This is to certify that the

thesis entitled

Forecasting Agricultural Prices in an Underdeveloped Open Economy:
The Case of Maize in Mozambique

presented by

Pedro Arlindo

has been accepted towards fulfillment
of the requirements for

Masters' degree in Agricultural Economics

Major professor

Date February 20, 2001

FORECASTING AGRICULTURAL PRICES IN AN UNDERDEVELOPED OPEN ECONOMY: THE CASE OF MAIZE IN MOZAMBIQUE

By

Pedro Arlindo

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

Department of Agricultural Economics

2001

ABSTRACT

FORECASTING AGRICULTURAL PRICES IN AN UNDERDEVELOPED OPEN ECONOMY: THE CASE OF MAIZE IN MOZAMBIQUE

By

Pedro Arlindo

Mozambique's maize marketing system faces high levels of price uncertainty. Price uncertainty can be reduced if future price are forecasted accurately and used along with consistent policies. This research estimates alternative out-of-sample forecasting models for maize prices in Mozambique and evaluates the statistical and economic forecasting performance of the alternative models. The research uses an econometric time-series approach through the estimation of univariate ARMA and ARIMA models and multivariate VAR and VEC models for the northern and the central/southern regions of the country. The data includes monthly retail maize prices, from November 1992 through August 2000, and maize production data for 1992 - 2000. Preliminary tests indicate that all price series are nonstationary and the series within each region are cointegrated. The main conclusion is that the performance of the forecasting models is generally poor due to significative differences in price path between the model estimation period and the forecasting period. Nevertheless, statistical evaluation suggests that the estimated models have a potential to improve random walk models' forecasts, and economic evaluation suggests that the estimated models tend to have better results than a 'no-model' strategy, especially in the center/south.

To my Parents.
To Telvio and Tunelga.

ACKNOWLEDGMENTS

This research would have never been possible if many people had not given their time, patience and support. I would like to thank all those who made possible the completion of this thesis. My special thanks are for my committee members. Professor David L. Tschirley, my major professor and thesis supervisor, was the author of the idea of time series models for maize prices in Mozambique. His e-mails and phone conversations helped me think about the importance of research on maize price modeling in Mozambique. Professor Michael T. Weber was my major professor for most of my Masters' program, and was extremely important for my achievements. He helped me in careful course selection leading me toward the skills needed for the kind of analysis that I present in this thesis. His support and meticulous advice throughout the program was critical to the completion of the research. Professor Robert J. Myers was also extremely important for the completion of this thesis. By agreeing to examine and correct the methods and models that I have used, he transmitted to me the confidence I needed to face the challenge of the topic. Throughout the research, he patiently spent many hours working with me. Finally, Professor Jeffrey Wooldridge agreed to be a member of my committee, and gave me important contributions. His comments on the thesis were very important for corrections and improvements.

Many other people and institutions made this research possible. Special thanks go to the ATLAS Program for funding my Masters' program, and to the Food Security Project in the Department of Agricultural Economics at Michigan State University for providing technical assistance. On campus, three Mozambicans gave me special support. Initially Dr. Paulo N. Mole, and later Rui Benfica and Marcia Jeremias-Benfica. Dr. Cynthia Donovan also played a key role for the completion of this research. To them and to all the fellow graduate students in the Department of Agricultural Economics, my great thanks.

In Mozambique, special friendship and support came from Perpétua, who agreed to stay alone with our two very young children while I was studying. It was a very difficult decision. Thanks Perpétua. Thanks also for the pictures you always sent, for all the good work towards a better life for our children, and for permanently updating me with our kids' health and our relatives' lives. Many thanks. In all this, "Tio Lourenço" and "Tia Graça" helped a lot. My sincere thanks to both.

Paula Santos and Anabela Mabota worked with me at SIMA in Maputo and helped me understand that a good team always makes a good success. Higino Marrule, Nita Dengo and José J. Jeje were also there. With their experience in similar programs, they helped me be confident and make the decision to "try" the experience. Thank you all. Conceição, Simão, Chico, Abel and Reginaldo did a great work in helping, each one in his way. Danilo Abdula had timely and extremely helpful responses to my e-mails when I needed additional data. To all of them and all other colleagues at SIMA in Maputo, my sincere thanks.

To my extended family and friends, my deepest thanks. Special thanks go to João, Moisés, Anselmo, Lino, Jaime, Ermelinda and Dominica for helping me with everything they did. To my grandparents, parents, Mana Teresa and Belalva, my gratefulness for all the help.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION AND RESEARCH OBJECTIVES

1.1 Introduction

In the 1980s, several developing countries experienced profound transformations in their economic policy, including the agricultural sector. Mozambique is not an exception. After following a centrally planned economic policy since 1975, in 1987 the government of Mozambique (GOM) adopted a structural adjustment program (SAP), funded by the World Bank and the International Monetary Fund that was aimed at reverting the downward trend in economic growth rates. One important constituent of Mozambique's SAP was the liberalization of agricultural markets.

The agricultural sector plays a key role in Mozambique's economy, employing approximately 85% of the population and generating 35% of GDP. Maize is the most important staple in the country, both in production and consumption, though cassava is more important in some areas of the north. Maize production in Mozambique takes place in all provinces, but production potential is greater in the northern and central regions of the country. On the other hand, the most important consumption centers in the country are the three major cities, namely Maputo in the south, Beira in the center and Nampula in the north. Maputo, the capital city and the most important consumption center in the country, and most southern areas, are net maize consumers. Due to these geographic patterns, commercial linkages between southern and central Mozambique are considerable and, as a result, retail prices in the south, especially in Maputo, are highly correlated with prices in some of the most important producer areas in central Mozambique.

Linkages between southern and central Mozambique face several infrastructure constraints, which leads to high transportation costs. As a consequence, maize imports to the south from South Africa are notable and prices in the southern urban markets are expected to be highly correlated with the expected production in South Africa. Transport cost between northern and southern Mozambique are so high that maize trade between these two regions occur very infrequently.

In Mozambique, expected profits in the maize subsector, as well as in other food staples, are generally low compared to the profits obtained from cash crop staples and non-agricultural sectors. Farm production in the maize subsector is almost entirely performed by smallholder households, and trade between surplus and deficit areas is basically dominated by small scale informal traders, with little presence of large scale formal traders. Also, there is a high level of uncertainty on quantity and prices in the maize subsector, and producers and consumers assume individually the risk associated with the quantity and quality of product they produce, sell, buy or consume.

1.2 Problem Statement and General Objective

One of the most important characteristics of agricultural commodities in Mozambique is their price instability. Indeed, both short-term and long-term management decisions in the agricultural sector involve uncertainty, which is greater the more the agricultural production depends on climatic conditions. This is the case in Mozambique where, in addition to the individual risk to participants in the food system in determining what, when and how much to produce, store and sell, agriculture depends almost entirely on climate, such as the timing and levels of rainfall. This increases risk and uncertainty. High levels of risk associated with price uncertainty can be reduced if consistent price forecasting models are estimated and results are disseminated to all participants in the food system throughout the country. Price forecasting models might help reduce uncertainty if used judiciously in combination with knowledge about (i) patterns and levels of production in the country, (ii) patterns and levels of production in the region, and (iii) policy initiatives which might alter production levels and/or trade flows in the region. The general objective of this research is to see if existing data and techniques will support the development of models which might improve commodity price forecasts in Mozambique.

Mozambique has a consistent and regular data series on weekly maize prices from 1992 to the present. Mozambique's agricultural market information system (SIMA, from its acronym in Portuguese), with financial and technical assistance under the Food Security II project in Mozambique, has played a critical role in the development of a solid system of price data collection and analysis in the Ministry of Agriculture and Rural Development of Mozambique. However, like in many other developing countries, this data has been primarily used for short-term market information reports, intended to allow quick, appropriate decisions by producers, traders, consumers and the public sector.

Nonetheless, in addition to the importance of the short-term price reporting services, this price data can be used for other purposes like price forecasting._This research estimates and tests short-term forecasting models for maize prices in Mozambique.

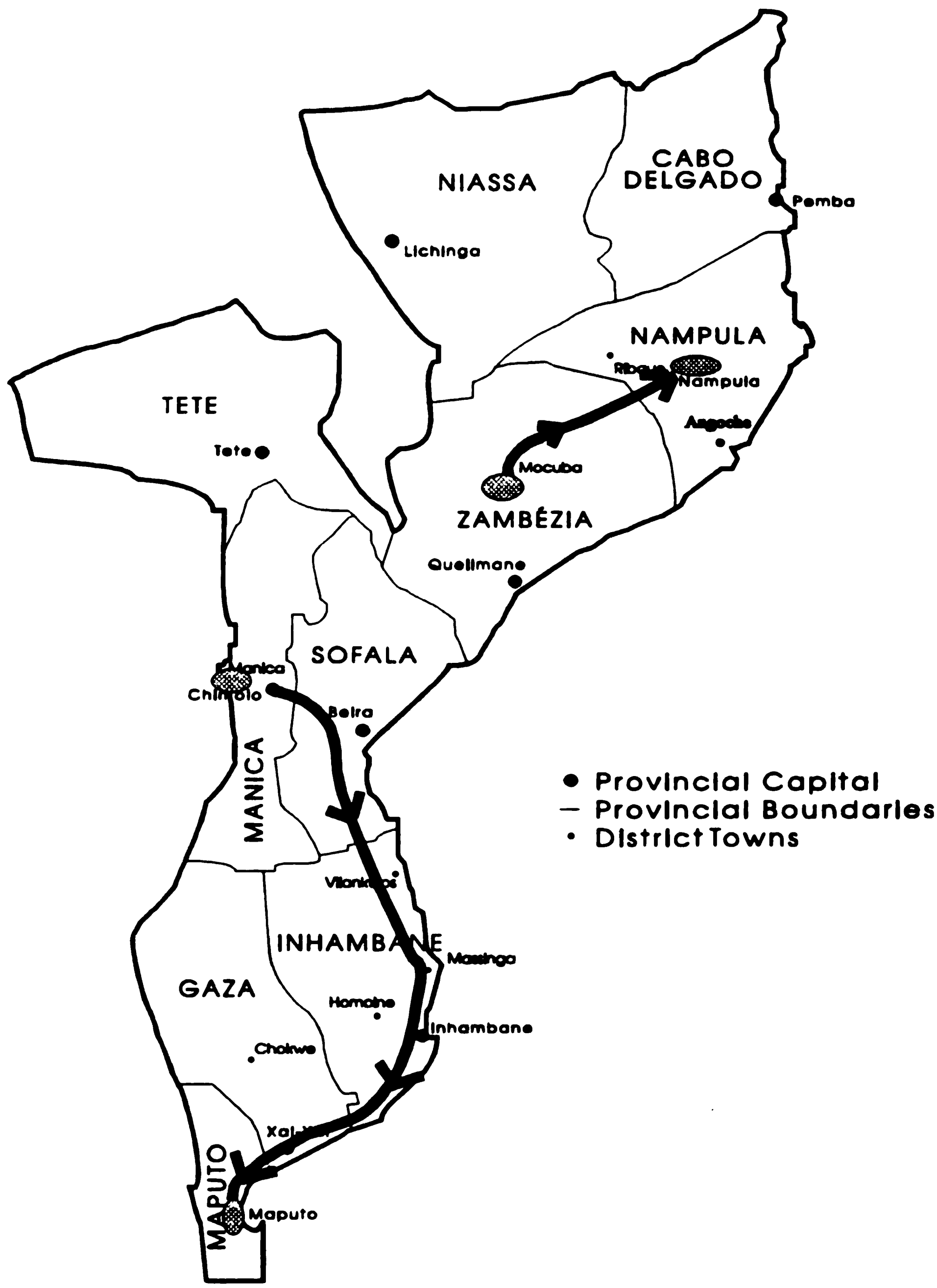
1.3 Modeling Maize Prices in Mozambique

Modeling maize prices in Mozambique requires taking into consideration that Mozambique has two regions with different characteristics in maize production, and without commercial connections. On the one hand, southern Mozambique is a poor maize producer, where most maize consumed is brought from central Mozambique. Except for large-scale millers and animal feed producers, who import nearly all their maize from South Africa due to better quality, large scale, and lower transportation costs, SIMA data indicate that during the last decade, about 95 percent of the domestic maize traded in informal retail and wholesaler markets in Maputo has been bought either in central Mozambique or in the producer districts within southern Mozambique. Indeed, the most important destination for maize produced in central Mozambique is southern Mozambique. Between 1992 and 2000, about 64 percent of the informal wholesalers in the area of Chimoio, central Mozambique, indicated that they would sell their product in southern Mozambique. The remaining product has been supplied to the net consumer areas within central Mozambique, among them Beira, where about 93 percent of the maize has been acquired within the central region of Mozambique.

In Figure 1.1, the typical maize trade flows from Mocuba to Nampula, and from Manica to Maputo are emphasized. By road, the distance between Mocuba and Nampula is approximately 400 km, and that between Manica and Maputo is 1,135 km.

On the other hand, northern Mozambique is commercially isolated from the center and south, regardless of its high potential to produce maize, due to poor and expensive road linkages, un-

Figure 1.1 Map of Mozambique With Principal Domestic Maize Trade Flows



operational and inefficient maritime connections, and nonexistence of north-south railway connections. Maize produced in the north is consumed within the region and exported to neighboring countries, especially Malawi. Indeed, Malawi has been shown to be an important destination of Mozambican maize when maize production drops in the southern African region. As an example, when maize production decreased 34 percent in Malawi in the 1997/98 season, imports from Mozambique were an important resource to Malawian traders. Total formal exports from northern Mozambique to Malawi were about 42,000 metric tons that season, and estimated to be approximately 100,000 metric tons including informal trade in the 1998/99 season (Santos and Tschirley, 1999).

These exports to Malawi had noteworthy effects on maize prices in northern Mozambique, but little or no effect in the south. Maize retail price in northern Mozambique increased between 13 percent in Nampula city and 21 percent in the rural district of Mocuba in 1997, and cash income earnings from maize sales had a total increment of about 3.5 million US dollars in Nampula and Zambezia in the 1998/99 marketing season as a result of trading to Malawi (Santos and Tschirley, 1999). No statistically significant price increase effects were found in southern markets. Based on these characteristics of maize production, trading and consumption in Mozambique, modeling maize prices in the country might imply dividing the country into two regions with different price generating processes: southern-central Mozambique and northern Mozambique.[1]

1

In Mozambique's political administrative division, northern Mozambique includes the provinces of Niassa, Cabo Delgado and Nampula, central Mozambique is composed by Zambezia, Tete, Manica

This research divides Mozambique into these two geographic units, and the cities of Nampula and Maputo are taken as the representative markets for the northern and the central/southern regions respectively due to their importance in the respective regions. Having this in mind, the research estimates and tests short-term forecasting models for these two cities, using a time series approach. Time-series models are widely used for forecasting purposes, do not require economic theory, have economical data requirements, have been found to give good forecasts, and are easy to implement and interpret. This research assumes that price forecasts in Mozambique are of interest primarily to producers and traders. Producers in Mozambique face high levels of this uncertainty about prices and earnings, thus accurate and well diffused forecasts might reduce uncertainty. For traders, forecasts are important in that if accurate forecasts are available and timely disseminated, it can improve plans of volumes to trade. The government is also interested in forecasts because it needs to plan public investments and policies.

---

and Sofala provinces, and southern Mozambique includes the provinces of Inhambane, Gaza and Maputo. Agro-climatic characteristics and the pattern of agricultural trades, however, lead analysts at SIMA in the Ministry of Agriculture and Rural Development (MADER, from its acronym in Portuguese) to include Zambezia in the north. In fact, there is a high commercial linkage between northern Mozambique, especially Nampula province, and Zambezia. And maize prices seem to be driven by the same data generating mechanism in northern Mozambique and Zambezia. For instance, when northern Mozambique exported considerable quantities of maize to Malawi in 1997, prices in Zambezia were affected, but there was no statistical significance of price changes in central and southern Mozambique (Santos and Tschirley, 1999). In addition, about 86% of the maize marketed in informal markets of Nampula city between 1992 and 2000 has originated from northern Zambezia. In this study, therefore, Zambezia is considered as part of northern Mozambique.

1.4 Specific Research Objectives and Thesis Organization

In developing forecasting models for maize retail prices in the cities of Nampula and Maputo, this research is directed toward two specific objectives. First, to develop alternative quantitative, short-term forecasting models for retail prices of maize in northern and southern Mozambique. Second, to evaluate the forecasting ability of alternative models using statistical and economic criteria.

The thesis is organized as follows. Chapter 2 discusses the characteristics and organization of the maize subsector in Mozambique, with special emphasis on maize marketing channels, subsector participant characteristics, and problems of the subsector. Chapter 3 reviews the theoretical framework behind quantitative and qualitative forecasting techniques, and outlines the methods followed in the research. Chapter 4 presents a discussion of the preliminary data analysis. Based on the methods discussed in chapter 2 and findings from chapter 4, chapter 5 uses time-series econometric tools to identify and estimate the comparative univariate and multivariate forecasting models. Chapter 6 evaluates the alternative forecasting models, and chapter 7 presents the conclusions and draws recommendations for further work.

Chapter 2

STRUCTURE OF THE MAIZE SUBSECTOR IN MOZAMBIQUE

2.1 Introduction

This research develops price modeling for maize subsector in Mozambique. A solid understanding of the price data generating process is an important premise for a well-formulated price analysis and a better understanding of the model results. A way to understand the underlying price formulation mechanism in a given subsector is to examine the structure and organization of the subsector through the analysis of the characteristics of demand and supply. Particularly, by analyzing the structure and organization of the maize subsector in Mozambique, this research seeks to examine how the subsector is organized in terms of the participating agents and their functions.

The chapter is organized as follows. Section 2.2 reviews the methodological approach followed in this chapter, section 2.3 analyzes the basic conditions in Mozambique's maize subsector, section 2.4 presents a discussion on the maize supply chain and the characteristics and functions of the participants. Section 2.5 reviews the most important coordination problems. Section 2.6 discusses the indicators of performance in Mozambique's maize subsector, and section 2.7 has the major conclusions of the chapter.

2.2 Methodological Approach

In general, the organization and structure of a given subsector or food system can be addressed in two ways: the subsector approach or the industrial organization approach. Although these two approaches are similar in focusing on the performance of the subsector, the former seems to be appropriate for the specific objectives of this research as (i) it puts emphasis on aspects related to the transformation, value adding, and transactions that take place at every stage of production, from the stage of input supply to the stage of consumption of the final output; (ii) it focuses on the vertical coordination of the firms that add value to a product or related products; (iii) it analyzes how, through vertical coordination, all the participants in the subsector have incentives to participate in institutional arrangements that reduce fluctuations of the commodity supply and consequently reduce excessive fluctuations of prices and; (iv) it is focused on the channel coordination and specialization of all participants in the subsector.

As Shaffer et al (1983) and Holtzman (1986) argued, by focusing on such aspects, the subsector approach leads to a better comprehension of the characteristics and problems in the coordination of a food system. Through such an effort, the underlying price formulation process can be understood, and the food system can move from levels in which firms have low productivity, operate at a small scale, have poor levels of innovative capacity, and operate at high costs with little specialization into systems in which firms have higher levels of specialization and can explore economies of scale, minimize unit costs, reduce excessive fluctuations of commodity supply and prices, and improve

producers and consumers' welfare. This chapter uses the subsector approach to examine the structure of the maize subsector in Mozambique.

2.3 Basic Conditions in Mozambique's Maize Subsector

The most important basic conditions in the maize subsector in Mozambique are related to agro-ecological conditions, technological development, and the characteristics of demand. Agro-ecological conditions are determinant in the geographic pattern of maize production in Mozambique. Maize supply depends on rainfall. There is a single rain season and a single cropping season per year in most parts of the country. Furthermore, as Tschirley (1998) stated, areas north of the Zambezi river have a better potential of maize production than south of it, determined by better reliability of rainfall and much better fertility of soils.

Agricultural technology is rudimentary in Mozambique. As a result, most maize is cultivated with the use of rudimentary tools and few external inputs. A national survey by the Ministry of Agriculture and Rural Development of Mozambique estimated that, in 1996, hoes were the most important tool in farm production for 99.5 percent of the households. Only 3.2 percent used tractors and 1.2 percent used animal traction systems. In addition, 55 percent of the seed used was retained from previous production, and only 16.7 percent was acquired through the commercial channel. Only 0.1 percent of the households used fertilizer (Strasberg, 1997). Furthermore, although the National Institute for Agronomic Research (INIA) has some experimental field stations for new and improved varieties of maize seed in the country, most households use traditional varieties of seed. This situation is likely to have changed little up to the present date.

Finally, due to production lags and the use of rudimentary tools, maize supply is inelastic in the short-run. Due to both factors, households can barely adjust their levels of production to changes in demand. If demand decreases, households can hardly adjust in the short-run because of the asset fixity that characterizes the use of land and other production factors. If demand increases, households cannot, either, easily adjust the levels of production in the short run. Also, because maize is a basic component of the consumption patterns in the country, it is demand inelastic.

2.4 Organization of the Subsector and Participants' Characteristics and Functions

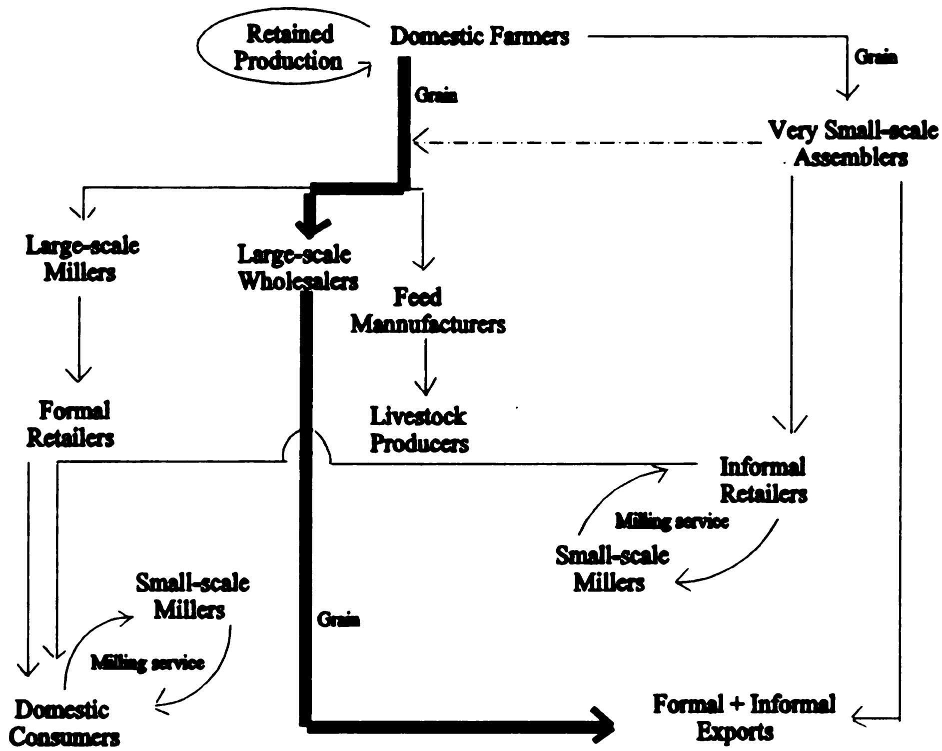
A comprehensive subsector analysis involves a review of the supply chain from the input market stage to the stage of consumption of the final product. By reexamining the basic characteristics of the supply chain and the organization, characteristics and functions of the participants, such a review helps understand the mechanisms that influence prices. This section reviews the characteristics of the different stages of price formation in Mozambique's maize subsector. The analysis is conducted first for northern Mozambique and then for central/southern Mozambique. In both cases, the stage of input supply is not analyzed. Figures 2.1 and 2.2 present the structure and organization of the maize marketing channels in the two regions and describe the linkages between the agents participating in each marketing channel. From top to bottom, these Figures show the different levels of transformation and transaction in the maize subsector and the coördination processes within the channels. The remained of this section reviews the marketing channels in the two regions.

2.4.1 Marketing Channels in Northern Mozambique

Nearly all maize consumed in northern Mozambique is produced within the region. As a food staple, most families produce maize primarily for on-farm consumption, farmers retain the majority of their production for on-farm consumption, and only the surplus is marketed through the marketing channels.

As Figure 2.1 indicates, the most important marketing channel in northern Mozambique, given by the thick arrows, involves domestic farm producers, large-scale wholesalers and exports. In addition to this main channel, there are other channels by which different participants deliver product either to domestic consumers or for export. These channels involve small scale assemblers, small-scale and large-scale millers, feed manufacturers, meal wholesalers, retailers, and final consumers. The level of specialization of the different groups of participants in the subsector varies across the different levels of production-transaction both in the vertical and in the horizontal coordination. This performance depends on the institutional arrangements at each stage.

Figure 2.1 Maize Marketing Channels in Northern Mozambique



At the different stages of production-transaction in these channels, the transactions involve different types and levels of product processing. The different types of output obtained from the different stages of transformation are: maize grain, refined maize meal milled in the country, hand-processed refined maize meal, and whole maize meal. The remained of this sub-section describes the linkages and functions of the participants in the maize marketing channels in northern Mozambique.

2.4.1.1 Farm producers

In Mozambique, nearly all maize is produced by smallholder households. The nationally representative rural household survey carried out by the Ministry of Agriculture and Rural Development of Mozambique in 1996 indicated that 80 percent of rural households in the country cultivated maize grain, but around 90 percent of the produced grain was consumed on the farm, and only the remaining 10 percent was sold. This semi-subsistence nature of the country's agriculture is emphasized by the fact that average households cultivate only 1.85 hectares, and average maize yield in the country is 717 kg per hectare (Strasberg, 1997). In northern Mozambique, farm producers mainly sell their surplus to large-scale wholesalers, but also sell to small-scale assemblers and large-scale millers.

2.4.1.2 Assemblers

In northern Mozambique, most of the grain sold by farm producers is bought by large-scale wholesalers working through small-scale assemblers or directly from farm producers. Compared to central/southern Mozambique, small-scale assemblers in northern Mozambique have little importance in trading maize. An important characteristic of small-scale assemblers in northern Mozambique is that they usually buy small quantities of maize, and trade other agricultural commodities in addition to maize. There is an emergent entry of large-scale wholesalers to the maize grain trading in northern Mozambique, especially for export or to sell to donors and/or large-scale millers within the country. This is contributing strongly to a recent and dynamic organization and consolidation of the assembling function in the north.

2.4.1.3 Large-Scale Grain Wholesalers

Grain wholesalers in Mozambique can be divided into two groups: the large-scale and the small-scale grain wholesalers. In northern Mozambique, basically there are only large-scale wholesalers. Wholesaler traders have considerable experience and usually trade other products in addition to agricultural commodities, taking advantage of profit opportunities in the market. Also, they usually have access to credit and adequate access to working capital. This sector of wholesale traders entered firmly into maize marketing in northern Mozambique when opportunities to export grain to neighboring countries, especially into Malawi, appeared (Tschirley, 1998).

2.4.1.4 Grain Processors/Manufacturers

In Mozambique, maize processors can be divided into two groups. One is composed of industrial millers, the most important of which are Companhia Industrial da Matola (CIM), in the Maputo city area, and MOBEIRA in Beira, and the second is composed by small-scale custom millers, who can be found throughout the country.

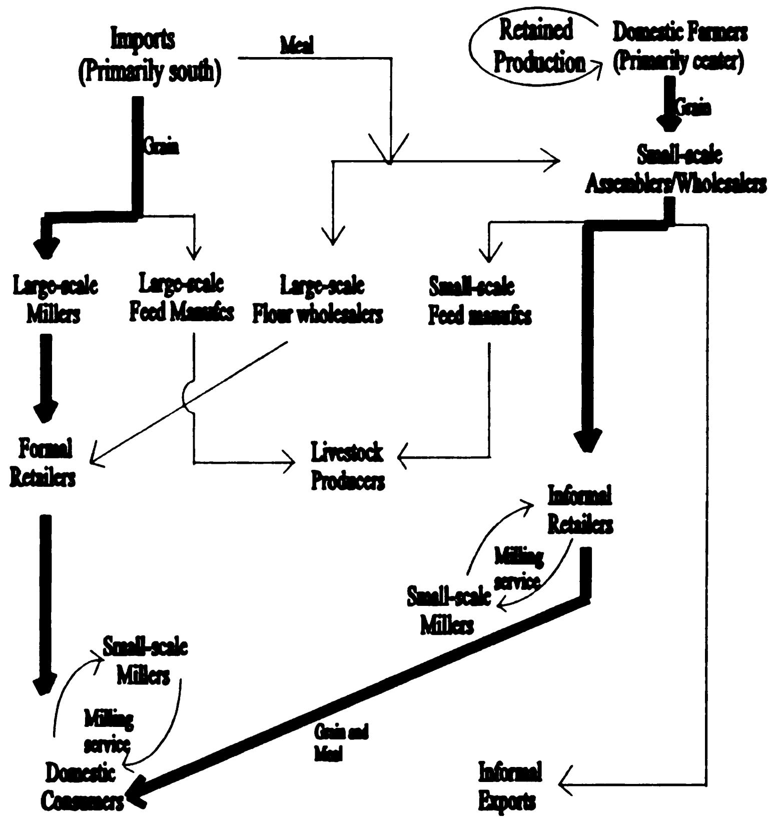
The structure of milling industry is changing in northern Mozambique. A new large-scale miller was installed in July 2000 in Nampula city, with a processing capacity of 70 metric tons per day, around 25 percent of the processing capacity of the large-scale milling industry in the country (SIMA-Mozambique, 2000). CIMPAN, the name of this new processing unit, expects to buy all its maize grain within northern Mozambique. Small-scale custom millers, on the other hand, rather than buying maize grain and selling maize meal, process grain belonging to small-scale wholesalers, informal retailers and consumers, and receive a payment for the milling service.

2.4.2 Marketing Channels in Central/Southern Mozambique

Unlike in the north, maize supply in central/southern Mozambique does not depend solely on domestic production. As Figure 2.2 indicates, there are two important maize marketing channels in central/southern Mozambique. First, most maize marketed from central Mozambique is channeled by small-scale assemblers/wholesalers and informal retailers to domestic consumers, especially to the cities of Maputo and Beira. A small part of it is exported by informal traders to neighboring countries. Second, the south, in addition to receiving delivers from central Mozambique, imports maize grain and maize meal from South Africa, a connection made especially by large-scale millers and formal retailers.

Such as in the north, participants involved in maize transformation and transaction at the different stages of production-transaction in the marketing channels in central/southern Mozambique have different levels of specialization. The characteristics and functions of these participants are discussed next.

Figure 2.2 Maize Marketing Channels in Central/Southern Mozambique



2.4.2.1 Farm Producers

Maize production is lower in southern Mozambique than in the rest of the country. In 1996, the three provinces of southern Mozambique had an average maize yield of 400 kg/ha, ranging between 484 kg/ha in Gaza province and 296 kg/ha in Inhambane province. In the same year, the average maize yield in central Mozambique was 758 kg/ha, and it was 862 kg/ha in northern Mozambique. The poor performance in farm production in southern Mozambique is likely to have worsened in considerable areas of this region and some areas of central Mozambique after the February/March 2000 floods, which were estimated to be the worst of the last 50 years.

2.4.2.2 Small-Scale Assemblers/Wholesalers

In opposition to northern Mozambique, there is a considerable consolidation of a group of small-scale assemblers and wholesalers in central/southern Mozambique, with a visible level of specialization. The major problem with this function is that, given that most maize grain in Mozambique is produced by smallholder households, assemblers usually acquire small quantities from many small farmers, thereby they need much time to accumulate quantities up to a level that allows profitable transactions. Also, nearly all the small-scale grain assemblers, generally from the major cities, perform this function, acting simultaneously as wholesalers. The emergent entry of large-scale wholesalers to the maize grain trading that is observed in the north has not been observed in central/southern Mozambique up to date. The absence of this new dynamic in this regions implies that maize assemblers will continue operating more actively only in the strategic areas of the most important producer districts in the country, like the area of Chimoio, Manica, and Sussundenga districts in Manica province, central Mozambique. From this area, assemblers supply the product to wholesalers from southern Mozambique and from the central city of Beira.

