markets that may be present in the set of markets examined in the study. We shall denote the set of suspect extreme observations identified in the earlier section and given as $L = \{65, 69, 13\ 68, 54, 89, 90, 17, 56, 46, 14, 83, 55, 47\}$ as $L = \{l_1, l_2, \ldots, l_\upsilon\}$ for the discussion in this section.

The data in the kth year has been represented as $\mathbf{X}_k = (\mathbf{X}_{1k}, \mathbf{X}_{2k}, \ldots, \mathbf{X}_{pk})$ on p variables. The representation of the data for specific portions of years is key in the detection procedure. In MATLAB, we will represent this portion of the data for the kth year as

$$\mathbf{X}_k = \mathbf{X}((k-1)*n+1 : k*n, 1 : p). \tag{4.1}$$

Thus, in each year, the size of the data is $n\ (= 91)$. The mean matrix of the data in Equation (4.1) is given by the $p \times r$ matrix

$$\bar{\mathbf{x}} = [\bar{\mathbf{x}}_1 \ \ \bar{\mathbf{x}}_2 \ \ldots \ \bar{\mathbf{x}}_k \ \ldots \ \bar{\mathbf{x}}_r], \tag{4.2}$$

where $\bar{\mathbf{x}}_k$ is the mean vector for the data in Year k. Now in the entire dataset of $nr \times p$ matrix, it is also important to determine the representation of the individual observation market. Observations on each market, ε also constitute a matrix which is represented as

$$\mathbf{X}_{p\times r}^{(\varepsilon)} = \left[\mathbf{X}(\varepsilon,:)' \ \mathbf{X}(\varepsilon+n,\ :)' \ \mathbf{X}(\varepsilon+2*n,:)' \ldots \mathbf{X}(\varepsilon+(r-1)*n,:)'\right]. \qquad (4.3)$$

Now, since the number of years is 5, the pooled sum of squares and cross-product matrix becomes

$$\mathbf{S} = \sum_{j=1}^{5}\left(\mathbf{X}'_j(\omega) - \bar{\mathbf{x}}_j(\omega)\mathbf{1}'\right) * \left(\mathbf{X}'_j(\omega) - \bar{\mathbf{x}}_j(\omega)\mathbf{1}'\right)' \qquad (4.4)$$

We now obtain the matrix of Mahalanobis distance of any observation $(\pi_j,\ j = 1, 2, \ldots, \upsilon)$ from $\bar{\mathbf{x}}$ given by the projection $\mathbf{Y} = \mathbf{X}_{nr\times p}\boldsymbol{\beta}_{\pi_j,p}$, where $\boldsymbol{\beta}_{\pi_j}$ is the 1-Outlier Displaying Component assuming π_j is a suspect outlier and given by $\boldsymbol{\beta}_{\pi_j} = \mathbf{S}^{-1}(\mathbf{x}_{\pi_j} - \bar{\mathbf{x}})$. As explained in the methodology, it is important to identify the exact location of the projection for a particular π_v in the matrix $\mathbf{Y}$, which is very large. This sub-matrix is given by

$$\mathbf{y}_{p\times r}^{(\pi_v)} = \left[\mathbf{Y}(\pi_v,\ 1)\ \mathbf{Y}(\pi_v+n,\ 2)\ \mathbf{Y}(\pi_v+2*n,\ 3) \ldots \mathbf{Y}(\pi_v+(r-1)*n,\ r)\right]$$

The projection of observations $\pi_j \in L$ based on the total sample variance-covariance matrix is given in Table 12 for twelve of them, whiles in Table 13, we have the projections based on the pooled variance-covariance matrix.

Table 12: Projected Values of Suspect Outlying Markets from the Mean Based on Total Sample Variance-Variance Matrix

No.	π_j	Year 1	2	3	4	5
1	65	13.9311	3.0370	13.8802	263.9890	47.7801
2	69	6.6609	23.7912	11.0015	13.9365	58.8302
3	13	1.9278	1.3535	19.6604	16.1751	27.4488
4	68	2.7572	21.6740	7.7632	13.0138	58.9821
5	54	3.9404	7.6739	16.7003	11.0493	37.3578
6	89	4.2535	3.7267	15.0472	25.1642	23.7623
7	90	4.7164	3.0395	14.4356	20.6649	49.0961
8	17	1.9614	0.6981	8.4387	14.5450	55.0455
9	56	5.2528	3.0540	24.3344	30.9917	64.3164
10	46	1.8222	6.4996	17.8794	23.3667	20.0537
11	14	4.9814	2.6910	16.4266	52.9580	138.6904
12	83	2.1387	1.8405	9.2228	3.4933	114.7280

Source: Ministry of Food and Agriculture, Ghana

From Table 12, we observe that Market 65 in Year 4 is the most extreme priced market in all the years. Markets 14 and 83 in Year 5 also show extremely large distances indicating high prices levels in that year. It is also noticeable that Market 17 is quite consistently among the lowest distances each year, an indication of consistently low price levels over the years, perhaps with exception of the last Year 5. It must be observed that the extreme disparity in the distances is obviously as a result of the wide variations in prices in particular years.

Table 13: Projected Values of Suspect Outlying Markets from the General Mean Based on Pooled Sample Variance-Covariance Matrix

No.	π_j	Year 1	2	3	4	5
1	65	0.0889	0.0147	0.0488	0.5605	0.1112
2	69	0.0320	0.0538	0.0076	-0.0096	0.1256
3	13	0.0011	-0.0058	0.0252	0.0040	-0.0263
4	68	0.0047	0.0402	-0.0093	-0.0251	0.0818
5	54	0.0085	0.0299	0.0276	0.0120	0.0411
6	89	0.0282	0.0273	0.0542	0.0829	-0.0465
7	90	0.0224	0.0222	0.0566	0.0743	0.1814
8	17	-0.0003	-0.0035	-0.0205	-0.0227	0.0642
9	56	0.0309	0.0221	0.0508	0.0721	0.2349
10	46	0.0132	0.0191	0.0784	0.0719	0.0201
11	14	0.0174	0.0054	0.0344	0.1245	0.3768
12	83	0.0109	0.0093	0.0155	-0.0005	0.3061
13	55	0.0151	0.0219	0.0624	0.1424	0.2570
14	47	0.0142	0.0175	0.0699	0.0818	0.1005

Source: Ministry of Food and Agriculture, Ghana

Table 13 shows that projected values have drastically reduced as a result of the effect of the pooled variance-covariance matrix. This is an indication that within each year, variations in prices are quite low compared to variations across years. In this table, it is clearer that the prices of items in Market 17 remain consistently lower than the average prices across all years. The exception is in Year 5. Markets 13 and 68 are the markets with next lower prices. Market 65 has also remained quite consistently high priced market.

The observation matrix $\mathbf{X}_{p\times r}^{(\pi_v)}$ is not extreme in all r levels in the vector $\mathbf{y}_{p\times r}^{(\pi_v)}$. Suppose that it is particularly extreme in the value $\mathbf{Y}(\pi_v+(k-1)*n,\ k)$ in the kth year. We then project this value among all values in that year using the Modified One Outlier Displaying Component, M1-ODC, given by $\boldsymbol{\beta}_{\pi_v} = \mathbf{S}_{(\pi_v)}^{-1}(\mathbf{x}_{\pi_v} - \overline{\mathbf{x}}_{(\pi_v)})$.

Since each observation in the data constitute a multivariate dataset, it is important to identify the general expression for the data matrix $\mathbf{x}_{(\pi_v)}$ with observation π_v deleted. We delete the outlying multivariate time-dependent observation of the form in Equation (4.3) from the entire dataset $\mathbf{X}_{nr\times p}$ by the following procedure: First, we set the original dataset $\mathbf{X}_{nr\times p}$ to $\mathbf{G}$, say. Then we delete every nth observation beginning from $\mathbf{x}_{\pi_v}$ as follows:

$$\mathbf{X} = \mathbf{G};$$

$$\mathbf{G}(\pi_v, :) = [\ ];$$

$$\mathbf{G}(\pi_v + n - 1, :) = [\ ];$$

$$\mathbf{G}(\pi_v + 2*n - 2, :) = [\ ];$$

$$\mathbf{G}(\pi_v + 3*n - 3, :) = [\ ];$$

$$\vdots$$

$$\mathbf{G}(\pi_v + (k-1)*n - (k-1), :) = [\ ];$$

$$\vdots$$

$$\mathbf{G}(\pi_v + (r-1)*n - (r-1), :) = [\ ];$$

The data is now reduced to a size of $(n-r)$ with each year containing $(n-1)$ multivariate data points. The reduced data $\mathbf{X}_{(\pi_v)}$ after deleting $\mathbf{x}_{\pi_v}$ from the kth year is given as

$$\mathbf{X}_{k(\pi_v)} = \mathbf{G}\big((k-1)*(n-1)+1 : k*(n-1), 1:p\big) \tag{4.5}$$

The mean vector of the reduced data is represented as

$$\bar{\mathbf{x}}_{(\pi_v)} = [\bar{\mathbf{x}}_{1(\pi_v)} \;\; \bar{\mathbf{x}}_{2(\pi_v)} \;\ldots\; \bar{\mathbf{x}}_{k(\pi_v)} \ldots \bar{\mathbf{x}}_{r(\pi_v)}]$$

We then compute the SSCP matrix from $\mathbf{X}_{k(\pi_v)}$ as

$$\mathbf{S}_{k(\pi_v)} = \left(\mathbf{X}'_{k(\pi_v)} - \bar{\mathbf{x}}_{k(\pi_v)} * \mathbf{1}'\right) * \left(\mathbf{X}'_{k(\pi_v)} - \bar{\mathbf{x}}_{k(\pi_v)} * \mathbf{1}'\right)'$$

where $\mathbf{1}$ is a column vector of ones of dimension $(n-1)\times 1$. Therefore, the pooled SSCP matrix similar to that in Equation (4.4) is given as

$$\mathbf{S}_{(\pi_v)} = \sum_{j=1}^{r}\left(\mathbf{X}'_{j(\pi_v)} - \bar{\mathbf{x}}_{j(\pi_v)} * \mathbf{1}'\right) * \left(\mathbf{X}'_{j(\pi_v)} - \bar{\mathbf{x}}_{j(\pi_v)} * \mathbf{1}'\right)' \; . \qquad (4.6)$$

The projection of suspect observation are obtained for all observation π_j $(j = 1, 2, \ldots, \upsilon)$ from $\bar{\mathbf{x}}_{(\pi_v)}$ as $\mathbf{Q} = \mathbf{X}_{(\pi_v)}\boldsymbol{\beta}^{(\pi_v)}_{\pi_v}$, where $\boldsymbol{\beta}^{(\pi_v)}_{\pi_v}$ is the 1-Outlier Displaying Component given by $\boldsymbol{\beta}_{\pi_j} = \mathbf{S}^{-1}_{(\pi_v)}\left(\mathbf{x}_{\pi_v} - \bar{\mathbf{x}}_{(\pi_v)}\right)$. The exact location of the projection of $\mathbf{x}_{\pi_v}$ in the matrix $\mathbf{Q}$ is the sub-matrix given by

$$\mathbf{q}^{(\pi_v)}_{p\times r} = [\mathbf{Q}(\pi_v, 1)\; \mathbf{Q}(\pi_v + n, 2)\; \mathbf{Q}(\pi_v + 2*n, 3) \ldots \mathbf{Q}(\pi_v + (r-1)*n, r)]$$

.

The projection of observations in $\mathbf{Q}$ based on the pooled sample variance-covariance matrix is given in Table 14 for all fourteen of them.

Table 14: Projected Values of Suspect Outlying Markets from Reduced Mean Based on Pooled Reduced Sample Variance-Covariance Matrix

		Year				
No.	π_j	1	2	3	4	5
1	65	0.0967	0.0161	0.0540	1.4144	0.1443
2	69	0.0341	0.0573	0.0090	-0.0085	0.1490
3	13	0.0010	-0.0061	0.0261	0.0039	-0.0293
4	68	0.0044	0.0422	-0.0107	-0.0267	0.0928
5	54	0.0089	0.0315	0.0287	0.0124	0.0460
6	89	0.0291	0.0281	0.0579	0.0900	-0.0471
7	90	0.0232	0.0232	0.0624	0.0837	0.2093
8	17	-0.0007	-0.0044	-0.0225	-0.0257	0.0704
9	56	0.0329	0.0234	0.0609	0.0789	0.2821
10	46	0.0136	0.0198	0.0831	0.0773	0.0224
11	14	0.0188	0.0056	0.0385	0.1574	0.5466
12	83	0.0115	0.0096	0.0155	-0.0010	0.3990
13	55	0.0156	0.0225	0.0799	0.1586	0.3415
14	47	0.0145	0.0181	0.0750	0.0891	0.1123

Source: Ministry of Food and Agriculture, Ghana

The values in Table 14 are slightly higher than those in Table 13. As explained in the methodology, the projections in Table 14, which is based on the pooled sample variance-covariance matrix and involves the mean of the data with the suspect outlying observation removed, gives more realistic representation of outlyingness.

A graphical representation of the projections in **Q** based on $\pi_v = 65$ is given in Figure 9. It is a simultaneous display of the projections based on Market 65 for all years in the same scale.

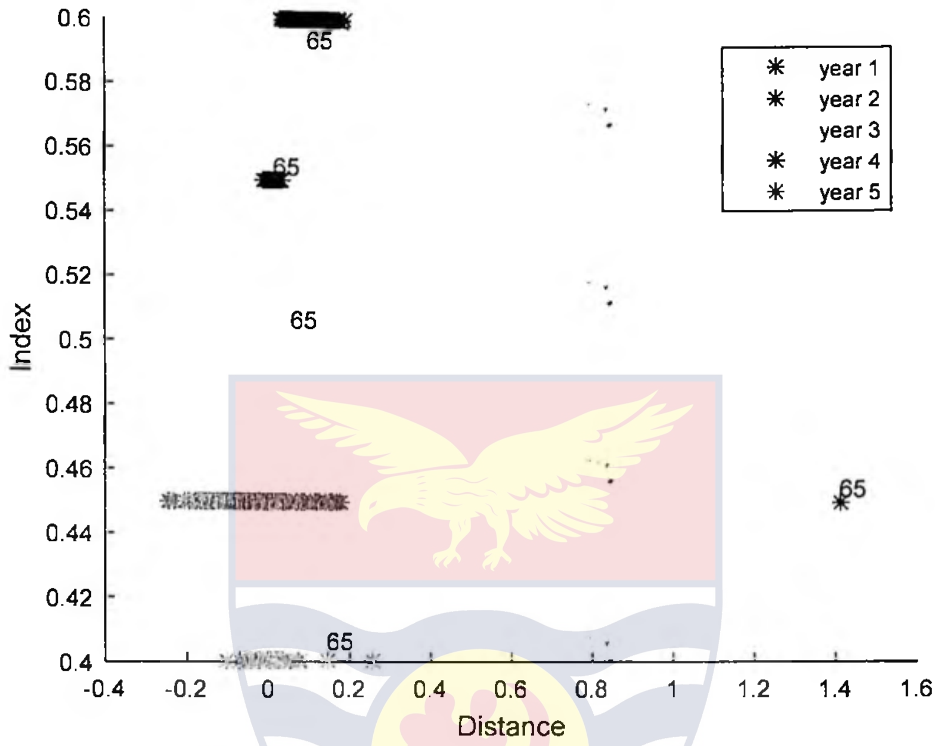

Figure 9: Simultaneous Projection of Prices in all Five Years Based on Market 65.

Figure 9 shows that the variation in prices across the market varies from year to year with reference to 65. It shows that the spread of the prices in Year 4 is much wider compared to the other years on the same scale. It identifies Market 65 as the most outlying in that year. However, that market is not extreme in the other years. It should be pointed out that the spread in the other years appears more compact as a result of the presence of 65 in Year 4, and the generally varied prices in that year. Secondly, the general prices in 65 is just about the same or slightly higher than those in the other years.

Similarly, Figure 10 Shows that the Simultaneous Projection of Prices Based on Market 17.