In central/southern Mozambique, small-scale grain wholesalers either started or fortified their activity in 1992, with the end of the civil war in Mozambique, or in 1994 after the first multiparty elections and the establishment of a trustworthy environment of peace and stability in the country. This group of wholesalers is composed of informal traders and is responsible for the linkages between most surplus and deficit areas in central/southern Mozambique, especially between the major rural producer areas in the center and the most important urban markets in the south.

2.4.2.3 Grain Wholesalers

There is a strong, organized and dynamic group of small-scale wholesalers in central/southern Mozambique. Also, contrary to northern Mozambique, there is not a large-scale sector trading maize grain from the producer areas into the consumer centers in central/southern Mozambique. CIM, the only large-scale milling industry in the south, has experimented purchasing some maize grain form large-scale wholesalers from northern Mozambique, but continues to rely almost entirely on imports.

2.4.2.4 Grain Processors/Manufacturers

The most important and traditional large-scale maize processors in central/southern Mozambique are MOBEIRA, which produces 40 percent of Mozambique's refined maize meal, and CIM, which processes about 35 percent. Maize meal processed by these two industrial units is mostly oriented to domestic markets, especially urban markets. In addition to these large-scale processing units, small-scale custom millers can be found throughout central and southern Mozambique. While CIM buys most of its maize grain in neighboring countries, especially in South Africa, MOBEIRA acquires most of its maize grain within central Mozambique.

Such as in the north, small-scale custom millers, on the other hand, usually do not buy grain. Instead, they process grain belonging to their customers, either small-scale wholesalers, informal retailers or consumers, and receive a payment for the milling service that they provide.

2.4.2.6 Retailers

In the maize marketing channel, both in northern and in central/southern Mozambique, the maize retail sales function is performed both by formal and informal traders. Formal traders are those who have a government licence to trade, and the group of informal retailers includes all those retailers who sell maize grain or maize meal with no governmental licence. In the past, during the colonial period in Mozambique, only formal retailers could participate in agricultural commodity trading. This policy was maintained after the country's independence in 1975, and it only was replaced in 1987 when the government started to liberalize agricultural markets. At present, food crops in Mozambique are traded mostly by informal retailers. Nevertheless, financial problems faced by these traders lead to high operational costs.

2.4.2.7 Maize Consumption

In Mozambique, there is a distinction between urban and rural consumers. On the one hand, urban consumers usually acquire most of the maize grain or maize meal in the markets. Urban consumers will buy maize grain and manually process it, instead of buying flour, if they want a high quality flour but face financial problems to directly buy refined flour. Alternatively, urban consumers might buy grain and take it to small-scale custom millers and have non-refined flour, instead of buying flour from retailers, if the price margin between grain and flour is much higher than the price paid for the milling service. The margin of prices between grain and maize meal can be higher than the prices paid for milling if there is no competition in maize flour marketing, and prices of flour are artificially high.

In general, the percentage of household expenditures used to buy agricultural products, maize included, are still low in Mozambique's urban areas. For instance, Donovan (1996) noted that only about 15 percent of the average monthly household expenditures in the provincial capitals were in maize consumption in Mozambique in 1993. In the area of Maputo city, the 20 percent poorest households spent 18 percent of their household income in maize consumption (Donovan, 1996). On the other hand, although rural households purchase grain especially during the planting season and in drought years, most maize consumed by rural consumers is mostly own-produced by the households.

2.5 Coordination Problems

The analysis of the maize subsector in Mozambique suggests the existence of coordination problems in the subsector. Coordination is one of the key concepts in subsector analysis. If the maize subsector faces few coordination problems, more clear functions of the participants and high levels of specialization should be expected.

As Boughton et al (1995) pointed out, two consecutive production levels in the production-distribution-consumption sequence are linked with a transaction, which must be performed by a specialized group of agents. If this does not happen or it happens with deficiency, the marketing channel is facing coordination problems and there are not enough incentives for individuals to perform transaction activities, which results in negative consequences to the subsector.

Mozambique's maize subsector faces several coordination problems, the most important of which are: the high cost of marketing from north to south, instability of returns to storage, uncertainty about market opportunities, difficult access to working capital, low yields, and inefficiency in the small-scale custom milling industry.

2.5.1 High Cost of Marketing from North to South

High costs to trade from north to south directly affect primarily large and small-scale wholesalers and consequently producers and consumers. Large-scale wholesalers are affected because the high costs they face transporting maize grain from northern or central

Mozambique to the south result in high price to consumers in the south. This situation is currently overcome with the introduction of substitute products in Mozambique's southern markets, especially from South Africa.

In fact, in southern Mozambique maize grain brought from South Africa is usually cheaper than that brought from northern Mozambique. In these conditions, Mozambican producers and traders from the north do not have a market in the south. However, consumers in southern Mozambique are likely to face the higher prices of the maize grain from northern Mozambique when the southern Africa region, particularly South Africa, faces a shortage in maize grain production. Moreover, if the whole southern Africa region or most of it faces a drought, Malawi and probably Zambia may be the preferred destination for traders from northern Mozambique due to the comparatively lower transportation costs. As a consequence, the reduced quantities that can be transported to the south in drought years are likely to be more expensive than they would be in a normal season due to the scarcity of the product.

High transportation costs from northern to southern Mozambique result from the geographical characteristics of the country and the poor development of the transportation system. First, although the country is bordered by the Indian Ocean from the north to the south with good natural conditions for efficient and active ports in most of the coast, a poor development of the maritime transportation sector and the resulting high transportation costs, added to the low operating scale and small operational capital for nearly all the participants in the sector lead to the exclusive use of roads for most traders.

Second, the railway system in the country has been conceived to serve inland neighboring countries. While the country is long from north to south, the most important and systematically maintained highways and railways link the country's three most important ports to inland neighboring countries. In fact, Mozambique's most important highways and railways have been conceived in the colonial period, in the context of Mozambique as an "economy of transportation services" to the neighboring countries. And this policy seems to be currently continued. Currently, the most important focus in road and railway development is on the three most important lines linking Mozambique to the inland neighboring countries. All three have been announced to be transformed into "development corridors" in the last five years.

Finally, road linkages between north and south of Mozambique are difficult due to road access problems. Both in a within province or in a inter-province context, nearly all the maize trade from the producer to the deficit areas is done via roadways, but the poor maintenance of the rural roads and the only road linking north to south leads to considerable transportation problems. According to Tschirley and Santos (1998), in 1997 the margin of transport and handling cost as percent of the selling price was between 31% and 34% from producer areas within Nampula province into Nampula city, and 38% from northern Zambezia to Nampula city. In turn, transporting product from the area of

Chimoio in Manica province to Maputo represented a margin of 37% of the retail price in Maputo. If maize trades occur between northern and southern Mozambique, these margin can probably double due to the difficult road linkages between these two regions.

2.5.2 Instability of Returns and High Costs of Storage

Participants in the maize marketing channel in Mozambique face the problem of unstable returns to storage. This directly affects primarily large traders and farmers. Because returns to storage are uncertain, there is difficulty in the organization and consolidation of a storage sector in the maize marketing channel. Three immediate consequences result from this. First, large-scale wholesalers, who buy large quantities of maize grain usually for exporting purposes, have their own storage infrastructures and assume individually the risk of storing the product for an undetermined period if the market is uncertain. Second, producers suffer the consequences of this situation in that the lack of export or domestic market implies uncertainty about sales in the subsequent seasons. Likewise, most producers store their product in household-owned rudimentary storage infrastructures, hence they individually assume the risk associated with the volume stored, and experience lack of market. Finally, small-scale assembler/wholesalers face extremely high storage costs, and therefore do not engage in this function. Grain prices in retail markets therefore show pronounced seasonal price swings.[2]

2

There is no available data series on storage cost for maize in Mozambique. However, the available data on storage cost indicate that informal wholesalers both in Nampula and Maputo face extremely high storage costs. For instance, in December 2000 the monthly storage cost of 428 Metical per kilogram both in Nampula and Maputo represented a margin of 27% of the retail price for Nampula

2.5.3 Uncertainty About Market Opportunities

Uncertainty about market opportunities directly affects primarily farmers and large-scale grain wholesalers. This problem is directly linked to that of instability of returns to storage, and has two main consequences. First, there are no incentives to buy considerable quantities from farmers unless a known market exists. Second, there exists a disincentive to producers. The supply channel does not work properly. Most large-scale traders are not willing to specialize in trading this product, and producers are uncertain about the market, which leads to disincentives in production.

2.5.4 Limited Access to Working Capital

Participants in the maize subsector in Mozambique face the problem of limited access to working capital. This affects primarily small-scale wholesalers and retailers linking surplus to deficit areas in the country. Limited access to working capital leads to high unit costs, in that if traders do not have enough operating capital they cannot buy considerable quantities of maize grain, and as a result, unit costs are very high.

2.5.5 Low Farm Yields

Farm yields in Mozambique's maize subsector are very low even when compared to southern African average levels. This primarily affects both farmers and consumers. For farmers, low yields lead to low returns to inputs, time and capital, and consequently low

---

and 17% for Maputo traders. In the same period, the purchasing price represented 71% of the selling price for Nampula traders and 44% for Maputo traders.

household income. To consumers, low yields at the farm level imply high prices. This happens for two reasons. First, the lower the quantity produced, the higher the price they will face given inelastic demand. Second, low yields at the farm level imply higher unit costs for traders.[3]

2.5.6 Inefficiency in the Small-Scale Milling Industry

Inefficiency in the small-scale milling industry has an impact on the price differential between maize grain and maize flour especially in rural areas. Areas with lower density of small-scale custom millers experience higher difference between price of maize grain and price of maize flour. More discussion on this problem is addressed in section 2.6.2.

2.6 Indicators of Performance in the Maize Subsector in Mozambique

Agricultural economists argue that the performance of a given industry or subsector is measured in several ways. According to Jesse (1978), many industrial organization economists state that measures of performance require the use of standards to which observed values should be compared. However, as Jesse underlines, it has often been difficult to find consensus regarding such standards.

3

High unit costs faced by traders (and consequently by consumers) when farm yields are low can be explained at least in two ways. First, traders need more time to accumulate minimum quantities of product necessary to ensure profitability. Second, when yields at farm level are low, full capacity of transport may not be used, leading to high unit transport costs, and consequently, higher price to consumers, affecting the subsector as a whole.

In evaluating performance in the maize subsector in Mozambique, the present section examines the index of availability of maize grain and maize flour in markets, marketing margins between prices of maize grain and maize meal, coordination of maize prices between related producer and consumer markets, and price stability over time. These measures of performance are not compared to any standard levels.

2.6.1 Index of Availability of Maize Grain and Maize Flour

Table 2.1 shows that product availability follows the same pattern throughout the country. First, it shows that a high index of availability of a given product in the cities of Maputo and Nampula corresponds to a high index of availability in the corresponding regions.[4]

Second, (i) in all regions, domestic maize grain, hand processed maize flour, and whole maize flour processed in small-scale custom mills are the products with the highest availability index; (ii) industrially processed (refined) domestic maize flour and industrially processed imported maize flour have a relatively higher availability index in southern Mozambique, but a lower availability index in northern Mozambique; and (iii) imported maize grain has a very small availability index in the whole country. The greater availability of imported meal in the south is clear in the numbers.

4

In this research, the index of availability of a commodity in markets is defined as the percentage of weeks in which the product was available in the market. Maize prices in Mozambique are collected weekly. Table 1 shows, for instance, that maize grain was available in Maputo 99.8 percent of the weeks between November 1992 and May 1999. Hence, the index of availability of maize grain in Maputo was 99.8 percent in the period.

The index of availability of products is important in the sense that the higher the index of availability of a given product, the better the access of consumers to the product, which indicates that there are good levels of production. Also, when the index is similar throughout the country, the whole system is functioning well. The product availability index can be improved by reducing problems in transport and by ensuring better infrastructure investment throughout the system. A recent major investment in a large-scale maize milling unit in Nampula should improve the availability of refined meal in the north while simultaneously providing a more stable source demand for northern farmers.

Table 2.1 Index of Availability of Product in Consumer Markets in Mozambique, 1992 - 1999

City/Region	Percentage of Weeks Products Were Available					
	Grain (domestic)	Grain (imported)	Refined meal (domestic)	Refined meal (imported)	Hand Processed meal	Whole meal
Northern Mozambique						
Nampula	99.4	5.6	1.2	1.2	100.0	99.8
All northern markets	90.3	0.9	0.9	4.2	62.7	99.0
Southern Mozambique						
Maputo	99.8	8.1	56.5	89.6	85.6	87.6
All southern markets	98.0	3.4	34.0	38.8	53.0	82.0

2.6.2 Marketing Margin Between Price of Maize Grain and Maize Flour

Table 2.2 shows that the marketing margins between real prices of maize grain and maize flour at the retail markets in Mozambique followed different patterns over the period of study. While in the north they increased slightly between 1993 and 1996 and decreased in 1999, in the south they tended to decrease from 1993 to 1996 and become stationary from 1996 onward. The path of the prices of maize grain and maize flour in the two regions was not the same. In the south (Maputo), while real prices of maize grain decreased between 1993 and 1999, maize meal prices did not decrease. In the north (Nampula), on the other hand, both prices (maize grain and maize meal) tended to maintain over time.

Table 2.2. Average Annual Real Prices of Maize Grain and Whole Maize Flour in Selected Retail Markets in Mozambique, 1993, 1996 and 1999 (May 1999 = 100)

Local	Year	Price of Maize Grain	Price of Maize Flour [1]	Marketing Margin [2]	Percentage Margin
Maputo	1993	4028	4515	487	12.09
	1996	3237	4440	1203	37.16
	1999	3185	4531	1346	42.26
Nampula	1993	2251	4731	2480	110.17
	1996	1756	4204	2448	139.41
	1999	2834	4892	2058	72.62

[1] The data refers to whole maize meal processed in small-scale mills. Table 1 showed that this is the quality of maize flour more widely available in Mozambique in both time and space dimensions.

[2] Refers to the marketing margin defined as the absolute difference between the price of whole maize meal and the price of maize grain.

Table 2.2 also shows that the marketing margin is higher in Nampula than in Maputo. This evolution of the maize marketing margins suggests the existence of problems in the small-scale milling industry in the country. A very high level of marketing margin may suggest a lack of competitiveness in the small-scale industry. This problem may be worse in the center/north than in the south, and in the rural areas compared to the urban areas. Finally, the high marketing margins in the south may suggest that prices paid for milling maize grain is increasing in this region, which may indicate the need for new investments in the small-scale milling industry.

2.6.3 Price coordination and stability

The evolution of retail maize prices in Nampula and Mocuba, and in Maputo and Manica is graphically shown in Figures 4.1 and 4.2 respectively. These graphs indicate that real retail prices in the two pairs of markets are highly correlated over time. Over the period, prices in the producer districts of Manica and Mocuba are lower than those observed in the consumer markets of Maputo and Nampula, a pattern that did not substantially change over time, and reflects that Maputo and Nampula are supplied with maize grain from Manica and Mocuba. Also, Figures 4.1. and 4.2. suggest that prices follow a consistent pattern of seasonal variations over time. For instance, nominal prices in Maputo increased about 70% from March to November 1998, and decreased about 50% from November 1998 to March 1999. Similar levels of variations were observed in Manica.

However, although seasonal variations of prices should be expected and are a sign of the abundance or scarcity of the product in the different areas of the country, it should not be very high. High price instability discourages producers for new investments in the sector as a result of price uncertainty, and it does not benefit consumers not only when prices are high but also because of budget planning problems.

2.7 Conclusions

The main conclusion of this chapter is that Mozambique's maize marketing channel faces coordination problems, which are reflected in the performance of the sector. However, regardless of these coordination problems, most retail markets have a high level of product availability and there is some level of product diversity, even though this varies across regions. Moreover, marketing maize in Mozambique from the producer to the consumer areas depends upon the season, geographic location, the availability and cost of transport and storing infrastructures, and the size of the harvest, both within the country and in the neighboring countries.

Better decisions by farmers and other participants about the maize subsector in Mozambique on when and where to sell maize, when and how to store, and whom to sell to and at what price, depend on better investments in agricultural sector in particular and in the rural economy in general. Roads and other infrastructures are particularly important to overcome the basic problems currently faced by the subsector.

Chapter 3

THEORETICAL FRAMEWORK AND METHODOLOGY

3.1 Introduction

A forecast is defined as a qualitative or quantitative estimate about the likelihood of future events based on current and past information (Pindyck and Rubenfeld 1991, Aldridge 1999). The importance of price forecasting lies in the fact that prices play an important role in guiding both production and consumption. The future is always uncertain, and uncertainty in future outcomes that results from present decisions can be reduced if accurate forecasts are available. The more accurate the predictions, the greater the ability that decision makers have in making appropriate and timely decisions (Holden 1990).

This chapter is organized as follows. Section 3.2 reviews forecasting techniques, with an emphasis on the quantitative techniques. Section 3.3 addresses the theory behind univariate time-series models, and section 3.4 discusses multivariate models. Section 3.5 presents a discussion on the methodology of time series analysis. Section 3.6 addresses the issue of statistical and economic evaluation of forecasting accuracy, and section 3.7 presents the univariate and multivariate models to be estimated.

3.2 Forecasting Techniques

A conservative definition would conceptualize forecasting as a set of tools which allow *ex ante* predictions of values of certain variables, outside the available sample of data, typically for future dates. However, in addition to *ex ante* predictions, there is also *ex post* forecasting. In *ex post* or in-sample forecasting, observations on both the endogenous and exogenous variables are known with certainty during the forecast period. *Ex post* forecasting is useful for the purpose of selecting, among different forecasting models, the best fit of historical data, by comparing the observed with the predicted values over the data period. On the other hand, *ex ante* or out-of-sample forecasting is a prediction of values beyond the period covered by the observed data, and it is the most important evaluation of the forecast accuracy of a model.

Forecasting can be qualitative, also known as subjective or implicit, or quantitative, also called model-based or explicit. Qualitative forecasting is not based on quantitative methods. Instead, predictions are made by different expert individuals and/or institutions based on their wisdom and experience regarding the structure of the industry, seasonal production, consumption patterns, and international events that are likely to affect the commodity. On the other hand, quantitative or model-based forecasts are based on quantitative data analysis methods.

Qualitative models are often better for long-term forecasts than quantitative technique are, because in the long term the structure of the economy tends to change, the data generating mechanism might also be different, and historical data might not suffice or may induce errors when performing a long-term forecasting (Aldridge, 1999). On the other hand, although both the qualitative and the quantitative methods are used for forecasting, quantitative forecasting is preferred when consistent data is available, and is commonly used for short-term predictions. The results from a quantitative model, however, may be subject to qualitative evaluations. Both forecasting methods incorporate information from similar data sources.

In modeling maize prices in Mozambique, this research uses quantitative methods and focuses on short term forecasting. There are two quantitative forecasting methods: causal or structural and non-causal. The casual method measures the structural relationships between variables using econometric techniques, and the non-casual technique is based on the evolution of the variable or variables to be analyzed with no concern about causal relationships.

Causal or structural models are based on the study of the structural relationship between variables, and appropriate economic theory has to be employed in the model building process. The aim is to explain the behavior of the endogenous variable based on the level and statistical significance of the parameters of the regression model. A linear single-equation structural model with *k* explanatory variables explaining the variations in the endogenous variable takes the form:

$$p_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \ldots + \beta_k x_{kt} + \varepsilon_t \quad (1)$$

If the results of the regression model are in accordance with economic theory, and are statistically significant, then the estimated parameters are used to predict the future behavior of the endogenous variable, based on a set of assumptions about the future behavior of the explanatory variables, and assuming that the relationship between the variables will continue.

As in all econometric models, building structural models for forecasting purposes implies a careful selection of the explanatory variables according to economic theory, requires using economic theory to formulate hypotheses on the signs of the parameters of the different variables, finding data on the variables in the model, estimating the parameters in accordance with appropriate econometric techniques, examining the model residuals especially for serial correlations and heteroscedasticity, examining the validity of the parameters based on the hypothesis previously formulated, and testing their statistical significance for purposes of extrapolation into the future.

One important advantage of structural models is that, because they are based on economic theory and identify the relationship between variables interacting in the price generating process, they may allow decision makers to evaluate the impact of different alternative policies. However, often there is lack of strong knowledge of the underlying economic structure or the required data for a structural model is not available. Furthermore, structural models are often complex, and require more data and time to build than non-causal models, therefore they are more costly (Donovan 1996; Aldridge, 1999). For their relative simplicity both in data requirements and modeling, non-causal models may be preferred.

Non-causal models are grouped into two categories: univariate and multivariate models. Univariate models have only one variable - the variable object of the analysis - and multivariate models have more than one time-series variable in the model.

3.3 Univariate Models

The most common univariate forecasting techniques that are used to predict time series data are the naive, the random walk models, trend extrapolation models, smoothing models, and autoregressive integrated moving average (ARIMA) models, also known as Box-Jenkins models. Each will be addressed in turns.

3.3.1 Naive and Random Walk Models

Naive models are the simplest form of time series econometric models. Building time series econometric models is based on the theory of linear stochastic difference equations, in which the difference equation describes the value of a variable as a function of its own lagged observations, exogenous variables and a disturbance term (Enders 1995, Aldridge 1999). The naive model then is:

$$\hat{p}_t = p_{t-1} \tag{2}$$

where $\hat{p}_t$ is a forecast of p_t conditional on information available at time *t-1*. This would imply a model for p_t of the form:

$$\hat{p}_t = p_{t-1} + e_t \tag{3}$$

where the stochastic term e_t is white noise.

The most important advantage of the random walk model is that it requires only a small amount of data and no parameter estimation. The only information needed is the current value of the variable. Moreover, because random walk models are stochastic, a standard error of forecast can be computed, which allows forecast confidence intervals to be constructed, if more .

3.3.2 Trend Extrapolation Models

Econometric trend models express the variable p_t as a function of time, and are particularly useful when the parameters describing the time series do not change over time, and when the time series data tends to consistently follow a movement in a particular direction, upwards or downwards. In the trend extrapolation methodology we can have linear or nonlinear models. The type of trend observed in past values is projected into the future, and one among several polynomial functions of time is estimated using regression methods or by forming moving averages of the time series (Aldridge 1999).

The most widely employed trend extrapolation models are the linear trend, exponential growth curve, and the quadratic, the autoregressive and the logarithmic curves. The model to be employed depends upon the forecaster's beliefs about the future evolution of the variable which is object of the study. One noteworthy disadvantage of the trend extrapolation models is that they often have large standard errors compared to some other non-causal forecasting models, and thus they only tend to be used as a quick and inexpensive way of formulating initial forecasts.

3.3.3 Exponential Smoothing Models

The exponential smoothing technique is a method that gives more weight to the more recent observations based on the use of a single or multiple smoothing parameters that are determined by smoothing equations. Like all time series forecasting techniques, there is no economic theory or statistical model supporting the exponential smoothing method. On the contrary, it is a simple technique of adaptive forecasting where the forecasts adjust based on past forecast errors of the type:

$$p_{t+1} = p_t + \alpha e_t \tag{4}$$

where p_{t+1} is the forecast price for period $t+1$, based on p_t, forecast price for period t and α, the smoothing or adjustment parameter. e_t is the forecast error in period t.

In practice, using the smoothing technique, past forecast errors are used to correct the next forecasts by adjusting them in a direction opposite to that of the past error (Makridakis 1978, Aldridge 1999).

An advantage of smoothing techniques is that they do not require a large amount of time series data, and are especially effective when the parameters describing the data change slowly over time (Bowerman 1993, Aldridge 1999). However, smoothing models are only suitable for one-step ahead forecasts, being less accurate when longer periods have to be included.

3.3.4 ARIMA Models

ARIMA models have been found to give good forecasts in a wide variety of situations and hence are one of the most popular forms of linear models which describe the data generating mechanism with no resource to structural models. They were popularized by Box and Jenkins, and as result they are also known as Box-Jenkins models. ARIMA models explain the movement of a time series by relating its present values to its own past observed values and/or to a weighed sum of the current and lagged random disturbances (Aldridge, 1999).

ARIMA models are estimated under the assumption that many time series are generated following either an autoregressive (AR) process, a moving average (MA) process, or some combinations of the two (ARMA) processes.

In an AR process current observations are assumed to have been generated by a weighed average of past observations. An autoregressive process lagged up to length p is represented as:

$$p_t = \mu + \Phi_{1p}p_{t-1} + \Phi_2 p_{t-2} + \ldots + \Phi_p p_{t-p} + \varepsilon_t \quad (5)$$

Likewise, in a MA process present observations of the variable are assumed to be generated following a random disturbance pattern. A moving average process of a weighed average of random disturbances going back *q* periods MA(*q*) is simply a linear combination of white noise error terms, and is presented as:

$$p_t = \mu + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \ldots - \theta_q\varepsilon_{t-q} \quad (6)$$

When the two components are present simultaneously in the time series data generating process, as often happens, the ARMA (*p, q*) process is represented as:

$$p_t = \mu + \Phi_1 p_{t-1} + \Phi_2 p_{t-2} + \ldots + \Phi_p p_{t-p} + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \ldots - \theta_q\varepsilon_{t-q} \quad (7)$$

where the auto-regressive component is the difference equation, and the moving average component is the white noise process.

In building AR, MA or ARMA models we assume that the underlying random process that generated the time series is stationary. However, in many time-series the processes are non-stationary, thus differences have to be taken in order to eliminate unit roots and transform them into stationary processes. The differences can be presented as:

$$w_t = \Delta p_t = p_t - p_{t-1} \tag{8}$$

After the differences have been taken, the processes become ARIMA(p,d,q), i.e., autoregressive integrated moving average processes, where I(d) denotes the number of differences that were taken from the original data until the time series became stationary. The following equation represents an ARIMA ($p, 1, q$) model:

$$w_t = \delta + \Phi_1 w_{t-1} + \ldots + \Phi_p w_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q} \tag{9}$$

3.3.4.1 Identifying ARIMA Models

The identification procedures employed in the specification of the data generating mechanism of ARIMA models involve examining the properties of the autocorrelation functions (ACF) and partial autocorrelation coefficients (PAC) of the time series. The autocorrelation functions are a sequence of correlation coefficients, where at each period we have the correlation coefficient of the current value of the series with the series lagged that number of periods. The partial autocorrelations at each point in time measure the additional (or partial) predictive power of the series at that period after taking into account the cumulative predictive power of all the values of the series with smaller lags.

In examining the ACF and the PAC, we have to to decide the number of AR, MA and integration terms to include in the ARIMA model. Plotting the ACF and the PAC is a simple way to examine the kind of data generating process that the time series follows.

The shape of both the ACF and the PAC suggests the for of the model. If the ACF declines geometrically and the PAC is zero after k lags, then we have an AR(k) process. Likewise, if the ACF were zero after low k lags and the PAC declined smoothly at a geometric rate, a MA(k) model would appear appropriate. If neither ACF nor PAC converge to zero, we have to try a lower order ARMA model following the parsimonious principle, i.e., starting from ARMA (1,1). If both the ACF and the PAC cut off immediately, it is a white noise or purely random process, also known as an identically independently distributed (iid) process with mean zero and constant variance. Finally, if the ACF appears to have a seasonal frequency of spikes or cyclical "waves", this could suggest the existence of seasonality in the ARIMA model.

Plotting the ACF and PAC also suggests other characteristics of the time series data. For instance, if the ACF and the PAC are very close to 1 at the first lag, and the ACF tends to be a straight line instead of decaying geometrically, it might suggest that the time series is stationary.

Finally, it should be noted that, in identifying the most suitable ARIMA model, the goal is to have a final model that is a parsimonious representation of the data generating mechanism. Hence, only the strictly necessary AR and MA terms should be included in the model. Furthermore, the innovations of the tentative ARIMA model should be checked for the existence of any serial correlations at every stage of the identification procedure. A common way of testing for serial correlations in the innovations is implementing Ljung-Box Q-Statistics tests.

3.4 Multivariate Models

In addition to the univariate forecasting models, multivariate models are also employed in forecasting Among different multivariate, non-causal forecasting models, the vector autoregression (VAR) and the vector error correction (VEC) models are the most commonly used. VAR models result from an extension of the univariate time series Box-Jenkins methodology.

VAR models are built when the time series involved are not cointegrated. If the two or more nonstationary time series involved in the model are cointegrated, then the VAR should be specified with error correction terms, and estimated as VEC models (see section 3.4.2). If the series are not cointegrated, then the system of nonstationary variables is run in first differences (Aldridge 1999).[5]

3.4.1 VAR Models

The VAR is a multivariate time series technique is focused on the study of the degree of association of two or more interrelated time series variables. VAR models can be viewed as reduced-form equations for structural systems, particularly useful when time series data is available for a multivariate model, but knowledge of the underlying economic structure

---

[5] Cointegration is a concept often referred to in time series analysis involving more than one variable. Engle and Granger (1987) indicated that two or more nonstationary time series might have a stationary linear combination, and the two or more nonstationary time series are said to be cointegrated if this stationary linear combination exists. A detailed discussion of cointegration can be seen in section 3.4.4.

is uncertain and therefore it might be difficult to conceptualize a structural relationship between the variables (Sims 1980; Fackler 1988; Myers, Piggott and Tomek 1990, Donovan 1996).

Although the variables in a VAR model are interrelated, there is no concern about any causality relationship between the variables nor about an economic theory supporting the structural relationship between the variables. VAR models are, therefore, non-structural models which explain the dynamic structure of the data generating mechanism assuming that the variables in the system follow an underlying relationship that does not change through time. The VARs are, therefore, processes for which there is no concern about the distinction between dependent and explanatory variables, and thus economic theory and identifying restrictions are not necessary. In the model building process, it is necessary only to identify and specify the variables that are assumed to interact.

Regardless of the non-structural relationship between the variables included in VAR models, these models have been criticized for interpreting statistical parameters as representing economic relationships, without knowledge of the price formation process (Harris 1979, Donovan 1996). As Faminow and Benson (1990) suggest, it is important to know the underlying market structure when determining which prices - across markets or across products - to select on the basis of what the statistical parameters can reveal. It is important to evaluate whether or not the markets are competitive in price setting process, and on the basis of this knowledge determine the appropriate models for price analysis (Donovan 1996).

VAR models express variables as linear functions of the lagged values of each variable and all other variables in the system. The standard VAR reduced-form equation is described as:

$$p_t = \delta + \varphi_1 p_{t-1} + \varphi_2 p_{t-2} + \ldots + \varphi_p p_{t-p} + \theta x_t + \varepsilon_t, \tag{10}$$

where p_t is a vector of k variables and p-lagged values of p_t, x_t is a vector of variables such as deterministic trends and seasonal dummies, δ are the intercepts, φ and θ are unrestricted matrices of coefficients to be estimated, and ε_t is a vector of individually serially uncorrelated innovations with zero means and constant variances.

The statistical importance of the VAR models is that, being descriptions of the dynamic interrelations between different time series variables in a vector, these descriptions can be extrapolated into the future and used for forecasting purposes (Aldridge 1999).

A two-variable VAR model with one lag, VAR(1), can be represented as follows:

$$p_{1t} = \delta_1 + \varphi_{11} p_{1,t-1} + \varphi_{12} p_{2,t-1} + \varepsilon_{1t} \tag{11}$$

$$p_{2t} = \delta_2 + \varphi_{21} p_{1,t-1} + \varphi_{22} p_{2,t-1} + \varepsilon_{2t} \tag{12}$$

While structural models have been found important for purposes of policy analysis and evaluation of relationships in commodity prices, VAR models have traditionally been used for forecasting systems of interrelated variables. In many cases, forecasts obtained from VAR models are better than those extrapolated from more complex structural models (Sims 1986, Donovan 1996, Aldridge 1999). In addition to avoiding the restriction problem that characterizes identifying structural models, VAR models are simpler and less costly than structural models. Usually, prior to the estimation of VAR models all deterministic trends and seasonal components are removed from the series or trend and seasonal variables are included directly in the VAR model (Donovan 1996).