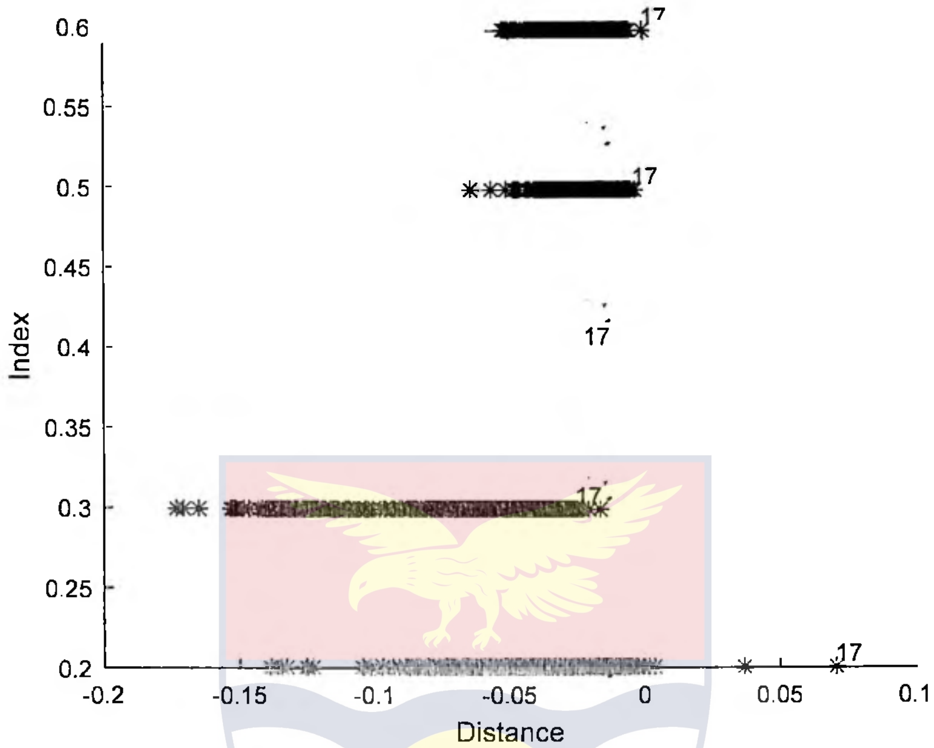

Figure 10: Simultaneous Projection of Prices in all Five Years Based on Market 17.

With reference to observation 17, the markets are more spread out in the same scale. The spread means that prices in 17 may be substantially lower than those of all other markets. It identifies observation 17 as becoming quite high in the last year.

The actual spread in each year may be ascertained by separate plots as in Figures 11 and 12.

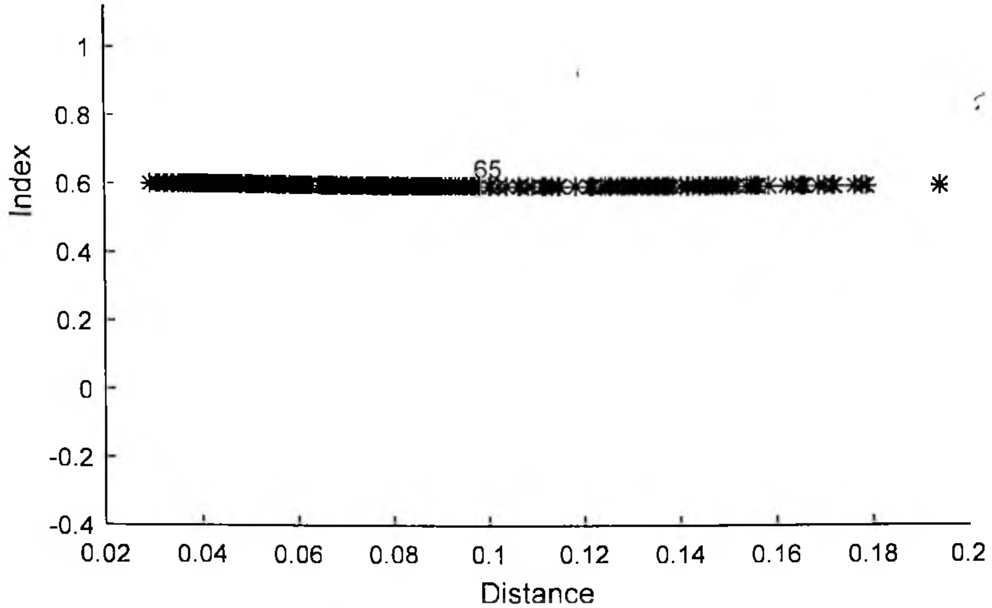

Figure 11: Projection of Prices in Year 1 Based on Market 65.

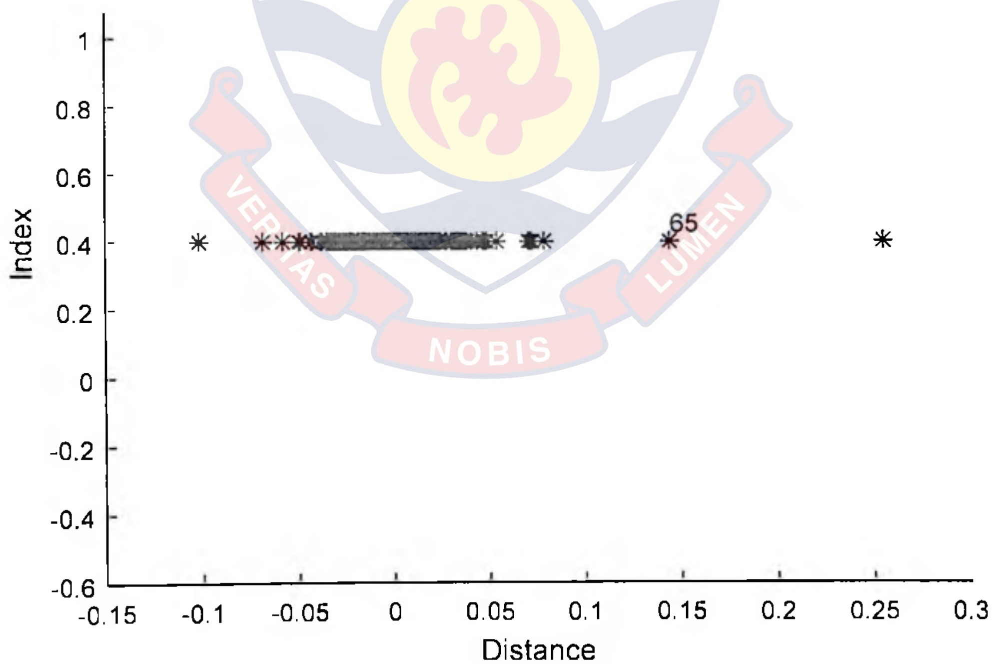

Figure 12: Projection of Prices in Year 5 Based on Market 65.

Extraction of Factors that Underlie Market Price Levels

The actual prices obtained for this study must be influenced by some latent factors in the markets. As explained in the introductory chapter, fifty-three indicators of these factors have been identified and measured on a five-point Likert scale. The data is obtained from nine markets that are determined to be typically of specific price level categorisation in the previous section. From each of the markets data is obtained from 40 respondents who are traders of different commodities. The selected markets with their features are given in Table 15. We know that each market is already known to have a designated categorisation. This type of data creates problem of complete separation or quasi-complete separation of data points. With such a problem, the model cannot be fit as the maximum likelihood estimates of parameters may not exist. To overcome this problem in the data processing, a few of the observations from each market is treated as though they came from some other markets with different descriptions. The table therefore contains a column for treatment for complete separation. For example, in the first market (79) in the table, three of the 40 respondents are designated as coming from other markets: 2 from a low priced (LP) market in Upper West (UW), and 1 from LP market in Upper East. This treatment is guided by the value of the combined factor for that respondent that will be used in the final structural model. In this case therefore, three respondents in Market 79 in Volta obtained a score that is similar to typical scores in Low priced market in UW and UE. It should be noted that only few of such modifications are permitted that are just enough to overcome the problem of complete separation or quasi-complete separation.

Table 15: Markets and their Description Selected for Factor Extraction.

SN	Market	Initial Mkt Type	Region	No. of Respts	Treatment for Complete Separation	Loc
1	79	Mod	Volta	37	2 LP UW; 1 LP UE	Urban
2	45	Mod	GAccra	35	5 HP Western	Urban
3	40	Mod	Eastern	35	2 LP, UE; 2 HP, BA; 1 LP, BA	Urban
4	13	Low	BA	40	-	Urban
5	17	Low	BA	32	7 LP, GA; 1 Mod UE	Urban
6	68	Low	UW	35	5 LP Volta	Rural
7	90	High	Western	37	3LP, GA	Urban
8	14	High	BA	36	4LP, ER	Rural
9	65	High	UE	38	2 LP, Volta	Rural

In order to determine meaningful labels of the factors, we first examine the distribution of the responses on these indicators. The percentage distribution of the responses is given in Table 16. It can be observed that indicators with high percentage distribution in the positive direction are largely in the set {2, 5, 6, 7, 8, 24, 38, 39, 43}. Thus, there is an overwhelming support for these considerations for fixing prices. These indicators therefore be described as popular indicators. On the other, there are others on which responses are largely in the negative. These are found in the set {12, 13, 19, 20, 30, 33, 34, 36, 41, 42, 46, 48, 51, 52}. These will constitute unpopular consideration for fixing prices. There is the third set of indicators that reflect almost as much support for the consideration as rejection against it. These are indicators that have almost equal distributions in the opposite directions. This set is constituted as {1, 18, 25, 31}. This is a class of controversial considerations. There is yet another group that has a semblance of controversial consideration, but is not clearly expressed as described. This group is usually manifest in large percentage of response on the middle of the scale. These are seen on the indicators {9, 14, 26, 44, 47, 50}. These four categories of considerations constitute about 62% of the indicators.

Table 16: Percentage Distribution of Responses on Indicator Variables.

	Indicator	SD	D	NS	A	SA
1	I consider the prices of traders of similar product in fixing my prices	11.9	31.1	5.8	27.5	23.6
2	I pay attention to the source from which I obtain the food items in my pricing	1.4	3.3	2.2	41.7	51.4
3	The prices of my items should enable me dispose of them as quickly as possible	0.8	6.1	9.2	57.5	26.4
4	In order to bring in fresh stock, I price my items a little low in order to re-coup my capital quickly	3.1	16.9	11.1	42.2	26.7
5	I always ensure that my items are fresh to enable me charge good price	1.7	5.0	6.9	50.6	35.8
6	I insist on selling products of highest quality to enable me keep my customers	0.6	4.4	3.9	46.7	44.4
7	I must not lose my customers due to high prices	1.7	5.0	9.2	52.2	31.9
8	I will go every length to obtain quality product for my customers	1.4	3.3	11.9	49.7	33.6
9	My customers are always prepared to pay any reasonable price for quality product	1.9	10.3	26.7	39.4	21.7
10	I usually price my items low if I obtain my items over a short distance	5.6	24.2	14.4	38.6	17.2
11	I fix my prices taking into consideration the price levels agreed on by market association	28.3	38.3	5.6	15.6	12.2
12	Market tolls and other market services I require for my business do push my prices quite high	31.9	27.8	6.4	24.7	9.2
13	Rent for market space makes me fix my prices a bit high	39.2	30.6	4.7	20.3	5.3
14	My prices are always among the lowest irrespective of where I obtain the items	10.0	28.6	18.1	35.8	7.5
15	I particularly make sure that prices of perishable items are always among the least priced in the market to enable me dispose of them as quickly as possible	2.5	20.0	11.4	43.9	22.2
16	I would want to keep my customers by making sure that my prices are low	3.1	13.6	11.1	56.9	15.3
17	Since I must have constant supply from source of production I make sure I dispose of my items quickly at competitive price	1.9	22.5	8.9	51.1	15.6
18	Since I do not have storage facility to protect the freshness of my products I am compelled to sell them at low prices	12.2	36.7	11.1	30.6	9.4
19	I do not really care about how customers react to my prices because I am the only person who sells such items	34.4	56.9	3.3	3.6	1.7

Table 16 Continued

	Indicator	SD	D	NS	A	SA
20	I am not really worried about how customers react to my prices because I am located at a convenient place for customers to reach	28.9	55.0	4.7	8.9	2.5
21	I am tactful in dealing with my customers since they will find the same items elsewhere at almost the same price	4.7	8.1	3.1	51.4	32.8
22	The prices of my items are supposed to be within some amount of the production price	1.7	13.9	10.0	56.4	18.1
23	My items are usually priced high if I obtain them from a long distance	3.3	18.3	12.8	38.1	27.5
24	I price my items quite reasonable so as to attract more customers	1.4	6.9	5.3	63.1	23.3
25	I consider the prices of similar items from different market around to fix my prices	12.5	34.7	11.4	23.3	18.1
26	My customers will always buy from me no matter how much I sell my items since they know they can get the best from me	4.7	18.9	23.6	38.1	14.7
27	I am not so much bothered about my profit margin as long as I do not run at a loss	7.8	23.1	8.6	47.5	13.1
28	I have to fix my prices to enable me stay in business	1.9	8.3	8.3	57.5	23.9
29	If items are generally in high supply in the market I usually price my items according to the prevailing prices around the market	4.2	15.3	8.1	45.3	27.2
30	My prices are fixed even if customers buy in bulk	15.8	51.7	6.4	15.6	10.6
31	I am willing to fix my prices much lower than usual regardless of profit margin when I am dealing with first time customers	4.7	37.8	10.6	37.2	9.7
32	I usually fix my prices according to the prevailing prices around the market when I am dealing with my regular customers	4.2	23.9	12.2	40.3	19.4
33	I am not too careful about my price levels when I am dealing with my regular customers	32.8	39.4	4.4	14.4	8.9
34	I always have in mind some particular customers when fixing the prices of my items	28.9	43.1	11.7	12.2	4.2
35	I do not usually have fixed prices for my items; the price depends on the customer I am dealing with	25.3	33.3	9.4	23.9	8.1
36	I really do not have fixed price; my prices normally depends on how the customer bargains	24.4	32.8	8.9	25.6	8.3
37	I sometimes have to vary my prices in order to meet prevailing market conditions	3.6	15.8	11.4	45.8	23.3
38	I usually price my items quite low if I sell in bulk	4.7	13.9	5.3	48.1	28.1

Table 16 Continued

	Indicator	SD	D	NS	A	SA
39	If items are generally in short supply in the market I usually price my items quite high	4.2	16.1	11.9	27.8	40.0
40	I sometimes sell out items virtually for free to some particular customers in order to maintain my relations with them	20.8	31.9	5.8	26.7	14.7
41	An important component of prices of my items should take care of my market bills	33.9	31.7	5.8	22.8	5.8
42	I include payment for market security in my pricing	48.9	33.9	5.6	7.2	4.4
43	One major component of my prices is the cost price at which I obtain my items from source	1.9	5.3	1.9	32.5	58.3
44	If the roads to the source where I obtain my items are so bad, then I have to increase my prices to enable me take care of huge transport cost	5.0	19.2	15.3	41.9	18.6
45	I make sure that my prices are usually fixed above just breaking even in order to attract more customers	3.3	15.6	15.0	54.2	11.9
46	I am not too bothered about my price levels since my items are always in high demand	11.1	58.1	12.2	15.6	3.1
47	Prices of my items are just the same as what is in the market	4.4	16.9	15.3	39.7	23.6
48	Since we pay substantially for cleaning the market, I am compelled to include this in fixing prices	34.4	40.6	5.6	14.2	5.3
49	My items are relatively low priced because I obtain them from production source which has moderate prices	7.8	25.3	14.4	41.9	10.6
50	I package my items very attractively so customers are not too careful about my price levels	6.9	16.9	21.4	45.3	9.4
51	If I anticipate many customers in the day, my prices are usually lower	15.6	54.2	14.7	13.6	1.9
52	I insist on the required prices irrespective of the perishableness of my items	15.8	44.7	10.8	21.4	7.2
53	I find it more prudent to sell off my items at reduced prices than to allow them to lose their quality over time	4.2	10.8	4.4	48.1	32.5

Source: Survey Data, 2017

The Kaiser-Meyer-Olkin's measure of sampling adequacy of the data on these indicators is found to be 0.726. This indicates that the data is quite factor-suitable. To proceed to determine the factor model, we are guided by two issues raised in the methodology regarding factor extraction that is necessitated by the structure of this particular data. Since the indicators are many, there is the need to examine different solutions that give plausible interpretation as well as 'non-significant' solution that fits the data. Therefore in Table 17, we examine a ten-factor solution, and in Table 19, we examine a twelve factor solution. In the ten-factor solution in Table 17, all the factors are influenced by at least two indicators. As pointed out in the literature, this will help avoid the incidence of one-indicator factors in the solution.

In the table, we notice that for the ten-factor model, the first three factors have four or more indicators, the fourth to the sixth have three indicators, whilst the seventh to tenth have exactly two indicators. Thus, in this model, 40% of the factors have two influential indicators.

Table 17: Ten-Factor Solution for Price Level .