Identification of VARs has been a source of considerable controversy. Early analysis focused on using assumptions regarding the contemporaneous relationship between variables to identify the system. More recent efforts by Blanchard and Quah (1989), Lastrapes (1992), and others use information on the long-run effects and the infinite moving average representation of the VAR to impose identifying restrictions on the system. When there are nonstationary series and the possibility of cointegrating relationships, the use of long-run identifying restrictions is key.

3.4.2 VEC Models

If in testing for cointegration among the variables involved in a multivariate model the results indicate that the variables are nonstationary and cointegrated, then there exists an error correction representation such that the differences respond to the previous period's deviation from long-run equilibrium (Aldridge 1999). As a result, estimating the nonstationary prices as a VAR in first differences is inappropriate and results in a misspecification error. If the prices are cointegrated, then we should use VEC instead of VAR.

VEC models are restricted VAR models that have cointegration restrictions built into the specification. In particular, the VEC restricts the long-run behavior of the endogenous variables to converge to their co-integrated relationship, while allowing for a wide range of short-run dynamics (EViews 1997, Aldridge 1999).

3.5 Methodology in Time Series Analysis

The basic steps in analyzing time series data involve investigating stationarity, testing for the existence of a trend component in the time series, testing for cointegration when more than one time series data is involved in the model, testing for autocorrelation and heteroscedasticity in the error terms, and investigating seasonality.

3.5.1 Unit Root Tests

Testing for a unit root or nonstationarity is one of the most important steps in time series data analysis. The data may be stationary around a constant mean, or it can be trend stationary. The data is said to be stationary around a mean when there is no statistical evidence that it grows with a trend component and it is stationary, and it is said to be trend stationary when there is statistical evidence of the presence of a trend component and the data is stationary after we account for the trend component. If the trend component is not significant in the series, we test for a unit root assuming the null of a unit root without drift against the alternative of stationarity around a constant mean. On the other hand, if the trend is statistically significant, we include the trend component and we test the null of a unit root with drift against the alternative of trend stationarity.

Price data is said to be stationary if it has a constant mean over time, the variance is time invariant, and the covariance between prices at different lags, say p_t and p_{t-j}, is always the same, depending solely on the lag length j. If the time series is stationary, then the sample mean, variance and autocorrelations can be used to estimate the parameters of the underlying data generating mechanism, and the model can be used to forecast the time path of prices and future values of the variable. The starting point in building ARIMA models is to test for stationarity to determine whether first differences should be taken to eliminate unit roots in the time series or if the model should be built using levels in the data (Aldridge 1999).

A first-order autoregressive model is stationary if the absolute value of the parameter of the lagged observation is smaller than 1, i.e., if $|\rho| < 1$ in:

$$p_t = \delta + \rho\, p_{t-1} + \varepsilon_t, \qquad (13)$$

If the original data has a unit root, it is said to be nonstationary, and first differences have to be taken to induce stationarity.

There are three primary reasons for wanting to know whether the series contains a unit roots. First, because in the presence of unit roots, the usual central limit theorem that underlines the asymptotic standard normal distribution for the *t*-statistic does not apply, the *t*-statistic does not have an approximate standard normal distribution even in large sample sizes, and therefore we cannot make statistical inferences about regression model parameters using the *t* and F statistics in the usual way[6] (Wooldridge, 2000).

Second, many time series may have time trends or seasonality and therefore regressing one series on other might show high R^2 and significant t-statistics even if they follow independent data generating mechanisms. Testing the variables for a unit root might help examine whether different variables are driven by the same data generating process, and avoid spurious regressions.

Finally, in many works involving time series data, people often want to perform sophisticated analysis such as testing whether two or more series are cointegrated or not. Such analysis cannot be performed if we do not know the order of integration of the variables.

6

The most well-known central limit theorem for time series data requires stationarity. The central limit theorem states that the average from a random sample for any population (with finite variance), when standardized, has an asymptotic standard normal distribution, i.e., the distribution of the estimator is collapsing around the parameter as the sample size gets large (Wooldridge, 2000).

Testing for stationarity is carried out using the Dickey-Fuller and/or the Phillips-Perron tests. Even though both are aimed at dealing with the potential existence of serially correlated residuals, they are found to have different features. While the Dickey-Fuller method is focused on allowing for the explicit presence of serial correlation in the models, the Phillips-Perron method adjusts the test statistics but allows the disturbances of the regression models in the procedure to be weakly dependent and heterogeneously distributed (Lai 1999), and uses nonparametric methods of controlling for higher-order correlations in ε_t to make corrections to the *t*-statistic of the coefficient from the AR(1) regression (EVIEWS 3, 1998).

The Phillips-Perron method has been found to be simple and broadly applicable, and therefore widely used. Nonetheless, it has also been demonstrated that this method suffers from severe size distortions when there are negative moving average errors (Phillips and Perron 1987; DeJong et al. 1992). On the other hand, in addition to easy implementation and interpretation, the Dickey-Fuller tests are thought to be more useful for practical purposes (DeJong et al. 1992; Perron and Ng 1996).

This research uses the Dickey-Fuller approach in testing for unit root. Formally, the Dickey-Fuller test for unit root is performed under the null of H_0: $\rho = 1$ in (13), that is, the time series has unit root, against the alternative of H_1: $\rho < 1$, i.e., the time series is stationary. In practice, however, the Dickey-Fuller test for unit root is carried out by estimating an equation in which p_{t-1} is subtracted from both sides of the equation (13), and this results in:

$$p_t - p_{t-1} = \delta + \rho\, p_{t-1} - p_{t-1} + \varepsilon_t, \tag{14}$$

which is equal to

$$p_t - p_{t-1} = \delta + (\rho - 1)p_{t-1} + \varepsilon_t\,, \tag{15}$$

which is equal to

$$\Delta p_t = \delta + \Phi p_{t-1} + \varepsilon_t \tag{16}$$

where $\Phi = \rho - 1$, and thus we test the null that H_o: $\Phi=0$, i.e., the time series contains unit root, against the alternative that H_1: $\Phi<0$, i.e., the time series is stationary[7].

The Dickey-Fuller test is applied solely in AR(1) models. If the innovations from the Dickey-Fuller test are found to be serially correlated, we implement the Augumented Dickey-Fuller (ADF) test to control for higher-order serial correlations in the innovations. The ADF test consists of adding lagged difference terms of the dependent variable to the right-hand side of the AR(1) regression model. The number of lagged first differences depends on the correction of the serial correlations in the innovations, and it does not affect the asymptotic distribution of the *t*-statistic test on the coefficient δ (Eviews 3, 1998).

7

If $\rho > 1$, P_t is said to be an explosive series. An explosive time series makes little economic sense, and therefore is not allowed under H_1.

3.5.2 Testing for Trend Component

Many time series grow over time, i.e., they contain a time trend. When drawing inferences on time series data based on the classical tests, we need to determine whether the data series contains a trend component or not. Often, various time series variables seem to be correlated solely because they all grow over time, following other, unobserved phenomena. By including a time trend variable, we can avoid concluding that two or more variables are related when they in fact are not, and we can avoid trusting in spurious regressions involving time variables that have a positive or negative trend. A widely used approach to address the problem of time trend when the data is stationary is to write the series, say p_t, as:

$$p_t = \beta_0 + \beta_1 t + \mu_t, \tag{17}$$

where μ_t is independently and identically distributed (i.i.d.), with $E(\mu_t) = 0$ and $Var(\mu_t) = \sigma_\mu^2$. In equation (18) β_1 can be interpreted as the change in P_t from one period to the next due to the time factor, *ceteris paribus*, and is statistically tested with the null of H_0: $\beta_1 = 0$, against the two-sided alternative of $H_1 \neq 0$, i.e., there is (either positive or negative) trend. If there is statistical significance of the existence of a trend component in the time series variable, we control for it by including a time trend variable in the regression model. When the data is nonstationary, first differences must be taken, and the evidence of a time trend component is tested through the drift term. If the drift term is positive, the expected value of the time series is positively growing over time. A significant positive drift (i.e., a significant constant in an equation in first differences) means series drift upward over time.

3.5.3 Checking for Seasonality

Seasonality can be defined as any consistent, periodic and natural movements in a time series that are repeated cyclically at the same phase of the period (Goetz and Weber 1986), and is independent of other factors affecting the fluctuations of the data over time. For instance, even though prices of a commodity are constantly affected by several demand and supply conditions, common sense suggests that, if the commodity is planted in October/November and harvested in April/May, we should expect that prices would usually be higher in January than in June or July, *ceteris paribus*. Commodity market data often exhibit seasonality. Usually, the higher the frequency of the data collection the higher the probability that the data contains seasonal effects. For instance, time series data that is observed with high frequencies, such as daily, weekly, monthly or quarterly, is very likely to exhibit seasonality.

Checking for seasonality can be conducted by plotting graphs on the original data and examining the frequency of persistent fluctuations. Another way is to use a correlogram of the autocorrelations and partial autocorrelations of the time series. If the commodity has a single marketing season per year - as is the case of maize in Mozambique - and the ACFs show spikes consistently near every 12 observations in monthly data, that may suggest the existence of a seasonal pattern in the data.

In ARIMA models, seasonality can be dealt with in two ways. The first is adjusting the time series for seasonality, i.e., removing the seasonal factors from it before they are modeled, and the second is modeling the seasonality. Even though some researchers prefer to deseasonalyze the data before it is modeled, Enders (1995), Davidson & Mckinnon (1993), and Aldridge (1999) argue against this approach, and point out that seasonality and ARIMA coefficients are better identified and estimated when included jointly in the model.

ARIMA models can be modeled with seasonality by including seasonal dummy variables or estimating trigonometric-seasonality models. Because of their ability at capturing periodic fluctuations, trigonometric functions have been used to describe seasonality in time series. Seasonal dummy variables have also been found to be an effective tool to capture seasonal patterns.

The seasonal pattern of maize grain production in Mozambique suggests that the price series should be investigated for seasonal patterns. The paper investigates seasonality by including monthly seasonal dummy variables, due to their simplicity in calculation, statistical interpretation and economic meaning. In practice, we test for seasonality after checking for the order of integration. If the series is I(0), and the data suggests the presence of a time trend component and seasonality, a time trend is included in equation (18) and both seasonality and time trend components are checked for at once.

$$p_{1t} = \beta_0 + \beta_1 t + \beta_2 d1 + \beta_3 d2 + \ldots + \beta_{12}\, d11 + u_t \tag{18}$$

where t is the time trend and $d1$ through $d11$ are monthly seasonal dummies.

3.5.4 Cointegration and Vector Error Correction (VEC) Models

In a multivariate model, the time series are said to be cointegrated if each series is stationary after differences at any order have been taken, say order d, I(d), and some linear combination of them is stationary with order d-1, I(d-1). Specifically, if two time series, say p_{1t} and p_{2t} are individually integrated of order 1, I(1), then in general $p_{1t} - \beta\, p_{2t}$ is also an I(1) process regardless of the value of β. However, if the two time series are driven by the same stochastic trend, then there is a $\beta \neq 0$, such that the linear combination $p_{1t} - \beta p_{2t}$ is stationary I(0). If such a β exists, we say that p_{1t} and p_{2t} are cointegrated, and we call β the cointegration parameter (Wooldridge, 2000).

3.5.4.1 VEC Models

If the variables p_{1t} and p_{2t} are both I(1) and are not cointegrated, we can estimate the VAR model:

$$\Delta p_{1t} = \theta_{10} + \theta_{11}\Delta p_{1,t-1} + \ldots + \theta_{1p}\Delta p_{1,t-p} + \gamma_{12}\Delta p_{2,t-1} + \ldots + \gamma_{1p}\Delta p_{2,t-p} + \mu_{1t}, \tag{19}$$

$$\Delta p_{2t} = \theta_{20} + \theta_{21}\Delta p_{1,t-1} + \ldots + \theta_{2p}\Delta p_{1,t-p} + \gamma_{22}\Delta p_{2,t-1} + \ldots + \gamma_{2p}\Delta p_{2,t-p} + \mu_{2t}, \tag{20}$$

and perform the multivariate time series forecasting process using this model.

If, instead, $p_{1,t}$ and $p_{2,t}$ are both $I(1)$ and are cointegrated, where the cointegration equation can be represented as:

$$p_{1t} - \mu - \beta p_{2t} = s_t, \tag{21}$$

then we include the contegration equation (22) in the VAR model, and we form a vector error correction (VEC) model. For instance, allowing the VAR model represented in equations (19) and (20) for a single autoregressive lag, and including the cointegration equation (21), yields the VEC model:

$$\Delta p_{1t} = \theta_{10} + \theta_{11}\Delta p_{1,t-1} + \gamma_{11}\Delta p_{2,t-1} + \delta_{11}s_{t-1} + \mu_t \tag{22}$$

$$\Delta p_{2t} = \theta_{20} + \theta_{21}\Delta p_{1,t-1} + \gamma_{21}\Delta p_{2,t-1} + \delta_{21}s_{t-1} + \mu_t \tag{23}$$

which can be rewritten as:

$$\Delta p_{1,t} = \theta_{10} + \theta_{11}\Delta p_{1,t-1} + \gamma_{12}\Delta p_{2,t-1} + \delta_{11}(p_{1,t-1} - \mu - \beta p_{2,t-1}) + \mu_{1,t} \tag{24}$$

$$\Delta p_{2,t} = \theta_{20} + \theta_{21}\Delta p_{1,t-1} + \gamma_{22}\Delta p_{2,t-1} + \delta_{21}(p_{1,t-1} - \mu - \beta p_{2,t-1}) + \mu_{2,t} \tag{25}$$

In this VEC model, $\delta_1(p_{1,t-1} - \beta p_{2,t-1})$ and $\delta_2(p_{1,t-1} - \beta p_{2,t-1})$ are error correction terms, and the parameters δ_1 and δ_2 measure the speed and the direction of adjustment. This model allows examining the short-run dynamics, while restricting the long-run behavior of the endogenous variables to converge to their cointegration relationships.

Finally, if the two series are nonstationary and not cointegrated, we forecast using either univariate ARIMA models or a multivariate VAR model, again on the first differences. To increase the forecasting robustness, we might use both ARIMA and VAR models, but if we have to choose one of the models we might prefer a VAR model because of its advantages. Regardless of their lesser parsimony compared to ARIMA models, VAR models have been found to do a good job of modeling many related commodity market series, and are easy to estimate in that although they are multivariate, there is no need to determine which variables are endogenous and which are exogenous.

3.5.5 Testing for Autocorrelation in the Error Term

Another important step in every analysis involving time series data is testing for autocorrelation in the error terms of the time series. It is usual in regressions involving time series that the residuals of a variable are serially correlated overtime. The existence of serial correlations in the residuals violates the standard assumption of the OLS regression theory that disturbances are not autocorrelated. If the residuals of a regression model are autocorrelated, OLS is no longer efficient among linear estimators, standard errors from OLS models are not correct, and are generally understated, and in presence of lagged dependent variables on the right-hand side, OLS estimates are biased and inconsistent.

A common approach in testing for serial correlation in the residuals of models involving time series data is the examination of the shape of the autocorrelations and partial autocorrelation functions of the residuals, together with the Ljung-Box Q-Statistics. If a model does not have serial correlation in the residuals, a correlogram should indicate that the autocorrelations and partial autocorrelations are nearly zero at all lags, and all Q-statistics should be insignificant with large p-values. The null hypothesis in these tests is that the residuals of the model are not serially correlated, against the alternative that there is autocorrelation.

3.6 Evaluation of Forecasting Accuracy

By definition, forecasting is a process of predicting future values of a variable. The confidence given to the predicted values lies on the assumption that the pattern identified in the time series and described by the model parameters will not change in the future (Aldridge, 1999). However, because of the stochastic characteristic of the price data generating process, there is a certain presence of error in every forecasting process regardless of the technique employed. Therefore, there is a need to evaluate the accuracy of every forecasting model.

3.6.1 Statistical Criteria

Statistical evaluation of forecasts is possible whenever we have actual values for the forecast period, either in an in-sample or in an out-of-sample forecasting. However, we should be especially concerned with the results of statistical evaluation criteria from out-of-sample forecasts, as forecasting is basically an out-of-sample problem.

Uncertainty in the residuals is the major source of the out-of-sample forecasting prediction errors in time series models. Therefore, even if a model provides a good fit in an in-sample forecasting context, it needs to show a good out-of-sample forecasting performance, in order to give an idea of what we would have to expect in practice if we did not yet know the future values of the variable (Wooldridge, 2000).

Several statistical forecast evaluation methods evaluate the sum of the squared residuals. Residual uncertainty exists because the unknown disturbances in the forecast period, present in the equation, are replaced by their expected value of zero, while the actual values are different from zero. The wider the deviation in the individual errors, the larger the overall error in the forecasts. In time series forecasts, residual uncertainty is explained by the fact that lagged dependent variables depend upon lagged disturbances.

The statistical criteria most widely used to evaluate and compare competitor forecasting techniques are the mean square prediction error (MSE), the root mean squared error (RMSE), the mean absolute error (MAE), the mean absolute percentage error (MAPE), Theil's inequality coefficient (TIC), and the turning point error (TPE).

If we assume that forecast users are risk averse, a loss function such as the mean square prediction error (MSE) is a convenient method because it penalizes large errors more heavily. Therefore, it is widely used.

By definition, the MSE is the sum of the squared differences between the observed and the predicted value of the variable at time t, divided by the number of observations in the sample:

$$MSE = 1/n\sum_{t=1}^{n}(\hat{p}_t - p_t)^2 \tag{26}$$

where p_t is the actual price at time t, $\hat{p}_t$ is the predicted price for time t, and n is the number of observation in sample.

The RMSE and MAE depend upon the scale of the dependent variable and are used as relative measures to compare alternative models in forecasting the same series. The decision rule is that the smaller the RMSE, MAE or MAPE, the better the accuracy of the model in its forecasting ability. The Theil criterion is based on the U-statistic, which varies from 0 to 1, being equal to 0 when the forecast is exact, and equal to 1 when the forecasting technique is no better than a naive method.

The TPE is another common evaluation tool, and indicates the ability of a model to predict changes in direction. This technique is especially used if the forecast user is particularly concerned with predicting changes in the direction of the time series.

3.6.2 Economic Criteria

Agricultural commodity forecasts are made with the aim of helping participants in the sector to make better decisions in the production process, for better investment planning and policy decisions. Yet forecasting models are typically subject only to statistical evaluation. An implicit assumption in the statistical evaluation of forecasting models is that the statistical criteria are consistent with, and optimal for, the subsequent use of the forecast in the decision making process (Aldridge 1999). However, model selection decisions based solely on the statistical evaluation criteria sometimes are not optimal from the standpoint of economic decision making, hence there is a need for further evaluation aimed at examining the extent to which forecast users are able to use the model to make more profitable decisions (Brandt and Bessler 1983; Parks 1989; Wright et al. 1986; Gerlow 1993; Aldridge 1999). For this purpose, some literature argues that forecasting models should be chosen in accordance with the preferences of the final users of the forecast, rather than for their statistical fit. That is, economic evaluation should prevail over statistical criteria (Brandt and Blessler 1983; Wright et al. 1986; Leitch and Tanner 1991; Gerlow 1993). This literature argues that sometimes forecasting models with poor statistical performance may be better than those with better statistical accuracy when evaluated using economic criteria. For example, Gerlow's (1993) study indicated ARIMA models have better economic results although their statistical accuracy is poor compared to structural econometric models.

Results in economic evaluation have suggested that more complex models do a better job than naive models in allowing profitable decisions in the production process. Aldridge, for instance, forecasted that the random walk model is dominated by all other strategies in terms of mean price received and average percent returns.

Usually, economic evaluation criteria are based upon the storage problem, where the decisions to sell or store under each model are compared to "no model" scenarios. The profitability of the decisions to sell or store under the different models' strategies is compared along with the "no model" or default strategies by the use of criteria such as the mean price received, net to storage cost or not, and the percentage of correct decisions.

In evaluating the economic performance of the different forecasting models, this research uses present value of the gross marginal revenue in addition to the mean price received criterion and the percentage of correct decisions. The default strategies and three economic criteria are presented and broadly discussed in chapter 6.

3.7 The Models

In the practice of price forecast modeling, it is usual to use prices at the producer level. In this study, however, retail prices are used instead of producer prices, because the latter are not available in a long time series in Mozambique. Also, checking for the correlation between first differenced producer and retail prices in Mocuba and Manica for the period in which both prices are available shows that in Mocuba the coefficient of correlation is

0.867, significant at any conventional significance level. In Manica, the coefficient of correlation between first differenced producer and retail prices is 0.918, also significant at any reasonable significance level. These results suggest that there is a high co-movement of producer and retail prices in both places, and therefore we could expect that the forecasted values from our models could be used to infer expected prices at the producer level. A graphical analysis involving producer and retail prices in Manica and Mocuba support this idea.

3.7.1 The ARIMA models for Nampula and Maputo

After taking differences in the price data if necessary, the tentative ARIMA models for Nampula and Maputo will be specified using the following identification procedures: First sample autocorrelation coefficients and partial autocorrelation coefficients will be plotted. Then, using the parsimonious principle, an ARIMA(1,d,1) model of the form:

$$w_t = \delta + \Phi_{1wt-1} - \theta_1\varepsilon_{t-1} + \varepsilon_t \tag{27}$$

will be the first tentative model for each time series. The final ARIMA (*m,d.n*) models will be determined according to the values of *m*, *d* and *n* after the models have been corrected for autocorrelation in the error term if necessary.

3.7.2 VAR Model for Nampula

In addition to the univariate ARIMA model, a VAR or VEC model will be estimated for maize prices in Nampula city, depending on whether the data follows an $I(0)$ or an $I(1)$ process. The estimation of the VAR or VEC model is intended to reflect the fact that the process generating maize prices in this city in northern Mozambique is strongly affected by maize exports, especially to Malawi. As stated in chapter 1, about 70 percent of the maize consumed in Nampula is brought from northern Zambezia, and considerable maize produced in Nampula and Zambezia province is exported to Malawi. Therefore, there is reason to expect that both demand for maize in the north of Mozambique and Nampula's maize price generating process involves both northern Zambezia prices and Malawian production estimates, and therefore these variables influence the process of maize price generation in Nampula city in particular and northern Mozambican in general. A VAR model with a single autoregressive component for Nampula prices, VAR (1), can be presented as:

$$p_{1t} = \delta_1 + \varphi_{11p}p_{1,t-1} + \varphi_{12}p_{2,t-1} + \varphi_{13}q_{3,t-1} + \varphi_{14}q_{4,t-1} + \varepsilon_{1t} \tag{28}$$

$$p_{2t} = \delta_2 + \varphi_{21}p_{1,t-1} + \varphi_{22}p_{2,t-1} + \varphi_{23}q_{3,t-1} + \varphi_{24}q_{4,t-1} + \varepsilon_{2t} \tag{29}$$

where p_{1t} represents maize price in Nampula at time t, p_{2t} represents prices of maize in Mocuba, northern Zambezia, at time t, q_{3t} represents actual quantity of maize production in Mozambique at time t, q_{4t} represents actual quantity of maize production in Malawi at time t; δ, λ, and φ are parameters to be estimated, and ε_{it} are the impulses.

3.7.3 VAR Model for Maputo

Like northern Mozambique, the price generating mechanism in the south of Mozambique involves some exogenous variables. Most maize consumed in southern Mozambique is either produced in central Mozambique or imported from South Africa. Therefore, both prices in central Mozambique and estimates of production in South Africa are important for the price generation mechanism in southern Mozambique. Maputo's model will therefore include both central Mozambique actual prices and predicted maize production in South Africa.

Because of its importance as one of the major sources of maize production and trade in central Mozambique, prices of Manica district are taken as representative of the region. A VAR (1,0) for Maputo prices is:

$$p_{1t} = \delta + \varphi_{11}p_{1,t-1} + \varphi_{12}p_{2,t-1} + \varphi_{13}q_{3,t-1} + \varphi_{14}q_{4,t-1} + \varepsilon_{1t} \tag{30}$$

$$p_{2t} = \lambda + \varphi_{21}p_{1,t-1} + \varphi_{22}p_{2,t-1} + \varphi_{23}q_{3,t-1} + \varphi_{24}q_{34t-1} + \varepsilon_{2t} \tag{31}$$

where p_{1t} represents prices of maize in Maputo at time t, p_{2t} represents prices of maize in Manica at time t, q_{3t} represents actual maize production in Mozambique at time t, q_{4t} represents predicted maize production in South Africa at time t; δ, λ, γ, and φ are parameters to be estimated, and ε_{it} are the impulses.

In both the univariate and the multivariate models, maize prices in each region will be forecasted out-of-sample. The research used ARMA, ARIMA, VAR and VEC models.

Production variables are used for the following reasons: (i) shocks in prices are expected to be explained by production fluctuations. A decrease in production in a given production year is expected to lead to an increase in the prices in the following marketing year. Hence, production variables are expected to improve the level of predictability of prices, and explain variations in prices. In addition to data on maize production in Mozambique, the research uses production data in Malawi and South Africa. Northern Mozambique is closely related to Malawi, if production in Malawi decreases prices in northern Mozambique are expected to increase as exports to Malawi are likely to occur. On the other side, if production in South Africa decreases, prices in southern Mozambique are expected to increase as this region considerably depends on imports from South Africa.

Chapter 4

PRELIMINARY DATA ANALYSIS

4.1 Introduction

This research uses data on retail maize prices and volume of maize production to estimate alternative time series models for forecasting maize prices in Mozambique. This chapter reviews the characteristics of the price data, and the implications of such characteristics for the results of the models. The chapter is organized as follows. Section 4.2 describes the data and data sources. Section 4.3 explains the methods used in dealing with missing data. Section 4.4 discusses the basic characteristics of the data. Section 4.5 addresses the deterministic time trend and seasonality issues. Section 4.6 discusses structural shocks in the data. Section 4.7 discusses on marketing margins between the pairs of markets involved in the bivariate VAR models, and section 4.8 discusses the implications of the data characteristics for the model results.

4.2 The Data

The data used in this research includes retail maize price in four Mozambican markets, and maize production figures in Mozambique, Malawi and South Africa. The data is monthly, covering the period from November 1992 to August 2000, and it was obtained from different sources. The data on maize prices is from the Agricultural Market Information System (SIMA) in the Ministry of Agriculture and Rural Development of Mozambique. These price series refer to the markets of Nampula and Mocuba in northern Mozambique, Manica in central Mozambique and Maputo in the south. In all cases, the prices are nominal and refer to the most important retail market in the area, in terms of volume of transactions.

The data on maize production in Mozambique and Malawi is the annual actual production for 1992 through 1999 and production estimates for 2000, and is from the Southern African Development Community database. For each country, every annual figure is repeated for the 12 months of the marketing year, from May to April, to match with the monthly price data.[8]

The data on South Africa's maize production estimates are forecasts from the South African Estimates National Committee (SAEC), under the Foreign Agriculture Services (FAS) of the United States Department of Agriculture (USDA). The SAEC does monthly updates of the agricultural production estimates. Every year, the first forecast estimates are made either in February or March, and the last estimates are made in August. From September through January or February, the August estimates are verified and used in the estimation of the final figure for actual annual production.

---

[8] In the southern African region, the average planting season occurs from September to November, and the harvest season starts around April. Thus it can be considered that the marketing year in Mozambique is from May of one year to April of the following year.

All prices are given in Metical per kilogram (sometimes presented as Mt/kg in this research). Metical is Mozambique's currency. Figures on maize production are given in thousands of metric tons. Table 4.1 presents the summary statistics of the prices.

Table 4.1 Summary Statistics

Variable[1]	Mean	Std Dev	Minimum	Maximum
Northern Mozambique				
Nampula Maize Price	1298	741.56	457	3896
Mocuba Maize Price	1111	727.64	243	3215
Southern Mozambique				
Maputo Maize Price	2276	889.78	695	3866
Manica Maize Price	1288	854.06	229	4643
Production Data				
Mozambique Maize Production	836	310.14	133	1196
Malawi Maize Production	1713	527.47	657	2478
South Africa Maize Production	8405	2,511.63	3683	13906
Number of Observations = 94				

[1] Price data are nominal prices in Metical per kilogram, and production data is in metric tons.

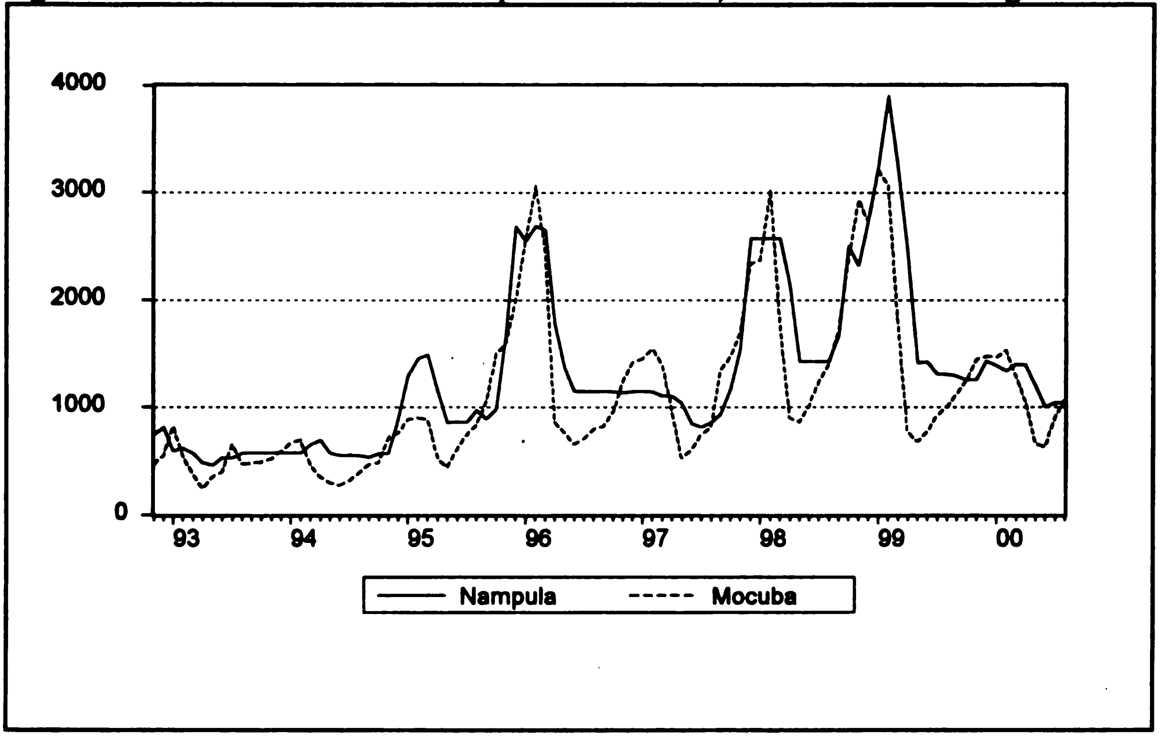
Table 4.1 shows that the average price in the producer areas of Mocuba and Manica is lower than that in the consumer areas of Nampula and Maputo respectively. Although these price differentials are an important condition for arbitrage activities to take place, too high a margin between prices in Manica and Maputo could be an indication that traders face problems in trading the commodity from the exporting market to the importing market. One of the weaknesses of Mozambique's database is the lack of systematic data on transportation cost, therefore no conclusive ideas can be drawn about this issue on the basis of the price differentials.