	Component									
	1	2	3	4	5	6	7	8	9	10
1	-.016	.028	.701	.017	.170	.057	.048	.019	-.017	.058
2	.505	-.167	.005	-.153	.012	-.129	-.069	.034	.038	.227
3	.330	.012	.187	-.036	.194	.006	.172	.231	.190	-.024
4	.075	.119	.048	.123	.673	-.059	.167	-.048	.033	.028
5	.659	-.126	.016	-.050	.134	.087	.230	.099	.232	-.128
6	.644	-.108	-.011	-.025	.018	.138	.124	-.077	.147	-.017
7	.597	.109	.157	.138	.111	.020	-.037	.155	-.151	-.059
8	.670	.045	.039	-.035	-.070	.098	.044	-.058	.096	.072
9	.264	.116	.179	-.118	.100	.383	.018	.193	.259	.060
10	.028	.020	.054	.063	.084	.394	-.104	.085	.489	.164
11	.034	.190	.674	-.199	-.062	.241	.134	-.029	.157	.077
12	.015	.790	.105	.070	.026	.086	.073	.047	.186	-.032
13	.137	.735	-.035	-.011	.081	.216	-.076	-.053	-.082	-.141
14	.073	-.288	-.191	.012	.263	.092	.151	.584	-.111	-.103
15	.034	.063	-.031	.104	.742	.202	.030	.032	-.033	-.033
16	.281	.239	-.031	.205	.387	.098	-.110	.369	-.060	.061
17	-.047	.205	.237	.104	.413	.081	-.222	.290	-.068	.108
18	.044	.007	.151	-.121	.630	.086	-.039	-.161	.192	.310
19	.099	.015	.088	-.065	.102	.007	.724	.067	.133	-.015
20	.084	-.035	.051	.069	.035	.054	.658	.075	-.128	-.002
21	.450	-.085	-.039	.060	-.080	.201	-.339	-.102	-.436	.024
22	.094	.297	.056	.228	-.037	-.229	-.034	-.064	.047	-.326
23	.328	.176	-.042	.137	-.115	.066	-.046	-.218	.604	.025
24	.550	.171	.178	.158	.171	-.088	-.153	.165	.083	-.140
25	.165	.073	.746	.095	.145	.036	.026	-.179	-.049	.167
26	.184	.037	-.035	-.182	-.037	.558	-.023	.276	-.177	-.072
27	-.060	.097	.230	-.067	.103	.243	-.219	.160	.088	.507
28	.404	.259	-.085	.141	-.077	.058	-.132	.243	-.227	.043
29	.247	-.022	.476	.191	-.092	.079	-.157	.051	-.001	-.044
30	-.166	.185	-.006	-.280	-.278	-.143	.323	.122	-.122	.397
31	-.067	.156	.117	.381	.197	.531	-.044	-.001	.135	-.043
32	-.021	.105	.488	-.029	-.012	.386	-.175	.262	.048	.165
33	-.055	.215	.108	.174	.017	.602	.099	-.102	-.044	-.072
34	-.057	.163	.017	.508	.072	.128	.127	-.091	-.213	.375

Table 17 Continued

35	-.030	.081	.012	.838	.083	-.040	.091	.093	.064	.005
36	.011	.189	-.074	.812	.036	.034	.024	.108	.088	.057
37	.204	.047	.127	.352	.248	.190	-.221	.012	.229	-.012
38	.414	-.036	-.080	.377	.218	.206	-.044	-.183	.093	-.198
39	.346	.190	-.195	.298	.165	.132	-.156	-.188	-.232	-.259
40	.142	.025	.014	.045	.158	.606	.047	-.154	.031	.279
41	-.134	.731	.128	.109	.085	.038	.015	.034	.242	-.019
42	-.087	.488	-.112	.128	.010	-.144	.475	-.033	-.102	.200
43	.410	-.256	.074	-.159	.035	-.232	-.270	.126	.082	-.015
44	.131	.193	-.200	.095	.020	-.100	-.146	-.020	.556	.078
45	.251	-.062	.156	-.045	.162	-.011	-.414	.108	.163	.297
46	-.198	.035	-.162	.139	-.239	.119	.348	-.028	-.242	.092
47	-.061	-.015	.614	-.117	-.092	-.244	.014	.016	-.188	-.148
48	-.073	.673	.109	.159	.128	.155	.043	-.181	.025	-.064
49	.083	-.183	-.018	.204	-.024	.047	-.066	.298	.212	.417
50	.106	-.100	.014	.089	-.180	-.006	.023	.704	-.026	.106
51	.017	-.197	.044	.186	.157	-.038	.077	-.156	.071	.587
52	-.143	.136	.186	-.092	-.361	-.199	.218	.444	.157	-.062
53	.402	-.006	-.204	.158	.386	-.057	-.091	-.245	-.148	.022

Table 18 shows a summary of the ten-factor solution. It gives the label or the interpretation of the factor and the indicators that influence their formation. It also attempts to provide a brief remark on the factor based on the composition of the indicators identified earlier in this section.

Table 18: Summary Interpretation of a Ten-Factor Model.

Factor	Factor Label	Indicators	Remarks
1	Special Considerations	2, 5, 6, 7, 8, 24	Special nature of the item/customer
2	Market Services	12, 13, 41, 48	For lack of it
3	Prevailing market conditions	1, 11, 25, 47	Controversial
4	Customer Considerations	34, 35, 36	Lack of it
5	Level of perishableness	4,15,18	
6	Market mix	26, 31, 33, 40	
7	(Lack of) Impersonality	19, 20	
8	Market share drive	14, 50	
9	Proximity	23, 44	
10	Sales orientation	27, 51	

It is observed that two of these factors in the ten factor model are not part of the initial factors that are suspected to underlie prices of items. These are the first and the seventh. The first factor considers the special nature of the item or the customer in determining the price of the item. The seventh is interpreted as the impersonality factor. This is actually the lack of it since the responses on the indicators suggest that they are rather unpopular considerations for price fixing.

Table 19: Twelve-Factor Solution for Price Levels.

	Component											
Indicator	1	2	3	4	5	6	7	8	9	10	11	12
1	.055	.099	.675	.157	.062	.013	.117	.004	.027	-.141	-.162	-.070
2	.525	-.177	.008	.042	-.096	-.148	.142	-.003	-.140	.054	.018	.111
3	.467	.096	.149	.161	.042	.027	-.020	.281	.025	-.273	.014	.007
4	.106	.124	.045	.684	.143	.023	-.098	.076	.085	-.090	.037	.040
5	.714	-.065	-.009	.105	-.043	.098	-.077	.157	.183	-.048	.135	-.113
6	.647	-.042	-.033	.014	-.028	.050	.118	-.105	.167	.140	.033	-.176
7	.512	.107	.169	.085	.114	-.029	-.033	.151	-.011	.354	-.095	-.084
8	.655	.051	.035	-.051	-.032	.072	.072	-.080	.024	.176	.082	.001
9	.248	.170	.173	.036	-.150	.227	.339	.169	.150	.100	.166	-.173
10	.076	.134	.025	.027	.057	.203	.522	.023	.038	-.090	.270	-.298
11	.104	.242	.650	-.072	-.175	.184	.188	-.061	.147	-.149	.022	.003
12	-.003	.783	.104	.010	.064	.062	.012	.039	.087	.070	.197	.089
13	.103	.769	-.046	.068	-.030	.161	-.049	-.032	-.090	.188	-.103	-.028
14	.088	-.295	-.179	.176	.013	.126	-.100	.688	.090	-.046	-.072	.028
15	.026	.130	-.041	.703	.085	.139	.094	.134	.070	.054	-.105	-.212
16	.119	.138	.030	.334	.126	.089	.037	.432	-.019	.418	.192	.034
17	-.118	.179	.266	.364	.079	.028	.151	.326	-.164	.168	.021	-.028
18	.047	.021	.156	.651	-.109	.012	.357	-.134	.027	-.011	.140	-.039
19	.120	-.011	.087	.095	-.054	.020	-.025	.069	.767	-.045	.114	.158
20	.105	-.043	.043	.036	.100	.054	-.030	.057	.647	.029	-.172	.156
21	.335	-.158	-.005	-.041	.003	.269	-.131	-.050	-.412	.392	-.231	.039
22	.068	.278	.057	-.040	.187	-.096	-.445	.005	-.078	.006	.150	-.048
23	.362	.176	-.050	-.090	.107	.150	-.007	-.190	-.056	-.136	.611	-.053
24	.473	.129	.203	.142	.104	-.016	-.222	.242	-.161	.208	.239	-.061
25	.225	.077	.735	.193	.139	.121	-.001	-.153	-.096	-.112	-.075	.107
26	.198	.116	-.055	-.105	-.193	.433	.173	.294	-.028	.100	-.296	-.123
27	-.052	.082	.246	.105	-.018	.121	.570	.081	-.200	.037	.050	.136
28	.226	.152	-.030	-.097	.076	.022	.008	.215	-.057	.541	.014	.090
29	.112	-.054	.506	-.125	.100	.016	.061	.006	.026	.341	.123	-.199
30	-.147	.042	.029	-.214	-.191	-.080	.123	.047	.208	-.039	-.003	.629
31	-.036	.217	.097	.162	.336	.561	.058	.093	-.071	-.080	.082	-.187
32	-.075	.050	.521	-.055	-.082	.370	.215	.282	-.117	.094	.165	.048
33	-.061	.227	.101	.008	.109	.650	-.020	-.015	.077	.028	-.025	-.050
34	-.102	.081	.042	.136	.535	.179	.170	-.126	.058	.201	-.102	.262

Table 19 Continued

35	-.042	.078	.013	.073	.833	.039	-.085	.099	.056	.045	.105	-.064
36	-.016	.194	-.073	.021	.806	.057	.033	.082	.017	.119	.110	-.077
37	.053	-.009	.170	.213	.219	.192	.057	.065	-.036	.308	.429	-.234
38	.364	.025	-.094	.203	.301	.199	-.089	-.123	.019	.188	.075	-.354
39	.135	.164	-.165	.152	.157	.076	-.151	-.161	.030	.597	-.055	-.302
40	.114	-.036	.036	.195	-.005	.666	.211	-.071	.014	.096	.138	.128
41	-.136	.740	.124	.064	.107	.007	.047	.024	.045	-.008	.225	.036
42	-.042	.385	-.100	.084	.211	.015	-.149	-.017	.270	-.077	-.008	.549
43	.425	-.231	.073	.019	-.137	-.247	-.019	.127	-.296	-.010	.058	-.099
44	.074	.120	-.167	.020	.035	-.049	.044	-.002	-.054	.019	.705	.003
45	.226	-.134	.192	.176	-.047	.045	.155	.148	-.464	.025	.302	.110
46	-.158	-.038	-.157	-.188	.176	.265	-.149	-.005	.178	-.083	-.166	.363
47	-.112	-.064	.634	-.097	-.144	-.199	-.230	.011	.061	.056	-.078	.033
48	-.049	.719	.085	.140	.169	.148	-.034	-.164	-.004	-.004	-.049	.000
49	.220	-.070	-.057	-.046	.343	-.120	.568	.158	-.132	-.190	-.085	-.023
50	.111	-.074	.018	-.276	.143	-.160	.277	.594	.059	.088	-.104	.013
51	.031	-.272	.066	.242	.253	-.014	.382	-.217	.027	.001	.111	.281
52	-.073	.184	.165	-.431	-.023	-.310	.095	.318	.253	-.160	.006	.049
53	.280	-.051	-.177	.426	.102	-.046	-.060	-.197	-.039	.371	-.007	-.081

In Table 19, we notice that for a twelve-factor model, the first three factors have four or more indicators, the fourth to the seventh have three indicators, whilst the eighth to the twelfth have exactly two indicators. Thus, in this model, 41.7% of the factors have two influential indicators.

In Figure 13, we have the scree plot and the parallel analysis. The intersection of the two plots is just about the tenth component. This suggests that a ten factor model may be suitable. However, this model may only be suitable as an exploratory model. This may be seen from the plot itself since after the intersection, the scree plot does not level off sharply. This means that the remaining factors contribute substantially to the correlation matrix, even though each of these factors contributes marginally to the correlation matrix. In the next section therefore, we examine a confirmatory model that provides a significant factor solution from the data.

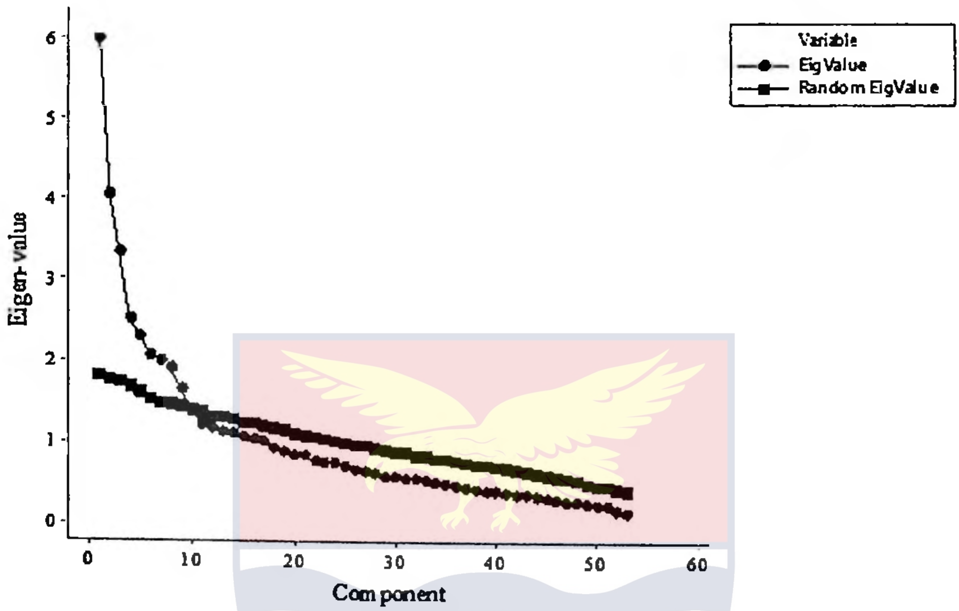

Figure 13: Graph of the Scree Plot with Parallel Analysis.

Confirmatory Test of Factor Model

As pointed out in the introductory chapter, the intended model for the study requires the latent factors of price levels as part of the input. It is therefore necessary to determine the salient factors, out of the fifty-three indicators, that significantly determine price levels. In this section, we perform various confirmatory tests to obtain the factor model that is plausible for the study. It will be recalled that the determination of the extreme prices were mainly based on the first three factor components. However, it will be relevant to assess the significance of the solution using the usual factor extraction heuristics. The test will therefore begin from 10-factor solution.

For correlation matrix, **R**, generated on the indicator variables, $|\mathbf{R}| = 7.698 \times 10^{-9}$, and the sample size $n = 360$. The test statistic of the test of hypothesis

$$H_o : \underset{p\times p}{\mathbf{R}} = \underset{p\times m}{\mathbf{L}}\ \underset{m\times p}{\mathbf{L}'} + \underset{p\times p}{\mathbf{\Psi}} \quad ; \quad m = 10,\ p = 53,$$

is given by

$$\left[n - 1 - \tfrac{1}{6}(2p + 4m + 5)\right]\ln\frac{|\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\psi}|}{|\mathbf{R}_n|} = 333.8333 \ln\frac{0.0304}{7.698\times10^{-9}}$$
$$= 333.8333 \times 0.2789$$
$$= 1576.432,$$

which is almost the same as the value in Table 20. In the table, we have the results of the test of significance of factor solutions in the hypothesis for $m = 10, 12, 17, 20(1)25$ generated in SPSS. The table shows that each of up to twenty-two factor solutions does not fit the data. The minimum solution that gives a 'non-significant' factor model (for which H_o is not rejected) is twenty-three factor model.

Table 20: Goodness-of-Fit Test of Various Factor Solutions.

Model	Chi-Square	df	Sig.
10	1576.432	893	0.000
12	1339.662	808	0.000
17	842.715	613	0.000
20	617.824	508	0.001
21	555.662	475	0.006
22	505.129	443	0.022
23	442.511	412	0.144
24	-	-	-
25	350.351	353	0.530

In the table, result for $m = 24$ is not possible as the solution does not have a local minimum.

Remarks 4.1

The result of the confirmatory test obviously does not provide a parsimonious solution that could give meaningful interpretation. The design of the indicators actually contained twelve initial dimensions. However, for reasonable assessment of the final model which will be determined at the latter part of the chapter, we are compelled to make use of a factor solution that is determined to be significant. This reduces the subjectivity of the reasonableness of the final model.

Assessing the Influence of Factors on Price Levels

In order to assess the factors in the twenty-three factor solution, we will have to obtain a univariate transformation of the matrix of factor solution, so that it is in a form that is usable in the final model. We will explain how the methodology is used to obtain a univariate equivalent of the factor solution and hence, the factor score matrix. The assessment will be done by first examining the distribution of the scores, and then its performance in the final model.

Determination of the factor effect

The matrix of factors of twenty-three-factor solution will typically look like that in Table 17 (or Table 19) which is a matrix of dimension 53×23. Since there are 360 observations, we transform this original dataset by the matrix of factor solution, **F**, so that the resulting data now captures the effect of the factors and will corresponds in row-dimension to that of the original for the 360 observations. This is done by obtaining the factor scores given by $\mathbf{V} = \mathbf{DF}$, where $\mathbf{D}$ is the 360×53 matrix of original responses on the indicator variables obtained from traders.

The second phase is to obtain a univariate projection

$$T_{y_i} = \mathbf{v}_i'\boldsymbol{\beta}_i;\ i = 1, 2, \ldots, n, \tag{4.7}$$

for a measure T_{y_i} of the price level y_i of the market using the One-outlier Displaying Component (1-ODC), $\boldsymbol{\beta}$, derived from the transformed data of factor scores, $\mathbf{V}$, and given by

$$\boldsymbol{\beta} = \mathbf{S}_\mathbf{V}^{-1}(\mathbf{v} - \overline{\mathbf{v}}).$$

Now, since $\mathbf{V}$ is generally of the form

$$\mathbf{V} = \mathbf{ZRL}^{-1},$$

where $\mathbf{R}$ is the correlation matrix of the data, $\mathbf{D}$, and $\mathbf{L}$ is the matrix of factor score coefficients, the variance-covariance matrix of $\mathbf{V}$, $\mathbf{S}_\mathbf{V}^{-1} \approx \mathbf{I}$, an identity matrix. Also, the mean vector, $\overline{\mathbf{v}} \approx \mathbf{0}$, a zero vector.

Thus, we obtain a special case of the projection explained in the methodology in which $\mathbf{T}_\mathbf{y} = \text{diag}(\mathbf{VV}')$, since $\boldsymbol{\beta} \approx \mathbf{v}$. Thus, the effect of the combined factor solution is simply the sum of squares of the factor scores for each of the twenty-three factors obtained for any respondent. The resulting vector $\mathbf{T}_\mathbf{y}$ is placed in Appendix B.

Remark 4.2

The special case of the transformation in Equation (4.1) introduces a certain effect into the categorisation of the price levels. Since a quantitative measure of the price level is given by

$$T_{y_i} = \sum_{j}^{m} \mathbf{v}_{ij}^2, \quad i = 1, 2, \ldots, n,$$

it is now difficult to identify the level y_i that represents either a low-priced or high-priced market. That is, the extremes of the price levels are captured in the sum of the squares. The categorisations are now collapsed into just two: extreme and moderate priced.

Distribution of the factor effect

In Figure 14, we have the distribution of the univariate factor scores (Appendix B) obtained by a vectorisation of the matrix of factor scores. The graph thus shows the effect of the combined extracted factors on the level of prices in a given market. As explained in the methodology, the formation of the univariate vector scores would merge the two extreme price levels, which are low and high-priced levels. Consequently, in the graph, we see only two price levels: namely, extreme and moderate. The factor effect distribution is shown for both locations of rural and urban markets where primary data were obtained from traders. It can be observed that factors that influence the determination of prices could be extreme in markets that have extreme price levels, in both rural and urban centres.

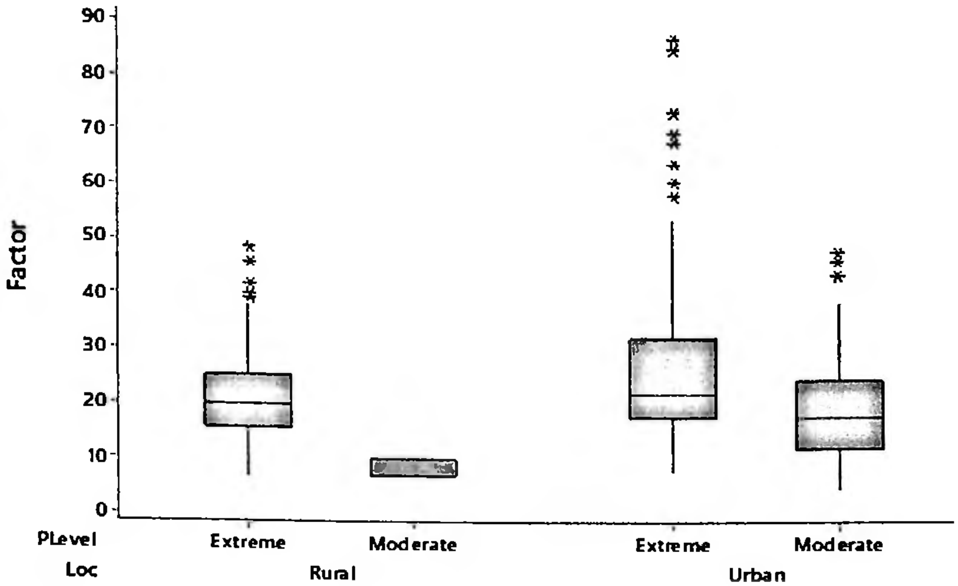

Figure 14: Boxplot of Univariate Factor Scores Showing Extreme Factor Effect on Moderate and Extreme Priced Levels.

It is also observed that the effect could be extreme even in moderate priced markets, which is particularly associated with urban markets. However, for rural moderate priced markets, the factors are almost uniform. The graph therefore shows that effect of factors may (or may not) be influential in the determination of price levels.

Model for Determining Price Levels

In order to assess the real effect of the factor in the determination of the price levels of the markets, we will obtain a model for determining the price level. In the methodology, it has been explained that the model for determining the price levels is of the form

$$\eta_i = \gamma \mathbf{M} + \alpha f, \tag{4.8}$$

where f is the effect of the combined factors identified under a confirmatory factor analysis in Table 17, **M** is a matrix of observations on some market characteristics, and η_i is a link function that connects the price level, y_i, of the particular market which are originally in three categories: high-priced, low-priced and moderate priced . In this study, the preliminary analysis shows two main market features that influence price levels. These are the 'Region' and the 'number of days' trading activities is undertaken in the market (by the trader). The matrix **M** will therefore contain observations on these two characteristics. In Table 20, we have extracted twenty-three factors as significant in determining prices from the perspective of the traders in these markets. In Equation (4.1), however, the factors are represented by a single value, *f*. This means that the twenty-three factors have been reduced to a single vector, **f**, using the technique of vectorisation of a matrix. The vector **f** is already obtained as in Appendix B. The effect of the factor solution now conforms to a matrix of dimension $n \times 1$, where $n = 360$.

Now, since the categories are reduced to two, we will make use of the link function

$$\eta_i = \log\left(\frac{p}{1-p}\right), \tag{4.9}$$

so that p is defined as the probability of a market having a particular price level. Suppose we specifically define the event of interest to be a market that has a moderate price level.

Table 21 shows the model for the log(odds) of the event of a moderate priced market in terms of the composite factor effect, the 'Region' and the 'number of days' a trader trades in the market. There are seven of the regions considered in the model. These regions are those that showed to be typical of the three categorisations of the markets. The regions include Brong Ahafo

(BA) which in the model is used as the reference region. Brong Ahafo Region is used as reference region because it showed the most typical of all the categorises. The region BA is therefore not explicitly found in the model.

Table 21: Model Coefficients

Term	Coef	SE Coef	VIF
Constant	-2.770	1.180	
Factor	0.034	0.036	2.53
Mktdays	-1.353	0.252	2.12
Region			
Eastern	6.804	0.879	2.11
GAccra	7.600	1.380	1.79
UE	5.840	1.420	1.97
UW	0.700	1.030	1.77
Volta	5.810	1.040	3.40
Western	5.730	1.780	1.53

Source: Combined MoFA Data and Survey Data

In Table 22, we have the deviance which also shows the significance of each of the predictor variables in the model in Table 21.

Table 22: Model Deviance

Source	DF	Adj Dev	Adj Mean	Chi-Square	P-Value
Regression	8	338.012	42.2515	338.01	0.000
Factor	1	0.858	0.8582	0.86	0.354
Mktdays	1	86.614	86.6141	86.61	0.000
Region	6	243.449	40.5748	243.45	0.000
Error	351	120.278	0.3427		
Total	359	458.290			

We note in particular, that the factor effect is not significant. However, the 'Region' and the 'number of days' of trading are significant. This shows that even though the factors are found to constitute a significant factor solution for determining (unobserved) price levels, they are not significant when used to determine the actual price levels in the presence of the other predictors. The performance of the model is what is presented as shown.

Table 23: Model Summary

Deviance R-Sq	Deviance R-Sq(adj)	AIC
73.76	72.01%	138.28

The deviance residual plot based on the model is given in Appendix E. It is clear that though the model could be good, it is highly influenced by extreme observations, which is already anticipated. As pointed out in the methodology, the lack of outliers is a condition for such a model. Consequently, the model is not intended for prediction. It will enable us, however, to determine the relevance of the explanatory variables in the assessment of price levels.

In Tables 24 and 25, we have the odds ratios for the continuous and categorical predictor variables, respectively. In Table 24, the odds ratio for the 'factor' is approximately 1. This means that for a unit increase in the factor, the odds of being moderate priced market is almost the same. Thus, the probability of being a moderate priced market is almost the same as extreme market even for a unit increase in the factor. In fact, this explains why influence of price-determining factor is not significant in the model. However, the odds ratio of the 'number of market days' (Mktdays) is very low. This means that for a day increase in the number of trading days, the odds of a market being moderate priced is 0.2583 times that in BA. Thus, it is much less probable to be moderate priced compared to a market in BA for an increase in the market day.

Table 24: Odds Ratios for Continuous Predictors.

	Odds Ratio	95% CI
Factor	1.0343	(0.9644, 1.1092)
Mktdays	0.2583	(0.1577, 0.4234)

Table 25 gives the odds ratios of being a moderately priced market for each level (Level A) of the categorical predictor (Region) relative to a particular one of them (Level B). The odds ratios are particularly high when BA is the reference. This means that the odds ratio of a market in Greater Accra, for example, being moderate, is up to almost 30,000 times that of BA. Thus, it is much more likely to be moderate priced market in Greater Accra than it is in BA. In Upper West region, the odds are about twice that of BA, and could increase to as much as 15 times. Thus, generally, BA markets are predominantly far from being moderate. Particularly, BA markets are predominantly extreme markets, and in most cases, specifically low priced markets.

Table 25: Odds Ratios for Categorical Predictors

Level A	Level B	Odds Ratio	95% CI
Region			
Eastern	BA	901.6954	(160.9497, 5051.6058)
GAccra	BA	2006.3273	(134.5101, 29925.9931)
UE	BA	342.2162	(21.2407, 5513.5661)
UW	BA	2.0194	(0.2669, 15.2804)
Volta	BA	334.1674	(43.5979, 2561.3142)
Western	BA	307.7443	(9.4714, 9999.2579)
GAccra	Eastern	2.2251	(0.2129, 23.2533)
UE	Eastern	0.3795	(0.0334, 4.3164)
UW	Eastern	0.0022	(0.0004, 0.0142)
Volta	Eastern	0.3706	(0.0692, 1.9853)
Western	Eastern	0.3413	(0.0104, 11.1813)
UE	GAccra	0.1706	(0.0107, 2.7104)
UW	GAccra	0.0010	(0.0001, 0.0150)
Volta	GAccra	0.1666	(0.0133, 2.0928)
Western	GAccra	0.1534	(0.0033, 7.2221)
UW	UE	0.0059	(0.0003, 0.1028)
Volta	UE	0.9765	(0.0626, 15.2243)
Western	UE	0.8993	(0.0221, 36.6044)
Volta	UW	165.4745	(27.7097, 988.1663)
Western	UW	152.3902	(3.0846, 7528.6893)
Western	Volta	0.9209	(0.0163, 52.0977)

Relative to Eastern, Greater Accra, and Upper East regions, the odds ratio of a market being moderately priced in Upper West are much smaller than 1. This indicates that in each of these three regions, markets are mostly moderate priced compared to those in Upper West.

It is possible to derive the separate models for each region from the single model in Table 18. These separate models are given as follows in Table 26.

Table 26. Separate Models for Selected Regions

Region			Equation
BA	η	=	-2.772 + 0.034 Factor - 1.353 Mktdays
Eastern	η	=	4.033 + 0.034 Factor - 1.353 Mktdays
GAccra	η	=	4.832 + 0.034 Factor - 1.353 Mktdays
UE	η	=	3.064 + 0.034Factor - 1.353 Mktdays
UW	η	=	-2.069 + 0.034 Factor - 1.353 Mktdays
Volta	η	=	3.040 + 0.034 Factor - 1.353 Mktdays
Western	η	=	2.958 + 0.034 Factor - 1.353 Mktdays

The equations may be represented in matrix/vector form as

$$\boldsymbol{\eta} = \begin{pmatrix} -3.772 & -1.353 & 1 \\ 3.033 & -1.353 & 1 \\ 3.832 & -1.353 & 1 \\ 2.064 & -1.353 & 1 \\ -3.069 & -1.353 & 1 \\ 2.040 & -1.353 & 1 \\ 1.958 & -1.353 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ Mktdays \\ \text{Region}_{BA} \\ \text{Region}_{ER} \\ \text{Region}_{GA} \\ \text{Region}_{UE} \\ \text{Region}_{UW} \\ \text{Region}_{Volta} \\ \text{Region}_{W} \end{pmatrix} + \begin{pmatrix} 0.034 \\ 0.034 \\ 0.034 \\ 0.034 \\ 0.034 \\ 0.034 \\ 0.034 \end{pmatrix} Factor$$

in the form

$$\underset{g\times 1}{\boldsymbol{\eta}} = \underset{g\times n_{MC}}{\boldsymbol{\gamma}} \underset{n_{MC}\times 1}{\mathbf{M}} + \underset{g\times m}{\boldsymbol{\Gamma}} \underset{m\times 1}{\mathbf{F}} + \underset{g\times 1}{\delta},$$

where $g=7$, $n_{MC}=2$, $m=1$ and expectation of δ is zero. In the models, therefore, the market characteristics of covariates are 2, and γ and $\mathbf{M}$ are augmented to account for the intercept of the models.

In the models, the Region locations are defined as dummy variables as follows:

$$\text{Region}_{BA} = \begin{cases} 1, & \text{if market is in BA} \\ 0, & \text{otherwise} \end{cases};$$

$$\text{Region}_{ER} = \begin{cases} 1, & \text{if market is in Eastern Region} \\ 0, & \text{otherwise} \end{cases};$$

$$\text{Region}_{GA} = \begin{cases} 1, & \text{if market is in Greater Accra} \\ 0, & \text{otherwise} \end{cases};$$

$$\text{Region}_{UE} = \begin{cases} 1, & \text{if market is in Upper East} \\ 0, & \text{otherwise} \end{cases};$$

$$\text{Region}_{UW} = \begin{cases} 1, & \text{if market is in Upper West} \\ 0, & \text{otherwise} \end{cases};$$

$$\text{Region}_{Volta} = \begin{cases} 1, & \text{if market is in Volta} \\ 0, & \text{otherwise} \end{cases};$$

and

$$\text{Region}_{W} = \begin{cases} 1, & \text{if market is in Western} \\ 0, & \text{otherwise} \end{cases}.$$

The equations show that the effect of factor and number of market days is the same for all regions. However, if only categorical predictor (Region) variable is considered, the odds would differ. For example, the odds of being moderate priced will be the least in BA, whilst it is highest in Greater Accra, if market days and factor are not considered. That is, it is least likely to be moderate priced in BA, whilst it is most likely to be moderate priced in Greater Accra.

In the above equations, the predicted probabilities, p, may be derived from the equation

$$P(ModeratePricedMarket) = \frac{\exp(\eta)}{1+\exp(\eta)}.$$

Using this equation, we can obtain the predicted probabilities of being moderate priced market as perceived by the trader. Figure 15 shows a plot of these probabilities against the number of market days.