Second, Figures 4.1 and 4.2 suggest that all the time series have a clear seasonal pattern, where prices are in general higher between October and March and lower from April to September. Also, the presence of some spikes in the series suggests that structural chocks have affected the normal seasonal pattern in 1995, 1997 and 1998. Indeed, there was huge drop in production in 1994/95, and large amounts of maize were exported from northern Mozambique to Malawi in 1997 and 1998.

Third, these Figures suggest that Mozambique's maize price data should be investigated for trend, as should be expected with nominal price series. Figures 4.1 and 4.2 suggest that the rural markets are subject to greater seasonal price swings than are the urban markets. The lower prices are observed during the harvest season and the higher, in the planting and growing seasons.

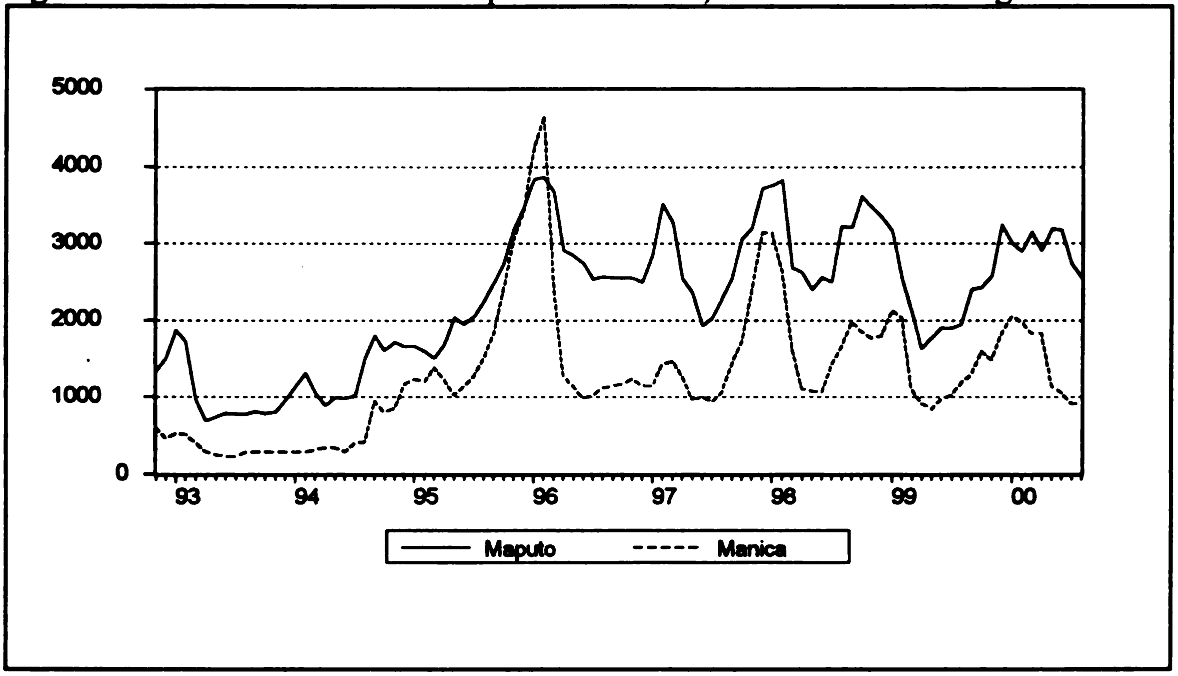
Finally, Table 4.1 indicates that prices in both the rural and the urban areas are higher in the center/south than in the north. Comparing Graphs 4.1 and 4.2 indicates that prices in the north are systematically lower than in Manica and Maputo. This pattern reflects the perpetual surplus conditions in the north, perpetual deficit in the south, and the proximity of the center to the south, making trade possible and strengthening prices in the center.

Figure 4.1 Retail Maize Prices in Nampula and Mocuba, November 1992 - August 2000



Figures 4.1 and 4.2 also show that prior to 1994, maize prices in Mozambique were relatively stable, suggesting either production stability or poor performance of the markets. Indeed, prior to October 1992, Mozambique was under a civil war and markets were nearly isolated from each other. Between 1992 and 1994, the country experienced a period of transition from war to peace and economic stability. For the first time in more than 10 years, traders started trading agricultural commodities between the producer areas and the consumer markets. This activity, however, did not become intense and solid until late 1994 and early 1995, when the first democratic elections were held in the country, peace and stability took place, and markets started operating more normally.

Figure 4.2 Maize Retail Prices in Maputo and Manica, November 1992 - August 2000



4.3 Missing Data

There were a few missing values in maize prices in some of the markets of study. For Manica and Nampula, for instance, price data were missing for the first two observations, which suggests that the data collection in these two markets started two months later compared to Maputo and Mocuba. The missing values were completed using regression models. For Nampula, a regression of Nampula prices on Mocuba was run and the estimates used to predict the missing values in the Nampula series. Likewise, Manica price series was regressed on Beira, the closest major market, and the estimates from this model were used to fill in the missing values in Manica. This approach was chosen for two reasons. First, it was not possible to regress retail prices on prices at other transaction levels either in Nampula or in Manica because the collection of data on other transaction levels started later compared to the retail level. Second, other statistical approaches such as average prices of the surrounding months were not appropriate because the missing values were the first observations in the series.

Before the regression models were estimated in order to find the estimates for the missing values, each one of the time series involved in the regression models was tested for stationarity and the innovations were checked for serial correlations.

Mocuba had missing values in February and December 1993 and in August 1994. An examination of the prices immediately before and after the missing values suggested that alternative statistical approaches to regressing this price series on another price series could be followed. In this case, for each month with missing data, a simple average of the prices of the immediately preceding month and the immediately following month was used to fill in each missing data point. This procedure was used for two reasons: there was not a complete series related to Mocuba which could be used to estimate the missing data in this series, and, a visual inspection indicated a consistent increase or decrease of the prices surrounding each missing value.

4.4 Characteristic Time Series Properties of Commodity Prices

This section reviews the most important characteristics of maize prices in the country. Specifically, this analysis involves the stochastic trend or unit root property, and the property of time-varying volatility.

4.4.1 Unit Root or Stochastic Trends

Many commodity price series appear to share several stochastic properties, such as stochastic trends, price comovement and volatility (Myers 1994). For proper modeling of the underlying data generating process and a better understanding of the results of the time series models, it is of critical importance to examine these stochastic properties of time series data prior to building forecasting models with time series methods. The challenge then is to estimate time series models that can make efficient use of the essence of the data generating process by examining the stochastic properties of the data.

As it has been noted in chapter 3, out-of-sample forecasts obtained from time series models can be highly divergent over time. These forecast errors can become larger over time if they contain a stochastic trend - or a unit root - which is not accounted for. As Myers (1994), and Asche, Bremes and Wessells (1999) noted, several empirical studies of commodity prices have found evidence of unit roots.

As noted in chapter 3, the Dickey-Fuller, the Augmented Dickey-Fuller and the Phillips-Perron tests are the most common tests used to test the hypothesis of a stochastic trend.

In this research, Dickey-Fuller tests are used to test for unit roots. When the results reveal that the innovations are serially correlated, the ADF was performed. The null hypothesis in these tests, as indicated in chapter 3, is that the data contains a stochastic trend. The results of the tests for stochastic trend are presented next.

4.4.1.1 Unit Root Tests

The Dickey-Fuller and Augmented Dickey-Fuller tests can be carried out with the inclusion or not of seasonal dummies and a time trend variable. Visual inspections of the data series used in this research suggests that they could contain seasonality and upward time trends. The DF tests are then conducted with the inclusion of time trends and monthly seasonal dummy variables. The number of lags in the DF tests was chosen so that the highest lag significantly different from zero is included in the test for the levels of the prices, thus white noise residuals are generated. Table 4.2 presents the results of the DF tests for unit root for the four maize price series.

Table 4.2 Unit Root Tests for Maize Prices in Mozambique, with trend and seasonal dummies

	Coefficient of The Unit Root Test	
	Levels	First Differences
Northern Mozambique		
Nampula	-2.18	-6.31
Mocuba	-1.74	-6.96
Southren Mozambique		
Maputo	-0.77	-4.86
Manica	-2.24	-5.48

With a critical value at a 5% level of -3.41 (due to the inclusion of a trend variable), Table 4.2 shows that the null hypothesis of nonstationarity cannot be rejected for the price series in levels, but it is rejected for all the series in first differences. Hence we conclude that all the price series are nonstationary in levels and are integrated of order one, $I(1)$. In order to have white noise innovations, lags were added to the DF tests for all the prices series when testing for unit root in levels. There was no need to add lags in the Nampula and Mocuba price series when testing for unit root in first differences.[9]

Recall that structural changes can affect the Dickey-Fuller tests. As Asche, Bremes and Wessells (1999) noted, a data series with structural change may seem nonstationary if the structural change is not taken into account, but stationary if it is accounted for. Tests for structural shocks in the data used in this research, whose results are presented and discussed in section 4.7, indicate that this data does not contain significant structural changes.

4.4.2 Time-Varying Volatility

In addition to not having a constant mean, many commodity prices are highly volatile, with implications to producers and consumers, and to the economy as a whole, especially in countries where the gross domestic product and exports depend mainly on primary commodities (Myers 1994). This high volatility may indicate an overall market inefficiency and risk management problems particularly when all the necessary institutions for a well-functioning market are not in place. In addition, volatility of commodity prices

---

9

Time series literature indicate that Dickey-Fuller tests should be compared to the critical values of -2.86 (when there is no a time trend variable), or -3.41 (when a time trend component is included in the test), at 5% level. Comprehensive Tables with the DF critical values also include 1%, 2.5% and 10% levels.

tends not to be uniform over time. As Myers (1994) noted, several studies have found that usually, periods of relative tranquility, where small changes in prices are followed by other small changes, are often followed by periods of high volatility, in which large changes in prices are followed by other large changes.

High volatility alone does not imply specific statistical problems. The concern with high volatility is to understand the reasons behind this problem and how to minimize the consequences. Advanced methods have been found to deal with the time-varying volatility problem. Such methods involve the use of the autoregressive conditional heteroscedasticity (ARCH) models and generalized autoregressive conditional heteroscedasticity (GARCH) models. Although Mozambique maize price data suggests the existence of high volatility, visual inspection does not suggest that this volatility is time-varying. This research does not estimate models with either ARCH or GARCH effects.

4.5 Deterministic Time Trend and Seasonality

Section 4.4.1 discussed the stochastic trend or unit root problem, and indicated that many time series of commodity prices contain stochastic trends. However, it did not discuss the issue of deterministic time trend. Indeed, a special case of a potential violation of the stationarity assumption is when the data grows with a trend, common in many time series. Because deterministic trends have a permanent and constant effect in the long term and lead to trend-stationary time series, while stochastic trends or unit roots lead to continuous changes in the conditional mean or intercept of the forecasts, it is important to determine whether the trend is deterministic or stochastic.

The nature of trend is examined when the data is tested for unit root. If it is found that the time series is trend stationary, then we are in presence of a deterministic trend, and this kind of trend can be controlled for by estimating equations with a polynomial trend such as described in equation (18). If, on the other hand, differences have to be taken in order to control for the trend component, then we are in presence of a stochastic trend, and the process is called difference-stationary.

The tests on the unit root presented in section 4.4 indicated that all the series of price data used in this research have unit root in levels and are all $I(1)$. When differences of the data are taken, any linear trend is automatically removed from the series.

Because the tests on the unit root indicated that all the price series used in this research are $I(1)$, this research tests jointly for drift and seasonality in each price series, in first differences. Table 4.3 has the results of these tests.

Table 4.3 Drift and Seasonality in First Differences in Maize Prices in Mozambique

Northern Mozambique	Coefficient		
	Drift[1]	F Statistic for Seasonal Dummies	
		Model F Statistic	Critical Value
Nampula	423.17 (109.69)	5.35	$F_{(8, 80)} = 2.04$
Mocuba	150.92 (98.78)	8.90	$F_{(6, 81)} = 2.20$
Central/Southren Mozambique			
Maputo	172.35 (112.18)	5.26	$F_{(11, 78)} = 1.94$
Manica	177.25 (160.05)	3.34	$F_{(13, 78)} = 1.94$

[1] Numbers in parenthesis are standard errors

The results in Table 4.3 suggest that Mozambique's maize prices have a positive but generally insignificant drift term, and a significant seasonality in the drift. In fact, at a 5% level critical value of 1.96, the *t* statistic tests for the drift terms, fail to reject the null hypothesis of no drift for Mocuba, Manica and Maputo. Nampula's price series is the only one with a significant positive drift at 5% level of significance. Similarly, the tests for the joint significance of the seasonal dummies indicate that, at the 5% level, there is a significant evidence of seasonal growth rates in all the time series. The model *F* statistics are higher than the critical values for all the time series.

4.6 Structural Shocks

In time series data, it is important to check for data stability if there are reasons to believe that structural shocks could have affected the data. Controlling for such structural shocks is a way to obtain better understanding of the results from the models. A common approach to control for structural shocks is by applying Chow's Breakpoint Tests.

In the application of Chow Breakpoint tests, the sample is divided into the number of sub-samples suggested by the data path. The number of observations in each sub-sample should be higher than the number of coefficients in the equation being estimated. Dividing the sample into sub-samples and running Chow Breakpoint tests gives coefficients that are used to investigate for stability in the data. In practice, the Chow Statistic test for a time series with a single breakpoint is run by estimating the sample equation for the two sub-samples. The coefficients from the two sub-samples are compared under the null hypothesis that they are not different. A rejection of the null means that there is structural shock with a strong impact on the data.

The practical procedure to perform a Chow test for a structural change is based upon the restricted and the unrestricted models. Summing the squared residuals from the two sub-samples gives the unrestricted residual sum of squares. The equation is then fitted to the complete set of sample observations, which yields the restricted residual sum of squares. An F statistic is calculated as follows: $F=[(SSR_T-(SSR_1+SSR_2)]/(SSR_1+SSR_2)*[n-2(k+1)]/(k+1)$, where SSR_T is the sum of squared residuals from the whole sample's equation, SSR_1 and SSR_2 are the sum of squared residuals from sub-samples one and two respectively, n is the number of observations, and k is the number of explanatory variables. The null hypothesis is tested by comparing the estimated F statistic to critical F values.

The data is said to be stable if the coefficient is constant over the sub-samples. If the data is stable, it can be concluded that the data generating mechanism was the same before and after the structural shock.

These sudden changes in the prices may cause difficulties in estimating price prediction models, especially when using time-series models which do not take into account these structural changes. Problems in modeling data with these kinds of shocks could be minimized if more information were available and structural models could be estimated. However, there is limited availability of consistent historical data on maize in Mozambique and Malawi. In this research, the existing price data was then used to test for stability by the use of Chow Breakpoint tests. November 1995 was considered the breakpoint. The results of the test indicate an $F(66, 14)$ statistic of 1.64, leading to a failure to reject the null hypothesis of stability at any reasonable critical value.

In southern Mozambique, in turn, prices only had an extraordinarily high increase in 1995 after the period of stability prior to this date. As pointed out in chapter 1, southern and central Mozambique are highly linked by comparatively less precarious roads. Because maize production in central Mozambique has been considerably stable except for 1995, maize supply to southern Mozambique and to the net consumer areas in central Mozambique has also been stable most of the period of study.

Moreover, because southern Mozambique, especially Maputo, is a considerable consumption market, if there is a sudden drop in maize supply from central Mozambique, maize grain or maize flour can be imported from South Africa and other neighboring countries. As a result, prices in Maputo are expected to be more stable than those in northern Mozambique, unless unexpected, strong events occur, like the February/March 2000 floods. Indeed, prices in southern Mozambique increased unexpectedly high in March and April 2000, against the usual pattern of low prices from March through September every year. With the floods, the southern region of Mozambique was isolated from the rest of the country for the next two to three months and, because considerable maize consumed in Maputo is supplied from central Mozambique, this isolation of the southern portion of Mozambique led to this unusual price spike in March and April in 2000.

This research had intended to run a Chow Breakpoint test to capture this event, but because the data used in this research goes only through August 2000 - there are only 6 observations between March and August 2000 - this test could not be run. Future research, however, could be able to capture this event with the Chow Breakpoint test approach. A Chow Breakpoint for 1995 indicated no evidence of structural shock. Using

October 1995 as the breakpoint, an F(62, 16) of 1.60 was found, with leads to a failure to reject of the null hypothesis of stability.

4.7 Marketing Margins

When prices of a commodity involve two or more different levels of transaction or two or more geographically separated markets, one common concern is that of marketing margins between the levels of prices. This research uses related price time series to estimate multivariate time series models. As such, the issue of marketing margin is of interest.

Marketing margins can be defined, simply, as the difference in the price of a commodity between two stages of transaction of the commodity. As Tomek and Robinson (1987) noted, marketing margin can be viewed either from the perspective of price difference between the two stages of production, usually between the producer and final consumer levels, or from the standpoint of the prices of the marketing services, the most important of which, from the agricultural commodities point of view, are the costs involved in packing, transporting, handling and storing the commodities.

Figure 4.3 shows the margins between maize retail prices in Nampula and Mocuba. This Figure indicates that the margin of maize prices between Nampula and Mocuba is positive in most of the period (i.e., Nampula prices are generally higher than Mocuba prices). Prices in Mocuba tended to be less variable before 1995 than from 1995/96 onward. Both the lowest price ratio of 0.79 and the highest value of 3.2 were observed between 1998 and 1999. Also, negative marketing margins between Nampula and Mocuba tend to occur during the growing months of November, December and January. This pattern is in accordance with the finding that the seasonal pattern is more accentuated in the rural areas (Mocuba) compared to the urban areas (Nampula), given that the urban areas have better storage infrastructure than the rural areas. Furthermore, urban markets are supplied with product from different origins, and the markets are bigger and more stable.

Figure 4.3 Marketing Margin for Maize Prices, Nampula - Mocuba, Nov 1992 - Aug 2000

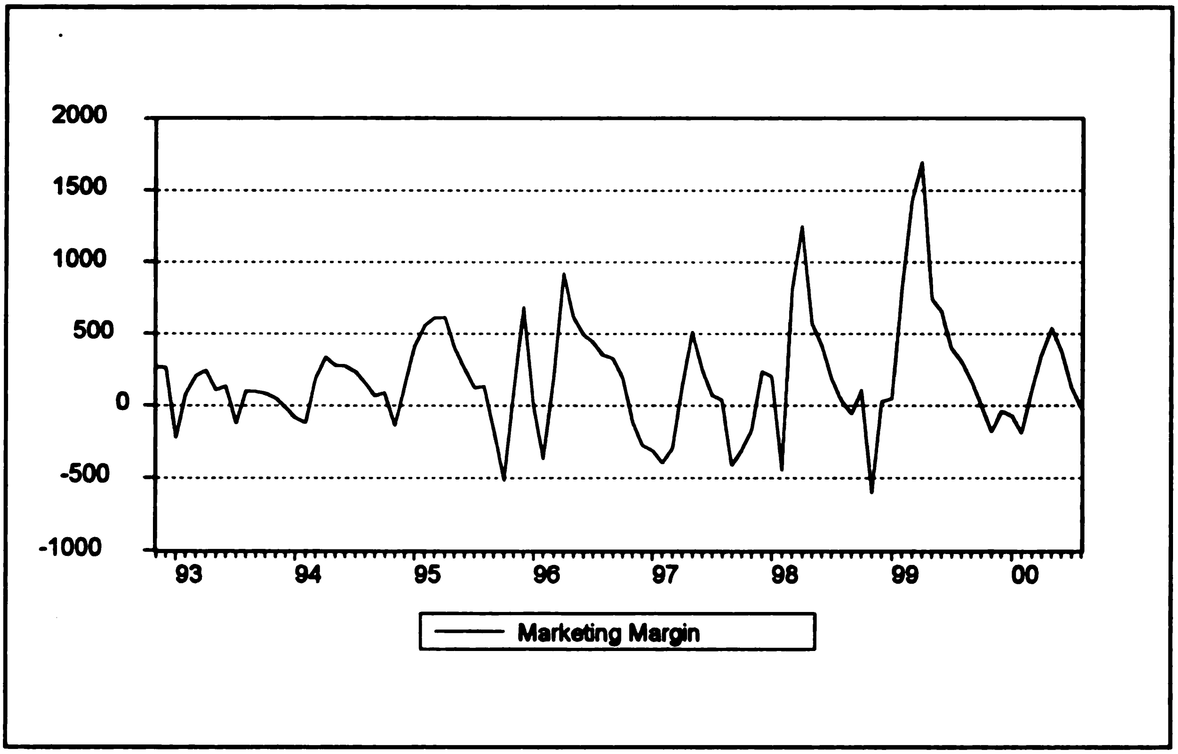


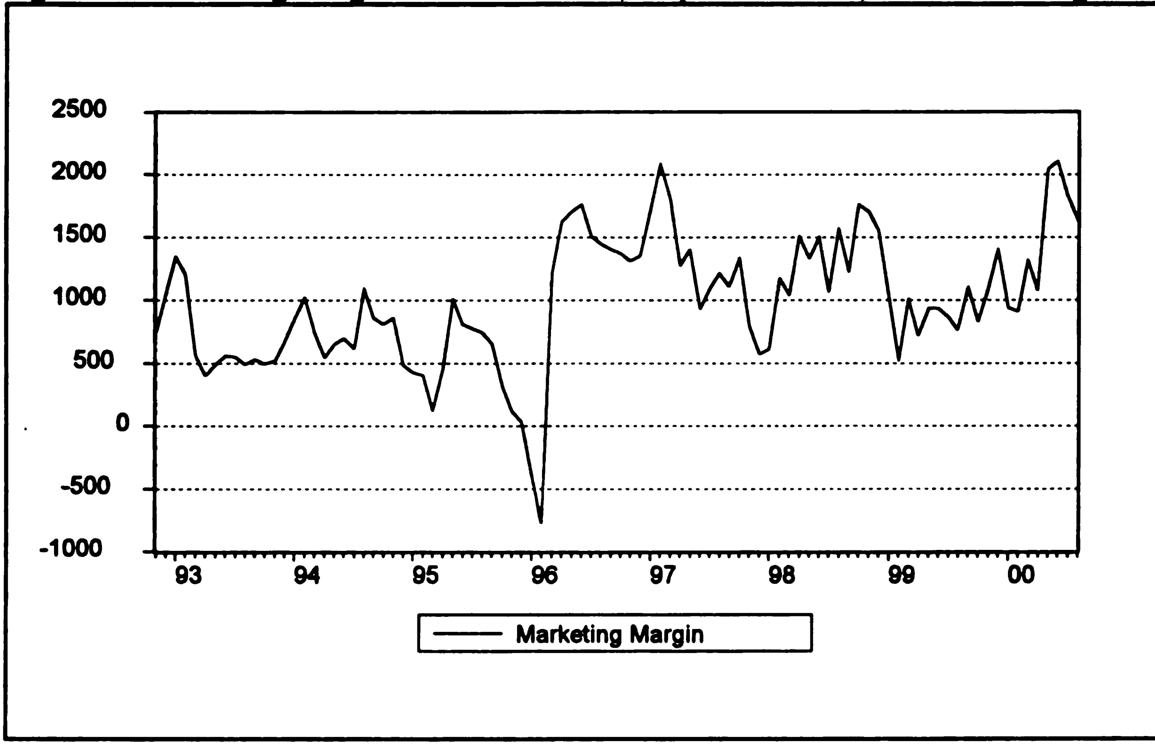
Figure 4.3 also suggests that these marketing margins may have increased over time. A simple linear regression of the marketing margins on time trend indicates, however, that

this positive tendency of the marketing margin is not significant at any reasonable significance level. The coefficient on the time trend of 0.92, with a *t* statistic of 0.77 after ensuring that the innovations are white noise suggests that there is no evidence to reject the null hypothesis of no trend in this marketing margin.

The marketing margin of nominal maize prices between Maputo and Manica is positive except in January and February of 1996. As explained in section 5.4., a considerable drop in production took place in Mozambique in 1995, and prices in the producer areas in northern and central Mozambique increased strongly in late 1995 and early 1996. This structural shock did not affect southern Mozambique, especially Maputo, where imported yellow and white maize, either through commercial imports or through food aid programs, was available.

In the remaining of the period, prices in Manica were higher than those in Maputo although prices in Manica also tended to have some clearly accentuated seasonal fluctuations. The long distance between Manica and Maputo that partly explains the high price differential between these two markets, is responsible for the systematic positive marketing margin between the two markets, in addition to Maputo's better access to alternative sources of maize supply than other Mozambican markets.

Figure 4.4 Marketing Margins for Maize Prices, Maputo - Manica, Nov 1992 - Aug 2000



Furthermore, the marketing margin of maize prices between Maputo and Manica is more accentuated than that between Nampula and Mocuba, and seems to be increasing. A linear regression of the marketing margin on time indicates that the coefficient of 2.93 on the time variable, with a t statistic of 2.01, white noise innovations ensured, is statistically significant at 5% level.

4.8 Implications of the Characteristics of the Data

When analyzing the results of any modeling based upon time series data, the characteristics of the data generating process have to be taken into consideration. A first step is to determine whether the price data is stationary or has unit root. Stationary data is different from nonstationary data in important dimensions. As Gujarati (1995) indicated, if there is a unit root in the data at any period t, the time series fluctuates not only as a result of shocks to the transitory component but also to the trend component, altering permanently their level, and resulting in weak inferences. This chapter has examined the stochastic trend problem, and it has been found that all the price series have a unit root.

The second concern when using Mozambique's maize data is related to the presence or not of a deterministic time trend and seasonality. Because the data series are nonstationary, this chapter has also investigated for drift and seasonality in first differences, and has found that there is no evidence of a drift in Mocuba, Maputo and Manica, but_there is evidence of seasonal drifts in all the time series. Note that a significant drift in differenced data is equivalent to a positive deterministic trend in levels. These results suggest that there is a need to include seasonality when modeling maize prices in Mozambique.

Finally, the three series on production data that are used in addition to the price data have also been tested for unit root and they all were found to be $I(1)$. Recall that the initial interest in this research was to use data on production estimates for the three countries, but lack of this kind of data for Mozambique and Malawi leads to the use of data on actual production instead.

Chapter 5

ESTIMATION OF THE ALTERNATIVE FORECASTING MODELS

5.1 Introduction

The main objective of this research is to forecast maize prices in Mozambique. In this chapter, several forecasting models are developed and results presented. In chapter 6, forecasting performance is compared by the use of statistical and economic criteria. The unit root tests presented in the previous chapter indicated that maize prices in Mozambique are $I(1)$ thus models in first differences should be estimated. In this research, however, in addition to estimating models in first differences, other models are estimated in levels, i.e., with no regard to whether the data is stationary or nonstationary, and the forecasting accuracy of all models is compared. This is a common practice when the final goal of estimating univariate and multivariate time series models is to perform forecasts. Even though the distinction between stationary and nonstationary data is crucial for time series data analysis and inferences, it is not of major importance when the final objective is short-term forecasting. Because the objective is forecasting and not hypothesis testing, the main criterion for model selection should be out of sample forecasting performance, not the results of unit root tests. For this reason, we examine models both with and without unit root and cointegration restrictions and compare their out of sample forecasting performance.

This chapter is organized as follows. Section 5.2 presents the specification of the univariate models, and the forecasts from these models. This discussion is presented by region, first for northern Mozambique, and then for southern Mozambique. Section 5.3 presents the model estimation and forecasting for multivariate models. Like in section 5.2, the discussion in this section is organized by region. The estimation of the multivariate models in section 5.3 is preceded by Cointegration tests. Finally, section 5.4 presents the conclusions of the chapter.

5.2 ARIMA Model Specifications

For the Nampula and Maputo price series, ARIMA models are estimated and used for a total of twelve one-step-ahead forecasts. These twelve monthly forecasts are obtained as follows. First, for each price series, an ARIMA model is estimated for the model estimation sample of November 1992 through August 1999, and a price for September 1999 is forecasted. Next, the model estimation sample is increased to September 1999, the ARIMA model is re-estimated, and a price for October is forecasted. This procedure is repeated until an ARIMA model with the model estimation sample of November 1992 through July 2000 is estimated and a price for August 2000 is predicted.

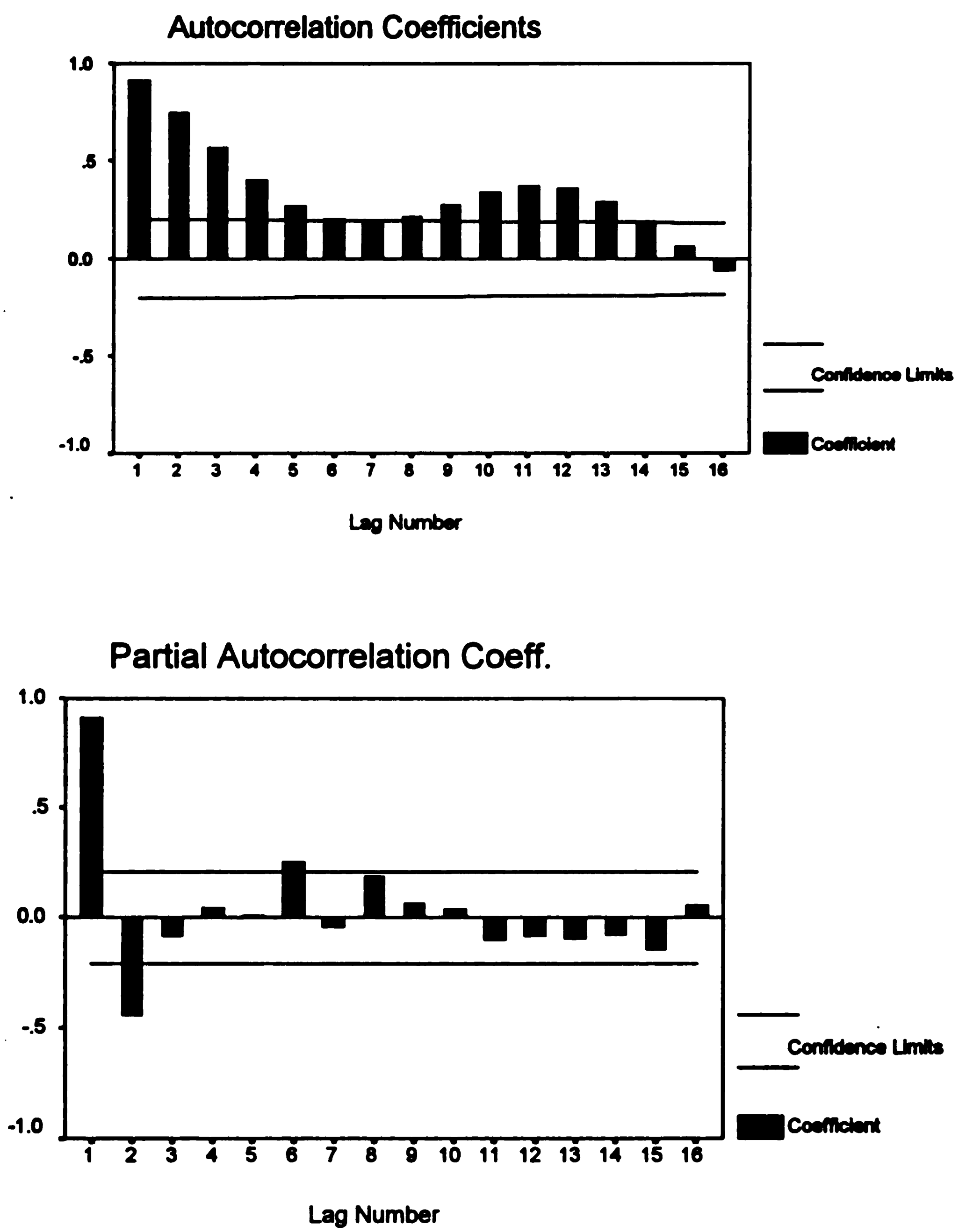
Two specifications of ARIMA models are used for each series. First, as the results of the test for trend and seasonality suggested the evidence of seasonality in each of the price time series, an ARIMA with monthly seasonal dummy variables is estimated for each series. Second, even though there is evidence of seasonality, ARIMA models without seasonality are estimated. These models are estimated for two reasons. First, because not all the months seemed to have significant coefficients when the models are checked for seasonality. Second, when estimating the ARIMA with monthly seasonal dummy variables, there is an indication that the forecast values tended to overestimate. The results from these ARIMA models with seasonal dummies are then compared to those from ARIMA models without monthly seasonal dummies. Random walk models are used as a point of comparison for more sophisticated models are estimated.