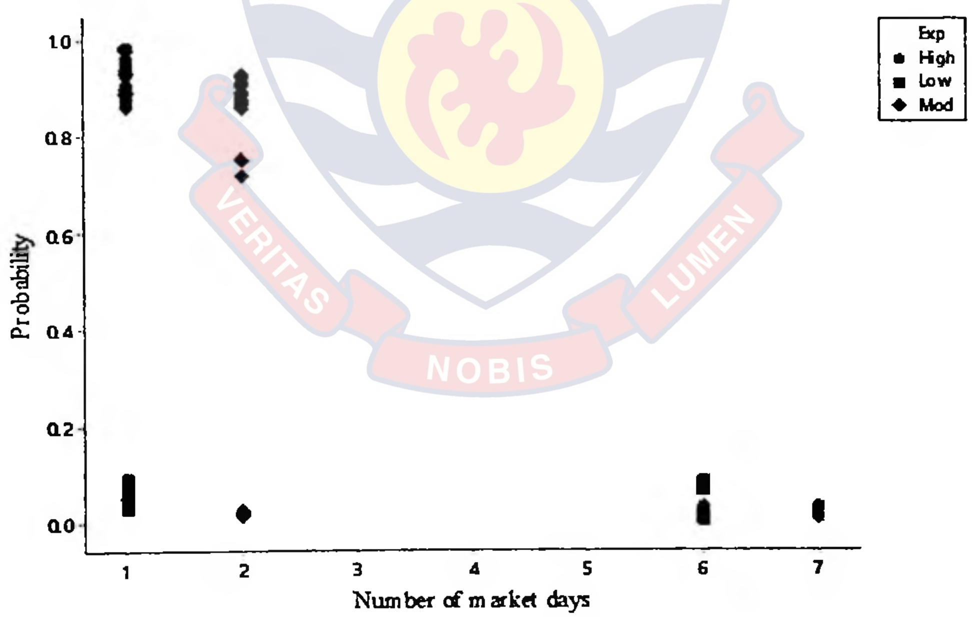

Figure 15: Probability Plot of Moderate Price for Given Number of Market Days for Various Market Price Levels.

Two observations are clear from Figure 15: They are: (1) Extreme markets are predominantly associated with large number of market days. Particularly, for high priced markets, trading goes on all week round. On the other hand, moderate priced markets are associated with one or two days of trading in the market. There are some who view low priced markets as moderate but with low probability; and (2) Extreme priced markets are associated with very low probability of being seen as moderate. However, moderate priced markets are mostly seen as such even though there could be those that have low probability of being moderate.

Remarks 4.3

In the model in Table 21, we notice that the combined effect of the factors that underlie market price levels is not significant even though the factor solution is significant. This necessitates further examination of the model to determine how this scenario arises. Now, if we drop the 'Region' or the 'number of market days' from the model, we obtain the following reduced models in terms of only two predictor variables.

Table 27: Odds Ratios for Continuous Predictors in Reduced Model Containing Two Predictors

Reduced Model	Continuous Variable	Odds Ratio	95% CI
1	Factor	1.0142	(0.9777, 1.0521)
2	Factor	0.9666	(0.9407, 0.9931)
	Mktdays	0.5656	(0.4731, 0.6763)

In Table 27, Reduced Model 1 contains the factor and the region. It is noticed that the 95% confidence interval contains 1, an indication that the factor is not significant in that reduced model. In Reduced Model 2, it does not contain the Region as a predictor. It is noticed that in this case, the 95% confidence interval does not contain 1, an indication that the factor effect is significant in that model. Thus, the effect of the factor is not significant in the presence of the Region. This shows that the Region is a dominant variable in determining price levels.

Discussion

A few of the findings in this study are in line with some results in the literature. In particular, the study has revealed that one of the most low-priced markets is Market 68 (Bugubelle), in the Upper West region. This market is found to be consistently the lowest priced in Maize in the country over the entire study period. Gage et al. (2012) rather report that most of the Maize reported in Ghana is produced in the northern part of the country. This study has identified the specific location where the commodity may be obtained at the most economic price level.

Furthermore, in this study, it is found that Maize is a major influential indicator of the dimensions that have been determined in the form of time-dependent principal components in identifying the price levels on the markets. This means that a major component of the expenditure of consumers of local food items is taken by Maize. Maize is found to be an influential indicator in the formation of the second principal component in 2008 and 2009, and the most dominant indicator on the first principal component in 2012 and 2013. However, it does not feature significantly on any component in 2015. (See

Appendix D.) The relative importance of Maize as a component of the local

food items consumed is therefore evident, and buttresses the findings on the importance of the item in the literature. The break in 2015 on the importance of Maize is interestingly consistent with the break in the movement in prices of all staple products reported by Cudjoe et al. (2008). The reported movement in opposite directions of prices of Maize and Rice around the period of 2015 and 2016 is however not observed in this study.

Chapter Summary

In this chapter, we have examined some preliminary features of the dataset used in this study. The descriptive analysis reveals two main markets as consistently being among the least priced markets. These are Markets 68 and 17. Market 68 is consistently the least priced market for Maize throughout the period, and among the least priced markets in two other items. Market 17 is among the least priced in five items. On the other hand, two markets are also identified as consistently being among the highest priced. These are Market 65 and 69. Market 65 is among the highest priced in five commodities, whilst Market 69 is among the highest priced in four commodities over the period.

The techniques outlined in Chapter Three have been applied to study the data in more detail. The techniques of time-dependent principal components and outlier displaying components are employed to determine the price levels of the markets using the secondary data that cover the actual prices of the food items obtained over the five non-consecutive years. It is observed that extremely high variations exist in the prices of some commodities. As a result, the principal components are determined in terms of loadings rather than the actual weights in order to avoid the influence of these high variations in these commodities on the results. The results based on the time-dependent components also highlight Markets 65, 68 and 69 as extreme markets. Interestingly, Market 65 is found to be both high and low priced along components of different orders in Year 1 and 4, in particular. It is observed that in most cases, beyond three-component solution, each of the remaining components is influenced by few items, including single-item components. Since this type of components is not useful in the general solution, component solutions up to order five are considered. Therefore, examing all component plots for the first five components in addition to the preliminary results gives the set of suspect outlying markets as {65, 69, 13, 68, 54, 89, 90, 17, 56, 46, 14, 83, 55, 47}. The results based on the principal components thus show that only a handful of markets in the country may be classified as extreme among several markets examined.

The extreme markets identified by the preliminary analysis and application of principal components constitute the set of suspect observations for the application of the Outlier Displaying Component. The Outlier Displaying Component based on the total SSCP identifies Market 65 in Year 4 as the most extreme market in all the period under study. Market 17, however, is found to be consistently low priced over the years considered, but outstandingly low in the last year (2015). Since the total variance-covariance matrix is affected by the extreme variations, the results may be overly highlighted or masked by this procedure. As a result, the Outlier Displaying Component based on the pooled variance-covariance matrix is also used. It drastically reduces the distance measures involved, and highlights clearly some high and low markets. It shows that Market 17 is still consistently low over the period, and that Markets 13 and 68 are the next low markets over the period. The extent of extremeness of Market 65 is now reduced drastically. To improve the results based on the Outlier Displaying Component, it is further applied based on the pooled reduced sample variance-covariance matrix. In this application, the mean and variance-covariance matrix are determined from the data after deleting the suspect outlier, which has been shown to provide a better projection of the outlier. The results now clearly buttresses those of the Outlier Displaying Component based on the pooled variance-covariance matrix. Market 65 appears to be the most consistently high priced over the period. The next is Market 14, particularly in the last two years over the study period. The results based on the Outlier Displaying Component further shows that yet very few markets in Ghana may be classified as belonging to the class of extremely high and low priced markets.

Having identified the price levels of the markets, latent factors have been determined that underlie the price levels. Based on primary data obtained from respondents from selected markets that represent all the price levels, twenty-three factors have been extracted from a confirmatory factor analysis as constituting a significant factor solution. Confirmatory factor analysis is used as the objective here is to obtain a significant solution rather than a parsimonious one so that the resulting factor model meets the condition for its further use in subsequent modelling. This primary data is generated on Likert scale that reflect traders assessment of indicators that influence pricing of items in the market.

To determine the combined effect of the twenty-three factors on the level of prices, the factors have been reduced to one based on a transformation of their factor scores obtained from a projection on the outlier displaying component. This transformation preserves the information on the factor model in the single factor vector. Using the combined factor vector and two other relevant covariates on market features, a structural model has been obtained for determining the price levels. The covariates are the number of trading days in the market and the location given by the region. The structural model involves a logit link function that connects the probability of the price level to the independent variables of market characteristics and the factor. It is observed that the factor effect is not significant. However, the 'Region' and the 'number of days' of trading are significant. This indicates that even though the factors constitute a significant factor model by themselves, they are not significant when used to determine the actual price levels in the presence of the other predictors. The factor effect is however significant when prices are viewed without reference to regions.

Based on the model, it is found that extreme markets are predominantly associated with large number of market days. In particular, for high priced markets, trading goes on all week round. On the other hand, moderate priced markets are associated with one or two days of trading. The model also shows that Extreme priced markets are not likely to be seen as moderate. Predominantly, markets have high probability of being regarded as moderate priced.

Since the data potentially contains outliers, the model may not be suitable for purposes of prediction. Rather, the model has been mainly used to provide assessment of the effect of the various variables that could influence price levels.

CHAPTER FIVE

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

Overview

This chapter presents the summary of the entire work done in the previous four chapters of the thesis. It presents the main objectives of the research and the procedures that have been followed to achieve these objectives. Particularly, the chapter will point out the nature of the level of prices of the Ghanaian local food items on markets across the country in recent years. It also seeks to highlight the performance of the techniques that have been utilised in the study. The assessment of the techniques is relevant as they are an extension of the original techniques to make them suitable for the data problem. From the summary, we will draw conclusions and make relevant recommendations that hopefully have national interest.

Summary

The study is largely motivated by some four main issues regarding local food prices in the markets across the entire country. These are common factors that influence price determination, effect of movements in prices in recent past around the globe, the complex nature of the multiple multivariate data that characterise studies on such subject, and how appropriate techniques may be formulated for the analysis of such data. Dominant among the prevailing factors that influence prices are the poor nature of roads to production centres that leads to high cost of transporting food commodities. Others include the proximity to production centres and general prevailing market conditions. It is observed that prices of items have basically been on the rise over the few years prior to 2008. The consequences of the price rise have been felt by almost every class of the society with incomes of individuals and nations being seriously eroded. The most vulnerable group is the hardest hit of the negative impact. The situation shows that a persistent high food prices deepens poverty. The study also identifies effort at monitoring price levels by government agencies and finds that academic corroboration in this regard has been scanty. It has highlighted the need for the use of appropriate techniques in the study of market statistics as the results could be affected if the technique is not specifically designed for the study. It has also stressed that the composition of the markets in such studies could be crucial in arriving at the kind of results obtained.

Consequently, the statement of the problems has focused on three key issues: the appropriateness of the techniques that have been used in this area of study and the lack of application of identified suitable methods for the study of the subject; the complex nature of the data problem usually involved in this type of studies; and the need for academic research into market statistics as it could generate useful market information for efficient market organisation as a national strategy for poverty-reduction. Subsequently, three main objectives have been derived that broadly assess the price levels of local food items in Ghana.

The study has provided in-depth description of the data problem and the research design. It has identified that the techniques employed in this research are not originally designed for the purpose, and that attempt would be made to justify their use and provide suitable extensions. The extensions are found necessary as in this study, a single observation in the data constitutes a matrix, rather than a vector for which these techniques are originally designed. The new and complex structure of the multivariate observation has arisen as a result of the time-dependent nature of the study data. The proposed techniques are therefore broadly of the class of time-dependent displaying components and structural equation modelling. This is because, it requires that we identify suitable dimensions along which to assess the main influence of the price data over a time-period whilst highlighting possible extreme prices simultaneously. To study fully the price levels, it also requires that we examine multiple sets of data-generating variables. Efforts made at adopting the techniques to suit the data problem are also informed by the literature. The principal component analysis, in particular, has enjoyed widespread applications on the subject. The technique, however, is identified in various authority texts as being suitable for preliminary studies, rather than the main technique. Structural Equation Modelling (SEM) is another major technique found in the studies. The studies that use SEM show that the technique encompasses various concepts that are intended in this study, which includes Confirmatory Factor Analysis.

In the literature, we find some considerations for fixing prices for food items. These include the perishableness or sale proneness of the item, and promotions, prevailing conditions of global changes and market impacts. Predominantly, these impacts are due to uncertainties during production and spikes in world food prices in the period of 2006 – 2008, rapid urbanisation and fast income growth, and high energy prices. The literature also shows that prices have been controlled in various ways. These may be by exercising some authority of single or group of individual persons or companies, and by organised information systems that seek to create transparency about prevailing prices. It also notes artificial price distortions, particularly among the poor population, that are as a result of perturbed competition that controls supply and entry of traders into the market. The effects of such control avenues have been found generally harmful, but in some few cases, it has helped to improve prices to the benefit of members of associations in some parts of the world.

Variations and movements in prices could also be influenced by the composition of diet. In this case, prices will likely be stable in communities whose diet is largely based on non-tradable staple foods, such as cassava and sorghum. An observation in the literature which may explain the structure of the study data is that changes in average prices may not be connected to changes in variability of domestic prices. This phenomenon may be due to imposition of constant tariffs on some imported items that generally affects domestic prices. Variations have also been assigned to seasonality of the items. It is found rather ironical in the literature that prices have reached record high over the period of 2007/2008 during a time of equally record abundance. This suggests that variations are not likely to be attributable to shortage in production. This phenomenon has been attributable to growing demand for meat in increasing-wealthy nations of China and India with extended growing demand for cereals, and demand for corn-based ethanol in America. It is therefore not surprising that cereals appear to be the most affected in variations in prices of items that have been reported, and has also gained considerable attention for research.

It is the concern of some literature that swings in food prices has broad and far-reaching macro-economic effects, which could potentially deepen poverty, as food prices constitute a key component for the determination of standard of living. Quiet unlike the motivation for the study, the literature rather emphasises the need for the inclusion of a variety of products as a key component in the assessment of market prices rather than the use of a large number of outlets and covering more regions.

The study finds rather little academic research conducted on local food prices in Ghana, and these have targeted few specific commodities. There appears to be none that adopts the approach used in this study.

To effective research the problem, the study has examined the main techniques. The displaying components comprise the principal component and the outlier displaying component. The review has made the necessary extensions that would make them suitable for multiple multivariate data employed in this study. The techniques have been redefined to be used as time-dependent displaying components as the data problem covers some time-periods. In the developments, treatments have made for conditions for generating component scores that could potentially be problematic for interpretations. These conditions are informed by the literature as well as the wide variations that exist in the data. The applications based on Outlier Displaying Component (ODC) focused on only the 1-ODC, since by highlighting a single extreme observation, other potentially extreme observations would lie in its neighbourhood. Techniques of general factor analysis and Confirmatory Factor Analysis (CFA) have been examined. The techniques are found useful in identifying independent latent variables that could explain the observable dependent variables created from the original item variables. The structural equation modelling is identified as an appropriate technique that could bring together the dependent variable of price levels and the factor model, in addition to other covariates of market characteristics. In the structural model used in the study, it is found necessary to obtain a combined factor effect in the model. The matrix of factors therefore has to be reduced to a single factor vector. A vectorisation procedure has been developed that would convert the factor matrix to a single column vector that retains all the properties of the original matrix. It is expected therefore that effect of this single factor vector will reflect the effect of the original factor solution in the structural model.

The implementation of the methods is preceded by assessment of some preliminary features of the dataset used in this study. In the preliminary studies, we find that two main markets are consistently among the least priced markets. These are Markets 68 and 17. Market 68 is consistently the least priced market for Maize throughout the period, and among the least priced markets in two other items. Market 17, on the other hand, is among the least priced in five items. However, two markets are also identified to be consistently among the highest priced. These are Market 65 and 69. Market 65 is among the highest priced in five commodities, whilst Market 69 is among the highest priced in four commodities over the period.

The application of time-dependent principal components examines the secondary data that cover the actual prices of the food items obtained over the five non-consecutive years. It is initially observed that extremely high variations exist in the prices of some commodities. As a result, the principal components are determined in terms of loadings rather that actual weights in order to avoid the influence on the results of these high variations in these commodities. The results based on the time-dependent components also highlight Markets 65, 68 and 69 as extreme markets. Interestingly, Market 65 is found to be both high and low priced along components of different orders in Year 1 and 4, in particular. This is consistent with the observations from the descriptive statistics. It is observed that in almost all years, each of the remaining components beyond three-component solution is influenced by few items, including single-item components. Since this type of components is not useful in the general solution, component solutions up to order five are considered. By examining all component plots for the first five components in addition to the preliminary results gives the set of suspect outlying markets as {65, 69, 13, 68, 54, 89, 90, 17, 56, 46, 14, 83, 55, 47}. The results thus show that only a handful of markets in the country may be classified as extreme.

The extreme markets identified by the preliminary analysis and application of principal components constitute the set of suspect observations for the application of the Outlier Displaying Component (ODC). The Outlier Displaying Component (ODC) based on the total SSCP identifies Market 65 in Year 4 as the most extreme market in all the period under study. Market 17, however, is found to be consistently low priced over the years considered, but outstandingly low in the last year. Since the total variance-covariance matrix is affected by the extreme variations, the results are found to be either overly highlighted or masked by this procedure. To improve the results based on the Outlier Displaying Component (ODC), it is further applied based on the pooled reduced sample variance-covariance matrix. In this application, the mean and variance-covariance matrix are determined from the data after deleting the suspect outlier, a procedure that has been shown to provide a better projection of the outlier. The results now clearly buttresses those of the Outlier Displaying Component (ODC) based on the pooled variance-covariance matrix. Market 65 appears to be the most consistently high priced over the period. The next is Market 14, particularly in the last two years. The results based on the Outlier Displaying Component (ODC) further shows that yet very few markets in Ghana may be classified as extremely high and low priced markets. It is only these few markets that appear to have significantly extreme prices.