5.2.1 ARIMA Specification for Northern Mozambique

The procedures just described are applied first for Nampula maize prices. To have a good specification of the ARIMA model, a correlogram with the autocorrelation coefficients and partial autocorrelation coefficients for the Nampula price series is checked. As indicated in chapter 3, investigating the shape and behavior of both the ACF and PAC gives a plausible indication of the tentative ARIMA model. Figure 5.1 is the correlogram with the autocorrelation coefficients and partial autocorrelation coefficients for Nampula maize prices.

The correlogram on Nampula nominal prices indicates that the ACF tends to decay geometrically and the PAC goes to zero after 3 lags. Together, the path of the ACF and PAC suggest that the Nampula price time series data could be an AR(3) process, integrated of order 1, $I(1)$, with a presence of some orders of an MA component -

Figure 5.1 Autocorrelation and Partial Autocorrelation Coefficients for Nampula Prices



ARIMA(2,1,n). Additionally, the fact that the ACF has a seasonal frequency of cyclical "waves" strengthens the idea that this time series follows a monthly seasonal pattern.

When these identification procedures are applied to estimate the ARIMA model for Nampula retail maize prices, an ARIMA (1,1,0) model is identified with the inclusion of monthly seasonal dummies. This ARIMA model is then estimated and the results are presented in Table 5.1a. An exercise of adding and/or reducing orders of the AR and the MA components suggests the (1,1,0) model is a good, parsimonious model.

When the same estimation procedures are applied in searching for a parsimonious model with white noise innovations, an ARIMA (5,1,0) is found to be a good representation of the data generating mechanism when seasonal dummies are not controlled for. The results of this model are also in Table 5.1a.

In addition to the joint significance of the seasonal dummies described in chapter 4, the ARIMA specifications presented in Table 5.1a suggest that price changes in Nampula are significant at 5% level for all months except October. In the model without seasonal dummies, the second, third and fourth AR components are not significant. The fifth AR component, however, is significant and ensures white noise innovations. Table 5.1b presents the results of the Ljung-Box Q statistics test for serial correlation in the innovations of both models.

Table 5.1a ARIMA Models for Nampula Maize Prices

	ARIMA (1,1,0) with Seasonal Dummies		ARIMA (5,1,0) with no seasonal dummies	
	Coefficient	t statistic	Coefficient	t statistic
Constant	463.55	(4.13)	18.98	(0.67)
Jan	-370.20	(-2.85)		
Feb	-321.63	(-2.19)		
Mar	-558.33	(-3.68)		
Apr	-814.48	(-5.33)		
May	-845.72	(-5.53)		
Jun	-514.87	(-3.36)		
Jul	-484.37	(-3.17)		
Aug	-437.19	(-2.86)		
Sep	-429.18	(-2.73)		
Oct	-263.46	(-1.72)		
Nov	-312.07	(-2.29)		
AR(1)	0.27	(2.29)	0.36	(3.23)
AR(2)			-0.05	(-0.45)
AR(3)			-0.08	(-0.69)
AR(4)			0.02	(0.17)
AR(5)			-0.43	(-3.60)

Table 5.1b Correlogram for the Nampula ARIMA Models

	ARIMA (1,1,0) with Seasonal Dummies				ARIMA (5,1,0) with no Seasonal Dummies			
Lag	AC	PAC	Q Statistic	p-value	AC	PAC	Q Statistic	p-value
1	0.02	0.02	0.02		-0.02	-0.02	0.03	
2	-0.11	-0.11	0.95	(0.33)	-0.12	-0.12	1.11	
3	-0.06	-0.05	1.22	(0.54)	-0.06	-0.06	1.37	
4	0.05	0.04	1.39	(0.71)	-0.06	-0.08	1.70	
5	-0.28	-0.30	8.22	(0.08)	0.00	-0.02	1.71	
6	0.00	0.02	8.22	(0.15)	0.08	0.06	2.21	(0.14)
7	-0.09	-0.17	8.98	(0.18)	-0.13	-0.14	3.69	(0.16)
8	0.02	-0.01	9.03	(0.25)	-0.06	-0.06	4.03	(0.26)
9	-0.05	-0.07	9.27	(0.32)	-0.03	-0.07	4.12	(0.39)
10	0.17	0.08	12.00	(0.21)	-0.02	-0.05	4.15	(0.53)
11	0.06	0.06	12.36	(0.26)	0.01	-0.03	4.17	(0.65)
12	-0.14	-0.22	14.24	(0.22)	0.14	0.11	5.89	(0.55)
13	-0.08	-0.03	14.92	(0.25)	-0.04	-0.04	6.08	(0.64)
14	0.12	0.03	16.42	(0.23)	0.08	0.10	6.67	(0.67)
15	-0.13	-0.13	18.18	(0.20)	-0.12	-0.13	8.10	(0.62)

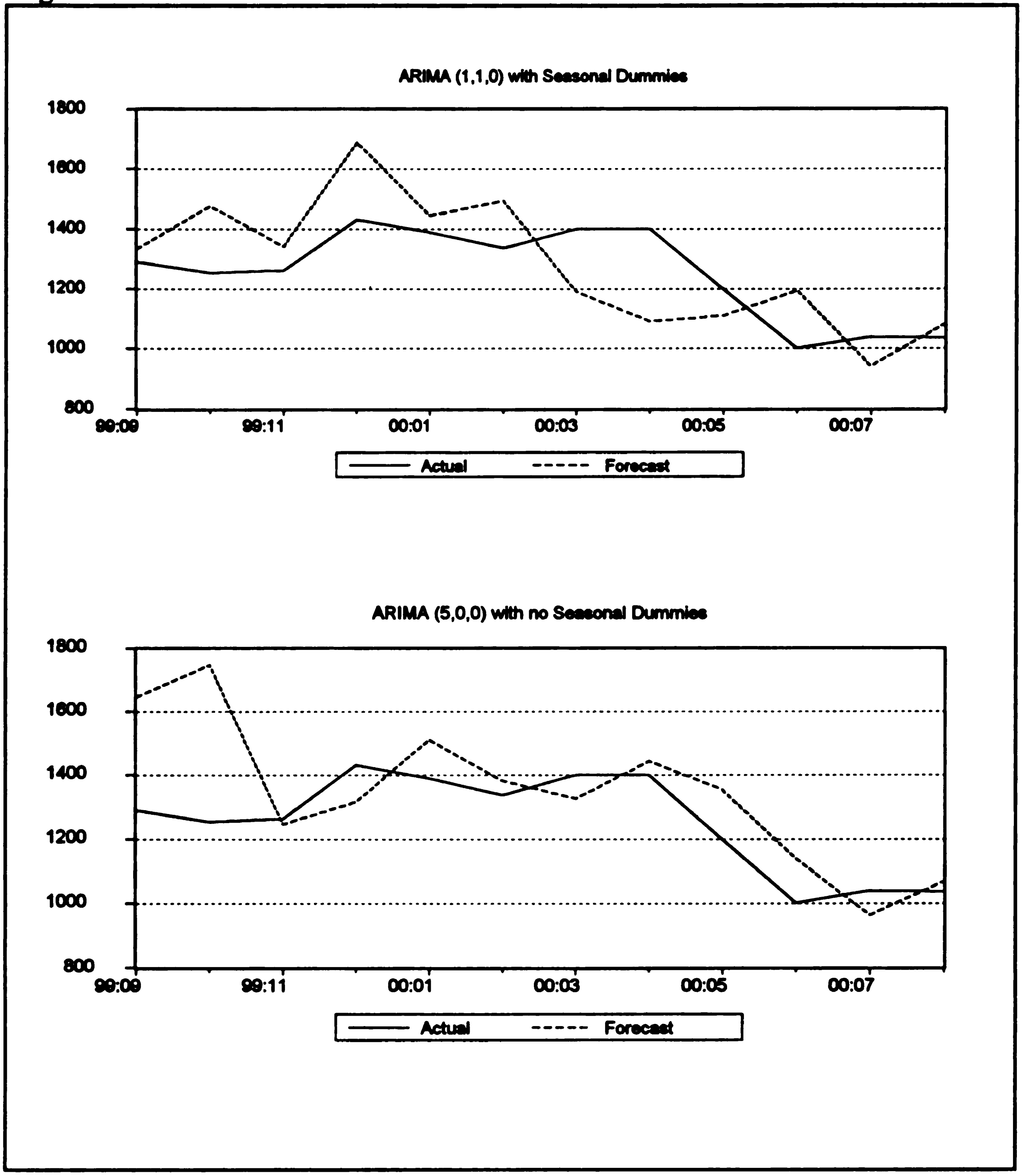
The results in Table 5.1b indicate that both ARIMA models for Nampula prices do not have serial correlations in the innovations. EViews adjusts for the AR and MA terms included in the ARIMA estimation, thus it does not report the p-values associated to the Q statistics for a number of lags equal to the number of AR and MA terms. For instance, in Table 5.1b, p-values associated to the first lag in the ARIMA model with seasonal dummies and the first 5 lags in the ARIMA model without seasonal dummies are not reported. For the lags where the probability values are reported, the statistic tests indicate that, at 10% level, we reject the null of no serially correlated innovations for all lags in the two models. Econometric packages such as GAUSSX enable calculating the p-values associated to X^2 and z-statistic for all lags. An inspection of those results confirms the conclusions drawn from the Q tests outlined in this research.

5.2.2 ARIMA Forecasting for Northern Mozambique

The two ARIMA models are used to perform short-term out of sample forecasts. Figure 5.2 shows the graphs with the ARIMA forecasts from the two Nampula price models.

Except for the first and second observations, the out-of-sample forecasts from the two comparative ARIMA models suggest that the model without seasonal dummies tends to predict better than the model with seasonal dummies. Comparing the two models suggests that the seasonal dummies appear to overestimate the actual price changes. This finding, however, seems not to be in accordance with what could be expected from the

Figure 5.2 Out-of-Sample Forecasts for Nampula Prices, ARIMA Models, Sep 1999 - Aug 2000



two models. Indeed, when the data was checked for seasonality, there was found evidence of seasonality thus the ARIMA model with seasonal dummies should do a better job in predicting future values of the variable than the model without seasonal dummies. The fact that this is not observed indicates that, either there is a problem in the specification of the ARIMA model with seasonal dummies, or that the pattern of seasonality had changed considerably between the model estimation sample and the forecasting sample. Because including higher AR components worsens the forecasting ability of the model, the mis-specification hypothesis is rejected, and the possibility that the pattern of the seasonal dummies is different between the model estimation period and the forecasting period is considered. Figure 4.1 suggested this difference in the seasonal price movements.

5.2.3 ARMA Models for Northern Mozambique

As stated earlier, when the primary objective of the research is forecasting, we can estimate forecasts without regard to whether the original data contains a unit root or not. In this work, even though there was evidence that the Nampula price series is nonstationary when testing for a unit root in the original data prices, ARMA models are estimated in addition to the ARIMA models, and the forecasting performance of all models is compared.

Table 5.2a ARMA Models for Nampula Maize Prices

	ARMA (2,0) with Seasonal Dummies		ARMA (2,0) with no Seasonal Dummies	
	Coefficient	t statistic	Coefficient	t statistic
Constant	1844.05	(5.06)	1334.17	(5.27)
Jan	73.17	(0.72)		
Feb	191.08	(1.21)		
Mar	72.41	(0.37)		
Apr	-300.47	(-1.36)		
May	-702.08	(-2.99)		
Jun	-770.34	(-3.20)		
Jul	-805.83	(-3.38)		
Aug	-792.12	(-3.48)		
Sep	-773.88	(-3.79)		
Oct	-589.24	(-3.54)		
Nov	-451.54	(-4.19)		
AR(1)	1.22	(10.62)	1.33	(13.20)
AR(2)	-0.31	(-2.70)	-0.46	(-4.60)

Table 5.2b Correlogram for the Nampula ARMA Models

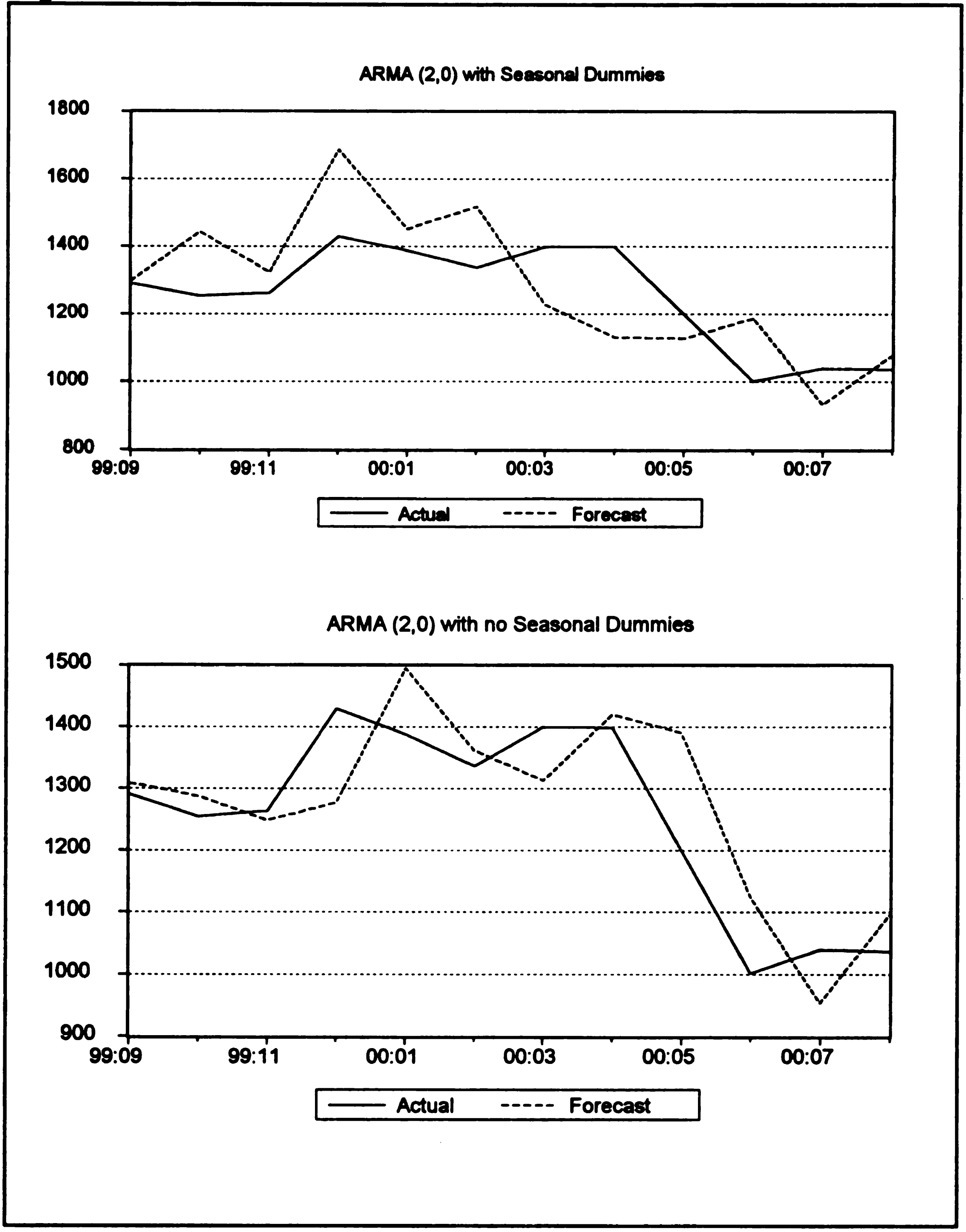
	ARMA (2,0) with Seasonal Dummies				ARMA (2,0) with no Seasonal Dummies			
Lag	AC	PAC	Q Statistic	p-value	AC	PAC	Q Statistic	p-value
1	-0.01	-0.01	0.00		-0.02	-0.02	0.04	
2	-0.07	-0.07	0.45		0.05	0.05	0.25	
3	-0.01	-0.01	0.46	(0.50)	-0.01	-0.01	0.26	(0.61)
4	0.11	0.11	1.50	(0.47)	0.10	0.10	1.20	(0.55)
5	-0.23	-0.23	5.93	(0.12)	-0.25	-0.25	6.77	(0.08)
6	0.07	0.09	6.35	(0.17)	0.08	0.07	7.32	(0.12)
7	-0.05	-0.09	6.58	(0.25)	-0.09	-0.08	8.11	(0.15)
8	0.06	0.06	6.90	(0.33)	0.02	0.01	8.15	(0.23)
9	-0.04	0.00	7.02	(0.43)	-0.06	-0.01	8.49	(0.29)
10	0.19	0.14	10.36	(0.24)	0.17	0.11	11.30	(0.19)
11	0.07	0.12	10.85	(0.29)	0.08	0.14	11.91	(0.22)
12	-0.14	-0.18	12.59	(0.25)	0.18	0.13	14.88	(0.14)
13	-0.08	0.00	13.20	(0.28)	0.00	0.02	14.88	(0.19)
14	0.13	0.06	14.89	(0.25)	0.16	0.11	17.30	(0.14)
15	-0.14	-0.11	16.77	(0.21)	-0.10	-0.06	18.25	(0.15)

Similar to the ARIMA models, two ARMA models are estimated, one with monthly seasonal dummy variables, and the second without seasonal dummies. An ARMA(2,0) with seasonal dummies and an ARMA(2,0) without seasonal dummies are estimated. Table 5.3a has the results of the models, Table 5.2b has the correlogram with the ACF and PAC, and Figure 5.3 presents the out-of-sample forecasts of the two ARMA models.

Like in the ARIMA models case, the significance of the seasonal dummies described in chapter 4 is confirmed by the individual coefficients of the monthly seasonal dummies in the ARMA (2,0) model. However, compared to the ARIMA model, the ARMA model indicates that from January to April the seasonal dummies are not significant at any conventional significance level. In both ARMA structures, with and without monthly seasonal dummies, the AR components are significant at 5% level, and the innovations are white noise as indicated in Table 5.2b.

The probability values associated to the Q tests in Table 5.2b indicate that, except for lag 5 of the ARMA model with no seasonal dummies which is statistically significant at 10% level but is not significant at 5% level, all the other lags for the two models are not significant at any reasonable significant level, which is an indication of no serial correlation in the innovations of the two models. With these results, the two ARMA models are used for forecasting purposes. A graphic view of the forecasts from the two models is presented in Figure 5.3.

Figure 5.3 Out-of-Sample Forecasts for Nampula Prices, ARMA Models, Sep 1999 - Aug 2000



As in the case of the ARIMA models, Figure 5.3 suggests that the seasonal dummies overestimate the monthly changes in the prices in the forecast period, and hence the ARMA model with seasonal dummies seems to predict higher changes in prices than the actually observed variations. As a result, the ARMA model with no seasonal dummies appears to generate better forecasts. Similar explanation to that of the ARIMA models seems to be valid for the ARMA models.

5.2.4 ARIMA Specification for Central/Southern Mozambique

This section estimates ARIMA and ARMA models for maize prices in Maputo. Like in the Nampula case, a correlogram with the ACF and PAC of Maputo maize prices is used as an initial indication of the tentative ARIMA model for this price series.

The geometrical decaying of the ACF of Maputo prices and the tendency of the PAC to be within the confidence limits after 2 lags shown in Figure 5.4 suggest that Maputo maize price could be an ARIMA model with at least 3 AR components. Furthermore, the slow geometrical decay of the ACF suggests the presence of some MA components. The less pronounced "waving" of the ACF compared to that in Nampula suggests a weaker seasonality in Maputo. However, there still is evidence of seasonality. Thus ARIMA models with seasonal dummies are investigated. Additionally, the fact that the ACF and PAC are very close to 1 at the first lag is in accordance with the conclusion from the unit root tests that this process is nonstationary. Following these indications, tentative

Figure 5.4 Autocorrelation and Partial Autocorrelation Coefficients for Maputo Prices

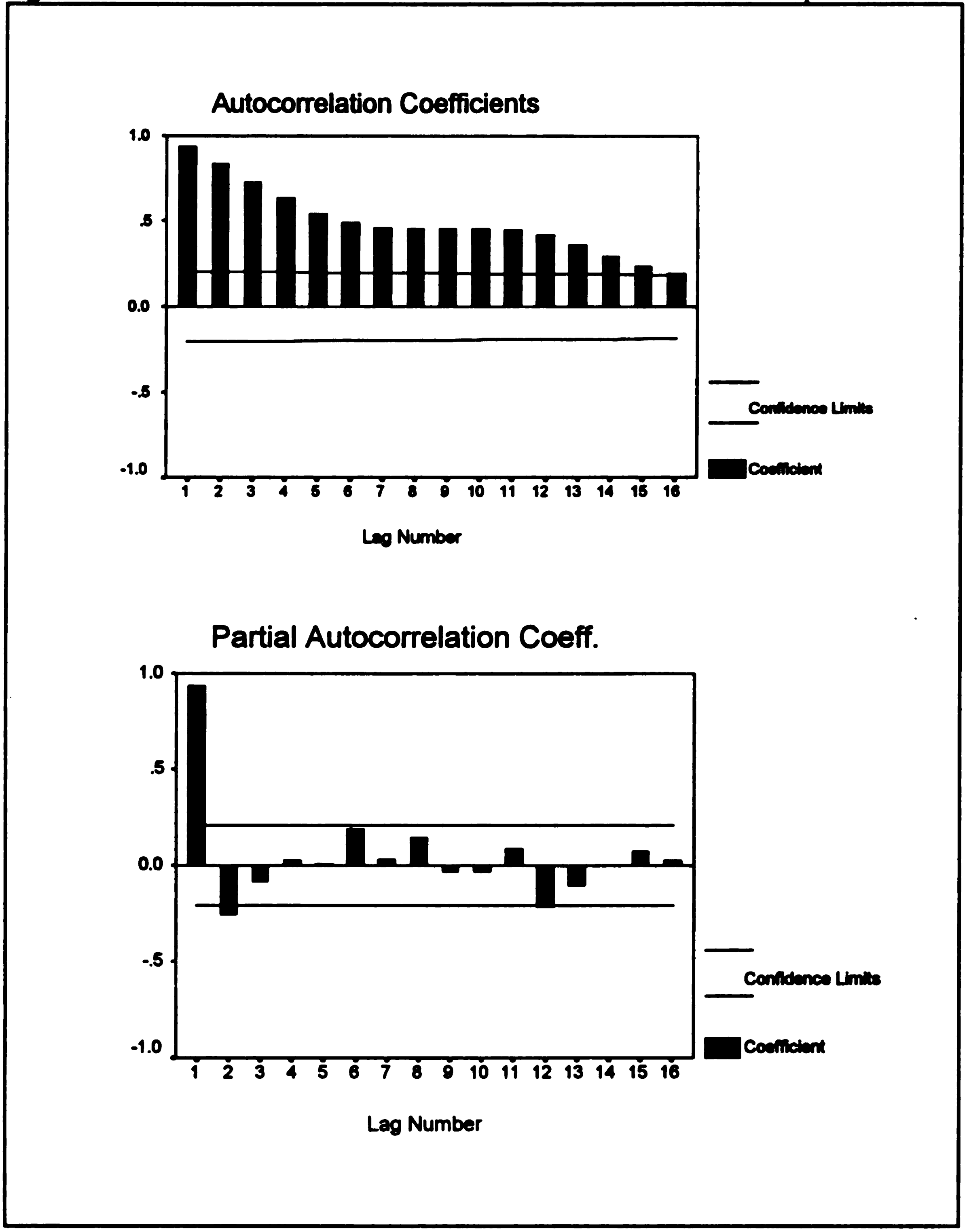


Table 5.3a ARIMA Models for Maputo Maize Prices

	ARIMA (2,1,1) with Seasonal Dummies		ARIMA (3,1,1) with no seasonal dummies	
	Coefficients	t statistics	Coefficients	t statistics
Constant	112.84	(1.00)	19.12	(0.82)
Jan	23.14	(0.16)		
Feb	-95.38	(-0.75)		
Mar	-564.70	(-3.47)		
Apr	-438.29	(-3.14)		
May	-90.25	(-0.55)		
Jun	-154.86	(-1.09)		
Jul	-121.66	(-0.75)		
Aug	138.37	(0.99)		
Sep	11.48	(0.07)		
Oct	40.11	(0.31)		
Nov	-18.14	(-0.12)		
AR(1)	-0.39	(-1.48)	0.90	(4.74)
AR(2)	0.35	(3.00)	-0.04	(-0.22)
AR(3)			-0.28	(-2.43)
MA(1)	0.56	2.11	-0.72	(-3.90)

Table 5.3b Correlogram for the Maputo ARIMA Models

	ARIMA (2,1,1) with Seasonal Dummies				ARIMA (3,1,1) with no seasonal dummies			
Lag	AC	PAC	Q Statistic	p-value	AC	PAC	Q Statistic	p-value
1	0.04	0.04	0.14		0.00	0.00	0.00	
2	0.03	0.03	0.21		-0.01	-0.01	0.01	
3	-0.14	-0.15	1.96		0.00	0.00	0.01	
4	-0.14	-0.13	3.69	(0.23)	0.13	0.13	1.33	
5	-0.17	-0.16	6.10	(0.11)	-0.13	-0.13	2.78	(0.12)
6	-0.10	-0.12	7.05	(0.20)	0.01	0.02	2.79	(0.27)
7	0.02	-0.02	7.08	(0.32)	-0.08	-0.08	3.30	(0.40)
8	0.18	0.12	9.83	(0.12)	0.09	0.08	4.05	(0.50)
9	0.08	0.01	10.46	(0.19)	0.06	0.09	4.39	(0.39)
10	-0.01	-0.07	10.47	(0.27)	-0.15	-0.18	6.50	(0.45)
11	0.21	0.24	14.77	(0.09)	0.11	0.16	7.69	(0.22)
12	-0.16	-0.13	17.07	(0.08)	0.21	0.17	11.94	(0.06)
13	-0.12	-0.10	18.55	(0.07)	-0.04	-0.05	12.12	(0.09)
14	-0.02	0.11	18.58	(0.09)	-0.14	-0.10	14.14	(0.09)
15	-0.01	-0.01	18.60	(0.12)	0.08	0.02	14.73	(0.12)

ARIMA models are estimated, and a final ARIMA (2,1,1) with seasonal dummy variables is identified. The results are presented in Table 5.3a.

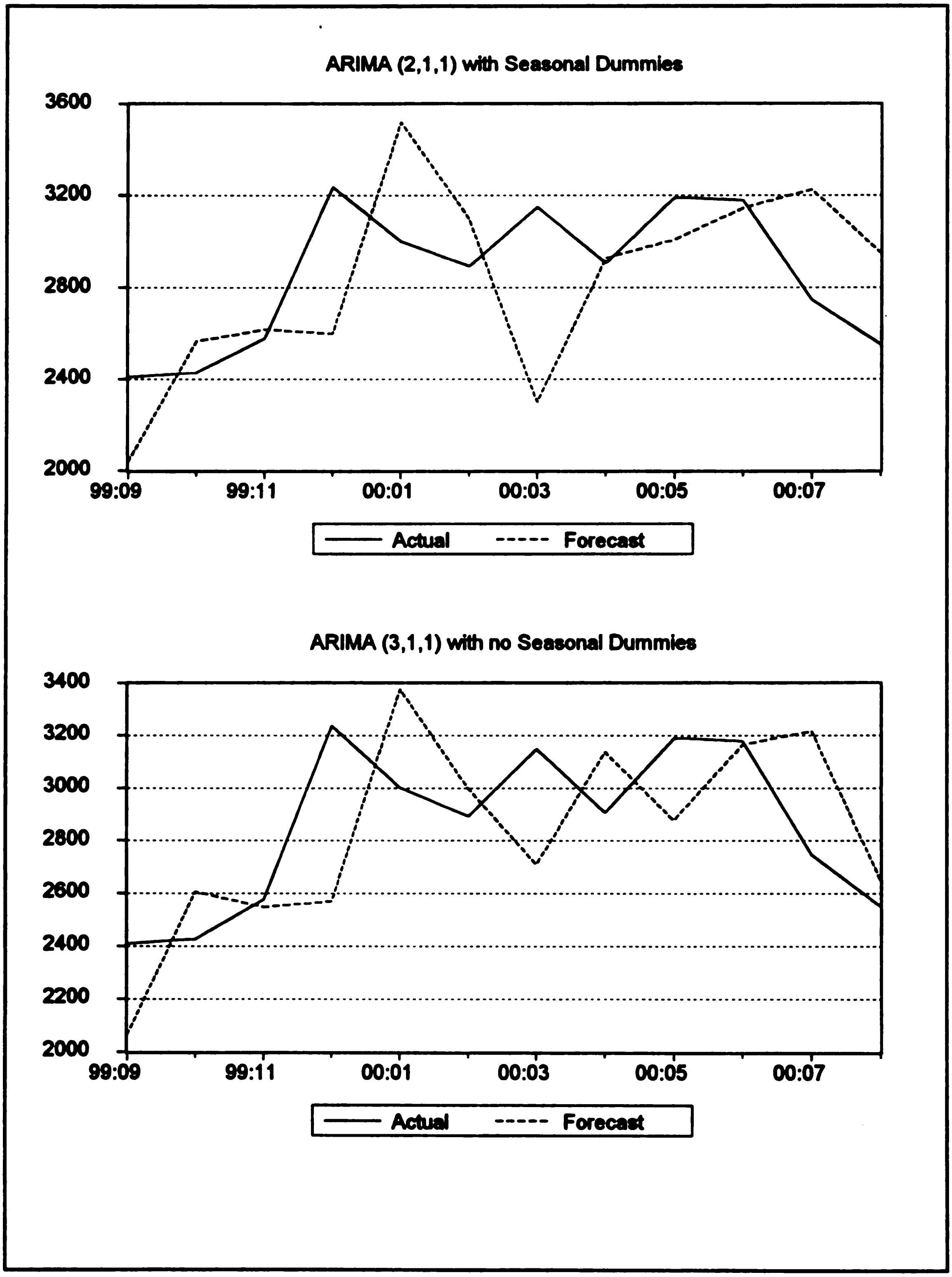
Similar to Nampula, as the unit root tests indicated that Maputo is an $I(1)$ series, an ARIMA model without monthly seasonal dummy variables is estimated and the forecasting performance of both ARIMA models is compared to that of the other forecasting models to be estimated later on in the chapter. When the traditional identification procedures are applied, an ARIMA (3,1,1) model without monthly seasonal dummies is identified for the Maputo prices.

The results of this ARIMA model are also presented in Table 5.3a, and the Q statistics for the ACF and PAC of both ARIMA models are presented in Table 5.3b. Table 5.3b shows that, in addition to the individual significance of the AR and MA components in both ARIMA models, these ARIMA models have no serial correlation. Indeed, the probability values associated to the Q statistics suggest a non-rejection of the null hypothesis of no autocorrelated innovations at the 5% level. Recall that the absence of probability values for the first 3(4) lags for the ARIMA model with(without) seasonal dummies indicates the inclusion of 2(3) AR and 1 MA component in the models. With these results, these ARIMA models can then be used to predict future prices for Maputo.

5.2.5 ARIMA Forecasting for Central/Southern Mozambique

The forecasting ability of the two ARIMA models for maize prices in Maputo, presented in Figure 5.5, suggest that, even though the two models exhibit similar pattern in predicting the future prices, the model with seasonal dummies appears to overweight the path of seasonality. For instance, in February 2000, the two models predict that the price in March will be lower, but while the ARIMA model without seasonal dummies forecasts a decrease from the actual price of 2,892 Mt/kg February to 2,711 Mt/kg in March, the ARIMA model with seasonal dummies predicts that the price in March will be 2,263 Mt/kg. Like in the Nampula case, a comparative examination of the price path between the model estimation and the forecasting sample indicate that the smooth seasonality observed over the model estimation sample is not repeated in the forecasting sample, thus the seasonal dummies appear to predict a seasonal pattern that is not observed in the out-of-sample forecasting sample.

Figure 5.5 Out-of-Sample Forecasts for Maputo Prices, ARIMA Models, Sep 1999 - Aug 2000



5.2.6 ARMA Models for Central/Southern Mozambique

Similar to the case of Nampula, two ARMA models are estimated for the Maputo price, in addition to the ARIMA models. Using the identification procedures and ensuring that the most parsimonious models should be chosen among the tentative ARMA models, a final ARMA (3,1) model with seasonal dummies and an ARMA (1,3) without seasonal dummies are identified and estimated. The results of both models are presented in Table 5.4a.

As indicated by the values of the t statistics, all the AR and MA components are significant at 5% or higher level of significance. Additionally, Table 5.4b indicates that there is no evidence of serial correlation in the innovations at 5% level of significance in both models. Hence these two ARMA models seem to be good models to predict the out-of-sample path of maize prices in Maputo. Similar to Nampula, there is no consistency in the ARMA and ARIMA models specifications in central/southern Mozambique.

With the above conclusions, the two ARMA models are used to perform out-of-sample forecasts for maize prices in Maputo. Figure 5.6 presents the forecasts obtained from the two ARMA models.

The out-of-sample forecasts from the two comparative ARMA models suggest that the model without seasonal dummies tends to predict better than the model with seasonal dummies. Examining the two models suggests that the seasonal dummies appear again to overestimate the forecast values. The procedures used to identify the ARMA models have been observed, and the

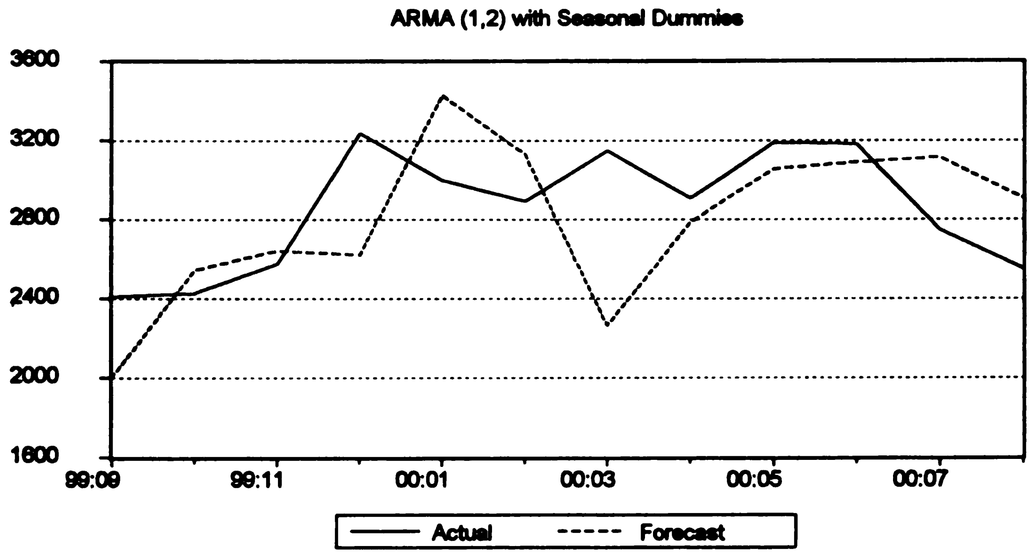
Table 5.4a ARMA Models for Maputo Maize Prices

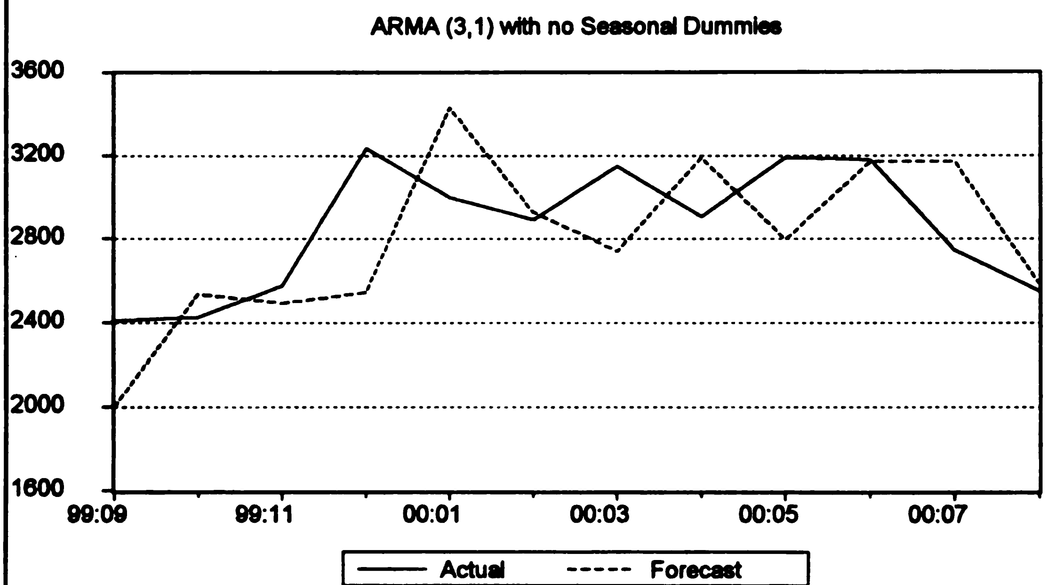
	ARMA (1,2) with Seasonal Dummies		ARMA (3,1) with no Seasonal Dummies	
	Coefficient	t statistic	Coefficient	t statistic
Constant	2612.90	(5.18)	2211.83	(5.53)
Jan	141.73	(1.47)		
Feb	145.89	(1.00)		
Mar	-309.33	(-1.64)		
Apr	-649.10	(-3.02)		
May	-632.06	(-2.75)		
Jun	-685.99	(-2.91)		
Jul	-700.32	(-3.01)		
Aug	-458.60	(-2.08)		
Sep	-345.12	(-1.78)		
Oct	-207.23	(-1.39)		
Nov	-102.69	(-1.05)		
AR(1)	0.91	(17.20)	0.52	(2.04)
AR(2)			0.72	(2.81)
AR(3)			-0.38	(-3.44)
MA(1)	0.27	(2.22)	0.71	(2.79)
MA(2)	0.36	(3.02)		

Table 5.4b Correlogram for the Maputo ARMA Models

	ARMA (1,2) with Seasonal Dummies				ARMA (3,1) with no Seasonal Dummies			
Lag	AC	PAC	Q Statistic	p-value	AC	PAC	Q Statistic	p-value
1	-0.03	-0.03	0.09		0.03	0.03	0.09	
2	0.02	0.02	0.12		0.02	0.02	0.13	
3	-0.10	-0.10	1.06		-0.01	-0.02	0.15	
4	0.07	0.06	1.46	(0.23)	-0.02	-0.02	0.18	
5	-0.18	-0.18	4.47	(0.11)	-0.16	-0.16	2.45	(0.12)
6	-0.04	-0.07	4.64	(0.20)	-0.04	-0.03	2.62	(0.27)
7	0.03	0.04	4.71	(0.32)	-0.06	-0.05	2.96	(0.40)
8	0.21	0.18	8.77	(0.12)	0.07	0.07	3.37	(0.50)
9	0.01	0.03	8.77	(0.19)	0.14	0.14	5.21	(0.39)
10	-0.01	-0.04	8.79	(0.27)	-0.08	-0.12	5.79	(0.45)
11	0.23	0.26	13.86	(0.09)	0.20	0.20	9.43	(0.22)
12	-0.13	-0.15	15.57	(0.08)	0.24	0.24	15.03	(0.06)
13	-0.14	-0.10	17.47	(0.07)	0.03	0.03	15.11	(0.09)
14	0.01	0.11	17.49	(0.09)	-0.11	-0.08	16.36	(0.09)
15	0.05	-0.03	17.73	(0.12)	0.06	0.07	16.77	(0.12)

Figure 5.6 Out-of-Sample Forecasts for Maputo Prices, ARMA Models, Sep 1999 - Aug 2000





preliminary data analysis suggested the evidence of seasonality in this price series. Thus, the poor forecasting performance of the ARMA model with seasonal dummies compared to the ARMA model without seasonal dummies is an indication that the pattern of seasonality had changed considerably between the model estimation period and the forecasting period.

5.3 Multivariate Models

The forecasting ability of the univariate models thus far estimated for the two regions is compared to that of multivariate models. Because the multivariate models include related price and production series to the Nampula and Maputo maize prices, it is hypothesized that these models will improve the forecasting performance of the univariate models.

Recall that this research has found that the Mocuba and Manica price series used to model price forecasting in, respectively, Nampula and Maputo, follow $I(1)$ processes. Therefore estimation of the multivariate models is preceded by cointegration tests. Specifically, cointegration tests are applied to Nampula and Mocuba price series in northern Mozambique, and similar test is performed for the Maputo and Manica prices for the central/southern Mozambique models. If the two series in each region are found to be cointegrated, VEC models are estimated.

5.3.1 Cointegration Models

Testing for cointegration is preceded by checking whether deterministic trends and intercepts should be included in the cointegration tests. EVIEWS, the econometric package used in the estimation of the models in this research, allows for five specifications of the deterministic trend. Table 5.5 has the five specifications and the results of checking for these two components. The Akaike Information criterion and the Schwarz Criterion are used in these tests. The bold numbers in the Table are the smallest values under each criterion.[10]

Table 5.5 Cointegrating Equations and Deterministic Trend Assumptions

Data Trend:	None	None	Linear	Linear	Quadratic
Cointegrating Equation	No Intercept No Trend	Intercept No Trend	Intercept No Trend	Intercept Trend	Intercept Trend
Northern Mozambique					
Akaike Information Criterion	21.727	21.667	**21.648**	21.665	21.687
Schwarz Criterion	22.057	**22.025**	22.034	22.078	22.128
Central/Southern Mozambique					
Akaike Information Criterion	22.227	22.194	**22.172**	22.194	22.212
Schwarz Criterion	**22.559**	22.646	22.646	22.739	22.739

10 The Akaike Inromation Criterion is based on the sum of squared residuals, and guides in selecting the number of coefficients or the length of a lag distribution in an equation. Smaller values of the AIC are preferable. The Schwarz criterion is an alternative to the AIC, with basically the same interpretation, but a larger penalty for extra coefficients (EVIEWS 3, 1998).

Two lags are incorporated in these tests. These results suggest that test for cointegration in maize prices between Nampula and Mocuba could be performed considering either that the data has a determinist trend or it has not. The Akaike Information Criterion suggests that the data has a deterministic trend component, which the Schwarz Criterion contradicts. Preliminary analysis of this data has indicated that both the Nampula and Mocuba prices have a slight positive linear trend but it is not significant in Mocuba, and in Nampula it is not significant at 5% but significant at 10% significant level. Thus, cointegration tests are run considering that the data has no significant deterministic trend. This assumption is also held in the estimation of the VEC model for northern Mozambique.

A similar conclusion is drawn for Maputo and Manica prices. While the Akaike Information Criterion suggests that the data has a deterministic trend component, the Schwarz Criterion suggests that it has neither a deterministic trend nor an intercept. Note that the values under the assumption of intercept and no linear trend and intercept and linear trend are very similar for each criterion, and the Schwarz Criterion imposes larger penalty for additional coefficients compared to the Akaike Information Criterion. Thus while the Akaike Information Criterion does not penalize the linear trend, the Schwarz Criterion does, which does not mean any contradiction between the two criteria, and both cases could be considered as having an intercept but with little evidence of a linear trend.

With these conclusions, cointegration tests are performed for the two pairs of prices. In these tests, the null that the two price series are not cointegrated is tested against the alternative that they are cointegrated. Table 5.6 has the results of these tests.

Table 5.6 Cointegration Tests for Mozambique Maize Prices

Cointegration Equation	Likelihood Ratio	5%Critical Value	1% Critical Value
Nampula - Mocuba	38.3598	15.41**	20.04**
Maputo - Manica	33.56111	15.41*	20.04*

*(**) denotes rejection of the hypothesis at 5%(1%) significance level

The Johansen's cointegration test indicates that the Likelihood Ratio for the cointegration equation under the assumption of no linear trend assumptions is 38.36 for northern Mozambique and 33.56 for central/southern Mozambique. These Likelihood values are higher than the adjusted critical values of 15.41 at 5% level or 20.04 at 1% significance level. Therefore it can be concluded that there is evidence of cointegration between Nampula and Mocuba maize prices, and between Maputo and Manica maize prices.

The Engle and Granger's two-step cointegration test are run for the two pairs of prices, without a linear trend but with the inclusion of monthly seasonal dummy variables. The conclusion drawn from the Johansen tests is confirmed. The results indicate that, at a critical value of -3.34 at 5% level, we reject the null of no cointegration in favor of the alternative that the prices are cointegrated in both cases. The cointegration coefficient for Nampula-Mocuba is -4.62, and the coefficient for Maputo-Manica is -3.84.

5.3.2 VEC Specifications and Forecasting

The Granger causality tests just carried out indicated that there is evidence that Mocuba and Manica Granger cause Nampula and Maputo prices respectively. Additionally, the cointegration tests indicated that there is evidence of a long term linear relationship between the exporting and the importing market in each region. This section uses these results to specify and estimate multivariate models for the two regions of Mozambique.

5.3.2.1 Northern Mozambique

The Granger causality and cointegration tests for northern Mozambique indicated that Mocuba prices are expected to help explain variations in the Nampula prices, and that the price series have a significant linear relationship in the long run. Mocuba prices are modeled as the second endogenous variable in the VEC model for maize prices in northern Mozambique in addition to the Nampula maize prices. Actual production in Mozambique and Malawi are two exogenous variables in this model, in addition to monthly dummy variables aimed to control for the already identified significant seasonality in the two price series. Two lags were included in the cointegrating equations based on EVIEWS' suggestion, and are also included in the VEC model for northern Mozambique. This VEC model is calculated as follows:

$$\Delta p_{1t} = \alpha_1(p_{1,t-1} - 1.44p_{2,t-1} + 293.12) + \gamma_{10} + \gamma_{11}\Delta p_{1,t-1} + \gamma_{12}\Delta p_{1,t-2} + \gamma_{13}\Delta p_{2,t-1} + \gamma_{14}\Delta p_{2,t-2} + \beta_1 x_t + \varepsilon_{1t} \quad (32)$$

$$\Delta p_{2t} = \alpha_2(p_{1,t-1} - 1.44p_{2,t-1} + 293.12) + \gamma_{20} + \gamma_{21}\Delta p_{1,t-1} + \gamma_{22}\Delta p_{1,t-2} + \gamma_{23}\Delta p_{2,t-1} + \gamma_{24}\Delta p_{2,t-2} + \beta_2 x_t + \varepsilon_{2t} \quad (33)$$

where p_1 is maize price in Nampula, p_2 is maize price in Mocuba, α_1 and α_2 are coefficients on the error correction terms, γ_{11}, through γ_{24} are coefficients to be estimated on the autoregressive terms, x is a set of deterministic or exogenous variables, which include Mozambique and Malawi production data and 11 monthly seasonal dummies, β_1 and β_2 are coefficients on the exogenous variables, and ε_{1t} and ε_{2t} are the innovations, assumed to be serially uncorrelated within an equation, but correlated across equations.[11]

Lower or higher orders of autoregressive lags were tried, and they did not improve the significance of the individual coefficient nor the forecasting ability of the VEC (2,0) model. Thus, a model with two autoregressive lags was identified as the final VEC model. Table 5.7 presents the results of the VEC model for maize prices in northern Mozambique.

Alternative to the VEC (2,0) with seasonal dummies just described, another VEC, also with two autoregressive lags but without seasonal dummies is estimated and the forecasting ability of the two models is compared to all the models for the Nampula prices. Table 5.8 indicates that all the parameters in the cointegration equation are significant at 5% level in both VEC models. However, three of the four lags of both ΔP_1 and ΔP_2 do not improve much the significance of the two models. Among the maize production variables, Q_1 improves only in the model without seasonal dummies. Recall that the

11 The selection of the lag order for VAR/VEC models is somehow arbitrary. Too small lag lengths may not ensure white noise residuals, and too large lag lengths might result in imprecise estimates.

production variables are annual figures, and they are repeated for the 12 months of each marketing year, thus they have little monthly variability.

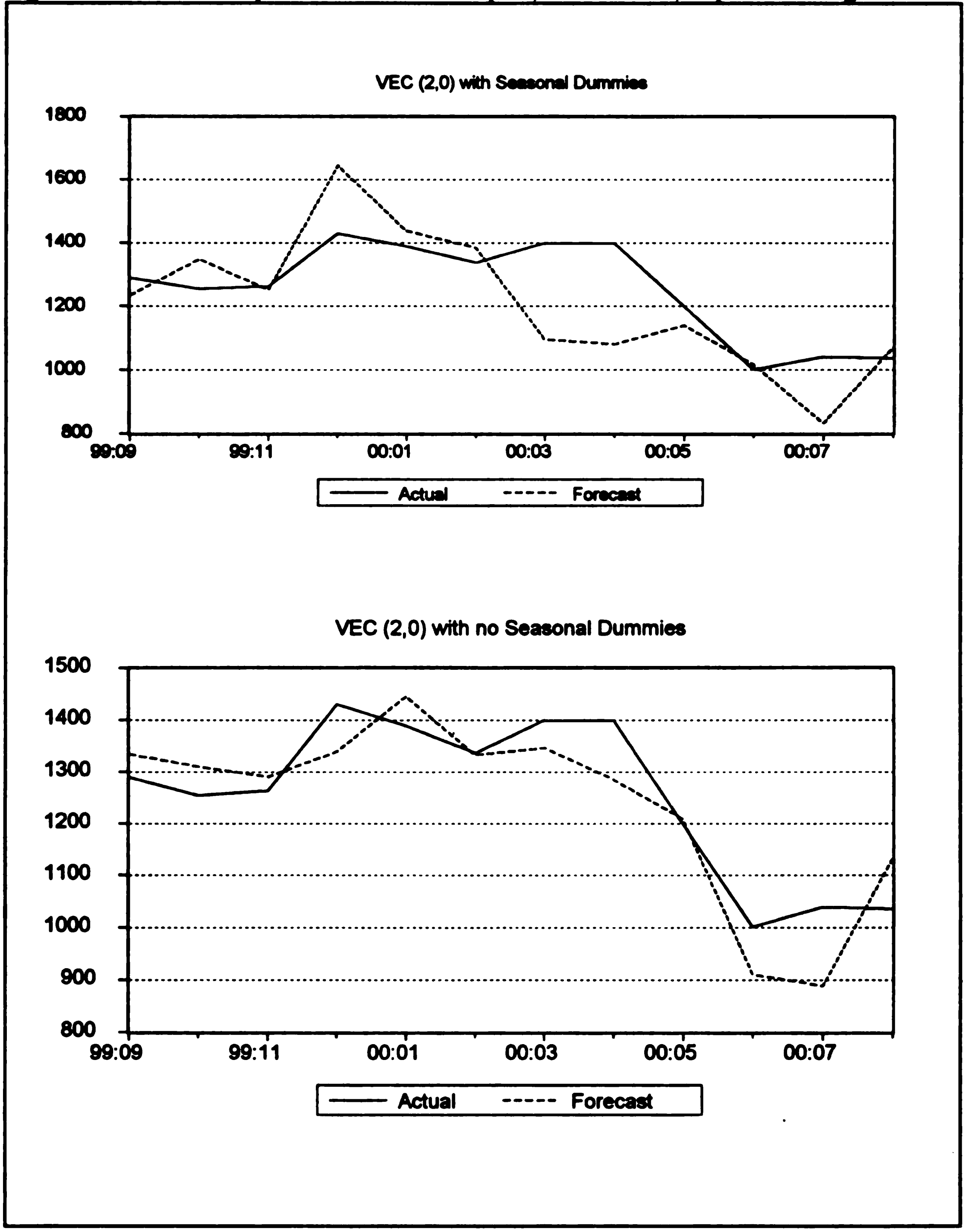
The two VEC (2,0) models are next used to forecast Nampula prices outside the model estimation sample. Figure 5.7 has the graphics of the forecasts of the two models.

Figure 5.7 suggests that, like in the case of the univariate models, the two models do not differ much in their power of predicting the path of the prices outside the estimation sample, but it seems that the model without seasonal dummies predicts with better accuracy the changes in prices when high increases or decreases are expected to happen, due to the different seasonal path in the prices between the model estimation period and the forecasting period.

Table 5.7 VEC (2,0) Model for Maize prices in Northern Mozambique

Variable	With Seasonal Dummies				With no Seasonal Dummies			
	ΔP_1		ΔP_2		ΔP_1		ΔP_2	
	Coeff	t stat	Coeff	t stat	Coeff	t stat	Coeff	t stat
Cointegration Equation	-0.23	(-2.21)	0.27	(2.16)	-0.02	(-0.42)	0.32	(4.73)
$\Delta P_{1,t-1}$	0.14	(1.08)	0.03	(0.18)	0.04	(0.28)	0.12	(0.75)
$\Delta P_{1,t-2}$	-0.20	(-1.614)	0.01	(0.09)	-0.14	(-1.18)	0.21	(1.43)
$\Delta P_{2,t-1}$	0.08	(0.54)	0.34	(1.94)	0.43	(4.21)	0.66	(5.27)
$\Delta P_{2,t-2}$	0.05	(0.36)	0.09	(0.54)	0.14	(1.18)	0.07	(0.49)
Constant	559.41	(3.38)	145.03	(0.72)	99.36	(0.69)	-426.59	(-2.43)
Q_1 (Mozambique production)	-0.12	(-0.97)	0.09	(0.58)	0.01	(0.08)	0.47	(2.60)
Q_2 (Malawi production)	-0.08	(-1.00)	-0.01	(-0.08)	-0.06	(-0.745)	0.03	(0.25)
Jan	-338.82	(-2.28)	-6.99	(-0.04)				
Feb	-283.27	(-2.02)	-52.80	(-0.31)				
Mar	-594.46	(-4.27)	-662.08	(-3.93)				
Apr	-595.33	(-3.76)	-724.12	(-3.78)				
May	-428.92	(-2.42)	-240.59	(-1.12)				
Jun	-228.84	(-1.33)	-187.12	(-0.90)				
Jul	-322.44	(-1.96)	-176.80	(-0.89)				
Aug	-281.25	(-1.87)	-251.75	(-1.38)				
Sep	-297.74	(-1.98)	-64.11	(-0.35)				
Oct	-190.86	(-1.31)	-1.98	(-0.01)				
Nov	-304.87	(-2.11)	9.00	(0.05)				
Log likelihood	-535.90		-550.93		-550.97		-566.71	
Akaike AIC	11.21		11.59		11.31		11.72	
Schwarz SC	11.78		12.16		11.55		11.95	
Model Statistics								
Log Likelihood	-1073.52				-1100.81			
Akaike Information Criteria	22.51				22.65			
Schwarz Criteria	23.71				23.19			

Figure 5.7 Out-of-Sample Forecasts for Nampula, VEC Model, Sep 1999 - Aug 2000



5.3.2.2 Central/Southern Mozambique

Like in the case of northern Mozambique, two autoregressive lags are used to estimate the tentative VEC model for central/southern Mozambique. Maputo and Manica price series, which have already been used in the Granger causality tests and in the cointegrating equations, are the two endogenous variables in the VEC model for central/southern Mozambique. In addition to the two variables, data on volume of maize production in Mozambique and South Africa are used as exogenous variables. Monthly dummy variables are also included as exogenous variables in the VEC (2,0) model, which has the following representation:

$$\Delta p_{1t} = \alpha_1(p_{1,t-1} - 1.09p_{2,t-1} - 862.99) + \gamma_{10} + \gamma_{11}\Delta p_{1,t-1} + \gamma_{12}\Delta p_{1,t-2} + \gamma_{13}\Delta p_{2,t-1} + \gamma_{14}\Delta p_{2,t-2} + \beta_{11}x_t + \varepsilon_{1t} \quad \textbf{(34)}$$

$$\Delta p_{2t} = \alpha_2(p_{1,t-1} - 1.09p_{2,t-1} - 862.99) + \gamma_{20} + \gamma_{21}\Delta p_{1,t-1} + \gamma_{22}\Delta p_{1,t-2} + \gamma_{23}\Delta p_{2,t-1} + \gamma_{24}\Delta p_{2,t-2} + \beta_{21}x_t + \varepsilon_{2t} \quad \textbf{(35)}$$

where p_1 is maize price in Maputo, p_2 is maize price in Manica, α_1 and α_2 are coefficients on the error correction terms, γ_{11}, through γ_{24} are coefficients to be estimated on the autoregressive terms, x is a set of deterministic or exogenous variables, which include Mozambique and South Africa production data and 11 monthly seasonal dummies, β are coefficients on the exogenous variables, and ε_{1t} and ε_{2t} are the innovations, assumed to be serially uncorrelated within equation, but correlated across equations. Table 5.8 presents the results of the VEC (2,0) model for maize prices in northern Mozambique.

Table 5.8 shows that all the parameters in the cointegrating equation of the two VEC (2,0) models are significant at 5% significance level, and most times the lags of the differences of the price series are not significant at 5% level. However, all of them are significant at the 10% level of significance. The exogenous production variables are generally more significant here than in northern Mozambique.

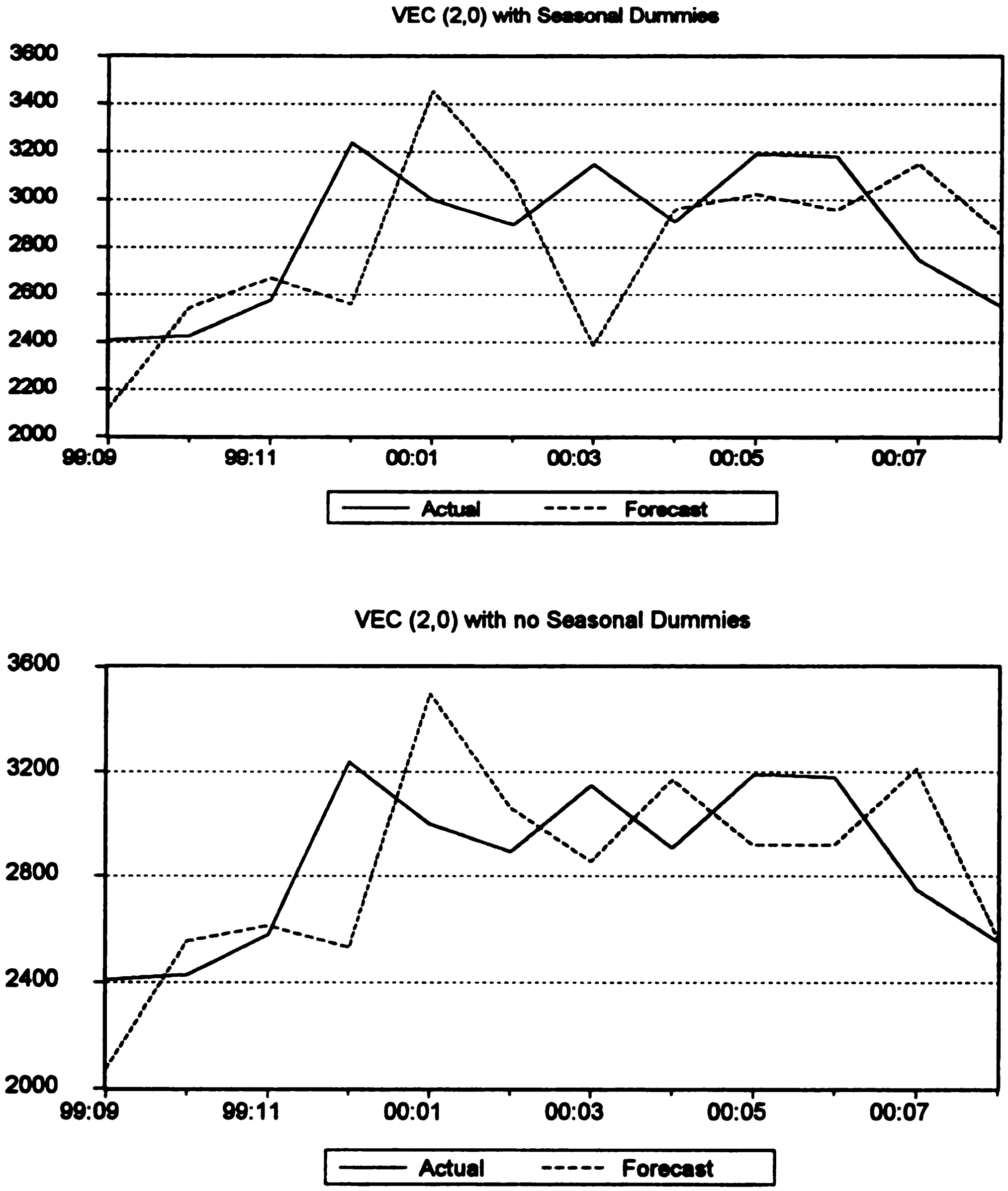
Inclusion of higher orders of autoregressive terms in both VEC models did not improve the significance of the parameters of these models, neither did it increase the forecasting power of the models. Hence, the two VEC (2,0) models are used to forecast prices maize prices in Maputo outside the model estimation sample. Figure 5.8 has the graphs with the forecasts from the two models.

The main conclusion drawn from these graphs is that, similar to the univariate models, the multivariate VEC models for Maputo prices indicate that the monthly seasonal dummies appear to overweight the monthly changes of the prices over the out-pf-sample forecasting period when considerable changes are expected to occur. Overall there is not much difference in the forecasting performance of the two VEC models.

Table 5.8 VEC (2,0) Model for Maize prices in Central/Southern Mozambique

Variable	With Seasonal Dummies				With no Seasonal Dummies			
	ΔP_1		ΔP_2		ΔP_1		ΔP_2	
	Coeff	t stat	Coeff	t stat	Coeff	t stat	Coeff	t stat
Cointegration Equation	-0.14	(-2.15)	0.22	(2.75)	-0.16	(-2.04)	0.25	(2.68)
$\Delta P_{1,t-1}$	0.14	(1.06)	0.29	(1.89)	0.21	(1.61)	0.31	(2.04)
$\Delta P_{1,t-2}$	0.16	(1.31)	-0.06	(-0.39)	0.06	(0.50)	-0.10	(-0.66)
$\Delta P_{2,t-1}$	0.20	(1.98)	0.43	(3.60)	0.29	(2.81)	0.51	(4.14)
$\Delta P_{2,t-2}$	-0.17	(-1.61)	-0.21	(-1.66)	-0.25	(-2.25)	-0.20	(-1.57)
Constant	-64.51	(-0.34)	685.62	(3.03)	-224.69	(-1.33)	433.83	(2.16)
Q_1 (Mozambique production)	0.06	(0.56)	-0.26	(-1.92)	0.18	(1.38)	-0.23	(-1.46)
Q_2 (South Africa production)	0.01	(0.49)	-0.04	(-2.55)	0.01	(0.70)	-0.03	(-1.88)
Jan	18.15	(0.13)	19.51	(0.11)				
Feb	-88.44	(-0.64)	-229.29	(-1.39)				
Mar	-501.55	(-3.59)	-674.85	(-4.05)				
Apr	-184.51	(-1.12)	-185.70	(-0.94)				
May	84.73	(0.51)	-224.47	(-1.14)				
Jun	-40.51	(-0.27)	-249.30	(-1.39)				
Jul	-41.35	(-0.30)	-103.06	(-0.62)				
Aug	205.28	(1.48)	-73.87	(-0.44)				
Sep	77.29	(0.53)	-50.59	(-0.29)				
Oct	34.37	(0.24)	-155.16	(-0.91)				
Nov	45.03	(0.31)	26.25	(0.15)				
Log likelihood	-536.00		-550.00		-553.12		-566.64	
Akaike AIC	11.21		11.57		11.37		11.71	
Schwarz SC	11.78		12.14		11.61		11.95	
Model Statistics								
Log Likelihood	-1082.17				-1108.56			
Akaike Information Criteria	22.73				22.84			
Schwarz Criteria	23.93				23.38			

Figure 5.8 Out-of-Sample Forecasts for Maputo Prices, VEC Models, Sep 1999 - Aug 2000



5.3.3 VAR Specifications and Forecasting

Similar to the univariate models, even though the preliminary data searching indicated that the VEC models are the most appropriate multivariate models for both northern and central/southern Mozambique, VAR models are also specified and estimated for both regions, and the forecasting power of both the VEC and the VAR models is compared to that of the univariate ARIMA and ARMA models. The remainder of this chapter presents the VAR specifications and estimations, and the forecasting. First, VAR models for northern Mozambique are estimated, followed by VAR models for central/southern Mozambique.