Having identified the price levels of the markets, latent factors have been determined that underlie the price levels. Twenty-three factors have been extracted from a confirmatory factor analysis as constituting a significant factor solution. The factor model is based on primary data obtained from traders from selected markets that represent all the price levels.

To determine the combined effect of the twenty-three factors on the level of prices, the factors have been reduced to just one based on a transformation of their factor scores obtained from a projection on the outlier displaying component. This transformation preserves the information on the factor model in the single factor vector. Using the combined factor vector and two other relevant covariates on market features, a structural model has been obtained for determining the price levels. The covariates are the number of trading days in the market and the location market given by the region. The structural model, which involves a logit link function, connects the probability of the price level to the independent variables of market characteristics and the factor. It is observed that the factor effect is not significant, even though the factor model is initially found to be significant and therefore meets the condition for its inclusion in the structural model. However, the 'Region' and the 'number of days' of trading are significant. This indicates that even though the factors constitute a significant factor model by themselves, they are not significant when used to determine the actual price levels in the presence of the other predictors. The factor effect is however significant when prices are viewed without reference to regions.

Based on the model, it is found that extreme markets are predominantly associated with large number of market days. In particular, in high priced markets, trading activities are undertaken all week round. On the other hand, moderate priced markets are associated with very few days of trading. The model also shows that extreme priced markets are not likely to be seen as moderate, and generally, markets have high probability of being regarded as moderate priced.

Since the data potentially contains outliers, the model may not be suitable for purposes of prediction. Rather, the model has been mainly used to provide assessment of the effect of the various variables that could influence price levels.

Conclusions

The study has been largely motivated by a number of issues on local food prices in markets across Ghana. These issues cover common factors that influence price determination, effect of rising movements in prices in recent past in the country and around the globe, the complexity of the multiple multivariate data that are associated with studies on the subject, and how appropriate techniques may be formulated for the analysis of such data. Among the prevailing factors that influence prices are predominantly the poor nature of roads to production centres that leads to high cost of transporting food commodities. Others include the proximity to production centres, the sale proneness of the item, promotions, and general prevailing market conditions of global changes and market impacts. The study finds that academic corroboration in the study of market statistics has been scanty, and fears that important information could escape notice of government agencies that have the mandate to monitor price levels using their routine methods. It has highlighted the need for the use of appropriate techniques in the study of market statistics as the results could be affected if the technique is not specifically designed for the study. It has also stressed that the composition of the markets in such studies could be crucial in arriving at the kind of results obtained.

Consequently, the study has focused on three key issues: the appropriateness and extent of application of the techniques useful for the study; the complex nature of the data problem usually involved in these studies; and the need for academic research into market statistics for efficient market organisation as a national strategy for poverty-reduction. Subsequently, the main objectives derived are broadly focused on assessing the price levels of local food items across the country.

The complex structure of the multivariate observation has arisen as a result of the time-dependent nature of the study data. The techniques that are proposed are therefore broadly of the class of time-dependent displaying components and structural equation modelling. This is because, it requires that we identify suitable dimensions along which to assess the major influence of the price data over a time-period and highlight possible extreme prices simultaneously. It also requires that we examine multiple sets of data-generating variables that may be put together in a single model. The displaying components comprise the principal component and the outlier displaying component. The study has provided necessary extensions that would make the components suitable for the study as the data problem covers some time-periods. The applications based on Outlier Displaying Component (ODC) focused on only the 1-ODC, since by highlighting a single extreme observation, other potentially extreme observations would lie in its neighbourhood. The technique of Confirmatory Factor Analysis (CFA) is also found useful in identifying independent latent variables that could explain the observable dependent variables created from the original item variables. In the structural model, we have rather included a combined factor effect of the initial factor model. The matrix of factors therefore has been reduced to a single factor vector such that the factor vector retains all the properties of the original factor solution. It is expected therefore that effect of this single factor vector will reflect the effect of the original factor solution in the structural model

Prior to the application of the techniques, preliminary studies find that two main markets are consistently among the least priced markets. These are Markets 68 and 17. Market 68 is consistently the least priced market for Maize throughout the period, and among the least priced markets in the spices and pulses. Market 17, on the other hand, is among the least priced in five items: Root and tubers, vegetables, local rice, and banana. However, two markets are also identified to be consistently among the highest priced. These are Market 65 and 69. Market 65 is among the highest priced in five commodities: Root and tubers (excluding Yam), vegetables, fruits and eggs. Market 69 is among the highest priced in four commodities covering fruits, spices and fishes, over the period.

The principal components are determined in terms of loadings rather that actual weights to avoid the influence of high variations in the commodities. Examining the component plots for the first five components in addition to the preliminary results gives the set of suspect outlying markets as {65, 69, 13, 68, 54, 89, 90, 17, 56, 46, 14, 83, 55, 47}. Using this set, the M1-ODC is applied based on the pooled reduced sample variance-covariance matrix, a technique that removes the negative effect of the suspect outlier, and hence, provides a better projection of the outlier. Market 17 is clearly identified as the most consistently low priced over the period. Market 65 appears to be the most consistently high priced over the period. The next is Market 14, particularly in the last two years. The results based on the Outlier Displaying Component (ODC) further shows that very few markets in Ghana may be classified as significantly high and low priced markets.

A twenty-three factor solution is then extracted from a confirmatory factor analysis based on primary data obtained from traders from selected markets that represent all the price levels. A combined factor effect has been obtained by reducing the initial factor model to include just one based on a projection on the outlier displaying component. This transformation preserves the information on the factor model in the single factor vector. Using the combined factor vector and two other relevant covariates on market features, a structural model has been obtained for determining the price levels. The structural model, involves a logit link function, and connects the probability of the price level to the independent variables. It is observed that the factor effect is not significant. However, the covariates, which are 'Region' and the 'number of days' of trading, are significant. This indicates that even though the factors constitute a significant factor model by themselves, they are not significant in the model that contains the other predictors. The factor effect is however significant when prices are viewed without reference to regions. The model shows that extreme markets are predominantly associated with large number of market days. On the other hand, moderate priced markets are associated with very few days of trading. The model also shows that whilst markets generally have high probability of being regarded as moderate priced, extreme priced markets are not likely to be seen as moderate.

Recommendations

The results show that markets that are low priced are predominantly those whose prices are low in cereals, root and tubers, vegetables and fruits. In contrast to this, markets that have been identified as high priced have high price levels in Root and tubers (excluding Yam), vegetables, fruits, fishes and the spices. What is not common in the two groups is cereals and spices. This means that a step towards food sufficiency and security should target the production of cereals and increase production of other food groups, particularly, spices. The literature shows the demand for and hence the price of cereals could continue the upward swing as the economy and that of other nations expand. The country could therefore gain competitive edge if cereals in particular are given priority attention across the country. At the moment, markets predominantly in the Upper West region have influential productivity in the commodity. This can easily be extended to almost all parts of the country as the Ghanaian soil could support these foods almost everywhere.

The study also identified wide variations in price of spices across the markets. This potentially has effect on the general price level of the markets, and hence the standard of living in these communities. Efforts at increasing food production could also target spices. The distribution of these items across the country could also be managed as this type of crop is usually concentrated in few parts of the country. Whilst increasing the production of these foods in areas where they already abound, concentration could also be directed in communities where markets are high priced in these items.

A major issue that has been identified to influence high price levels is the nature of poor roads leading to production centres. It is apparent that an improvement in the condition of these roads could reduce the distribution cost of the commodities and stabilise the disparities in price.

As the data has been found to contain outliers, it could violate the assumption regarding the use of the structural modelling. The model has therefore been used mainly to assess the significance of the effect of the independent variables that are found relevant to influence price levels.

The techniques that have been extended in this study have revealed that the usefulness of the pooled variance-covariance, in the ODC in particular. It has shown that such studies could potentially be affected by widespread variations and techniques that are not robust to such variations. It therefore recommends the use of the M1-ODC based on the pooled variance-covariance as it has proven to provide reliable results.

This study has focused mainly on the applications of the proposed extended techniques. However, the organisation of the items in the market could influence the price levels. This aspect of the study has not been reported. Subsequent research could focus on this area, in addition other market features that have the potential of influencing prices.

REFERENCES

Adger, W. N. (1999). Socal Vulnerability to Climate Change and Extremes in Coastal Vietnam. *World Development*, 3, 249-269.

Aggarwal, P., & Vaidyanathan, R. (2003). Use it or lose it: Purchase Acceleration Effects on Time-Limited Promotions. *Journal of Consumer Behaviour*, 2(4), 393-403.

Alem, M., & Soedermbom, M. (2012). Household-level Consumption in Urban Ethiopia: the Effect of Large Food Price Shock. *World Development*, 40(1), 146 – 162.

Anderson, T. W. (2003). *Introduction to Multivariate Statistical Analysis*, New Jersey: Prentice Hall.

Armeanu, D., & Lache, L. (2008). Application of the Model of Principal Components Analysis on Romanian Insurance Market. *Theoretical and Applied Economics*, No. 6, 525.

Barnett, E. (2008). "*High Food Prices: Challenges and solutions*" www.worldchanging.com/archives/007762.html

Barnett, V., & Lewis, T. (1994). *Outliers in Statistical Data*. (3rd Ed.). New York: John Wiley and Sons.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51, 1173-1182.

Bellemare, M. F. (2015). Rising Food Prices, Food Price Volatility, and Social Unrest. *America Journal of Agricultural Economics*, 97(1) 1-21.

Benyi, J. (2018). A Study of Some Data-based Problems in Factor Analysis. MPhil Thesis, Univeristy of Cape Coast.

Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York: John Wiley and Sons.

Caprihan, A., Pearlson, G. D., & Calhoun, V. D. (2009). Application of Principal Component Analysis to Distinguish Patients with Schizophrenia from Health Controls Based on Fractional Anisotropy Measurements. *Neuroimage*, 42(2), 675 – 682.

Chabane, N. (2002). An Evaluation of the Influences on Price and Production in the Maize Market following Liberalization, Trade and industrial policy strategies. *2002 Annual Forum at Glenburn Lodge, Mulders drift.*

Chatterjee, S., & Kapur, D. (2016). Understanding Price Variation in Agricultural Commodities in India: MSP, Government Procurement and Agriculture Markets. NCAER India policy Forum, July 12-13, 2006.

Chen, S. F. S., Monroe, K. B., & Lou, Y. C. (1998). The Effects of Framing Price Promotion Messages on Consumers' Perceptions and Purchase Intentions. *Journal of Retailing*, 74(3), 353-372.

Chun, Y. H. (2003). Optimal Pricing and Ordering Policies for Perishable Commodities. European Journal of Operations Research. 144, 1, 68 – 82.

Chung, J., & Li, D. (2013). The Prospective Impact of a Multi-Period Pricing Strategy on Consumer Perceptions for Perishable foods. *British Food Journal*, 115(3), 377-393.

Clapp, J. (2015). Food Security and the International Trade: Unpacking Disputed Narratives. *Food and Agriculture Organization of the UN.* Background Paper Prepared for the State of Agricultural Commodity

Markets, 2015-16.

Claro R. M., & Monteiro, C. A. (2010). Family Income, Food Prices, and Household Purchases of Fruits and Vegetables in Brazil. *Rev Saude Publica*. 44(6), 1014-20.

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Cudjoe G., Breisinger, C., & Diao, X. (2008). Local Impacts of a Global Crisis: Food Price Transmission and Poverty Impacts in Ghana. *Local Impacts of a Global Crisis*, 3, 249-269.

DiStefano. C., Leighton, E., Ene, M., & Monrad, D. M. (2015). An Examination of Predictors and Outcomes Related To School Climate Using Latent Class Analysis. *Structural Equation Modeling*, Nova: Science Publishers, Inc.

Ellis, F. (1992). Agricultural Policies in Developing Countries. Cambridge: Cambridge University Press.

FAO (2008). Soaring Food Prices: Facts, Perspectives, Impacts and Actions required. Paper presented at High–level conference on World Food Security: The Challenges of Climatic and Bio-energy, Food and Agriculture Organization, Rome, 3-5 June.

Fabrigar, I. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4, 272-299.

Ferris, R.S.B. (1999). Stakeholders report for the formation of a market information service in Uganda. USAID Report. 126.

Ferris, R. S. B., Legg J., Bua A., Agona A., & Whyte, J. (2000). *Dissemination and Utilisation of Mosaic Resistant Cassava in Uganda*, 8th Quarterly USAID Report. 159.

Fischer, G., M. Shah, F. N. Tubiello, H., & Velhuizen, V. (2005). Socio-economic and climate change impacts on agriculture: an integrated assessment, 1990–2080. *Philosophical Transactions of the Royal Society of London*, 360, 2067–2083.

Frempong, M., Nkansah, H., & Nkansah, B. K. (2017). Multivariate Assessment of Industrial Attachment by Students of Tertiary Institutions in Ghana: *Mathematical Theory and Modelling*, 7(6), 139 – 156.

Gage, D., Bangnikon, J., Abeka-Afari, H., Hanif, C., Addaquay, J., Antwi, V., & Hale, A. (2012). The market for Maize, Rice, Soy and warehousing in Northern Ghana. *United States Agency for International Development*. Enabling Agricultural Trade (EAT).

GSS (2009). Food Price Database. *Accra, Ghana:* Ghana Statistical Service

GSS (2008). Food Price Database. *Accra, Ghana:* Ghana Statistical Service

Gilbert, C. L., & Morgan, C. W (2010). Recent Trends in World Food Commodity Prices: Costs and Benefits. Past and future Trends in World Food Prices. *American Journal of Agricultural Economics*, 80(4), 738 – 749.

Gilbert, D. C., & Jackaria, N. (2002). The efficacy of Sales Promotions in UK Supermarkets: Consumer View. International Journal of retail & Distribution Management, 30(6), 315 – 322.

Gnanadesikan, R. (1977). *Methods for Statistical Data Analysis of Multivariate Observations*. New York: Wiley.

Gnanadesikan, R., & Kettenring, J. R. (1972). Robust estimates, residuals and outlier detection with multiresponse data. *Biometrics*, 28, 81 – 124.

Gordor, B. K., & Fieller, N. R. J. (1994). How to display an outlier in multivariate datasets. *Journal of Applied Sciences and Technology*, 4(2), 50-62.

Gorsuch, R. L. (1988). Exploratory factor analysis. In J.R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (2nd Ed.). 231-258. New York: Wiley.

Gulumbe, S. U., Dikko, H. G., & Suleiman, H. (2014). Comparing two Approaches of Data Reduction Techniques in Principal Component Analysis. *International Journal of Innovative Research in Technology & Science*. 3(1), 1- 4.

IFAD (2011). Rural Poverty Report. Policy Report by G8 in June, 2011. *Price Volatility in Food and Agricultural Markets: Policy Responses*. A Report Co-ordinated by FAO and OECD.

Imad, E., Abdul, K. Y., & Safar, H. A. (2014). Effects of high food prices on consumption pattern of Saudi consumers: A case study of AL Riyady City. *Journal of the Saudi Society of Agricultural Sciences*, 13(2), 169-173.

James, L. R., & Brett, J. M. (1984). Mediator, Moderators, and tests for mediation. *Journal of Applied Psychology*, 69, 307-321.

Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis* (6th ed.). New Jersey: Prentice-Hall Int., Inc.

Kline, R. B. (2005). Principles and Practices of Structural Equation Modeling, 2nd ed. New York: Guilford Press.

Konuk, F. A. (2015). The Effects of Price Consciousness and Sale Proneness on Purchase Intention Towards Expiration Date-Based Priced Perishable Foods. *British Food Journal*, 117(2), 793 – 804.

Ledermann, T. & Macho, S. (2015). Assessing Mediation in Simple and Complex Models. *Structural Equation Modeling*, Nova: Science Publishers, Inc.

Li, X., Feng, X., Lu, B. & Song, X. (2015). The Determinants of Capital Structure Choice for Chinese Listed Companies Based On Structural Equation Modeling Approach. *Structural Equation Modeling*, Nova: Science Publishers, Inc.

Lyon, F. (2003). Trader Associations and Urban Food Systems in Ghana: Institutionalist Approaches to Understanding Urban Collective Action. *International Journal of Urban and Regional Researh* 4(3), 230 – 259.

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. New York: Erlbaum.

Meyer, S., & Yu, X. (2013). The Impart of Production Uncertainties on World Food Prices. *China Agricultural Economic Review*, 5(4), 459 – 472.