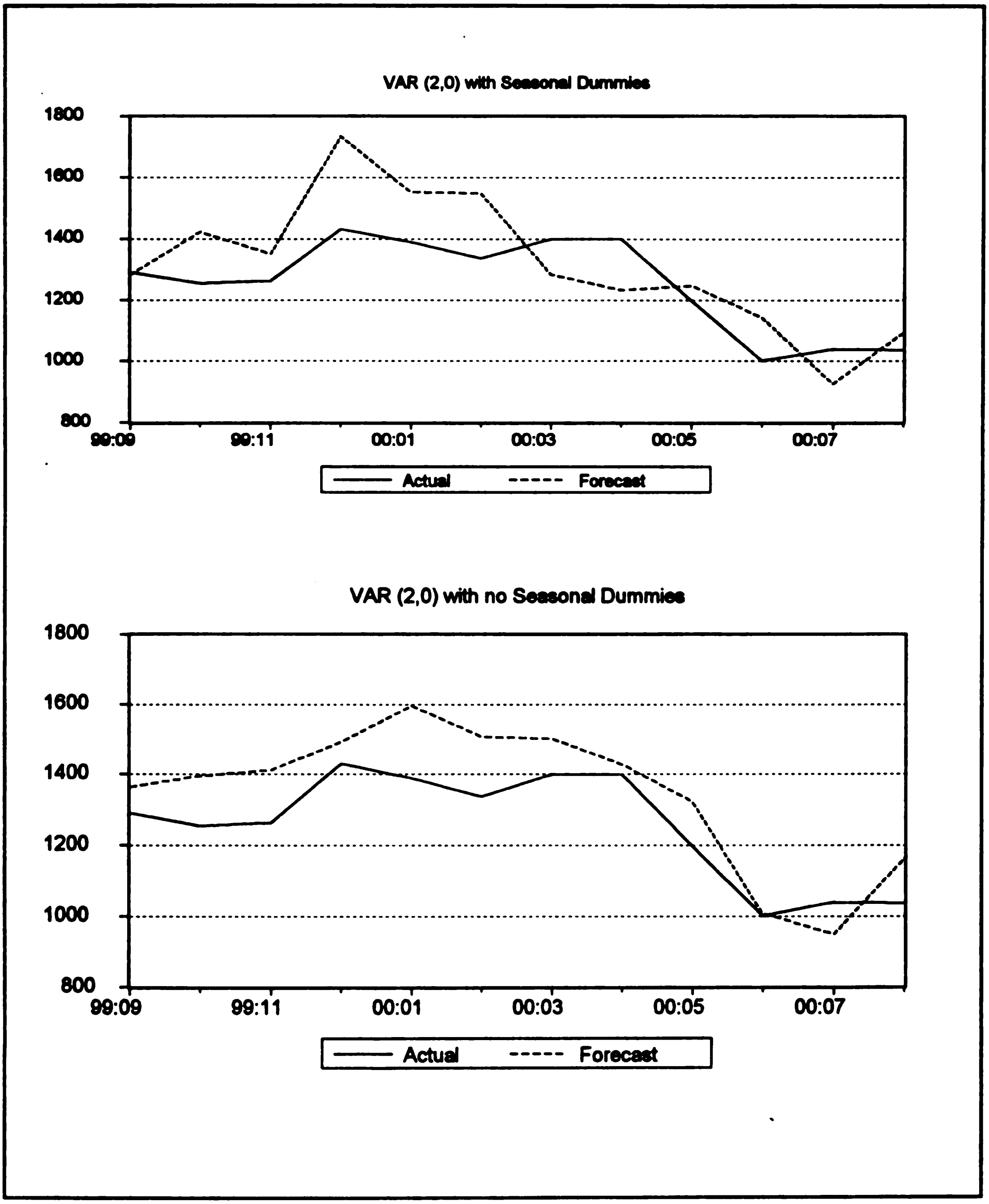
5.3.3.1 Northern Mozambique

With two autoregressive lags, two VAR models are estimated for maize prices in northern Mozambique. Both models have maize prices in Nampula and Mocuba as endogenous variables, and the data on maize production in Mozambique and Malawi as exogenous variables. The difference between the two models is that one of them has monthly seasonal dummy variables while the other does not. Table 5.9 presents the results of both VAR models, and the forecasts from both are presented in Figure 5.9, and the two models' out-of-sample forecasts are shown in Figure 5.9.

Table 5.9 VAR (2,0) Model for Maize prices in Northern Mozambique

Variable	With Seasonal Dummies				With no Seasonal Dummies			
	P_1		P_2		P_1		P_2	
	Coeff	t stat	Coeff	t stat	Coeff	t stat	Coeff	t stat
$P_{1,t-1}$	0.82	(5.87)	0.22	(1.31)	0.76	(5.61)	0.29	(1.59)
$P_{1,t-2}$	-0.25	(-0.12)	-0.16	(-1.14)	-0.21	(-1.99)	-0.26	(-1.82)
$P_{2,t-1}$	0.36	(3.10)	0.92	(6.58)	0.49	(5.42)	1.09	(8.73)
$P_{2,t-2}$	0.00	(-0.02)	-0.25	(-1.59)	-0.15	(-1.41)	-0.45	(-3.12)
Constant	375.52	(2.57)	221.08	(1.25)	50.86	(0.47)	-24.56	(-0.16)
Q_1 (Mozambique production)	0.25	(1.48)	0.42	(2.05)	0.34	(2.43)	0.55	(2.90)
Q_2 (Malawi production)	-0.06	(-0.94)	-0.04	(-0.45)	-0.06	(-0.89)	-0.03	(-0.32)
Jan	-302.23	(-2.24)	44.91	(0.27)				
Feb	-207.19	(-1.54)	50.56	(0.31)				
Mar	-439.51	(-3.23)	-553.37	(-3.35)				
Apr	-489.38	(-3.22)	-639.88	(-3.47)				
May	-402.20	(-2.51)	-268.67	(-1.38)				
Jun	-193.18	(-1.26)	-252.07	(-1.35)				
Jul	-287.97	(-1.98)	-237.51	(-1.34)				
Aug	-308.09	(-2.20)	-303.64	(-1.78)				
Sep	-326.74	(-2.32)	-117.21	(-0.69)				
Oct	-218.28	(-1.59)	-54.45	(-0.33)				
Nov	-305.00	(-2.27)	-20.62	(-0.13)				
Log likelihood	-538.65		-554.13		-548.36		-573.33	
Akaike AIC	11.08		11.47		11.05		11.67	
Schwarz SC	11.61		12.00		11.25		11.88	
Model Statistics								
Log Likelihood	-1083.44				-1109.82			
Akaike Information Criteria	22.31				22.42			
Schwarz Criteria	23.38				22.84			

Figure 5.9 Out-of-Sample Forecasts for Nampula Prices, VAR Models, Sep 1999 - Aug 2000



The main finding is that the two VAR models appear to be good representations of the path of the data prices for northern Mozambique. Even though the second lag is not significant at 5% level, it appears to increase slightly the forecasting properties of these VAR models, Mozambique production is significant in three of the four equations.

5.3.3.2 Central/Southern Mozambique

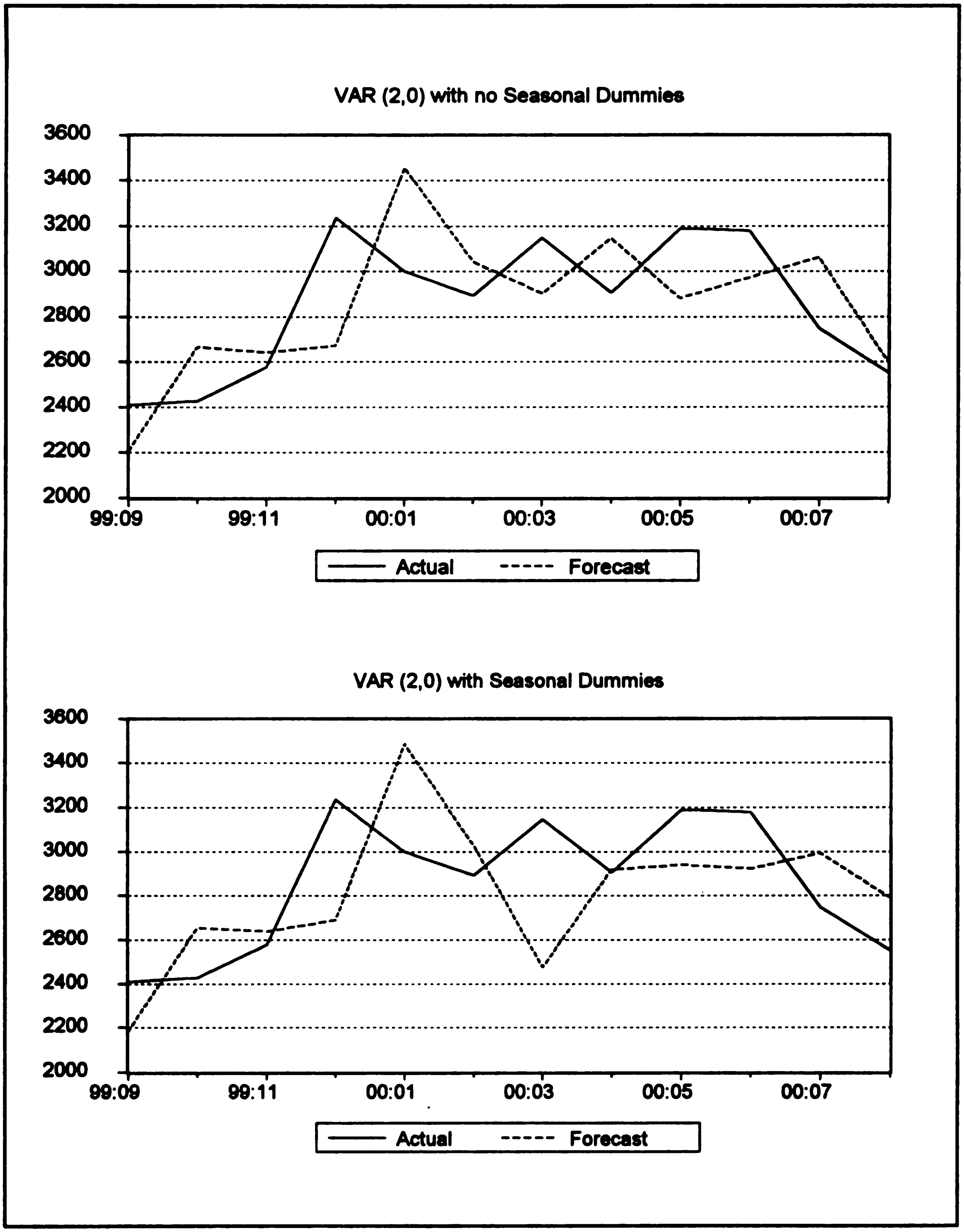
Two VAR models are estimated for maize prices in central/southern Mozambique. Both models have two autoregressive lags each, and the difference between the two models is that one of them has monthly seasonal dummy variables while the other does not. Both models have maize prices in Maputo and Manica as endogenous variables, and the data on maize production in Mozambique and Malawi as exogenous variables. Monthly seasonal dummies are additional exogenous variables in the model that controls for seasonality. Table 5.10 has the results of both models, and the forecasts from both are presented in Figure 5.10.

The autoregressive components of Maputo and Manica prices in the two alternative VAR models for central/southern Mozambique are significant at 5% level. Mozambique production is significant in both Maputo equations, while South Africa production is significant in both Manica equations. Because Maputo is closer to South Africa and is the market which effectively imports maize from South Africa compared to Manica, it should be expected that South African production is more significant in explaining Maputo prices than Manica prices. However, this did not happen, and the reason seems to be in the different sizes of Maputo and Manica markets, reflected in the extent to which the two markets are affected by regional droughts. While Maputo can import maize, Manica is so small that does not attract importers. This happened in 1995, when Maputo prices did not reflect the drought conditions while South Africa production and Manica prices did. Also, even if Maputo imports from South Africa, these imports are so small compared to South African production that they do not affect South African market conditions.

Table 5.10 VAR (2,0) Model for Maize prices in Central/Southern Mozambique

Variable	With Seasonal Dummies				With no Seasonal Dummies			
	ΔP_1		ΔP_2		ΔP_1		ΔP_2	
	Coeff	t stat	Coeff	t stat	Coeff	t stat	Coeff	t stat
$\Delta P_{1,t-1}$	0.88	(7.13)	0.39	(2.82)	1.00	(8.06)	0.45	(3.15)
$\Delta P_{1,t-2}$	-0.09	(-0.79)	-0.28	(-2.08)	-0.22	(-1.77)	-0.33	(-2.38)
$\Delta P_{2,t-1}$	0.27	(2.79)	1.06	(9.78)	0.32	(3.31)	1.11	(10.06)
$\Delta P_{2,t-2}$	-0.15	(-1.46)	-0.35	(-3.12)	-0.25	(-2.52)	-0.43	(-3.86)
Constant	303.86	(1.71)	578.99	(2.91)	139.83	(0.93)	330.17	(1.93)
Q_1	0.26	(1.81)	0.08	(0.51)	0.37	(2.45)	0.19	(1.06)
Q_2	0.00	(-0.41)	-0.04	(-2.87)	-0.01	(-0.54)	-0.04	(-2.68)
Jan	-71.37	(-0.53)	-85.95	(-0.57)				
Feb	-170.24	(-1.26)	-197.40	(-1.31)				
Mar	-488.42	(-3.57)	-603.60	(-3.94)				
Apr	-357.24	(-2.35)	-242.11	(-1.42)				
May	-78.08	(-0.53)	-283.73	(-1.72)				
Jun	-223.91	(-1.55)	-279.90	(-1.730)				
Jul	-204.43	(-1.43)	-196.43	(-1.23)				
Aug	20.52	(0.14)	-176.79	(-1.11)				
Sep	-41.98	(-0.30)	-136.85	(-0.86)				
Oct	-80.97	(-0.58)	-206.70	(-1.33)				
Nov	-88.79	(-0.64)	-98.95	(-0.64)				
Log likelihood	-610.51		-620.52		-559.01		-569.80	
Akaike AIC	11.29		11.51		11.31		11.58	
Schwarz SC	11.79		12.01		11.52		11.79	
Model Statistics								
Log Likelihood	-1227.98				-1121.21			
Akaike Information Criteria	22.73				22.70			
Schwarz Criteria	23.73				23.12			

Figure 5.10 Out-of-Sample Forecasts for Maputo Prices, VAR Models, Sep 1999 - Aug 2000.



Finally, similar to northern Mozambique, the two VAR models seem to be good representations of the path of the data prices for southern Mozambique. These two VAR (2,0) models are then used to perform twelve one-step-ahead forecast for Maputo prices, re-estimated at every step, and their forecasting ability is used in the forecasting evaluation along with the other multivariate and univariate models. The results of the two VAR models are presented in Figure 5.10.

5.4 Conclusions

The main conclusion of this chapter is that, in each class of models (ARMA, ARIMA, VAR, and VEC), the models with seasonal dummies overestimate the seasonal movements. The reason for this is that seasonal price movements were much less pronounced during the forecasting period than they typically were during the estimation period.

Second, the multivariate models seem to perform slightly better than the univariate models. The quality of the data, especially the production data for Mozambique and Malawi, could be the reason that the multivariate approach did not deliver greater improvements in performance.

Chapter 6

FORECAST EVALUATION

6.1 Introduction

This chapter evaluates the competitor forecasting models' ability to perform out-of-sample forecasts. Along with the univariate ARMA and ARIMA models and the multivariate VAR and VEC models, random walk models' forecasts evaluated. This chapter is organized as follows. Section 6.2 advances a discussion of basic characteristics of the prices that are expected to help explain the results of the forecasting models. Section 6.3 presents and discusses statistical evaluation criteria, section 6.4 addresses a discussion on the economic evaluation criteria for the competitor forecasting models, and section 6.5 presents the conclusions.

6.2 Preliminary Data Search

Comparing the ability of alternative models in accurately forecasting prices requires understanding the data itself. Performing statistical evaluation of different time series models without understanding the behavior of the prices over time and the reasons of the given behavior may lead to misleading conclusions. A description of several properties of the price time series used in this research has been presented in chapter 4. This chapter, however, revisits some of the issues discussed in chapter 4, to allow better understanding of the findings on the statistical and economic evaluation criteria for the competitor forecasting models.

Figure 6.1 Maize Retail Prices in Nampula, November 1992 - August 2000

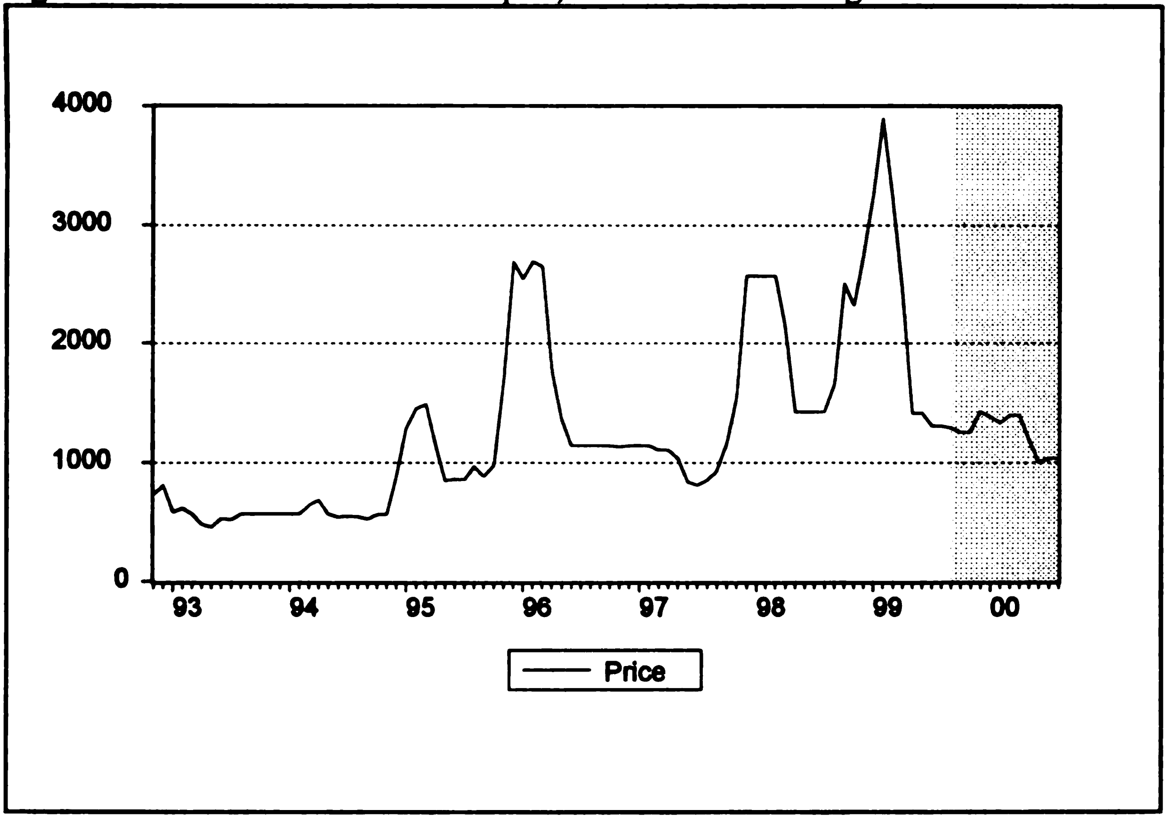
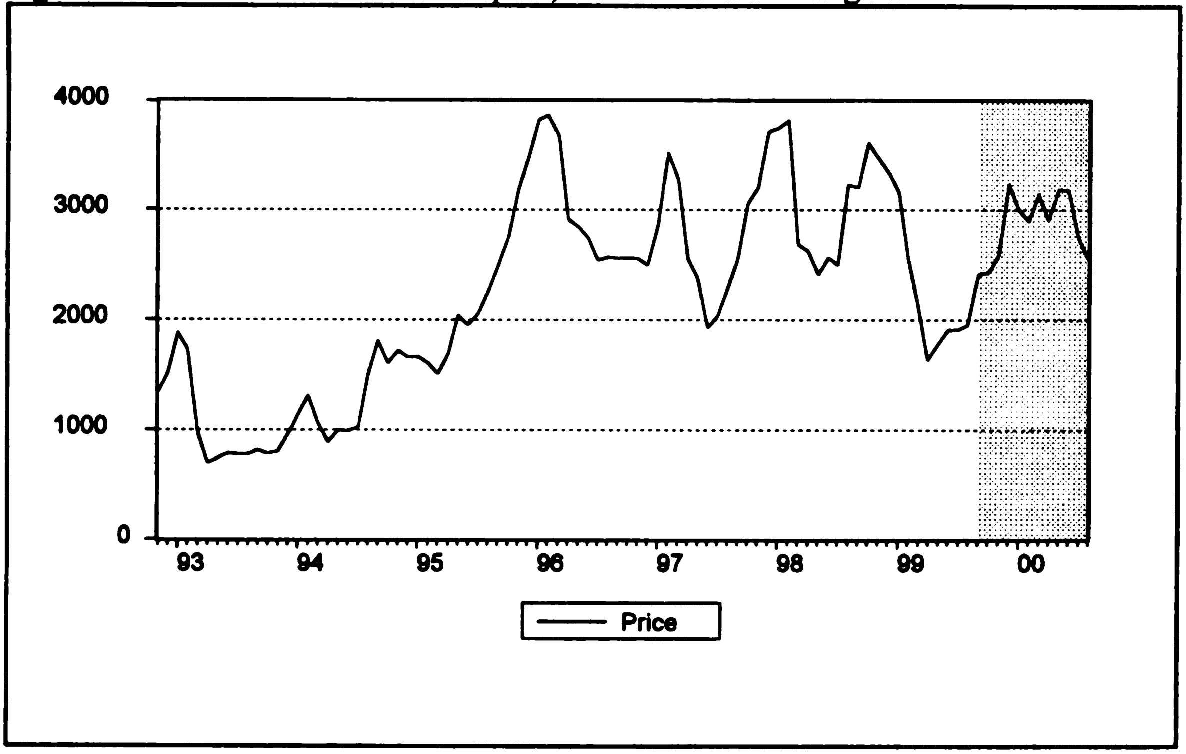


Figure 6.1 shows nominal retail prices for maize in Nampula. The shadowed area shows the forecasting period. This graph shows that prices in Nampula were stable until late 1994. Between 1995 and mid 1999 they were unstable, with periods of some stability and others of strong spikes. This behavior is not observed during the forecasting period of mid 1999 through August 2000.

Figure 6.2 shows the Maputo case. This figures indicates that in Maputo, similar to Nampula, maize prices tended to remain somewhat stable at low levels until early 1995. Prices in Maputo, however, were not as stable as those in Nampula in this period.

Figure 6.2 Maize Retail Prices in Maputo, November 1992 - August 2000



Like in Nampula, high spikes are observed in Maputo maize prices between 1995 and 1999, even though they were comparatively less accentuated.

Also, the behavior of these prices in Maputo over the sample estimation period of November 1992 through August 1999 is different from that of the out-of-forecast period of September 1999 through August 2000. In fact, prices tended to follow the normal seasonality between 1999 and 2000, but in 2000 (i) seasonal peak occurred in October, was very early compared to January, when it usually happens, (ii) seasonal decline started in June, instead of the normal decline in April, and (iii) there were monthly or bimonthly variations in the prices between October 1999 and March 2000, not observed in the same months over the model estimation period. This could be expected to be a source of "out-of-normal" forecasting errors, especially for the univariate models.

6.3 Statistical Evaluation

Having in mind the price characteristics just outlined, this section presents and discusses the comparative forecasting models' performance based on the results from statistical criteria. As suggested by theory, the different forecasting models employed in this research are expected to give different forecasts. Three statistical criteria are used, namely the root mean squared error(RMSE), the mean absolute percentage error (MAPE), and the turning point error (TPE). While the RMSE depends on the scale of the variable being forecast, the MAPE and the TPE do not. The performance of the different models is compared under the rule that the lower the value of the RMSE, the MAPE or the TPE, the better the model's ability to forecast the time series.

6.3.1 Nampula Prices

The fact that seasonal price behavior changed over the out-of-sample forecast period suggests we would expect a poor forecasting performance of the models used in this research, especially the univariate models. The ability of the alternative models' in forecasting maize prices in Nampula is presented in Table 6.1.

Table 6.1 Statistical Evaluation of the Price Forecasting Competitor Models for Nampula

Model		Statistical Criterion		
		RMSE	MAPE	TPE (%)
ARMA Models	ARMA (2,0) with no seasonal dummies	94.83	0.06	83.00
	ARMA (2,1) with monthly dummies	157.37	0.11	83.00
ARIMA Models	ARIMA (1,1,0) with no seasonal dummies	195.10	**0.04**	83.00
	ARIMA (1,1,0) with monthly dummies	170.30	0.06	50.00
VAR Models	VAR with no seasonal dummies	120.35	**0.04**	83.00
	VAR with seasonal dummies	152.33	0.10	**33.00**
VEC Models	VEC with no seasonal dummies	**78.339**	0.06	50.00
	VEC with monthly dummies	159.30	0.09	67.00
Random Walk Model		99.40	0.05	100.00

The results in Table 6.1 suggest three things. First, the three criteria do not lead to a similar conclusions regarding the "best" model. Overall, the multivariate models do a somewhat better job than the univariate model do. This should be expected as the univariate models are based solely on the variable to be forecast, thus they should be expected to have limitations in predicting their own future behavior especially when there is much unpredictability in the fluctuations of the time series. The multivariate models, on the other hand, involve the production variables in addition to the prices in Mocuba , which are thought to help explain the behavior of the Nampula price, and hence to control for most of the fluctuations in the Nampula prices.

Second, among the univariate models, the ARMA models seem to do a better job than the ARIMA models, while there is not a substantial difference between the VAR and the VEC models. In fact, it is often true that the ARMA and VAR models, which are based on price levels - used, in this case, with no regard on the nonstationarity of the price series in levels - forecast better than the ARIMA and VEC models, estimated with the first difference-stationary prices, when the time series are not long enough.

Finally, among the ARMA models, the VAR models and VEC models, the models with no seasonal dummies appear to do a better job in forecasting compared to the models with seasonal dummies. This is not surprising as Figure 6.1 showed that the extraordinarily high peaks observed several times over the model estimation sample are not repeated in the forecast sample, thus the seasonal variables, while controlling for these peaks in the model, predict similar behavior over the forecast period.

6.3.2 Maputo Prices

As in Nampula, examining the behavior of the time series between the model estimation period and the forecast period shows that the behavior in the two sub-samples is not the same, thus the models could have problems predicting the path of future prices. Table 6.2 has the results of the statistical criteria used to compare the ability of the alternative forecasting models.

Table 6.2 Statistical Evaluation of the Price Forecasting Competitor Models for Maputo

Model		Statistical Criterion		
		RMSE	MAPE	TPE (%)
ARMA Models	ARMA (3,1) with no seasonal dummies	345.04	0.10	**80.00**
	ARMA (1,2) with seasonal dummies	396.24	0.11	**80.00**
ARIMA Models	ARIMA (3,1,0) with no seasonal dummies	331.58	0.09	100.00
	ARIMA (2,1,0) with seasonal dummies	410.15	0.11	100.00
VAR Models	VAR with no seasonal dummies	**290.10**	**0.01**	,100.00
	VAR with seasonal dummies	336.25	0.04	100.00
VEC Models	VEC with no seasonal dummies	343.45	0.10	100.00
	VEC with seasonal dummies	378.80	0.03	**80.00**
Random Walk Model		311.18	0.09	100.00

As in Nampula, the three statistical criteria lead to different conclusions concerning the "best" model in performing out-of-sample forecasts for Maputo price series. In general, however, the multivariate models appear to perform better than the univariate models. Excluding the values of the RMSE for the VEC model with seasonal dummies, the RMSE and the MAPE lead to the conclusion that the multivariate models forecast better than the univariate models. This is in accordance with the economic theory as the multivariate models have better explanatory power of the variations of the forecast variable. The multivariate models include price of maize and maize production figures in Mozambique and South Africa in addition to the Maputo maize prices.

Secondly, because seasonal peaks observed in the model estimation period are not observed in the forecast period, and seasonality in Maputo maize prices is not strong in every single month even though the joint significance of the seasonal dummies is significant, the models that do not control for seasonality forecast better than those with seasonal dummy variables.

Thirdly, out of the three statistical evaluation criteria, the MAPE and the TPE indicate that the VAR with no seasonal dummies does the best job in forecasting this time series. It has the lowest RMSE and MAPE. It should be noted, however, that there is no single model that clearly outperforms the other models in forecasting the values for all observations in the period. Both in Nampula and in Maputo, all the models do a good job in forecasting some values, and they perform poorly in other occasions. Finally, all the models perform poorly in capturing the turning point errors when performing out-of-sample forecasts for Maputo maize prices.

6.4 Economic Evaluation

Under statistical evaluation criteria, models are examined from the standpoint of minimizing forecasting errors. Such criteria are not necessarily the most relevant in a decision-making framework, thus they do not tell how good the models are from the point of view of making profitable decisions based on the forecasts. Explicit economic evaluation criteria are needed to compare the models.

Often, economic performance of alternative forecasting models is evaluated based on the extent to which they lead to profitable sell or store decisions based on the signals generated by the forecast values.

To compare how well comparative models or lead to profitable decisions, sell or store signals, shown in equation (42), are calculated for each model and compared across models.

$$\textit{Sell when } P_t(1+r) > \hat{P}_{t+1}; \textit{ Store otherwise} \tag{36}$$

where P_t is the actual price at month t and $\hat{P}_{t+1}$ is the forecast price at month $t+1$.

Because each model gives different forecast values, sell/store signals could be different across models depending on the number of months in which, for each model, equation (37) suggests to sell or store. In this sense, each model generates a specific sell/store strategy.