Ministry of Food and Agriculture (2007, August). Food and Agriculture Sector Development Policy, FASDEP II.

Mitchell, D. (2008, July). A Note on Rising Food Prices. Policy Research Working Paper 4682, The World Bank, Washington D.C.

Nkansah, B. K. (2018). On the Kaiser-Meyer-Olkin's Measure of Sampling Adequacy. *Journal of Mathematical Theory and Modelling*. 8(7), 52 – 76.

Nkansah, B. K., & Gordor, B. K. (2012a). A Procedure for Detecting a Pair of Outliers in Multivariate Datasets. *Studies In Mathematical Sciences*, 4(2), 1 – 9.

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

Nkansah, B. K., &. Gordor, B. K. (2013). Discordancy in Reduced Dimensions of Outliers In High-Dimensional Datasets: Application of Updating Formula. *American Journal of Theoretical and Applied Statistics*, 2(2), 29 – 37.

OECD (2008). Rising Food Prices: Causes and Consequences, Working paper, Organization for Economic Co-operation and Development.

OECD (2012). Rising Food Prices: Causes and Consequences. Working paper, Organization for Economic Co-operation and Development.

Raykov, T., & Marcoulides, G. A. (2006). A First Course in Structural Equation Modeling , 2nd ed. New York: Lawrence Erlbaum Associates.

Roache, S. K. (2010). *What Explains the Rise in Food Price Volatility*? IMF Working Paper, WP/10/129 – Authorized for distribution by Thomas Helbling Research Department.

Robbins, P., & Ferris, R. S. B. (1999). A Preliminary Study of the Maize Marketing System in Uganda and the Design of a Market Information System. CTA/IITA. Contract No. 4-1-06-215-9. *23*.

Robbins, P., & Ferris, R.S.B. (2000). *Co-ordination of a Preliminary Study of the Maize Marketing System in Uganda and the Design of a Market Data*

Information System. Contract No. 4-1-06-215-9. Design of a Pilot Scheme for Testing of a Market Information System. *16*.

Robbins, P., Ferris, R. S. B. & Muganga, A. K. (2000). *Market information Services in Post-liberalized Uganda and Eastern Africa*. Phaction News.

Russel, D. W. (2002). In search of underlying dimensions: The use (and abuse) of factor analysis in *Personality and Social Psychology Bulletin*.

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting Structural Modeling and Confirmatory factor Analysis results: A Review. *The Journal of Educational Research*, 6, 323 – 338.

Scott-Joseph, A. (2009). The Nature of Rising Food Prices in the Eastern Caribbean. An Analysis of Food Inflation During the Period 2005 – 2008 in a Context of Household Poverty. *UNICEF Office for Barbados and the Eastern Caribbean*.

Seglah, F. (2014). Multivariate Analysis of Food Prices in Ghana. MPhil Thesis, University of Cape Coast.

Shah, R., & Goldstein, S. M. (2006). Use of Structural Equation Modelling in Operations Management Research: Looking Back and Forward. *Journal of Operations Management*, 24, 148 – 169.

Sharma, S. (1996). *Applied Multivariate Techniques*. New York: John Wiley & Sons.

Shepherd, A. W. (1997). *Market Information Services: Theory and Practices*, FAO.

Shepherd, A. W. (2001). *Farm Radio as a Medium for Market Information Dissemination*, FAO.

Shepherd, W. (1997). *The Economics of Industrial Organization*. (4th ed.). New Jersey: Prentice-Hall.

Shyam, M. J. S, Shanmugapriya, P. & Kumaran, N. (2016). Data Reduction Techniques for High Dimensional Biological Data. *International Journal of Research in Engineering and Technology*, 5, 319 -325.

Thompson, B. (2000). Ten Commandments of Structural Equation Modeling. In L. Grimm & P. Yarnell (Eds), Reading and Understanding more Multivariate Statistics. *American Psychological Association*, 1, 261-284.

Timmer, C. P. (1989). *Food Price Policy: The Rational for Government Intervention*. Washington, *D.C.*: Butterworth & Co Publishers Ltd.

Vickers, J., & Hay, D. (1987). *The Economics of Market Dominance*. Oxford: Basil Blackwell.

Von Braun, J. (2007). The World Food Situation, New driving Forces and Required Actions, Food Policy Report. IFPRI, Washington, D.C.

Warde, A. (1997). *Consumption, Food and Taste*. London, Sage Publications.

Wodon, A. (2008). Potential impact on poverty of higher food prices: Summary evidence from West and Central Africa. *Mimeo. Washington, D.C.: World Bank.*

Zimmerman, R. (1999). Trade Liberalization and Poverty: What are the Links? *The World Economy*, 25(9), 1339-1367.

APPENDICES

APPENDIX A

DESCRIPTIVE STATISTICS OF THE PRICE DATA

APPENDIX A-1: DESCRIPTIVE STATISTICS OF PRICE OF FOOD ITEMS FOR YEAR 1

Commodities	Mean	Standard deviation	Coef. of Variation	Minimum price	Name of market	Maximum price	Name of Market
Root and Tubers							
Yam white	96.90	31.94	0.33	30.00	Bawjiase	205.00	Agona
Cassava	17.25	10.26	0.59	3.64	D/Ahenkro	45.33	Kasseh
Plantain (Apentu)	4.43	1.98	0.45	1.80	Kukuom	9.29	Tamale
Gari	40.30	8.45	0.21	20.14	Berekum	65.96	Kumasi
Vegetables							
Tomato	49.44	17.55	0.35	15.20	Nsuta	91.67	Ashaiman
Garden egg	16.34	8.48	0.52	6.00	Asawinso	46.25	Fumbisi
Cereal							
Local Rice	100.95	21.79	0.22	40.65	Agogo	144.00	Asawinso
Imported Rice	62.69	9.98	0.16	40.67	Denu	119.20	Agormanya
Maize	52.59	11.56	0.22	30.00	Bugubelle	92.00	Denu

APPENDIX A-2: DESCRIPTIVE STATISTICS OF PRICE OF FOOD ITEMS

Commodities	Mean	Std	CV	Min price	Market	Max price	Market
Oil							
Palm oil	24.55	7.69	0.31	14.00	Asawinso	46.75	Zebilla
Fruit							
Orange	5.45	4.13	0.76	1.40	Ahoman	23.33	Fumbisi
Banana	3.15	2.36	0.75	0.80	Ahoman	14.13	Wa
Fishes							
Smoked herring	29.17	15.85	0.54	10.00	Bawjiase	66.34	Tumu
Koobi	45.66	19.96	0.44	17.00	Bibiani	130.00	Nkoranza
Egg	4.42	0.58	0.13	3.50	Yeji	7.67	Fumbisi
Spices							
Dried pepper	58.06	19.62	0.34	18.00	Nsuta	108.29	Suhum
Onion	97.08	28.20	0.29	47.78	Bugubelle	160.67	Kaneshie
Pulses							
Groundnut red	113.41	28.86	0.25	59.82	Bugubelle	185.00	S/Bekwai
Cowpea white	115.06	28.51	0.25	57.50	Gushiegu	194.22	Adidome

Commodities	Mean	Std	CV	Min price	Name of	Max price	Market

APPENDIX A-2 CONTINUED

					market		
Root and Tubers							
Yam white	122.07	32.90	0.27	55.00	Damongo	240.00	Adugyama
Cassava	21.08	10.97	0.52	5.86	Berekum	49.20	Fumbisi
Plantain (Apentu)	5.09	2.19	0.43	2.00	Winneba	11.01	Akatsi
Gari	50.13	13.98	0.28	15.17	Bekwai	92.40	A/Nkwanta
Vegetables							
Tomato	86.74	27.42	0.32	25.79	Nsuta	148.45	Ashaiman
Garden egg	23.09	8.35	0.36	7.50	Zebilla	58.33	Salaga
Cereal							
Local Rice	107.16	22.73	0.21	49.46	Agogo	171.14	Medina
Imported Rice	65.68	9.93	0.15	39.43	Gushiegu	94.80	Agormanya
Maize	54.31	9.00	0.17	33.76	Bugubella	75.21	Obuasi
Oil							
Palm oil	26.50	8.13	0.31	13.71	Assin Praso	52.68	Zebilla

APPENDIX A-2 CONTINUED

Commodities	Mean	Std	CV	Min price	Market	Max price	Market
Fruit							
Orange	5.72	3.09	0.54	2.27	Agogo	17.11	Fumbisi
Banana	3.60	2.56	0.71	0.82	Sunyani	11.23	Tumu
Fishes							
Smoked herring	31.80	21.38	0.67	6.00	Kute	90.00	Salaga
Koobi	56.44	23.07	0.41	18.00	Nsuta	103.42	Techiman
Egg	5.25	0.61	0.12	3.93	Agogo	7.50	Yendi
Spices							
Dried pepper	150.85	66.10	0.44	32.50	Damango	360.00	Tumu
Onion	131.25	31.23	0.24	67.63	D/Nkwanta	225.00	Agormanya
Pulses							
Groundnut red	121.54	21.26	0.17	76.84	Gushiegu	178.00	Tepa
Cowpea white	112.94	26.99	0.24	50.49	Bugubelle	162.00	A/ Nkwanta

APPENDIX A-3: DESCRIPTIVE STATISTICS OF PRICE OF FOOD ITEMS FOR YEAR 3

Commodities	Mean	Std	CV	Min price	Market	Max price	Market
Root and Tubers							
Yam white	218.02	60.65	0.28	63.00	Zebilla	387.50	Tema
Cassava	35.35	14.50	0.41	11.00	Yeji	75.80	Mallam
Plantain (Apentu)	9.86	3.87	0.29	4.92	Berekum	23.80	Kaneshie
Gari	80.61	23.11	0.21	44.51	Denu	161.36	Bogoso
Vegetables							
Tomato	138.27	45.33	0.33	38.50	Mampong	251.67	Damongo
Garden egg	39.01	10.10	0.26	16.89	Goaso	62.40	Mallam
Cereal							
Local Rice	146.92	30.57	0.21	77.78	Yeji	247.27	Agona
Imported	106.24	22.89	0.22	57.09	Denu	188.56	D/Nkwanta
Maize	90.65	20.51	0.23	48.69	Bugubelle	130.78	Nsawam
Oil							
Palm oil	41.36	10.38	0.25	14.00	D/Nkwanta	67.17	Mafi-Kumasi

APPENDIX A-3 CONTINUED

Commodities	Mean	Std	CV	Min price	Market	Max price	Market
Fruit							
Orange	10.78	7.02	0.65	4.00	Zebilla	37.83	Fumbisi
Banana	6.69	3.63	0.54	1.20	Obuasi	18.00	Nalerigu
Fishes							
Smoked herring	53.86	26.03	0.48	20.00	Dome	130.00	Navrongo
Koobi	84.44	36.67	0.43	26.00	Damongo	232.33	Adidome
Egg	8.10	0.73	0.09	6.29	Agona	9.75	Fumbisi
Spices							
Dried pepper	186.75	73.99	0.40	76.00	Tamale	405.45	Suhum
Onion	152.50	43.52	0.29	71.55	Zebilla	249.47	Mampong
Pulses							
Groundnut red	266.15	51.68	0.19	116.00	Yeji	427.85	Adugyama
Cowpea white	208.92	42.00	0.20	100.00	Agogo	312.92	Goaso

APPENDIX A-4:DESCRIPTIVE STATISTICS OF PRICE OF FOOD ITEMS FOR YEAR 4

Commodities	Mean	Std	CV	Min price	Market	Max price	Market
Root and Tubers							
Yam white	256.67	70.06	0.27	92.00	Fumbisi	500.00	Kukuom
Cassava	48.28	17.97	0.37	15.00	Yeji	88.13	Sekondi
Plantain (Apentu)	12.68	9.74	0.27	5.00	D/Ahenkro	95.10	Fumbisi
Gari	108.63	29.23	0.21	50.00	Yeji	208.86	Obuasi
Vegetables							
Tomato	185.36	67.89	0.37	53.60	Damango	311.67	Wa
Garden egg	55.63	20.40	0.37	20.93	Denu	124.00	Ashaiman
Cereal							
Local Rice	164.69	35.73	0.22	80.00	Yeji	269.00	Juaben
Imported	125.49	25.75	0.21	79.80	Tumu	206.67	D/Nkwanta
Maize	76.81	19.20	0.25	36.20	Bugubelle	120.00	Medina
Oil							
Palm oil	52.46	13.04	0.25	26.66	Obogu	93.64	Mpreaso

APPENDIX A-4 CONTINUED

Commodities	Mean	Std	CV	Min price	Market	Max price	Market
Fruit							
Orange	11.26	4.82	0.43	4.25	D/Nkwanta	28.80	Tumu
Banana	7.18	2.99	0.42	1.20	Obuasi	15.00	Nalerigu
Fishes							
Smoked herring	66.41	35.76	0.54	29.55	Bawjiase	221.17	Akatsi
Koobi	112.31	51.34	0.46	28.00	Winneba	275.50	Juaben
Egg	9.173	1.13	0.12	7.06	Agona	13.50	Salaga
Spices							
Dried pepper	202.12	83.55	0.41	74.00	Gusheigu	425.00	Sekondi
Onion	216.67	51.69	0.24	86.38	Denu	341.67	Nalerigu
Pulses							
Groundnut red	265.16	49.31	0.19	138.23	Denu	373.33	Medina
Cowpea white	223.66	51.13	0.23	91.13	Gusheigu	360.00	Kukuom

APPENDIX A-5: DESCRIPTIVE STATISTICS OF PRICE OF FOOD ITEMS FOR YEAR 5

Commodities	Mean	Std	CV	Min price	Market	Max price	Market
Root and Tubers							
Yam white	322.70	96.10	0.30	112.80	Asamankese	643.80	Kukuom
Cassava	55.18	23.05	0.42	19.33	Ejura	120.00	Yendi
Plantain (Apentu)	15.67	7.84	0.50	5.92	D/Ahenkro	39.17	Navrongo
Gari	117.75	33.98	0.29	53.08	Asamankese	224.30	S/Bekwai
Vegetables							
Tomato	292.70	105.60	0.36	72.20	S/Bekwai	566.90	Kasseh
Garden egg	85.13	26.35	0.31	30.00	Bimbilla	172.86	Dunkwa
Cereal							
Local Rice	255.21	79.33	0.31	100.00	D/Ahenkro	514.30	S/Bekwai
Imported	198.95	54.43	0.27	90.00	Winneba	340.00	Kintampo
Maize	144.84	29.86	0.21	80.78	Bugubella	231.25	Kukuom
Oil							
Palm oil	85.54	26.12	0.31	35.00	Bawjiase	167.34	Akoase

APPENDIX A-5 CONTINUED

Commodities	Mean	Standard deviation	Coef. of Variation	Minimum price	Name of market	Maximum price	Name of Market
Fruit							
Orange	25.18	18.99	0.75	5.68	Adugyama	94.25	Mafi-Kumasi
Banana	12.49	17.33	0.59	4.50	D/Ahenkro	38.89	Salaga
Fishes							
Smoked herring	148.6	96.40	0.65	35.00	Bawjiase	641.70	Juaboso
Koobi	237.8	111.30	0.47	30.00	Damango	644.90	Techiman
Egg	14.47	1.574	0.10	10.00	Mpreaso	18.67	Denu
Spices							
Dried pepper	263.8	99.30	0.38	60.00	Yeji	590.00	Dunkwa
Onion	277.29	77.81	0.28	91.30	Bogoso	432.00	Elmina
Pulses							
Groundnut red	398.39	76.36	0.19	50.00	Tepa	535.33	Juaben
Cowpea white	299.03	81.67	0.27	38.00	Agormanya	495.00	Kukuom

APPENDIX B

VECTOR OF TRANSFORMED UNIVARIATE FACTOR SCORES BASED ON TWENTY-THREE FACTOR SOLUTION

No.	Factor Score	No.	Factor Score	No.	Factor Score
1	6.2867	31	13.0232	61	14.2515
2	8.9920	32	10.2560	62	34.4907
3	8.2122	33	7.3375	63	18.0276
4	9.6211	34	24.3912	64	38.9369
5	9.7054	35	6.1111	65	24.6672
6	18.0622	36	15.3080	66	25.3261
7	8.4081	37	10.0504	67	16.5171
8	9.8343	38	10.8769	68	24.3025
9	8.3498	39	7.7695	69	14.3581
10	7.1758	40	12.2661	70	14.3581
11	10.0504	41	20.4549	71	43.8134
12	7.0233	42	22.3465	72	28.4106
13	7.0834	43	24.3249	73	33.2449
14	6.3715	44	13.3348	74	28.3278
15	9.2513	45	27.7189	75	22.8332
16	12.3385	46	33.4129	76	30.4318
17	8.1616	47	26.6544	77	17.5962
18	7.3469	48	19.0522	78	18.0701
19	17.5189	49	20.5347	79	18.9982
20	5.1584	50	22.1653	80	29.2298
21	4.8115	51	27.3187	81	20.4224
22	6.4041	52	22.3532	82	19.7477
23	7.7695	53	15.2025	83	17.2380
24	6.6475	54	16.8978	84	16.0279
25	8.1887	55	24.1924	85	22.2377
26	14.8700	56	17.2599	86	17.1566
27	10.0916	57	15.6938	87	14.1743
28	7.3010	58	18.5120	88	24.3313
29	8.2547	59	48.0724	89	22.3167
30	11.6129	60	13.6831	90	17.2736