In this research, the different strategies are compared to a default or "no model" strategy. This strategy consists of buying product in the harvest months of May, June and July, and selling equal amounts of 11.1% of product from August to April with no regard to sell/store signals. This is a reasonable strategy for three reasons. First, traders have expectation of peak prices in the hungry season thus they want to take advantage of the higher prices in later months instead of selling product according to the sell or store monthly signals. Second, traders are risk averse; they know that storing product for later months according to the sell/store signals involves risk. Since there is no certainty about the future price, thus selling equal amounts of product every month is less risky. Finally, traders need constant cash flows for operating capital, and selling product every month, whether prices are attractive or no, ensures operating capital. Under each model-generated strategy, the decision on whether the amount of product previously planned to be sold in month t is actually sold in that month or is stored into month $t+1$ depends on the sale/store signals given by the forecasts. When the forecast price for month $t+1$ is higher than the actual price in month t plus the opportunity cost of capital, the amount planned to be sold in month t is stored into month $t+1$. When, on the other hand, the price forecasted for month $t+1$ is lower than the actual price in month t plus the opportunity cost of capital, then the amount of product previously planned to be sold in month t is effectively sold in month t. Any amount of product available in April, either stored from previous months or simply the amount planned to be sold in that month is sold regardless of the sell or store signal.[12]

---

[12] The procedure to calculate the sell or store signals consists in using the annual lending interest rate of 22.8%, and obtaining the monthly compounded prices. The monthly interest rate is multiplied by the current price at each month t, and the compounded value of P_t for month $t+1$ is obtained. This indicates the value of the actual price in month t compounded into the next month. Next, the forecast price for month $t+1$ is compared to the compounded value of the actual price in month t. This comparison indicates whether selling the product at the actual price in month t is better or not than storing it into the next month, $t+1$, facing the storage costs. The opportunity cost of capital is used as a proxy to the physical storage cost due to lack of historical data on storage costs. If the actual price in month t plus the monthly opportunity cost of capital - i.e., the actual price in month t compounded into the next month - is higher than the forecast price for month $t+1$, then the decision is to sell in month t. Store otherwise.

In evaluating the profitability of the decisions derived from the comparative models , this research uses three economic criteria . The first is the mean price received, the second is the present value of the marginal revenue, and the third is the percentage of correct decisions.

The mean price received is the average price received when the sell/store signals generated by each strategy are observed, i.e., for the months in which product is sold under each strategy. If the strategy derived from a model predicts, every month, a lower price for the following month, then traders will never store additional product from one period into the next period, and the average price received will be the average monthly price for the whole period. If, on the other hand, a strategy whose one-step-ahead forecasts suggest to sell in March, June and September, and traders in fact only sell product in these three months, the mean price received by traders will be the average of the actual price in these three months.

The mean price received indicates how good or poor are the different strategies in leading to decisions which result in a high average price received. The higher the mean price received, the better the strategy generated by the model. In this regard, it is of special interest to compare the different strategies to the default strategy, and a strategy with a mean price lower than that obtained from the default strategy indicates incorrect sell/store decisions compared to selling equal amounts every month.

However, regardless of its importance in measuring future earnings, the mean price received does not take into account storage costs, thus it does not formally compare the profitability of the different strategies to buying and selling everything in July without incurring in storage costs. This is taken into account in the next criteria, the present value of the marginal revenue.

The present value of the marginal revenue is the difference between the discounted price at month *t+n* and July's price, weighted by the percentage of product sold in that month, where the proportion might change across months depending on the sell/store signals. If these signals indicate that the 11.1% planned to be sold in a given month should effectively be sold, then the actual price of that month is multiplied by 11.1% when calculating the gross marginal revenue. If, on the other hand, the sell-or-store signals suggest that the 11.1% of product previously planned to be sold in month *t* should be stored into month *t+1*, and they are effectively sold in the month *t+1* in addition to the 11.1% previously planned to be sold in that month, then the actual price of month *t+1* is weighted by 22.2%.

Summing the monthly present value of the marginal revenue under each strategy, gives present value of the marginal revenue for that strategy. Equation (38) represents this strategy.

$$\sum_{t=1}^{S}[P_{t+n}/(1+r)^{n} - P_{J}]*q \qquad (37)$$

where Pt, is the actual price in month t, $Pt+n$ is the forecast price for month $t+n$, n is the number of the month in the marketing year, being July = 1, s is the number of months in which product is sold under each strategy, t is July, (1+r) is a discount factor, and q is the proportion of the total product that is sold in each month where sales take place.[13]

The present value of the marginal gross revenue is a revenue above July's price, obtained from storing the product beyond July and facing positive storage costs, and indicates whether the strategy, overall, leads to profitable decisions or not compared to selling all the product in July. If the marginal revenue is positive, the marketing strategy leads to an overall profitable decision process compared to selling everything in July. Under the assumption that product can be sold in any of the 9 months of the marketing year, and that each strategy leads to a particular number of months in which product should be sold or stored, the values of the different economic criteria are expected to vary by strategy.

The percentage of correct decision criterion is a measure of accuracy of the strategies in predicting the right price direction change. A decision to sell or to store product is considered correct if the model predicts that the price in month $t+1$ will be higher than the actual price in month t plus the opportunity cost of capital, and this indeed happens. If the

13 The procedure to calculate the present value of the marginal revenue involves: (i) discounting the actual price for each month $t+n$ in which product is sold to obtain the present value of P_{t+n} at the harvest month P_{t}; (ii) calculating the difference between the present value of the actual price for each month $t+n$ with the actual price in July, to obtain the present value of the marginal revenue for each month. This is then weighted with the proportion of product sold in that month. In this research, prices are discounted at the opportunity cost of capital of 22.8%.

opposite happens, i.e., the model predicts one direction in the evolution of the price and the observed price in month $t+1$ is the in the opposite direction, then the model has led to an unprofitable sell or store decision.

6.4.1 Sell-or-Store Signals

Tables 6.3a and 6.3b have the results of how many months (and in which) sellers in Nampula and Maputo respectively would sell product based upon the sell and store signals given by each forecasting model or strategy.

Table 6.3a Post-Harvest Marketing Strategies Generated from Forecasting Models in Nampula

Strategy/Forecasting Model	Action-Sell in Nampula[1]	
ARMA (3,1) no seas. dummies	Aug, Sep, Oct, Nov, Jan, Feb, Mar, Apr	(1)
ARMA (1,2) with seas. dummies	Aug, Dec, Feb, Mar, Apr	(4)
ARIMA(3,1,0) no seas. dummies	Oct, Jan, Feb, Apr	(5)
ARIMA (2,1,0) w/seas. dummies	Dec, Feb, Mar, Apr	(5)
VAR no seasonal dummies	Apr	(8)
VAR with seasonal dummies	Aug, Sep, Oct, Nov, Jan, Feb, Mar, Apr	(1)
VEC no seasonal dummies	Sep, Dec, Jan, Feb, Mar, Apr	(3)
VEC with seasonal dummies	Sep, Oct, Dec, Jan, Feb, Mar, Apr	(2)
Random Walk	All 9 months	(0)

[1] Numbers in parenthesis are the number of months with no sales; product is stored into the next month.

Tables 6.3a and 6.3b indicate that the different strategies lead to different indications of the number of months in which selling product would be reasonable and those in which it would be preferable to store based on the buy and sell rule. For instance, the ARMA model with no seasonal dummies and the VAR with seasonal dummies forecast that prices will be lower in the following month for most of the months, while the VAR model with no seasonal dummies forecasts sufficient increases in prices to justify storage during 8 of the 9 months over the marketing year of 1999/2000. If traders in Nampula follow the sell or store signals given either by the ARMA model with no seasonal dummies or by the VAR with seasonal dummies, they will sell their product every month except in December, as these models suggest that the spot price in the following month will be lower than the actual price in the current month plus the opportunity cost of capital. On the other hand, if the least strategy is used (VAR with no seasonal dummies), sellers in Nampula will sell product only in April of 2000 over the marketing year as, in each month through February, this model, suggests that the price in month $t+1$ will be lower than Pt(1+r).

Table 6.3b Post-Harvest Marketing Strategies Generated from Forecasting Models in Maputo

Strategy/Forecasting Model	Action-Sell [1]	
ARMA (3,1) no seas. dummies	Nov, Jan, Feb, Apr	(5)
ARMA (1,2) with seas. dummies	Feb, Mar, Apr	(6)
ARIMA(3,1,0) no seas. dummies	Nov, Jan, Feb, Mar, Apr	(4)
ARIMA (2,1,0) w/seas. dummies	Nov, Feb, Mar, Apr	(5)
VAR no seasonal dummies	Jan, Feb, Mar, Apr	(5)
VAR w/seasonal dummies	Jan, Feb, Mar, Apr	(5)
VEC no seasonal dummies	Nov, Feb, Mar, Apr	(5)
VEC w/seasonal dummies	Nov, Feb, Mar, Apr	(5)
Random Walk	All 9 months	(0)

[1] Numbers in parenthesis are the number of months in which product is stored under each strategy.

In Maputo, four models generate strategies that lead to storing product through November, and three indicate that product should not be sold before 2000. Additionally, both in Nampula and in Maputo, the random walk models generate strategies which are similar to the default strategy as they suggest that product should be sold every month. Every month, the random walk models predict lower prices for the following month, discounting factor accounted for.

6.4.2 The Evaluation Criteria

The profitability of the decisions based on the strategies are evaluated using the three economic criteria. The results for Nampula and Maputo are presented in Tables 6.4a and 6.4b respectively.

Table 6.4a Economic Results of Post-Harvest Marketing Strategies for Nampula

	Mean Price Received	PV of Total Weighted Marginal Revenue	Percentage of Correct Decisions
Default Strategy	1341	-86	78
ARMA (2,0) with no seasonal dummies	1330	-92	67
ARMA (2,1) with seasonal dummies	1375	-43	56
ARIMA (1,1,0) with no seasonal dummies	1345	-103	44
ARIMA (1,1,0) with seasonal dummies	1391	-55	44
VAR with no seasonal dummies	1399	-117	33
VAR with seasonal dummies	1330	-92	44
VEC with no seasonal dummies	1374	-62	67
VEC with monthly dummies	1360	-77	78

Table 6.4a indicates that, if the strategy generated by the ARMA model with no seasonal dummies is followed, traders in Nampula are expected to receive the mean price of 1,330.10 Mt/kg over the marketing year. If, instead, the strategy to follow is that generated by the ARMA model with seasonal dummies, the average price received over the marketing year by Nampula traders would be 1,374.97 Mt/kg. As indicated in Table 6.4b, the same two strategies would yield a mean price received of 2,905.16 Mt/kg and 2,982.82 Mt/kg, respectively, for traders in Maputo.

On the whole, this criteria indicates that the best strategy for traders in Nampula is that generated by the VAR model with no monthly seasonal dummies, followed by the strategy derived from the ARIMA model with seasonal dummies. The least profitable strategies are those given by the ARMA with no seasonal dummies and the VAR model with seasonal dummies.

Table 6.4b Economic Results of Post-Harvest Marketing Strategies for Maputo

	Mean PV of Price Received	PV of Total Weighted Marginal Revenue	Percentage of Correct Decisions
Default Strategy	2728	588	44
ARMA (2,0) with no seasonal dummies	2905	632	44
ARMA (2,1) with seasonal dummies	2983	672	**56**
ARIMA (1,1,0) with no seasonal dummies	2905	632	44
ARIMA (1,1,0) with seasonal dummies	2881	601	44
VAR with no seasonal dummies	**2987**	**766**	**56**
VAR with seasonal dummies	**2987**	**766**	**56**
VEC with no seasonal dummies	2881	601	33
VEC with monthly dummies	2881	454	44

These results suggest that, for maize traders in Nampula, the most profitable strategy would be to buy product in the harvest months of May, June and July and store it through April and sell in this month. In fact, Figure 5.1 showed that, between August 1999 and March 2000, prices did not increase much. The average of monthly maize prices in Nampula over this period was 1,334.01 Mt/kg, smaller than the observed price of 1,398.98 Mt/kg in April 2000. Note that the mean price received does not take into account storage costs, thus if prices of maize have a decreasing trend over the period and then increase in the last month, as it happened in this marketing year, storing the product over the period with zero storage costs is better than any strategy which involves facing the low prices prior to April. This could be true for farm producers and large scale wholesalers in northern Mozambique, who have their own storage infrastructures and do not pay for storage service.

For traders in Maputo, in turn, the mean price received criteria indicates that the best strategies to follow are those generated by the two VAR models, and the second profitable strategy is that derived from the ARMA model with seasonal dummies. The least profitable strategies for Maputo are those given by the no-model and the random walk model. These results suggest that for maize traders in Maputo, the best strategy for the 1999/2000 marketing year was to buy product in the harvest months pf May, June and July 1999, and sell it from January to April 2000.

All the strategies, except the ARMA with no seasonal dummies and the VAR model with seasonal dummies in Nampula, give better results when compared to the default or "no model" strategy.

Similar to the mean price received, this measure takes into account buying the product in May, June and July and making sell/store decisions according to equation (42)'s decision rule. Because the present value of the marginal gross revenue compares the average of the present value of the price received for each strategy to the price in the harvest month of July, it is expected to be either positive, null, or negative. A positive PV marginal gross revenue indicates that the strategy leads to a more profitable sale/storage strategy compared to the marketing mark-up given by buying and selling the product in July, with zero storage cost. A negative PV marginal gross revenue indicates that traders are better off buying and reselling everything in July (or May, June and July). Finally, a null PV marginal gross revenue could indicate indifference between buying and reselling everything in July and following the strategy. However, indifference could not happen as risk averse traders would prefer to sell everything in July and invest the capital in less risky activities such as to earn the opportunity cost of capital. Traders with a more speculative behavior, on the other hand, would probably store product and expect to gain from better prices in the future.

Table 6.4a indicates that all the strategies lead to unprofitable decisions in Nampula compared to buying and selling the product in July. However, the ARMA and ARIMA models with no seasonal dummies, and the two VEC models lead to better strategies than the default strategy. In fact, the fact that the price behavior was highly different between the model estimation period and the forecasting period led to unprofitable sell/store decisions. For instance, the sell/store signals indicated that the best strategy would be to store everything through April 2000, but when the opportunity cost of capital is included in the analysis, this strategy generates the worse results. Also, the percentage of correct decisions criterion indicates that, under this strategy, 67% of the decisions were not correct, i.e., out of the nine months, the decisions were correct only in three months.

Table 6.4b indicates positive marginal gross revenue, which suggests that all strategies are more profitable than buying and selling everything in July. Also, except the VEC model with seasonal dummies, all the other models give better results than the default strategy.

Finally, the percentage of correct decisions criterion indicates that the strategies leading to accurate decisions most of the times in Nampula are those originated from the VEC with seasonal dummies and the default model. The least accurate strategies in Nampula is that derived from the VAR model with no seasonal dummies. In Maputo, in turn, the ARIMA model with seasonal dummies and the two VAR models are those leading to correct decisions most times than the others.

Overall, the economic criteria suggest that the multivariate models lead to better sell/store decisions compared to the univariate models. Also, the models tend to do a better job in Maputo than they do in Nampula. This is not surprising as the difference in the behavior of the prices between the model estimation period and the forecasting period is substantially higher in Nampula than in Maputo, as suggested by Figures 5.1 and 5.2. Under such circumstances, time series models hardly could predict the path of the price.

6.5 Conclusions

The statistical evaluation has shown that, for each model, there are periods of good forecasts and others of major problems. The statistical evaluation criteria have shown that the multivariate models tend to do better forecasts over the harvest season compared to the univariate models both in Nampula and in Maputo. This conclusion is also met when the economical criteria are applied. However, the multivariate models do not seem to lead to much improvement in the forecasting ability of the univariate models. This could be related to two things. First, the data on production, which could be expected to explain the high fluctuations in the forecast prices do not seem to do this job. The data does not meet the quality requirements needed for it to improve the quality of the models. Second, these models could be better if more variables were added, variables thought of as helping explain the shocks to the forecast time series.

Also, it seems that models with seasonal dummies tend to have considerable forecasting errors over the months of high peaks, as the data tended to have considerable picks over the model estimation period, which the models expect to happen in the forecast period, but did not happen.

Finally, the negative mean gross marginal revenue under all strategies in Nampula suggest that, in this market, traders are not using the storage facilities or other ways of payment rather than cash payment are being used. The extreme case that no storage cost is faced or no product is stored for a month or more is the only situation which can explain that traders operate in this market.

Chapter 7

CONCLUSIONS AND IMPLICATIONS

7.1 Introduction and Research Questions

Price forecasting is an important tool to ensure that participants in the food system plan the quantity and quality of product to produce, store, sell or consume. Predictions of the path of agricultural commodities prices is critical in countries like Mozambique, whose economy basically depends upon agricultural commodities. The importance of such predictions in reducing uncertainty about returns to production and trade depend, however, on consistent public policies aimed at ensuring broad improvement of the whole set of conditions in the economy as a whole, and incentives to the private sector for solid investments in the agricultural sector and the rural economy. Government's role is very important in this context. Better physical infrastructure, consistent policies for better production and trade, and solid marketing institutions are critical for the success prediction models.

Maize production in Mozambique depends on natural conditions, and maize prices fluctuate considerably between the harvest and the planting seasons, and between producer and deficit areas. Nearly all maize is produced by smallholder households. Expected profits are usually low compared to cash crops, and producers assume individually the risk associated with the quality of product, and the decision of how much to produce, store and sell. The level of uncertainty can, however, be reduced if accurate and timely predictions of future behavior take place along with consistent policies by the public sector and incentives for solid investments by the private sector.

This research is a first step in building forecasting models for maize prices in Mozambique. The research has been directed toward two main objectives. The first objective was to estimate alternative univariate and multivariate models for short-term forecasts of maize prices in Mozambique, and examine how well they can improve forecasts as compared to random walk models. The second objective was to evaluate, using statistical and economic criteria, how the alternative forecasting models, built upon limited price and production data, do in giving forecasts outside the model estimation period.

7.2 Research Methods

The present research used monthly retail price data for four Mozambican markets, data on actual maize production in Mozambique and Malawi, and monthly maize production estimates in South Africa.

In modeling maize prices in Mozambique, two regions were identified, the northern region and the central/southern region. Retail maize prices for Nampula and Mocuba, and data on actual production in Mozambique and Malawi were used for the northern region. Retail maize prices for Maputo and Manica, and production data for Mozambique and South Africa were used for the central/southern region.

Time series techniques were employed to identify the data generating process for all the data series. Specifically, Dickey-Fuller and augmented Dickey-Fuller tests were used to perform unit root tests on the data. These tests indicated that the data generating mechanism for all the data series had a unit root, thus first differences were taken before other preliminary data search tools, such as checking for seasonality and linear trend, were used. The final univariate models were estimated based on the investigations of the autocorrelation functions and partial autocorrelation functions. Granger causality tests and cointegration tests were used as a basis for identifying the degree of relationship of the price series involved in the multivariate models.

Knowledge of the structure and organization of the subsector is key for well-formulated price analysis. As a way to understand the characteristics, structure and organization of Mozambique's maize subsector, this research reviewed the basic conditions in this subsector, including agro-ecological and technological conditions on the supply side, and the basic characteristics of demand. Detailed subsector maps were developed for each region, and the characteristics and functions of the participants in the supply chain were examined. Coordination problems and indicators of performance of the subsector were also reviewed.

Finally, several univariate and multivariate forecasting models were estimated and their forecasting ability compared along with the random walk models. Both statistical and economic evaluation criteria were used. The economic evaluation involved storage problems. The basic rule of the storage problem was that product would be stored until price is higher than today's unit price plus per unit storage cost. Because historical information on storage costs is not available, the opportunity cost of capital was used. The decision rule became, then, that product would be stored from period t into period $t+1$ if price at period t plus the opportunity cost of capital were lower than the predicted price for period $t+1$. The profitability of the different models were compared to that from a default model of selling equal amounts every month, and with the scenario of selling everything in July with zero storage cost.

7.3 General Findings

The different stages of data investigation and model estimation led to several important findings, systemized and grouped into 1) those related to the organization of the maize subsector in the country, 2) those related to the characteristics of the maize data, and 3) those regarding the comparative performance of the forecasting models.

7.3.1 Organization of the Maize Subsector

Agricultural technology is rudimentary in Mozambique. Nearly all maize is cultivated by smallholder households, with the use of hand tools, and around 90 percent of the produced grain is consumed on the farm. Due to normal lag in production and the use of rudimentary tools, maize in Mozambique is supply inelastic in the short-run.

These characteristics of farm level production have a relationship with the organization in the remaining stages of the subsector. For instance, maize assemblers in Mozambique acquire small quantities from many small farmers, thus assembling product takes longer that it would under other circumstances. The level of participation of formal and informal traders is different between the north and the center/south. While in the north, the informal sector is still very weak and large-scale wholesalers trade maize, there is a solid and specialized informal sector in the center/south, where formal traders import maize but do not trade domestic maize.

In addition, most of the informal maize grain wholesalers in Mozambique operate at a small scale, usually located in the most important consumer centers or geographically strategic areas, and without access to formal credit. Large-scale wholesalers, with access to working capital, entered into the maize marketing system in the north when opportunities to export grain to neighboring countries, especially into Malawi, appeared.

Finally, because maize is a basic food crop in Mozambique, it is demand inelastic. Furthermore, the maize subsector in Mozambique faces coordination problems, the most important of which are: the high cost of marketing from north to south, instability of returns to storage for large-scale traders and extremely high operating costs and low returns to storage for small-scale traders, limited information about market opportunities, difficult and high-cost access to working capital, very low farm yields, and inefficiency in the small-scale custom milling industry.

7.3.2 Characteristics of the Data

In this research, the prices were subject to various specifications of the Dickey-Fuller tests, and augmented Dickey-Fuller tests when the results revealed that the innovations were serially correlated. The results of these tests indicated that Mozambique's maize retail prices are stationary in first differences, i.e., contain a stochastic trend, and after first differences were taken, the results indicated that there is an upwards drift, but it is not significant in three of the four series. All the price data were found to have a significant seasonal component. Seasonal dummy variables were found to be jointly significant in all the time series, thus models with monthly seasonal dummy variables were estimated.

The data series were also checked for structural shocks, as it appears that there could be structural shocks affecting the normal fluctuation of maize prices in Mozambique. Nonetheless, there was no statistical evidence of structural shocks.

7.3.3 Forecasting Models Results

Although the distinction between data with and without unit root is of crucial importance in time series analysis, it is not of critical importance when the final objective is to perform short-term forecasts. This research estimated ARMA, ARIMA, VAR and VEC models, and found that, in each class of models, the models with seasonal dummies generally overestimated seasonal movements, because price movement during the forecasting period was much less than typical movement during the estimation period.

Second, the multivariate models slightly improved the forecasting ability of the univariate models. The univariate models were outperformed by the multivariate models because, while in the former models the time series itself explains variations in the price series and predicts future values of the price, in the multivariate models other variables, hypothesized to increase the power of explanation of the data generating mechanism of the series to be forecast are included. This improvement was not major because the poor quality of the additional data, especially the data on maize production in Mozambique and Malawi.

The statistical evaluation indicated that, for each model, there are periods of good forecasts and others with major problems. The statistical evaluation criteria also showed that the multivariate models tended to do better forecasts compared to the univariate models.

Finally, economic evaluation showed that, in Nampula, all the models led to unprofitable sell/store decisions compared to selling everything in July, while in Maputo they led to positive results. Prices in Nampula were very low during the forecasting period, compared to those in Maputo. Compared to the opportunity cost of capital, net returns in Nampula easily become negative while in Maputo can still profitable.

When compared to the default strategy, the different economic criteria indicate that, for Nampula, only the ARMA model without seasonal dummies and the VAR model with seasonal dummies lead to less profitable sell/store signals, while the default strategy gives better economic results than all the models except the VEC models under the present value of the marginal revenue, and is more profitable than all the models except the VEC model with monthly seasonal dummies under the percentage of correct decisions. In Maputo, the mean price received indicates that default strategy is outperformed by all the model-generated strategies, and this finding is not contradicted when the present value of the marginal revenue and the percentage of correct decisions criteria are considered. Only the VEC models (with seasonal dummies and without seasonal dummies, respectively) give less profitability than the default strategy.

7.4 Lessons and Implications for Further Research

Given the importance of price predictions for the stability of production and prices to the food system and to the economy as a whole, systematic and solid price analysis and modeling is needed in Mozambique. When farmers have knowledge of the predictions of the price path, they easily adjust volumes to produce, to store and sell, increasing their earnings and family income. Likewise, if traders have knowledge of the prediction of the prices, they plan better where to buy and sell their product, reducing the level of uncertainty. In addition, previous knowledge of the path of the prices improves consumers' budget planning and policymakers' national programs for the agricultural sector, food security, and rural development. However, knowledge of the prediction of price paths only can be useful to producers and consumers if the system as a whole has improved conditions. Access to information, credit, improved storage infrastructure and better roads are other variables needed for the aimed reduction of uncertainty.

In Mozambique, there is a consistent database on maize and other food crop staples prices at various transaction levels covering most of the rural and urban areas of the country. This database is, however, still weak. Weaker, yet, is the data on producer level price. For this reason, this research modeled retail prices instead of producer prices, the latter being more commonly used in this nature of research. If the systematic price data collection process, started by Mozambique's SIMA in 1991, and improved and extended over the decade, continues, models with producer prices can be estimated in the near future.

The quality of the data on maize production needs to be improved. Because monthly updates of production in Mozambique and Malawi were not available, annual figures were repeated monthly, which reduced the variability of these data series and hence their power of improving the multivariate models. Second, the data on Mozambique production does not show the drought of 1994/95, therefore it does not help explain the spike in the prices observed in 1995/96.

This research is but a first step in the use of data on Mozambique's maize price and production to develop forecasting models for the country. In addition to the need to continue the systematic and consistent work that Mozambique's SIMA has been doing, much more analysis can be done using SIMA data. The models developed in this research can be improved and updated with the use of additional data on the variables, and with the inclusion of other variables such as producer and wholesaler prices, monthly updates of production estimates for Mozambique and neighboring countries, and information on carry-out/carry-in stocks, especially at the farm level. Also, models with logarithmic variables can be examined and compared to models with the variables in levels.

The limitations of this research have implications for further analyses, both under the current and under other model specifications. In general, univariate econometric time-series models are unlikely to give good predictions if the price path changes frequently and substantial differences are observed between the model estimation period and the out-of-sample forecasting period. The results of this research suggest that under the current specifications, multivariate models are likely to overcome this limitation of the univariate models if higher quality data is available. Moreover, other model specifications, especially structural models, should also be examined in future work. In addition, the forecast performance of the models should be subject to a comparative examination involving years with different price evolution patterns, along with a permanent update of qualitative information.

References

Aldridge, Kimberly M. 1999. The Path of Grain Prices With Limited Data: A Case Study From Mali. Ph.D. Dissertation. Michigan State University.

Alavalapati, Janaki R.R., Wiktor L. Adamowicz, and Martin K. Luckert. 1997. *A Cointegration Analysis of Canadian Wood Pulp Prices*. American Journal of Agricultural Economics. 79 (August 1977): 975-986.

Asche, Frank, Helge Bremes and Cathy R. Wessells. 1999. *Product Aggregation, Market Integration and Relationship Between Prices: An Application to World Salmon Markets*. American Journal of Agricultural Economics. 81 (August 1999): 568-581.

Baulch, B. 1997. *Transfer Costs, Spatial Arbitrage, and Testing for Food Market Integration*. American Journal of Agricultural Economics. 79 (May 1997):477-87.

Barret, C. 1996. *Market Analysis Methods: Are Our Enriched Tools Well-Suited to Enlivened Market?* American Journal of Agricultural Economics. 78 (Aug 1996):825-29.

Boughton, Duncan, E. Crawford, J. Howard, J. Oehmke and J. Staatz. 1995. *A Strategic Approach to Agricultural Research Program Planning in Sub-Saharan Africa*. MSU International Development Working Paper No. 49. Department of Agricultural Economics/Department of Economics. Michigan State University.

Brandt, Jon A. and David A. Bessler. 1981. *Composite Forecsting: An Application with U.S. Hog Prices*. American Journal of Agricultural Economics. 63 (February 1981): 135-140.

Donovan, Cynthia. 1996. Effects of Monetized Food Aid on Local Maize Prices in Mozambique. Ph.D. Dissertation. Michigan State University.

Eviews User's Guide. 1998. Quantitative Micro Software. Irvine. CA, USA.

Ferris, John. 1998. Agricultural Prices and Commodity Market Analysis. WCB/McGraw-Hill, Boston, Massachusetts.

Goetz, Stephan and Michael T. Weber. 1986. *Fundamentals of Price Analysis in Developing Countries' Food Systems: A Training Manual to Accompany the Microcomputer Software Program 'MSTAT.'* MSU International Paper No. 29. Department of Agricultural Economics. East Lansing, MI.

Gujarati, D. 1995. Basic Econometrics - Third Edition. McGraw Hill, Inc. United States of America.

Harrison, Kelly, D. Henley, H. Riley and J. Shaffer. 1987. *Improving Food Marketing Systems in Developing Countries Experiences from Latin America*. MSU International Development Working Paper No. 9. Department of Agricultural Economics. Michigan State University.

Holtzman, John S. 1986. *Rapid Reconnaissance Guidelines for Agricultural Marketing and Food System Research in Developing Countries*. MSU International Development Working Paper No. 30. Department of Agricultural Economics. Michigan State University.

Intriligator, Michael D., Ronald D. Bodkin and Cheng Hsiao 1996. *Forecasting*. Chapter 15 in Econometric Models, Techniques, and Applications, 2nd Edition. Englewood Cliffs: Prentice Hall.

Jesse, Edward V. 1978. *Measures of Market Performance: Quantifying the Non-Quantifiable*. WP-15. Cooperative Service, U.S. Department of Agriculture/Department of Agricultural Economics, University of California. Davis.

Kennedy, Peter. 1998. A Guide to Econometrics. IJ International. United Kingdom.

Lai, Jing-Yi. 1999. Optimal Post-Harvest Grain Marketing Strategies in a Risky Market Environment. Ph.D. Dissertation. Michigan State University.

MOA/MSU Research Team. 1997. *A Disponibilidade de Produtos e o Poder de Compra dos Consumidores: Graos de Milho e Arroz*. Flash 7P. Ministerio da Agricultura e Pescas. Maputo.
MOA/MSU Research Team. 1997. *A Disponibilidade de Produtos e o Poder de Compra dos Consumidores: Farinhas de Milho e Arroz*. Flash 8P. Ministerio da Agricultura e Pescas. Maputo.

MOA/MSU Research Team. 1995. *Diagnistico da Estrutura, Comportamento e Desempenho dos Mercados Alimentares em Mocambique*. Relatorio de Pesquisa No. 19. Ministerio da Agricultura e Pescas. Maputo.

Myers, R. 1994. Time Series Econometric and Commodity Price Analysis: A Review of Marketing and Agricultural Economics 62 (August, 1994): 167-181

Santos, Ana Paula and David Tschirley. 1999. *The Effect of Maize Trade with Malawi on Price Levels in Mozambique: Implications for Development Policy*. Department of Agricultural Economics, Michigan State University. East Lansing.

Staatz, John. 1997. Notes on the Subsector Analysis as a Diagnostic Tool for Linking Industry and Agriculture. Department of Agricultural Economics. Michigan State University.

Strasberg, Paul. 1997. Notes on a Short Term Consultancy Regarding Data Validation and Analysis of 'Trabalho de Inquerito Agricola, 1996'. (A Memo to Domingos Diogo. - Not Published)

Tomek, William G. and Robinson, Kenneth. 1987. Agricultural Product Prices. Cornell University Press. Fourth Edition.

Tschirley, David and A. Santos. 1998. *Desafios Para Garantir a Concorrencia e Reduzir os Custos No Sistema Alimentar de Mocambique*. Relatorio de Pesquisa No. 28. Ministerio da Agricultura e Pescas. Maputo.

Tschirley, David. 1998. *Regional Trade and the Economic Development of Northern Mozambique*. Flash 13E. Ministry of Agriculture and Fisheries. Maputo.

Wooldridge, Jeffrey M. 2000. Introductory Econometrics: A Modern Approach. South-Western College Publishing, a division of Thomson Learning. USA.