No.	Factor Score	No.	Factor Score	No.	Factor Score
91	15.9864	121	19.8955	151	42.3921
92	22.1459	122	31.7433	152	27.6845
93	16.4533	123	24.0121	153	26.3197
94	20.2713	124	25.9645	154	22.7654
95	30.1526	125	44.3980	155	49.4495
96	29.5104	126	17.7392	156	52.6673
97	35.8124	127	15.4494	157	14.2539
98	37.0285	128	19.3634	158	31.9450
99	20.3581	129	12.6581	159	43.6153
100	30.7290	130	41.5222	160	47.7948
101	34.5935	131	19.8614	161	13.9783
102	16.4142	132	20.2889	162	20.0818
103	14.4023	133	26.8240	163	19.1153
104	23.5539	134	38.9795	164	17.5058
105	27.6393	135	31.6419	165	18.2405
106	34.2145	136	21.3243	166	17.6647
107	27.4505	137	19.9709	167	18.8228
108	38.2078	138	14.7785	168	22.7075
109	30.8070	139	20.5297	169	25.0923
110	23.7341	140	13.5012	170	44.9150
111	19.1482	141	22.3790	171	17.3680
112	19.7004	142	31.3555	172	22.3072
113	22.1825	143	32.3366	173	27.8401
114	18.6728	144	22.5377	174	19.0171
115	19.7679	145	15.6471	175	22.1536
116	29.7934	146	27.1589	176	17.3203
117	16.4700	147	22.9719	177	24.3201
118	17.7320	148	41.2885	178	21.9464
119	46.3947	149	15.1299	179	13.2620
120	25.1509	150	39.4917	180	29.8679

No.	Factor Score	No.	Factor Score	No.	Factor Score
181	18.2158	211	20.9067	241	64.5051
182	23.4457	212	9.1231	242	27.9630
183	12.6041	213	12.3657	243	58.7074
184	9.1068	214	14.3301	244	74.2116
185	17.5808	215	23.3566	245	70.0908
186	14.7073	216	15.3574	246	44.6397
187	14.7312	217	11.9866	247	87.6627
188	31.9079	218	11.0187	248	45.9658
189	45.6666	219	14.4416	249	28.4264
190	8.9624	220	18.5390	250	85.6657
191	16.5027	221	12.9038	251	26.7949
192	26.3306	222	20.9860	252	47.1032
193	22.7491	223	6.6804	253	36.6521
194	18.8718	224	17.3145	254	20.8040
195	20.2660	225	8.6030	255	15.7990
196	16.3680	226	15.2969	256	10.0172
197	19.0068	227	18.5521	257	11.5113
198	19.9757	228	20.2815	258	21.8683
199	17.4689	229	19.4552	259	17.3558
200	31.4493	230	18.6982	260	16.7869
201	18.1866	231	14.2718	261	20.6023
202	19.1836	232	15.6390	262	23.0442
203	18.8038	233	22.9616	263	15.4153
204	13.3581	234	23.8855	264	13.6395
205	13.6037	235	11.2496	265	26.7335
206	18.4539	236	18.6749	266	38.0984
207	15.1966	237	12.8834	267	34.0230
208	9.4998	238	18.1633	268	33.2348
209	10.1229	239	18.8038	269	23.6260
210	20.1722	240	10.5595	270	25.7524

APPENDIX B CONTINUED

No.	Factor Score	No.	Factor Score	No.	Factor Score
271	15.3155	301	41.8154	331	34.3014
272	16.8968	302	30.9199	332	21.2962
273	54.7062	303	25.2384	333	19.2041
274	44.1003	304	21.9693	334	28.3685
275	47.5720	305	31.1860	335	19.7659
276	68.6390	306	21.4626	336	18.8135
277	61.2179	307	24.6731	337	35.3463
278	32.4258	308	23.7743	338	14.7285
279	36.7664	309	24.4768	339	22.1129
280	15.1666	310	21.0562	340	30.2766
281	14.5278	311	21.1449	341	16.3715
282	25.7120	312	13.0531	342	19.9994
283	22.7146	313	22.0876	343	22.8824
284	20.0761	314	22.5925	344	19.3032
285	48.3264	315	25.1340	345	39.3483
286	23.9625	316	17.2938	346	27.0355
287	25.3361	317	19.6509	347	37.9919
288	21.9739	318	18.6058	348	28.8325
289	30.3819	319	16.8511	349	25.1286
290	19.3361	320	14.6653	350	30.1660
291	24.7765	321	17.6621	351	21.5607
292	12.0457	322	21.3232	352	10.9177
293	22.9933	323	19.2064	353	20.0845
294	20.0901	324	15.8492	354	24.0188
295	17.7776	325	39.9279	355	35.9954
296	13.0719	326	45.7395	356	33.9473
297	19.3574	327	28.3628	357	15.7425
298	20.7936	328	35.4502	358	32.8656
299	30.9171	329	26.4117	359	7.6875
300	28.8572	330	24.0886	360	16.6793

APPENDIX C

VARIANCE-COVARIANCE MATRIX OF DATA ON YEAR TWO

Item	Maize	WhYam	Cassava	Tomato	Gegg	DPep	RdGrnt	WhCow	PalmOil	Orange	Banana	SmkHerr
Maize	81.0137											
WhYam	145.1891	1082.5573										
Cassava	-22.0832	-37.2580	120.2562									
Tomato	135.2218	322.0090	15.9541	751.7395								
Gegg	14.7628	25.3847	17.2733	44.4016	69.7932							
DPep	-101.7188	11.1655	141.4899	438.8775	80.1114	4369.4309						
RdGrnt	117.7146	343.0233	-98.6964	258.5589	6.5431	107.4214	451.9239					
WhCow	143.3425	548.7759	-78.1399	409.7417	-2.9519	-7.6589	421.2758	728.6847				
PalmOil	-28.8135	-20.7043	33.7437	-14.3553	-1.5374	94.0975	-58.0337	-16.1422	66.1672			
Orange	-13.7734	-25.9385	20.5890	-25.8151	7.0782	25.4796	-38.3950	-43.1780	9.7496	9.5239		
Banana	-9.4665	-13.5824	20.8482	-18.2200	5.4168	36.1742	-29.6431	-31.2867	8.4773	6.3373	6.5625	
SmkHerr	-100.8233	-237.8172	120.8839	-229.3404	31.5695	58.6489	-262.1332	-276.6205	56.6654	50.9097	39.3070	456.9799
Koobi	-115.0901	-274.8473	126.9976	-135.0972	6.2102	194.0075	-310.0265	-256.6075	84.5406	46.4619	35.2919	311.6451
Onion	78.2250	377.5193	-64.5392	460.5353	22.5471	398.5169	301.1007	357.3542	-62.9214	-27.9693	-19.3383	-241.5371
Egg	-0.3137	0.9063	3.5599	3.9649	0.2637	9.9263	-1.5638	0.7460	2.1584	0.6391	0.7362	3.0733
Plantain	-5.3210	-5.7955	17.9480	-1.0245	4.5084	21.9251	-19.8862	-15.9272	9.2426	4.8090	4.4305	32.3541
Gari	50.9733	202.0996	-19.4046	84.8243	6.4721	-174.5562	106.2780	88.1155	-38.4347	-7.0589	-5.2999	-88.0784
LocRice	36.9479	303.1509	-42.0003	142.4977	27.1852	292.6520	190.7029	306.9043	19.6572	-13.4520	-8.7731	-109.1581
ImpRice	-6.9731	-55.2654	-16.5091	-1.3420	-10.3878	141.7299	10.4616	-3.2598	11.0747	-1.6041	-2.2288	5.6855

APPENDIX C CONTINUED

Item	Koobi	Onion	Egg	Plantain	Gari	LocRice	ImpRice
Koobi	532.2693						
Onion	-173.9956	975.1509					
Egg	5.3195	2.3378	0.3664				
Plantain	29.5853	-9.6412	0.6228	4.7805			
Gari	-84.4502	195.3029	0.0156	-4.2237	195.5396		
LocRice	-96.3616	180.0601	0.1053	-4.4389	-21.2035	516.4945	
ImpRice	-27.4165	2.4011	-0.1374	-0.4773	-27.1752	10.8509	98.6275

APPENDIX D

COMPONENT SOLUTION FOR DATA FROM EACH YEAR

APPENDIX D-1: QUARTIMAX ROTATION (2008)

Variable	Factor1	Factor2	Factor3	Factor4	Factor5
Maize	-0.104	-0.774	-0.002	0.021	-0.049
WhYam	-0.046	-0.573	0.212	0.240	0.342
Cassava	0.857	-0.104	-0.071	0.041	0.091
Tomato	0.013	-0.702	-0.430	0.063	0.100
Gegg	0.706	-0.129	0.068	0.174	0.059
DPep	0.048	-0.179	-0.713	-0.044	0.110
RdGrnt	-0.347	-0.798	0.019	0.146	-0.022
WhCowpea	-0.335	-0.817	-0.119	-0.173	-0.005
PalmOil	0.444	-0.030	-0.195	-0.672	0.219
Orange	0.774	0.258	0.205	0.092	-0.090
Banana	0.890	0.203	0.055	0.039	-0.050
SmkHerr	0.758	0.394	0.212	-0.121	-0.079
Koobi	0.565	0.388	-0.112	-0.210	-0.031
Onion	-0.350	-0.023	-0.761	0.131	0.039
Egg	0.665	-0.017	-0.339	-0.279	-0.076
Plantain	0.898	0.004	0.085	-0.056	-0.012
Gari	0.057	-0.209	-0.180	0.834	0.068
LocRice	0.239	-0.501	-0.079	-0.063	-0.295
ImpRice	-0.062	0.004	-0.213	-0.065	0.889
Variance	5.3961	3.4939	1.6894	1.4698	1.1097
% Var	0.284	0.184	0.089	0.077	0.058

Appendix D-2: Quartimax Rotation (2009)

Variable	Factor1	Factor2	Factor3	Factor4	Factor5
Maize	-0.408	0.613	0.120	0.276	0.195
WhYam	-0.137	0.594	0.435	0.339	0.059
Cassava	0.842	0.085	-0.047	0.053	0.019
Tomato	-0.054	0.802	0.084	-0.189	0.013
Gegg	0.232	0.171	0.109	-0.003	0.845
DPep	0.236	0.239	0.067	-0.741	0.152
RdGrnt	-0.557	0.590	0.276	-0.027	0.091
WhCowpea	-0.351	0.625	0.535	0.083	-0.123
PalmOil	0.583	-0.152	0.452	-0.176	-0.369
Orange	0.819	-0.279	-0.086	0.020	0.231
Banana	0.885	-0.144	-0.043	0.024	0.175
SmkHerr	0.723	-0.418	-0.053	0.002	0.187
Koobi	0.744	-0.273	-0.084	-0.044	-0.140
Onion	-0.165	0.744	-0.137	-0.193	0.040
Egg	0.666	0.403	0.049	-0.108	-0.318
Plantain	0.887	0.008	0.078	0.018	0.094
Gari	-0.148	0.586	-0.340	0.427	0.066
LocRice	-0.148	0.220	0.785	-0.149	0.157
ImpRice	-0.127	-0.098	0.031	-0.683	-0.099
Variance	5.6292	3.6931	1.5747	1.5392	1.2284
% Var	0.296	0.194	0.083	0.081	0.065

APPENDIX D-3: QUARTIMAX ROTATION (2012)

Variable	Factor1	Factor2	Factor3	Factor4	Factor5
Maize	0.844	-0.161	-0.073	0.038	-0.077
WhYam	0.617	-0.443	-0.013	-0.106	0.025
Cassava	-0.228	-0.258	-0.730	0.100	0.191
Tomato	0.596	-0.294	-0.376	0.159	0.093
Gegg	0.280	0.090	-0.758	0.156	-0.116
DPep	0.352	-0.715	0.013	-0.111	0.084
RdGrnt	0.318	-0.621	-0.105	-0.337	-0.350
WhCowpea	0.267	-0.492	-0.041	-0.562	-0.256
PalmOil	-0.244	0.075	-0.335	-0.567	0.177
Orange	-0.816	-0.103	-0.282	0.068	0.062
Banana	-0.691	0.287	-0.411	-0.007	-0.033
SmkHerr	-0.768	0.052	-0.076	-0.042	0.144
Koobi	0.012	-0.092	0.064	-0.632	0.080
Onion	0.157	-0.817	0.106	0.070	-0.197
Egg	-0.228	-0.061	-0.602	-0.351	0.035
Plantain	-0.173	0.197	-0.656	-0.199	-0.058
Gari	0.222	-0.385	-0.168	0.265	-0.589
LocRice	0.046	0.016	-0.111	-0.556	-0.549
ImpRice	0.082	-0.119	0.144	0.031	-0.687
Variance	3.8322	2.5062	2.5031	1.7883	1.4880
% Var	0.202	0.132	0.132	0.094	0.078

APPENDIX D-4: QUARTIMAX ROTATION (2013)

Variable	Factor1	Factor2	Factor3	Factor4	Factor5
Maize	0.804	0.076	0.054	-0.049	-0.144
WhYam	0.613	0.224	-0.052	-0.114	0.178
Cassava	0.112	-0.073	-0.853	0.139	-0.051
Tomato	0.694	-0.113	-0.223	-0.186	-0.241
Gegg	0.453	0.114	-0.658	-0.048	0.221
DPep	0.699	-0.233	-0.254	0.131	-0.019
RdGrnt	0.761	-0.134	0.191	-0.366	0.117
WhCowpea	0.717	-0.095	-0.053	-0.129	0.300
PalmOil	-0.035	-0.680	0.152	0.077	0.128
Orange	-0.432	-0.483	-0.148	0.210	0.369
Banana	-0.708	-0.396	-0.108	-0.164	0.014
SmkHerr	-0.408	-0.192	0.258	-0.183	0.245
Koobi	0.077	0.056	-0.038	-0.101	0.822
Onion	0.191	-0.107	-0.078	-0.649	-0.426
Egg	0.013	-0.734	-0.075	-0.022	-0.221
Plantain	-0.347	-0.455	-0.394	-0.126	-0.049
Gari	0.243	0.353	-0.326	-0.560	0.051
LocRice	0.085	-0.102	0.087	-0.562	0.074
ImpRice	0.022	0.207	0.098	-0.378	0.106
Variance	4.3866	1.9965	1.7342	1.5701	1.4080
% Var	0.231	0.105	0.091	0.083	0.074

APPENDIX D-5: QUARTIMAX ROTATION (2015)

Variable	Factor1	Factor2	Factor3	Factor4	Factor5
Maize	-0.465	0.443	0.257	0.248	-0.046
WhYam	-0.008	0.269	0.283	0.337	-0.413
Cassava	0.718	0.164	-0.061	0.003	0.149
Tomato	-0.110	0.823	-0.172	0.033	-0.010
Gegg	0.144	0.308	-0.231	0.454	-0.255
DPep	-0.159	0.065	-0.088	0.719	0.305
RdGrnt	-0.113	0.315	0.118	0.609	0.076
WhCowpea	-0.245	0.431	0.528	0.177	0.216
PalmOil	0.180	-0.146	-0.020	-0.015	0.723
Orange	0.404	-0.304	-0.029	-0.150	0.264
Banana	0.822	-0.049	-0.001	-0.022	-0.103
SmkHerr	0.520	-0.474	0.344	0.124	0.080
Koobi	-0.042	-0.298	0.590	0.085	-0.081
Onion	0.048	0.716	0.123	0.149	-0.064
Egg	0.179	0.081	0.100	0.068	0.710
Plantain	0.862	-0.084	0.018	0.039	0.205
Gari	0.109	-0.123	0.197	0.641	-0.249
LocRice	0.016	-0.009	0.766	0.031	0.131
ImpRice	0.069	0.380	0.608	-0.122	-0.236
Variance	2.7984	2.4735	2.0231	1.8007	1.7069
% Var	0.147	0.130	0.106	0.095	0.090

APPENDIX E

DEVIANCE RESIDUAL PLOT FOR PRICE LEVEL